

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Институт прикладной математики и механики

Высшая школа прикладной математики и вычислительной физики

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

Выполнил студент:
Козлов Борис
группа: 3630102/80301

Проверил:
к.ф.-м.н., доцент
Баженов Александр Николаевич

Санкт-Петербург
2021 г.

Содержание

	Страница
1 Постановка задачи	4
2 Теория	4
2.1 Рассматриваемые распределения	4
2.2 Гистограмма	4
2.2.1 Построение гистограммы	4
2.3 Вариационный ряд	5
2.4 Выборочные числовые характеристики	5
2.4.1 Характеристики положения	5
2.4.2 Характеристики рассеивания	5
3 Реализация	5
4 Результаты	6
4.1 Гистограммы и графики плотности распределения	6
4.2 Характеристики положения и рассеивания	8
5 Обсуждение	10
5.1 Гистограмма и график плотности распределения	10
5.2 Характеристики положения и рассеяния	10

Список иллюстраций

	Страница
1 Нормальное распределение (3)	6
2 Распределение Коши (4)	6
3 Распределение Лапласа (5)	7
4 Распределение Пуассона (6)	7
5 Равномерное распределение (7)	8

Список таблиц

		Страница
1	Нормальное распределение (3)	8
2	Распределение Коши (4)	9
3	Распределение Лапласа (5)	9
4	Распределение Пуассона (6)	9
5	Равномерное распределение (7)	10

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

$$E(z) = \bar{z} \quad (1)$$

Вычислить оценку дисперсии по формуле:

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

Представить полученные данные в виде таблиц.

2 Теория

2.1 Рассматриваемые распределения

Плотности:

- Нормальное распределение

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

- Распределение Коши

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

- Распределение Лапласа

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

- Распределение Пуассона

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

- Равномерное распределение

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

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

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (8)$$

- Выборочная медиана

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

- Полусумма экстремальных выборочных элементов

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

- Полусумма квартилей

Выборочный квартиль z_p порядка p определяется формулой

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

Полусумма квартилей

$$z_Q = \frac{z_{1/4} + z_{3/4}}{2} \quad (12)$$

- Усечённое среднее

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

2.4.2 Характеристики рассеивания

Выборочная дисперсия

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

3 Реализация

Лабораторная работа выполнена на языке Python в средах PyCharm и Jupyter Notebook с использованием следующих библиотек:

1. `scipy` (генерация выборок)
2. `matplotlib`, `seaborn` (визуализация, построение гистограмм)
3. `numpy` (вычисление ряда числовых характеристик)

