

Санкт-Петербургский Политехнический Университет Петра
Великого

Физико-механический институт

Высшая школа прикладной математики и вычислительной физики

**Отчет по лабораторным работам по №1-2 по
математической статистике**

Выполнил:	Андреев Даниил Витальевич
Группа:	5030102/10101
Преподаватель:	Баженов Александр Николаевич

Санкт-Петербург
2024 год

1 Описательная статистика

1.1 Постановка задачи

Для 5 распределений:

- Нормальное распределение $N(x, 0, 1)$
- Распределение Коши $C(x, 0, 1)$
- Распределение Стьюдента $t(x, 0, 3)$
- Распределение Пуассона $P(k, 10)$
- Равномерное распределение $U(x, -\sqrt{3}, \sqrt{3})$

Сгенерировать выборки размером 10, 50, 1000 элементов. Построить на одном рисунке гистограмму и график плотности распределения.

1.2 Теоретическая справка

Плотности классических распределений:

- Нормальное распределение $N(x, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
- Распределение Коши $C(x, 0, 1) = \frac{1}{\pi(x^2+1)}$
- Распределение Стьюдента $t(x, 0, 3) = \frac{\Gamma(2)(1+\frac{x^2}{3})^{-2}}{\sqrt{3}\pi\Gamma_{\frac{3}{2}}}$
- Распределение Пуассона $P(k, 10) = \frac{e^{-10}10^k}{k!}$
- Равномерное распределение $U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}} & , |x| \leq \sqrt{3} \\ 0 & , |x| > \sqrt{3} \end{cases}$

Графически гистограмма строится следующим образом: сначала 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.

Построение гистограмм используется для получения эмпирической оценки плотности распределения случайной величины. Они являются хорошим инструментом для исследования неизвестных распределений.

1.3 Реализация

Реализовывать гистограммы для этих распределений будем на языке Python3 с использованием пакетов `numpy`, `matplotlib`, `scipy`. Ссылка на Github-репозиторий: <https://github.com/Highlandet/MathStatistics-Labs>

