Санкт-Петербургский политехнический университет
Высшая школа прикладной математики и
вычислительной физики, ИПММ

Направление подготовки
01.03.02 «Прикладная математика и информатика»

Отчет по лабораторным работам №1,2
по дисциплине «Математическая статистика»

Выполнил студент гр. 3630102/80201
Кирпиченко С. Р.
Руководитель
Баженов А. Н.

Санкт-Петербург
2021

Содержание

Страница

1 Постановка задачи **5**

2 Теория **5**

2.1 Рассматриваемые распределения 5

2.2 Гистограмма 6

2.2.1 Построение гистограммы 6

2.3 Вариационный ряд 6

2.4 Выборочные числовые характеристики 7

2.4.1 Характеристики положения 7

2.4.2 Характеристики рассеивания 8

3 Реализация **8**

4 Результаты **8**

4.1 Гистограммы и графики плотности распределения 8

4.2 Характеристики положения и рассеяния 11

5 Обсуждение **13**

5.1 Гистограмма и график плотности распределения 13

5.2 Характеристики положения и рассеяния 13

Список иллюстраций

Страница

1 Нормальное распределение (3) 8

2 Распределение Коши (4) 9

3 Распределение Лапласа (5) 9

4 Распределение Пуассона (6) 10

5 Равномерное распределение (7) 10

Список таблиц

Страница

1 Нормальное распределение (3) 11
2 Распределение Коши (4) 11
3 Распределение Лапласа (5) 12
4 Распределение Пуассона (6) 12
5 Равномерное распределение (7) 13

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

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: $\overline{x}, \operatorname{med} x, z_R, z_Q, z_{tr}$. Повторить такие вычисления 1000 раз для каждой выборки и найти среднее характеристик положения и их квадратов:

$$E(z) = \overline{z} \tag{1}$$

Вычислить оценку дисперсии по формуле:

$$D(z) = \overline{z^2} - \overline{z}^2 \tag{2}$$

Представить полученные данные в виде таблиц.

2 Теория

2.1 Рассматриваемые распределения

Плотности:

- Нормальное распределение

$$N(x, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{3}$$

- Распределение Коши

$$C(x, 0, 1) = \frac{1}{\pi} \frac{1}{x^2 + 1} \tag{4}$$

- Распределение Лапласа

$$L(x, 0, \frac{1}{\sqrt{2}}) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|} \tag{5}$$

- Распределение Пуассона

$$P(k, 10) = \frac{10^k}{k!} e^{-10} \tag{6}$$

- Равномерное распределение

$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}} & \text{при} \quad |x| \leq \sqrt{3} \\ 0 & \text{при} \quad |x| > \sqrt{3} \end{cases} \tag{7}$$

2.2 Гистограмма

2.2.1 Построение гистограммы

Множество значений, которое может принимать элемент выборки, разбивается на несколько одинаковых интервалов, откладываемых на горизонтальной оси, над каждым из которых затем рисуется прямоугольник. Высота каждого прямоугольника пропорциональна числу элементов выборки, попадающих в соответствующий интервал.

2.3 Вариационный ряд

Последовательность $\{x_{(k)}\}_{k=1}^{n}$ элементов выборки размера n, расположенных в неубывающем порядке, называется вариационным рядом.

2.4 Выборочные числовые характеристики

2.4.1 Характеристики положения

- Выборочное среднее

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{8}$$

- Выборочная медиана

$$\operatorname{med} x = \begin{cases} x_{(l+1)} & \text{при } \ n = 2l+1 \\ \dfrac{x_{(l)} + x_{(l+1)}}{2} & \text{при } \ n = 2l \end{cases} \tag{9}$$

- Полусумма экстремальных выборочных элементов

$$z_R = \frac{x_{(1)} + x_{(n)}}{2} \tag{10}$$

- Полусумма квартилей

  Выборочный квартиль z_p порядка p определяется формулой

$$z_p = \begin{cases} x_{([np]+1)} & \text{при } \ np \text{ дробном,} \\ x_{(np)} & \text{при } \ np \text{ целом} \end{cases} \tag{11}$$

  Полусумма квартилей

$$z_Q = \frac{z_{1/4} + z_{3/4}}{2} \tag{12}$$

- Усечённое среднее

$$z_{tr} = \frac{1}{n-2r}\sum_{i=r+1}^{n-r} x_{(i)}, \quad r \approx \frac{n}{4} \tag{13}$$

2.4.2 Характеристики рассеивания

Выборочная дисперсия

$$D = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \tag{14}$$

3 Реализация

Лабораторная работа выполнена на языке Python 3.9 с использованием библиотек numpy, scipy, matplotlib, seaborn.

