

Санкт-Петербургский Государственный Политехнический университет
Высшая школа прикладной математики и вычислительной физики
Дисциплина "Математическая статистика"

Отчёт по лабораторной работе №1

Работу выполнил:
Крупница С.В.
Группа: 5030102/20101
Преподаватель:
Баженов А.Н.

Содержание

1	Формулировка задания	2
2	Ссылка на GitHub	2
3	Теоретические сведения	2
3.1	Распределения	2
3.2	Статистики	3
4	Данные	3
5	Графики	3
6	Вывод	4

1 Формулировка задания

Для 4 распределений:

- Нормальное распределение $N(x, 0, 1)$
- Распределение Коши $C(x, 0, 1)$
- Распределение Пуассона $P(k, 1, 0)$
- Равномерное распределение $U(x, -\sqrt{3}, \sqrt{3})$

Задания:

1. Сгенерировать выборки размером 10, 50 и 1000 элементов. Построить на одном рисунке гистограмму и график плотности распределения.
2. Сгенерировать выборки размером 10, 100 и 1000 элементов. Для каждой выборки вычислить следующие статистические характеристики положения данных: \hat{x} , $medx$, z_Q . Повторить такие вычисления 1000 раз для каждой выборки и найти среднее характеристик положения и их квадратов, вычислить оценку дисперсии, представить полученные данные в виде таблицы.

2 Ссылка на GitHub

https://github.com/Cegeria/Polytech_Statistics/new/main/lab_1

3 Теоретические сведения

3.1 Распределения

Нормальное распределение $N(0, 1)$: Плотность распределения:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Матожидание: $M(X) = 0$, дисперсия: $D(X) = 1$.

Распределение Коши $C(0, 1)$: Плотность распределения:

$$f(x) = \frac{1}{\pi} \frac{1}{1 + x^2}.$$

У распределения Коши не существуют математическое ожидание и дисперсия (в классическом смысле) из-за тяжёлых хвостов.

Распределение Пуассона $P(\lambda = 10)$: Функция вероятностей (pmf):

$$P(X = k) = \frac{10^k e^{-10}}{k!}, \quad k = 0, 1, 2, \dots$$

Матожидание: $M(X) = 10$, дисперсия: $D(X) = 10$.

Равномерное распределение $U(-\sqrt{3}, \sqrt{3})$: Плотность распределения:

$$f(x) = \begin{cases} \frac{1}{2\sqrt{3}}, & x \in [-\sqrt{3}, \sqrt{3}], \\ 0, & \text{иначе.} \end{cases}$$

Матожидание: $M(X) = 0$, дисперсия: $D(X) = 1$.

3.2 Статистики

Пусть x_1, x_2, \dots, x_n — выборка размера n . Тогда:

- Выборочное среднее:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

- Выборочная медиана (med) — центральный элемент упорядоченной выборки (либо среднее двух центральных при чётном n) (2)
- Полусумма квартилей (25% и 75%):

$$z_Q = \frac{z_{1/4} + z_{3/4}}{2}, \quad (3)$$

где $z_{1/4}$ — 25%-квантиль, $z_{3/4}$ — 75%-квантиль.

По результатам 10, 50 и 1000 повторных экспериментов для каждой статистики вычисляются её среднее значение и дисперсия:

$$E(z) = \langle z \rangle, \quad D(z) = \langle (z - E(z))^2 \rangle \quad (4).$$

4 Данные

В данной таблице представлены все необходимые данные для каждого распределения и каждой выборки:

	Размер выборки	\bar{x} (1)	$\text{med}x$ (2)	z_Q (3)	\bar{x}^2	$\text{med}x^2$	z_Q^2	$D(x)$ (4)
Cauchy	10	3.70	0.02	-0.01	$2 \cdot 10^4$	0.32	0.91	$2 \cdot 10^4$
	50	7.82	0.01	0.01	$3.4 \cdot 10^4$	0.05	0.10	$3.3 \cdot 10^4$
	1000	1.20	0.00	0.00	$1.3 \cdot 10^3$	0.00	0.00	$1.3 \cdot 10^3$
Normal	10	0.01	0.00	0.02	0.10	0.14	0.12	0.10
	50	-0.00	-0.01	-0.00	0.02	0.03	0.03	0.02
	1000	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Poisson	10	10.02	9.87	9.93	101.44	98.87	99.86	0.99
	50	9.98	9.82	9.89	99.87	96.73	98.07	0.19
	1000	10.00	10.00	10.00	100.01	99.97	99.93	0.01
Uniform	10	0.01	0.02	0.01	0.10	0.22	0.14	0.10
	50	0.00	0.01	0.00	0.02	0.05	0.03	0.02
	1000	-0.00	-0.00	-0.00	0.00	0.00	0.00	0.00

Таблица 1: Статистические характеристики для различных распределений и размеров выборок.

5 Графики

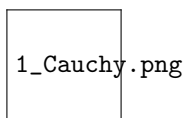


Рис. 2 График распределения Коши

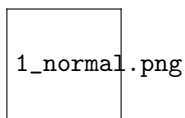


Рис. 3 График нормального распределения

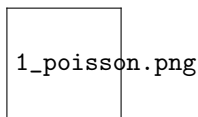


Рис. 4 График распределения Пуассона

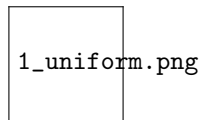


Рис. 5 График равномерного распределения

6 Вывод

По полученным данным наглядно видно, что распределение Коши не подчиняется центральной предельной теореме - с увеличением размера выборки дисперсия растет. При этом проблемы возникают именно из-за среднего квадрата - квадрат среднего не сильно влияет. Для остальных распределений четко видно, что дисперсия убывает с ростом размера выборки.