4 Результаты

4.1 Гистограммы и графики плотности распределения

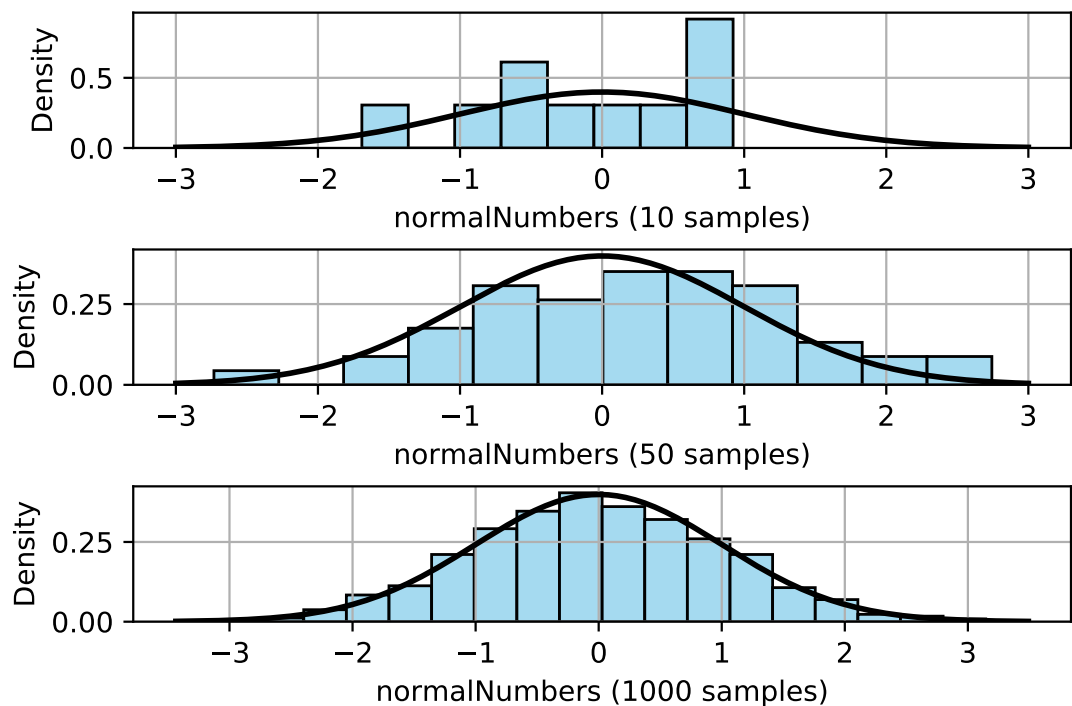


Рис. 1: Нормальное распределение (3)

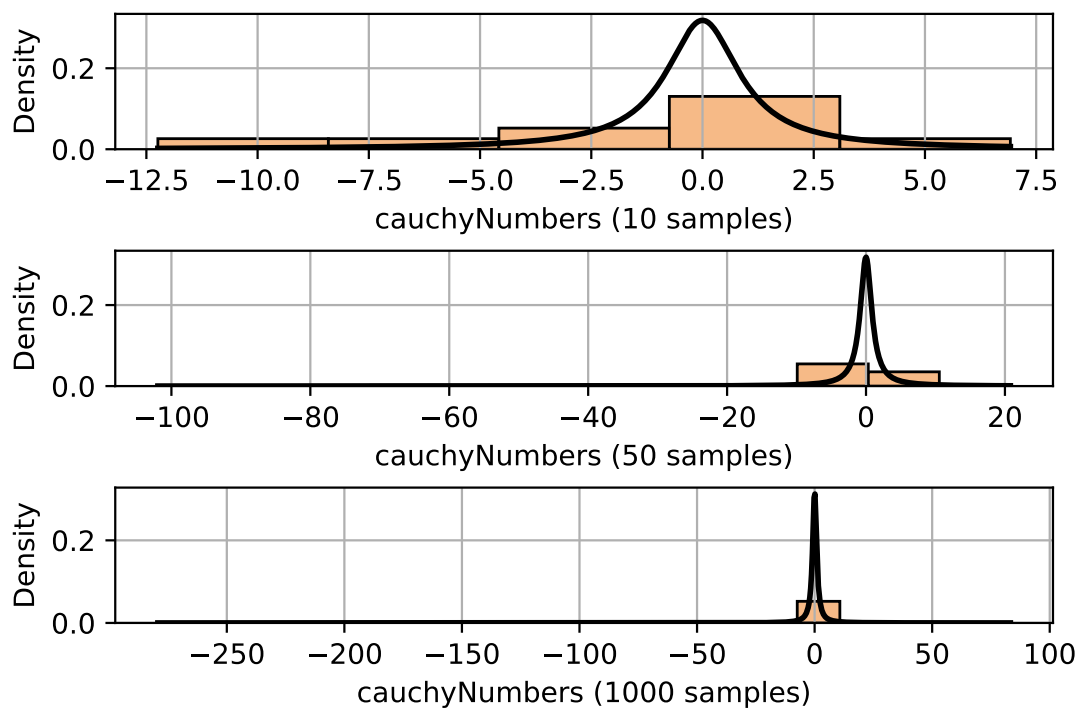


Рис. 2: Распределение Коши (4)