4 Результаты

4.1 Гистограммы и графики плотности распределения

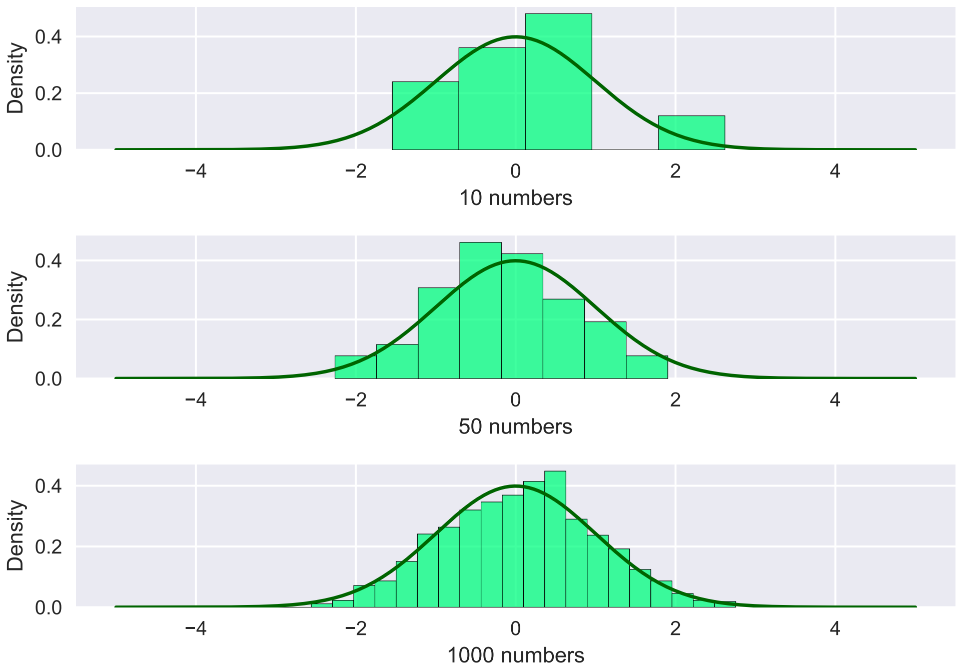

Рис. 1: Нормальное распределение (3)

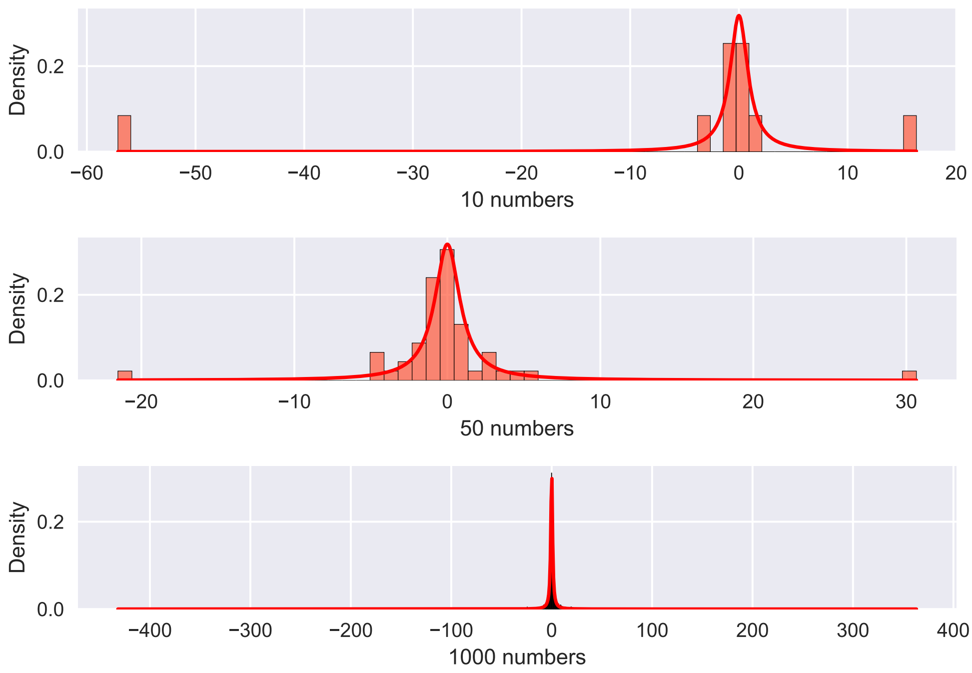

Рис. 2: Распределение Коши (4)

