Санкт-Петербургский политехнический университет Петра Великого

Институт прикладной математики и механики

Кафедра «Прикладная математика»

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

Выполнил студент: Колосков Александр
Группа: 3630102/80301
Проверил: к.ф.-м.н., доцент Баженов Александр Николаевич

Санкт-Петербург
2021 г.

Содержание

Список иллюстраций

Список таблиц

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: $\overline{x}, medx, z_R, z_Q, z_{tr}$. 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
$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}} & \text{при} |x| \leq \sqrt{3} \\ 0 & \text{при} |x| > \sqrt{3} \end{cases} \tag{7}$$

2.2 Гистограмма

2.2.1 Построение гистограммы

Множество значений, которое может принимать элемент выборки, разбивается на несколько интервалов. Чаще всего эти интервалы берут одинаковыми, но это не является строгим требованием. Эти интервалы откладываются на горизонтальной оси, затем над каждым рисуется прямоугольник. Если все интервалы были одинаковыми, то высота каждого прямоугольника пропорциональна числу элементов выборки, попадающих в соответствующий интервал. Если интервалы разные, то высота прямоугольника выбирается таким образом, чтобы его площадь была пропорциональна числу элементов выборки, которые попали в этот интервал

2.3 Вариационный ряд

Вариационным ряд - последовательность элементов выборки, расположенных в неубывающем порядке. Одинаковые элементы повторяются.

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

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

- 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

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

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 в среде PyCharm с использованием библиотек numpy, scipy.stats, matplotlib.pyplot.