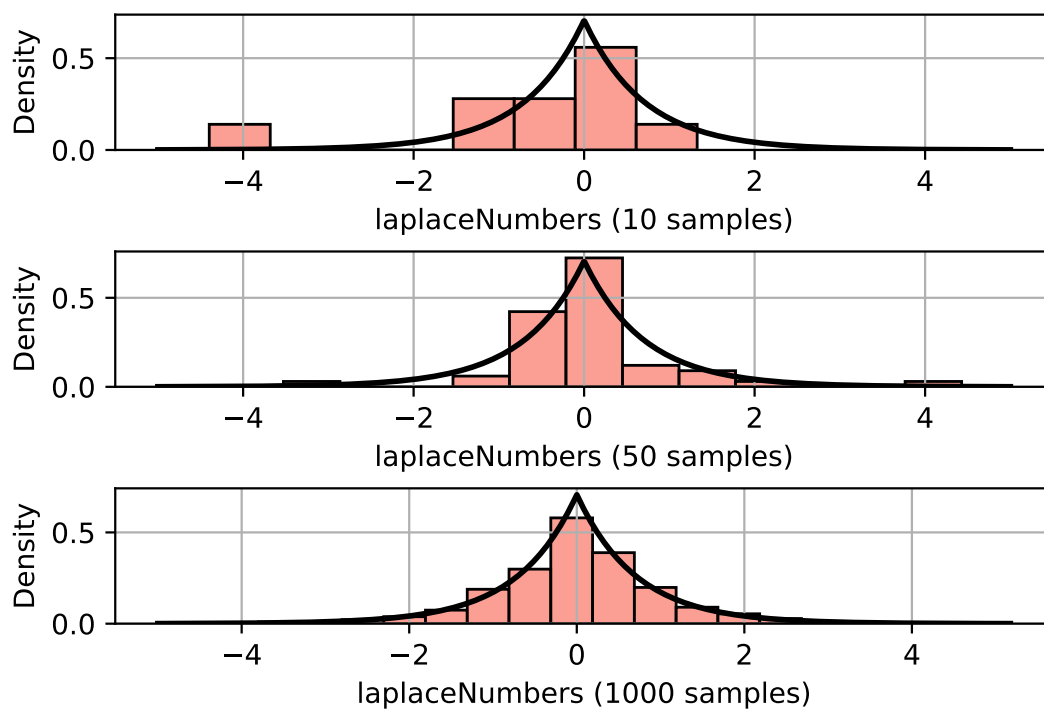


Рис. 3: Распределение Лапласа (5)