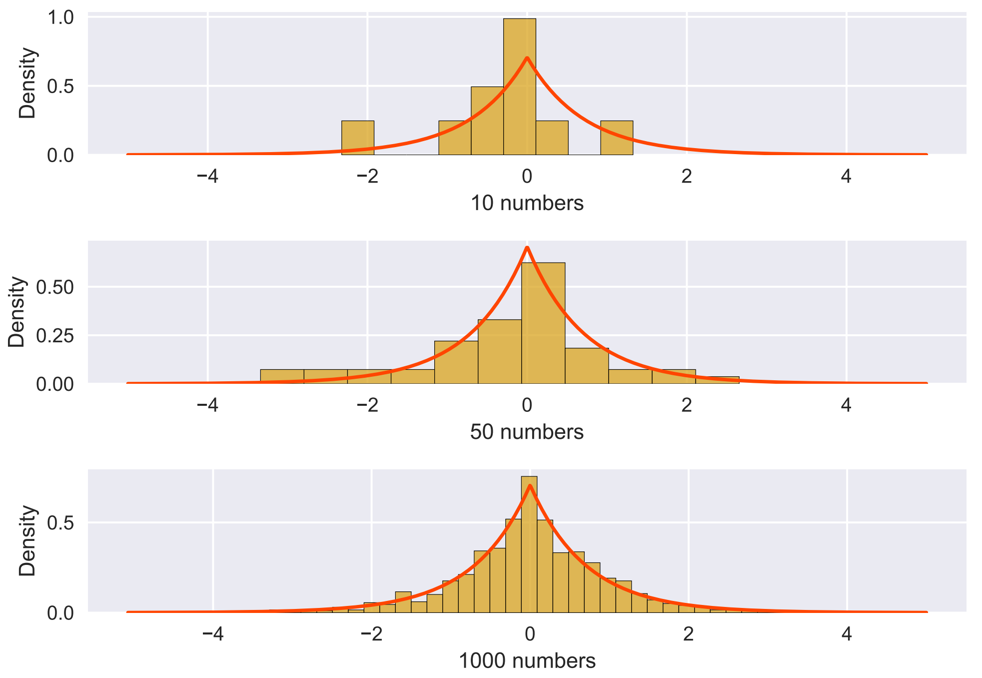

Рис. 3: Распределение Лапласа (5)

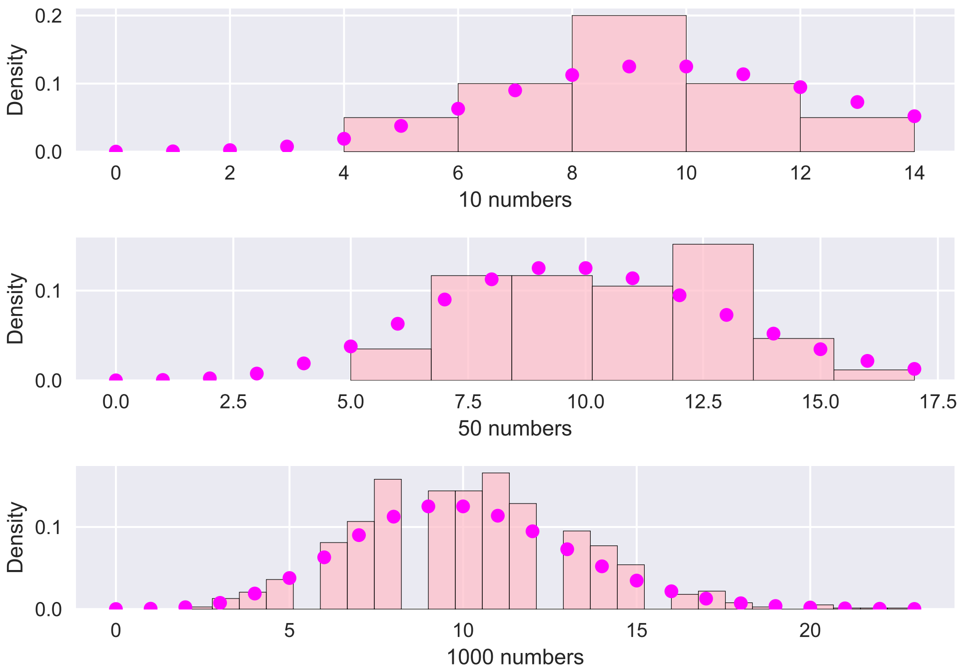

Рис. 4: Распределение Пуассона (6)