4 Результаты

4.1 Гистограммы и графики плотности распределения

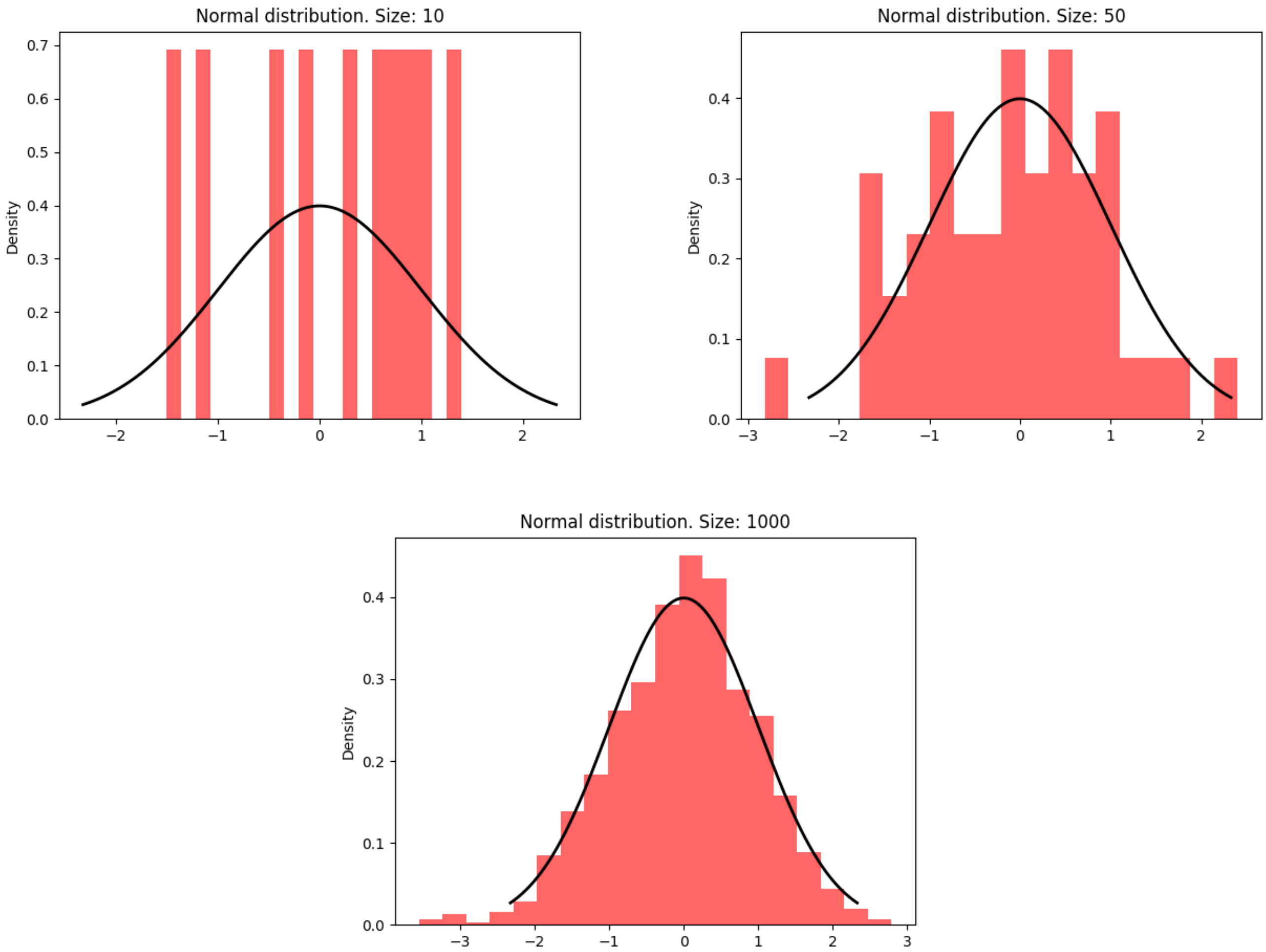

Рис. 1: Нормальное распределение (3)

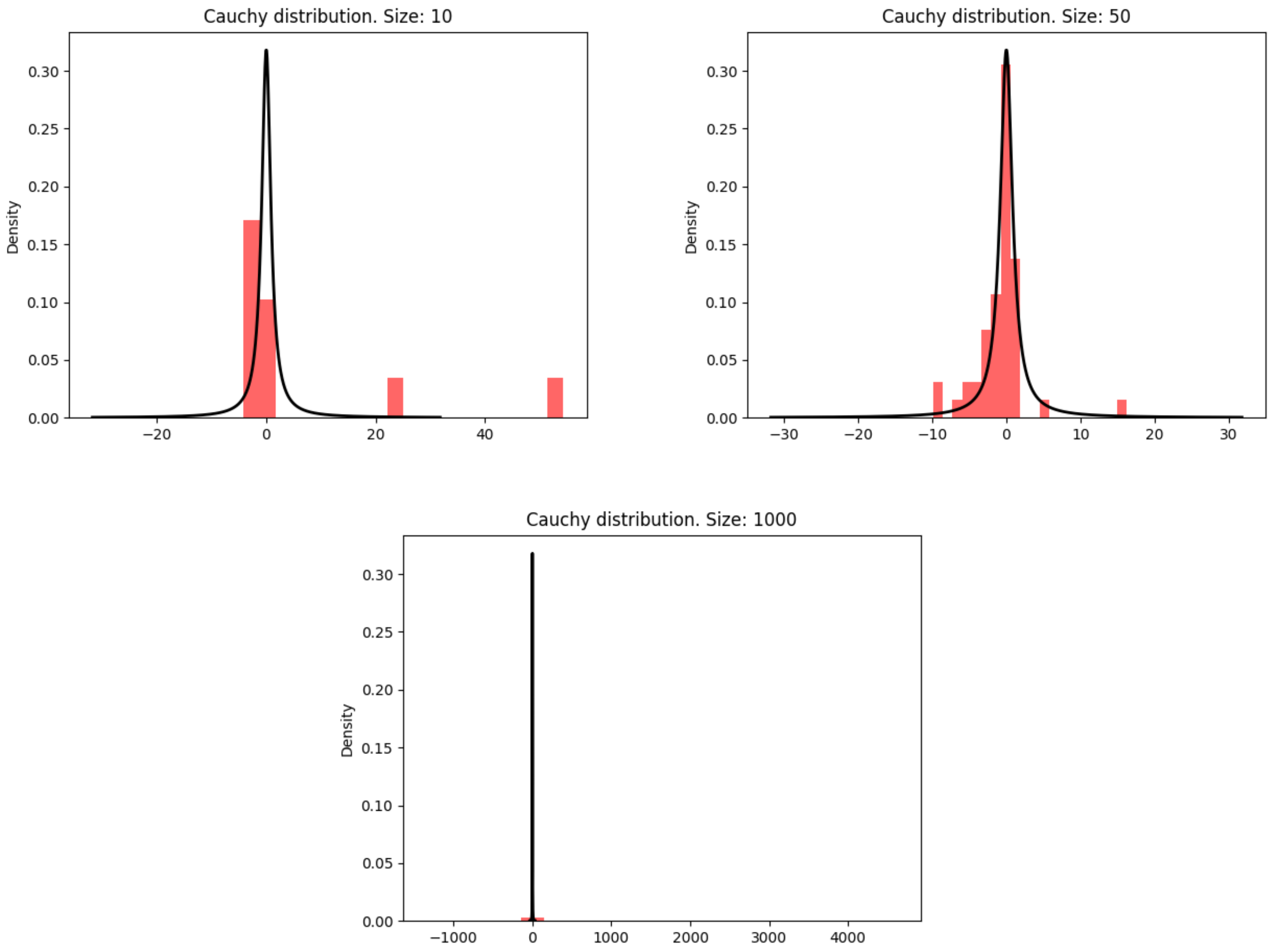

Рис. 2: Распределение Коши (4)