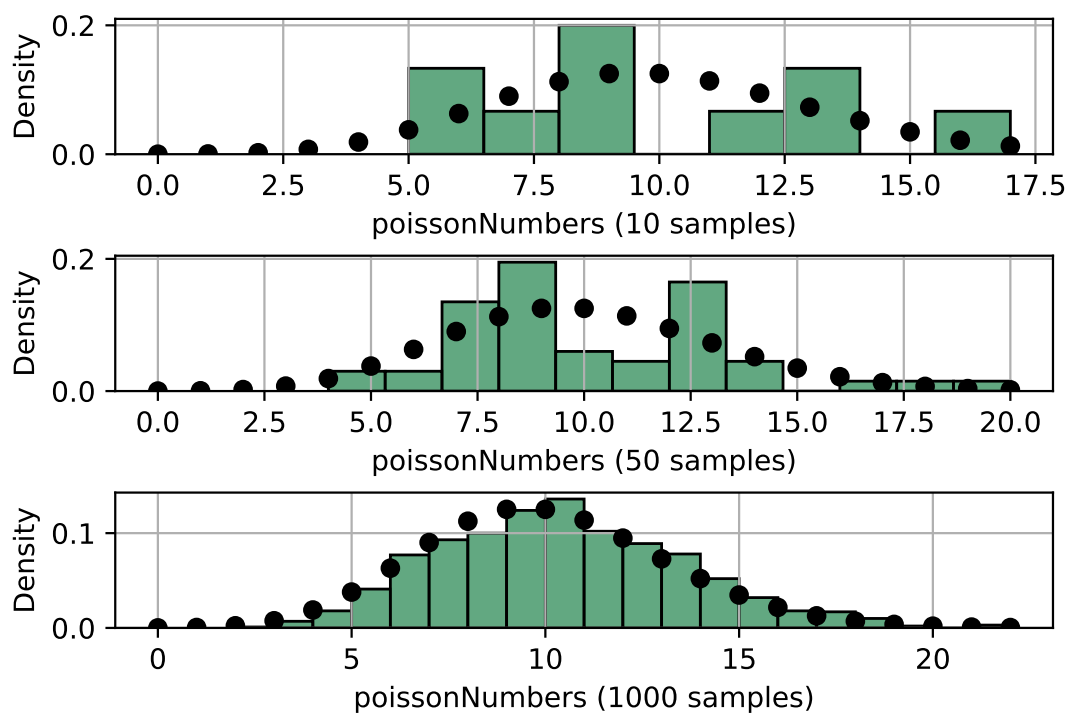


Рис. 4: Распределение Пуассона (6)

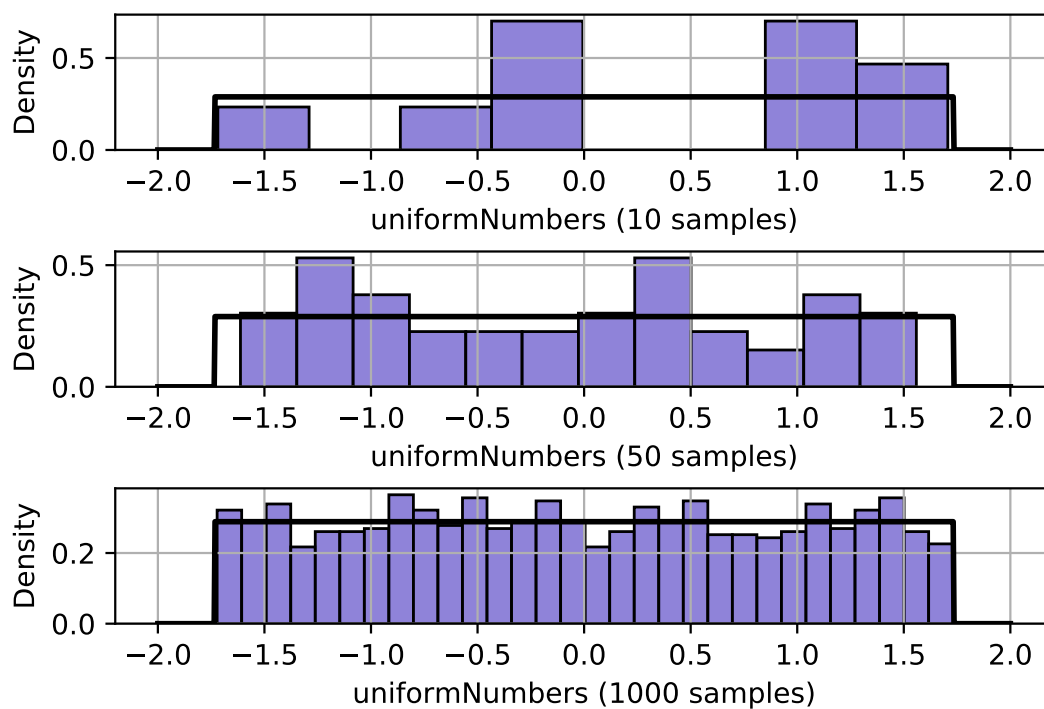


Рис. 5: Равномерное распределение (7)