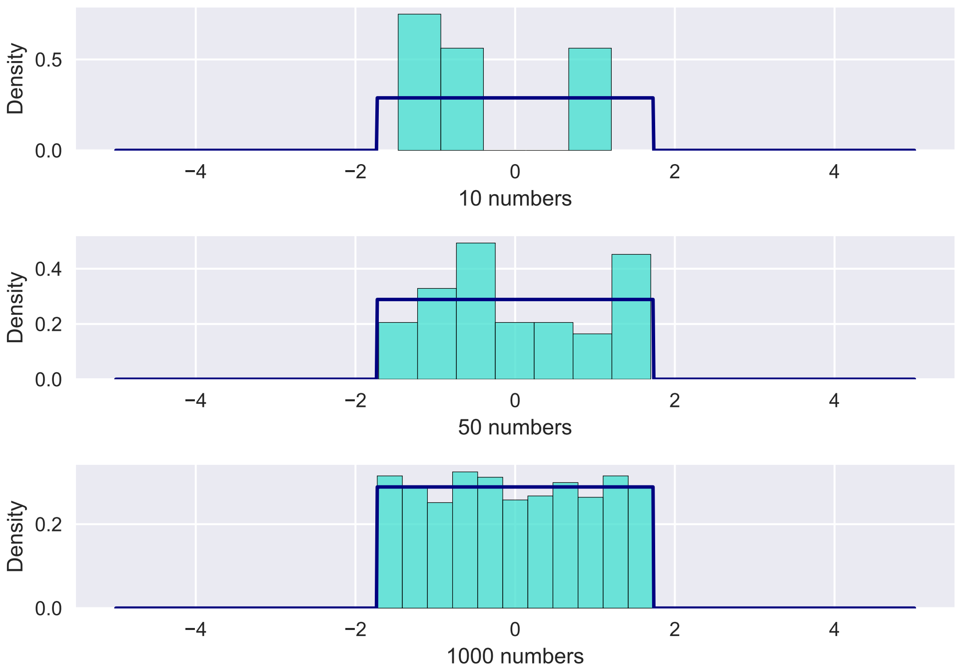

Рис. 5: Равномерное распределение (7)

4.2 Характеристики положения и рассеяния

$\hat{E}(z)$ - оценка, для ее вычисления нужно взять общие знаки в записи чисел $\hat{E}(z) = E(z) \pm \sqrt{D(z)} \cdot K_\alpha$. Вообще говоря, K_α - числовая характеристика конкретного закона распределения, но в данной работе будет использовано $K_\alpha = 1$ для всех вычислений.

normal $n = 10$	$\overline{x}$ (8)	med x (9)	z_R (10)	z_Q (12)	z_{tr} (13)
$E(z)$	-0.017122	-0.021506	-0.006723	0.287411	0.250703
$D(z)$	0.096772	0.137949	0.182028	0.125442	0.113678
$\hat{E}(z)$	0	0	0	0	0
normal $n = 100$	$\overline{x}$	med x	z_R	z_Q	z_{tr}
$E(z)$	0.002969	0.003794	0.010928	0.021474	0.030476
$D(z)$	0.009869	0.015401	0.09542	0.012189	0.01184
$\hat{E}(z)$	0	0	0	0	0
normal $n = 1000$	$\overline{x}$	med x	z_R	z_Q	z_{tr}
$E(z)$	-0.000267	-0.000109	0.008167	0.000739	0.002331
$D(z)$	0.000985	0.001575	0.066806	0.001219	0.001181
$\hat{E}(z)$	0	0	0	0	0

Таблица 1: Нормальное распределение (3)

cauchy $n = 10$	$\overline{x}$ (8)	med x (9)	z_R (10)	z_Q (12)	z_{tr} (13)
$E(z)$	0.366552	-0.032657	2.115823	1.06166	0.643818
$D(z)$	1893.549127	0.340705	47164.182247	4.010975	1.016329
$\hat{E}(z)$	0	0	0	0	0
cauchy $n = 100$	$\overline{x}$	med x	z_R	z_Q	z_{tr}
$E(z)$	-2.405841	0.001288	-123.15451	0.037977	0.043554
$D(z)$	12940.014738	0.024268	32301517.038689	0.04981	0.025628
$\hat{E}(z)$	0	0	0	0	0
cauchy $n = 1000$	$\overline{x}$	med x	z_R	z_Q	z_{tr}
$E(z)$	0.885873	-0.00089	450.266753	0.002554	0.002604
$D(z)$	964.067369	0.002516	236574408.487179	0.004797	0.002512
$\hat{E}(z)$	0	0	0	0	0