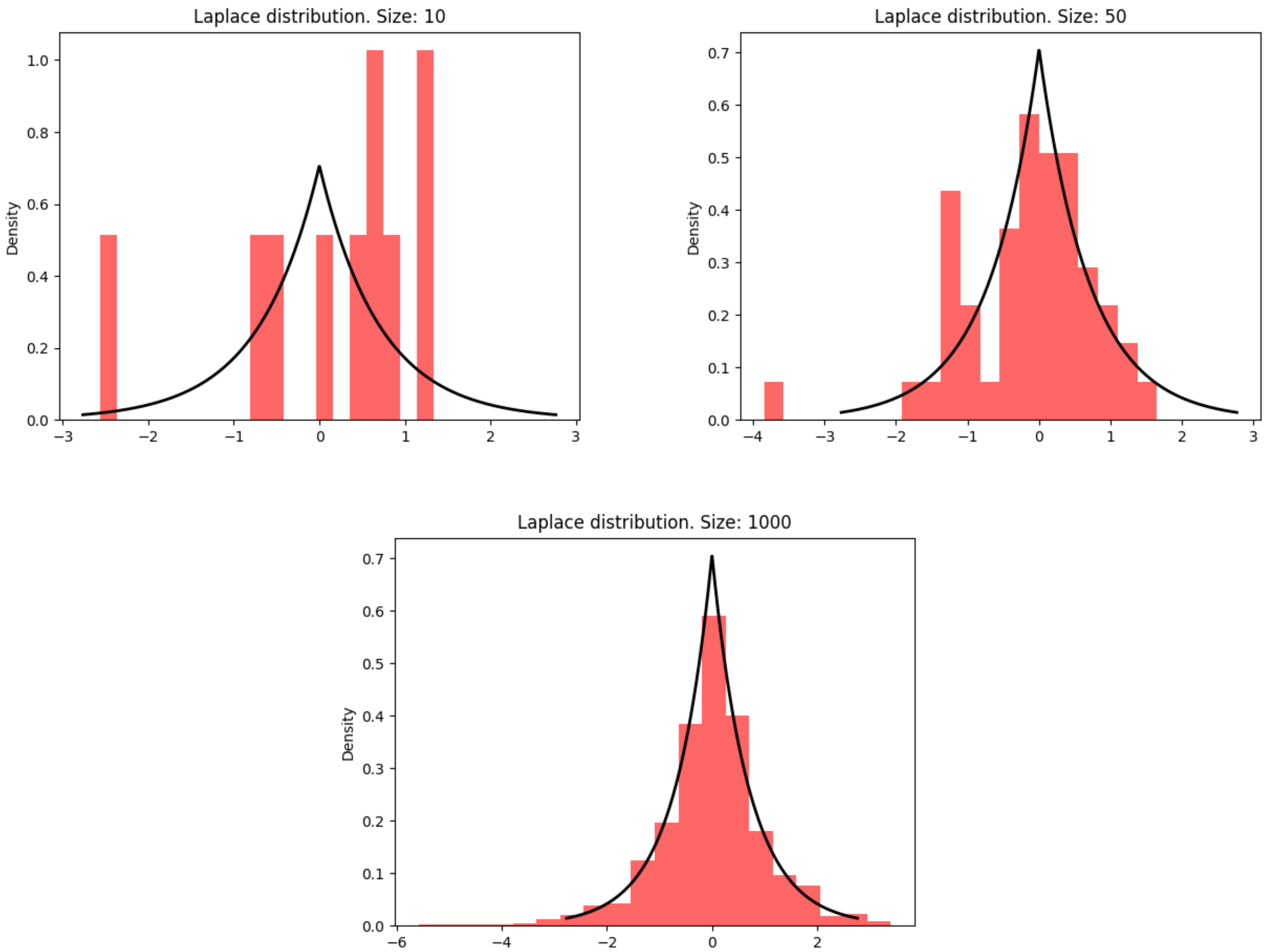

Рис. 3: Распределение Лапласа (5)