4.2 Характеристики положения и рассеивания

normal $n = 10$					
	\bar{x} (8)	$med\ x$ (9)	z_R (10)	z_Q (12)	z_{tr} (13)
$E(z)$ (1)	0.001643	-0.001715	0.011264	0.307549	0.274457
$D(z)$ (2)	0.105575	0.146178	0.174175	0.134176	0.123485
normal $n = 100$					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.002837	-0.006377	0.010113	0.011054	0.021504
$D(z)$	0.010176	0.01538	0.09277	0.013052	0.012069
normal $n = 1000$					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.002345	-0.001516	0.003679	-0.001195	0.000629
$D(z)$	0.001105	0.001723	0.063631	0.001349	0.001325

Таблица 1: Нормальное распределение (3)

cauchy $n = 10$					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	4.032778	0.024677	19.86405	1.339201	0.776474
$D(z)$	10523.431164	0.337081	262430.774818	15.269762	2.615929
cauchy $n = 100$					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-2.440452	-0.00389	-120.917099	0.037722	0.04242
$D(z)$	8567.668789	0.024353	21296287.803965	0.049615	0.02486
cauchy $n = 1000$					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-1.639749	-0.001563	-846.555835	0.000551	0.001988
$D(z)$	952.311641	0.002561	235973354.703035	0.005211	0.002642

Таблица 2: Распределение Коши (4)

laplace $n = 10$					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.001668	-0.004131	0.010891	0.289294	0.226136
$D(z)$	0.101372	0.073895	0.401826	0.123031	0.085186
laplace $n = 100$					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.001346	-0.001301	0.028134	0.013016	0.019805
$D(z)$	0.010855	0.005921	0.421155	0.01084	0.006523
laplace $n = 1000$					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.000328	-0.000592	-0.007745	0.00162	0.001773
$D(z)$	0.00091	0.000489	0.391542	0.0009	0.000573

Таблица 3: Распределение Лапласа (5)

poisson $n = 10$					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	10.0383	9.8805	10.359	10.9795	10.811167
$D(z)$	0.952043	1.31197	1.984619	1.36983	1.170536
poisson $n = 100$					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	10.00143	9.848	10.9655	9.9655	9.94214
$D(z)$	0.098683	0.205396	0.97106	0.15006	0.117383
poisson $n = 1000$					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	9.997855	9.994	11.6445	9.9935	9.864184
$D(z)$	0.010238	0.005464	0.65537	0.003208	0.011762

Таблица 4: Распределение Пуассона (6)

uniform $n = 10$					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.010428	0.007235	0.008858	0.333875	0.326122
$D(z)$	0.097559	0.218496	0.04188	0.126666	0.148893
uniform $n = 100$					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.008995	-0.014935	-0.000303	0.007003	0.021455
$D(z)$	0.009266	0.026934	0.000616	0.014319	0.018513
uniform $n = 1000$					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.000227	-0.000275	-3.3e-05	0.001693	0.003161
$D(z)$	0.001045	0.003089	6e-06	0.001582	0.002111

Таблица 5: Равномерное распределение (7)

5 Обсуждение

5.1 Гистограмма и график плотности распределения

Результаты проделанной работы указывают на то, что для каждого из распределений справедливо следующее замечание: при увеличении количества элементов выборки ее гистограмма становится ближе к графику плотности вероятности того закона, по которому распределены эти элементы. Чем меньше выборка, тем менее она репрезентативна и тем хуже по ней определяется характер распределения исследуемой величины.

В большинстве случаев максимумы гистограмм и плотностей распределения не совпали. В некоторых местах прослеживаются всплески гистограмм, наиболее отчетливо - на распределении Коши.

5.2 Характеристики положения и рассеяния

В полученных данных, приведенных в таблице, особый интерес представляет дисперсия характеристик рассеяния для распределения Коши, чьи значения можно назвать аномально большими. Ясно, что это результат выбросов, которые можно было наблюдать в результатах предыдущего задания.

Примечание

С кодом работы и отчета можно ознакомиться по ссылке: <https://github.com/Kozlov992/MS2021/tree/master/Labs1-2>