Таблица 2: Распределение Коши (4)

laplace $n = 10$	$\overline{x}$ (8)	med x (9)	z_R (10)	z_Q (12)	z_{tr} (13)
$E(z)$	0.006836	0.002158	0.006613	0.305288	0.236799
$D(z)$	0.10008	0.06789	0.409234	0.120073	0.081557
$\hat{E}(z)$	0	0	0	0	0
laplace $n = 100$	$\overline{x}$	med x	z_R	z_Q	z_{tr}
$E(z)$	-0.00058	0.000292	0.017743	0.015558	0.019848
$D(z)$	0.009482	0.005524	0.419544	0.009383	0.005949
$\hat{E}(z)$	0	0	0	0	0
laplace $n = 1000$	$\overline{x}$	med x	z_R	z_Q	z_{tr}
$E(z)$	-0.000706	-0.000469	-0.014639	0.001119	0.001636
$D(z)$	0.000912	0.000518	0.374504	0.000975	0.000594
$\hat{E}(z)$	0	0	0	0	0

Таблица 3: Распределение Лапласа (5)

poisson $n = 10$	$\overline{x}$ (8)	med x (9)	z_R (10)	z_Q (12)	z_{tr} (13)
$E(z)$	10.0068	9.889	10.248	10.9545	10.790833
$D(z)$	0.961674	1.412679	1.854996	1.29168	1.180888
$\hat{E}(z)$	0	0	0	0	0
poisson $n = 100$	$\overline{x}$	med x	z_R	z_Q	z_{tr}
$E(z)$	9.98998	9.851	10.9185	9.9585	9.94044
$D(z)$	0.095308	0.218299	0.989608	0.152028	0.116525
$\hat{E}(z)$	0	0	0	0	0
poisson $n = 1000$	$\overline{x}$	med x	z_R	z_Q	z_{tr}
$E(z)$	10.009293	9.9975	11.696	9.9965	9.875968
$D(z)$	0.01012	0.002244	0.647084	0.002738	0.011653
$\hat{E}(z)$	0	0	0	0	9

Таблица 4: Распределение Пуассона (6)

uniform $n = 10$	$\overline{x}$ (8)	med x (9)	z_R (10)	z_Q (12)	z_{tr} (13)
$E(z)$	-0.005577	-0.002646	-0.00692	0.310645	0.312055
$D(z)$	0.101244	0.226895	0.045378	0.129968	0.156618
$\hat{E}(z)$	0	0	0	0	0
uniform $n = 100$	$\overline{x}$	med x	z_R	z_Q	z_{tr}
$E(z)$	0.000136	0.001165	-0.0006	0.015422	0.033033
$D(z)$	0.01103	0.031839	0.000577	0.015796	0.02234
$\hat{E}(z)$	0	0	0	0	0
uniform $n = 1000$	$\overline{x}$	med x	z_R	z_Q	z_{tr}
$E(z)$	0.001722	0.002748	-1.5e-05	0.003489	0.005895
$D(z)$	0.00102	0.002849	6e-06	0.001552	0.001999
$\hat{E}(z)$	0	0	0	0	0

Таблица 5: Равномерное распределение (7)

5 Обсуждение

5.1 Гистограмма и график плотности распределения

Опираясь на проделанную работу, можно утверждать, что близость гистограммы к графику плотности вероятности закона, по которому была сгенерирована выборка, пропорциональна размеру этой выборки. На малых выборках характер распределения величины практически неузнаваем.

В большинстве случаев максимумы гистограмм и плотностей распределения не совпали. Иногда прослеживаются всплески гистограмм, например на распределении Коши.

5.2 Характеристики положения и рассеяния

По данным, приведенным в таблицах, можно заметить аномалию: дисперсия характеристик рассеяния для распределения Коши огромна даже на большой выборке. Это является результатом выбросов, которые можно было наблюдать в итогах предыдущего задания.

Примечание

С исходным кодом работы и данного отчета можно ознакомиться в репозитории `https://github.com/Stasychbr/MatStat`