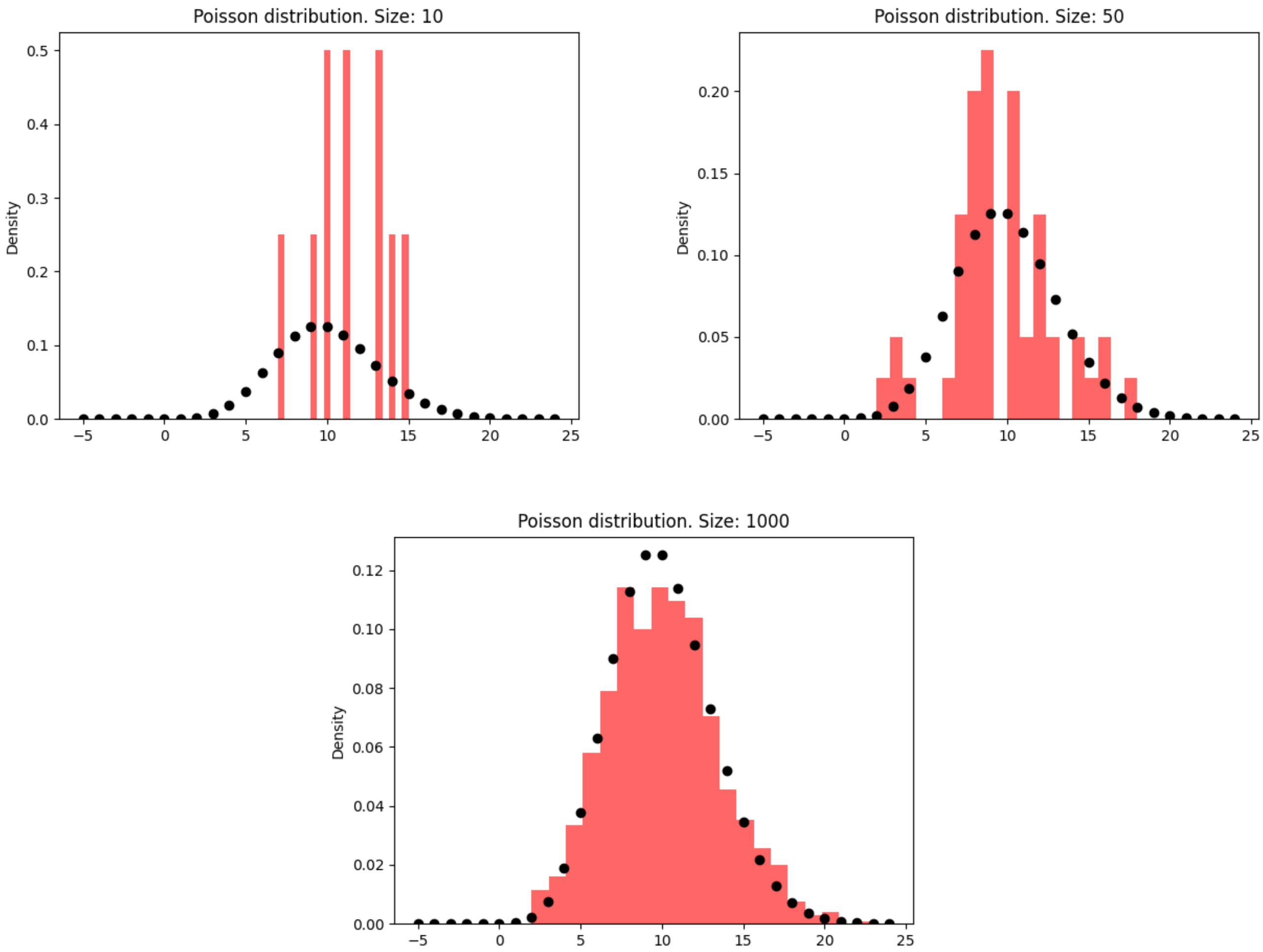

Рис. 4: Распределение Пуассона (6)

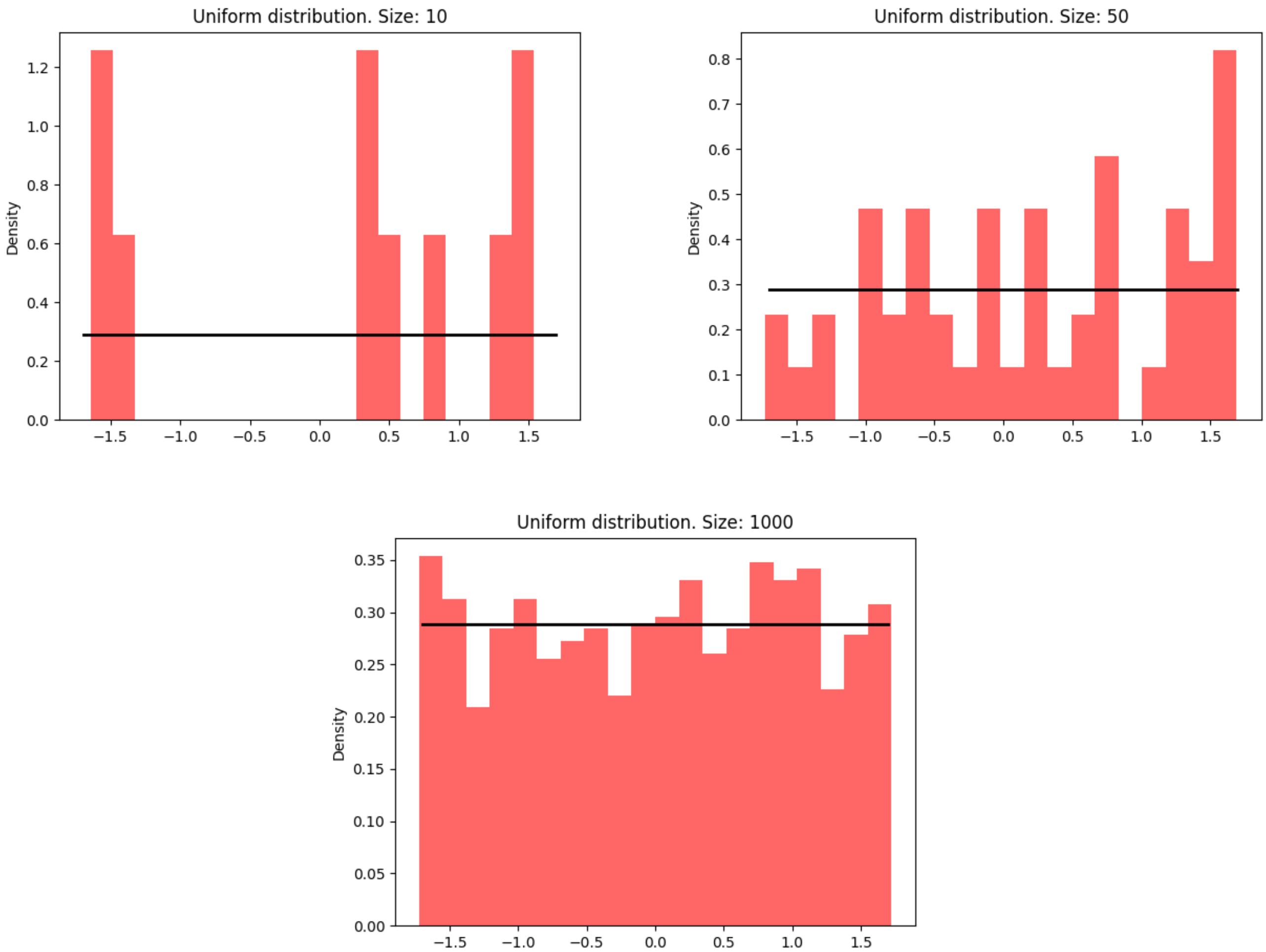

Рис. 5: Равномерное распределение (7)

4.2 Характеристики положения и рассеивания

Normal n=10					
	$\overline{x}$	$medx$	z_R	z_Q	z_{tr}
$E(z)$	0.01282	0.01272	0.0063	0.33148	0.28685
$D(z)$	0.10402	0.14234	0.20013	0.13261	0.11628
Normal n=100					
	$\overline{x}$	$medx$	z_R	z_Q	z_{tr}
$E(z)$	-0.00159	-0.0026	0.00587	0.01635	0.0261
$D(z)$	0.00975	0.01549	0.09409	0.01223	0.01152
Normal n=1000					
	$\overline{x}$	$medx$	z_R	z_Q	z_{tr}
$E(z)$	-0.00024	0.00043	-0.0043	0.00147	0.00269
$D(z)$	0.00109	0.00165	0.05826	0.00132	0.00129

Таблица 1: Нормальное распределение (3)

Cauchy n=10					
	$\overline{x}$	$medx$	z_R	z_Q	z_{tr}
$E(z)$	-0.5523	-0.00081	-2.85318	1.23021	0.72645
$D(z)$	402.07941	0.36718	9836.4676	12.12453	2.21432
Cauchy n=100					
	$\overline{x}$	$medx$	z_R	z_Q	z_{tr}
$E(z)$	-0.3056	3e-05	-15.48603	0.02223	0.03768
$D(z)$	246.86153	0.02562	582655.00488	0.05296	0.02614
Cauchy n=1000					
	$\overline{x}$	$medx$	z_R	z_Q	z_{tr}
$E(z)$	-4.01027	-0.00059	-2012.45348	0.0026	0.00391
$D(z)$	15589.37951	0.00243	3891875805.02797	0.00525	0.00255

Таблица 2: Распределение Коши (4)

Laplace n=10					
	$\overline{x}$	$medx$	z_R	z_Q	z_{tr}
$E(z)$	-0.00481	-0.0048	-0.01408	0.30002	0.23509
$D(z)$	0.10212	0.07575	0.37263	0.12141	0.08461
Laplace n=100					
	$\overline{x}$	$medx$	z_R	z_Q	z_{tr}
$E(z)$	-0.00371	-0.00414	-0.00244	0.00913	0.01472
$D(z)$	0.00993	0.00571	0.39463	0.01027	0.00637
Laplace n=1000					
	$\overline{x}$	$medx$	z_R	z_Q	z_{tr}
$E(z)$	-0.00012	0.00038	-0.00797	0.0014	0.00196
$D(z)$	0.00095	0.0005	0.37861	0.00095	0.00059

Таблица 3: Распределение Лапласа (5)

Poisson n=10					
	$\overline{x}$	$medx$	z_R	z_Q	z_{tr}
$E(z)$	10.00549	9.87013	10.26024	10.92408	10.78355
$D(z)$	1.08555	1.50162	1.88557	1.52371	1.36407
Poisson n=100					
	$\overline{x}$	$medx$	z_R	z_Q	z_{tr}
$E(z)$	10.00178	9.83616	10.96454	9.97253	9.94314
$D(z)$	0.10691	0.20243	0.9895	0.16433	0.12475
Poisson n=1000					
	$\overline{x}$	$medx$	z_R	z_Q	z_{tr}
$E(z)$	9.99931	9.99301	11.65834	9.99251	9.86493
$D(z)$	0.0112	0.00644	0.68147	0.00369	0.01245

Таблица 4: Распределение Пуассона (6)

Uniform n=10					
	$\overline{x}$	$medx$	z_R	z_Q	z_{tr}
$E(z)$	0.00636	0.01311	-0.00232	0.31991	0.32644
$D(z)$	0.09577	0.22741	0.0412	0.12049	0.14748
Uniform n=100					
	$\overline{x}$	$medx$	z_R	z_Q	z_{tr}
$E(z)$	-0.0011	-0.0031	0.00032	0.0145	0.03236
$D(z)$	0.01112	0.03042	0.00059	0.01602	0.02153
Uniform n=1000					
	$\overline{x}$	$medx$	z_R	z_Q	z_{tr}
$E(z)$	1e-05	0.0011	-1e-05	0.00178	0.00332
$D(z)$	0.001	0.00295	1e-05	0.00151	0.00196

Таблица 5: Равномерное распределение (7)

5 Обсуждение

5.1 Гистограмма и график плотности распределения

Результаты проведенных экспериментов подтверждают очевидное предположение о том, что при увеличении мощности выборки случайной величины форма ее гистограммы стремится к графику функции плотности распределения данной случайной величины. Обратно, при малых размерах выборки гистограмма мало совпадает с графиком заданной функцией плотности. В некоторых местах прослеживаются всплески гистограмм, наиболее отчетливо - на распределении Коши.

5.2 Характеристики положения и рассеяния

Сообразно результатам первого эксперимента, в результатах второго также имеется ясная линейная зависимость приближения точных аналитических характеристик заданных случайных величин их статистическими эквивалентами от мощности выборки для всех распределений кроме распределения Коши. На это, в частности, указывает линейное стремление к нулю дисперсии для заданных характеристик.

Поведение выборочной дисперсии в случае распределения Коши также можно объяснить ссылкой на аналитические свойства данного распределения - отсутствием математического ожидания и бесконечной дисперсией.

Репозиторий

`https://github.com/KoloskovAleksandr/MathStatLabs2021`