1.4 Результаты

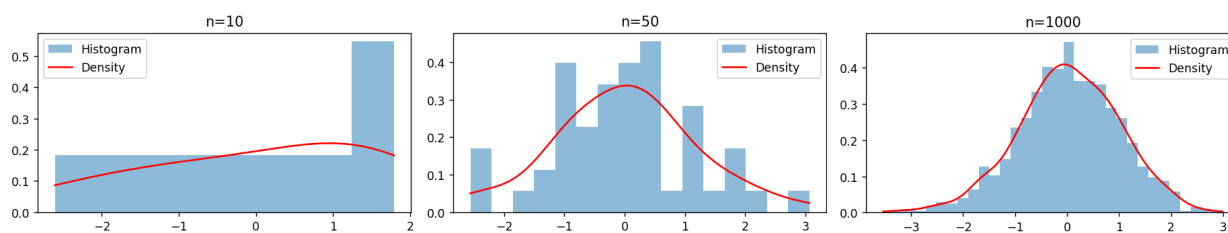


Рис. 1: Нормальное распределение $N(x, 0, 1)$

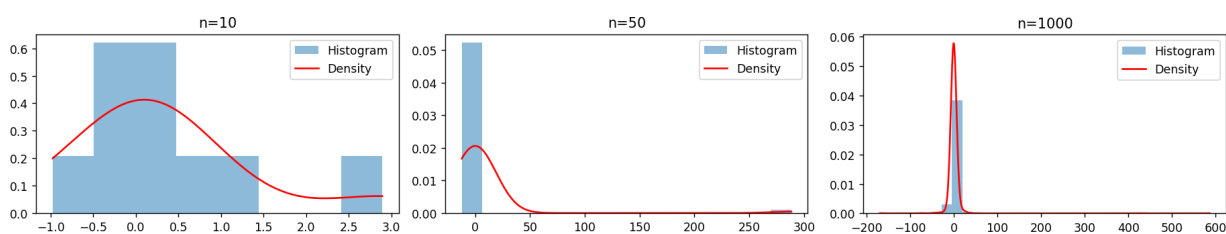


Рис. 2: Распределение Коши $C(x, 0, 1)$

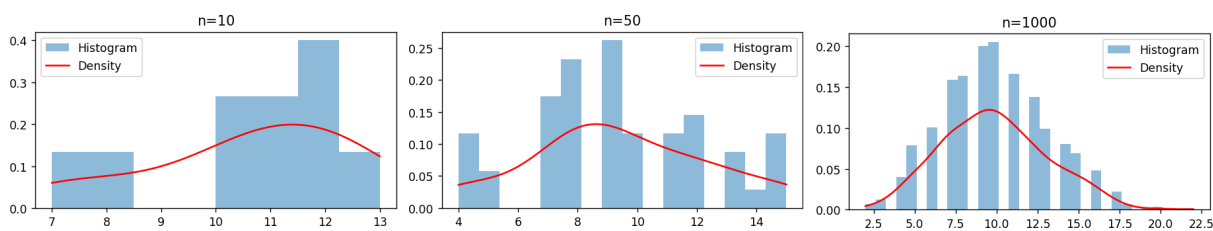


Рис. 3: Распределение Пуассона $P(k, 10)$

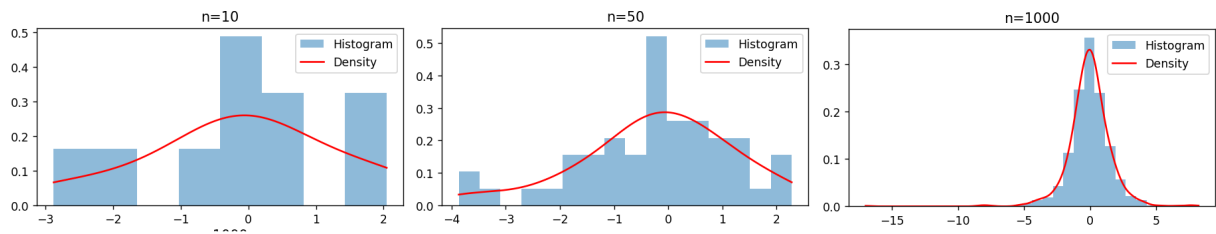


Рис. 4: Распределение Стьюдента $t(x, 0, 3)$

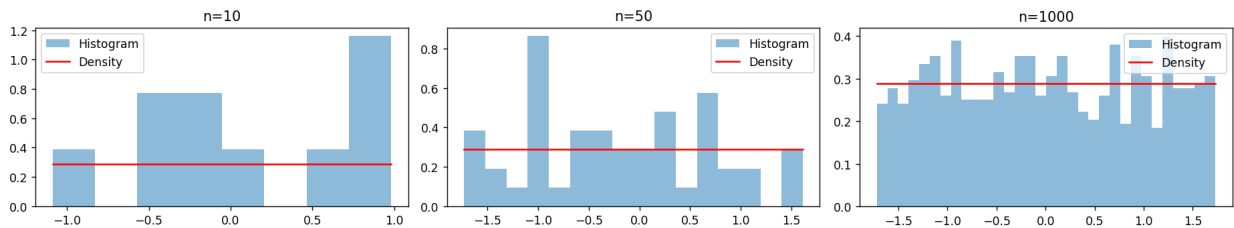


Рис. 5: Равномерное распределение $U(x, -\sqrt{3}, \sqrt{3})$

1.5 Вывод

Анализ результатов показывает, что чем больше выборка из распределения, тем больше соответствует гистограмма распределения графику плотности распределения. Аналогично: чем больше выборка, тем более заметен характер распределения.

2 Точечное оценивание характеристик положения и рассеяния

2.1 Постановка задачи

Сгенерировать выборки размером 10, 100, 1000 элементов. Для каждой выборки вычислить следующие статистические характеристики

положения данных: \bar{x} , $\text{med } x$, z_R , z_Q , z_{tr} . Повторить такие вычисления 1000 раз для каждой выборки и найти среднее характеристик положения и их квадратов: $E(z) = \bar{z}$, вычислить оценку дисперсии по формуле $D(z) = \overline{z^2} - \bar{z}^2$. Представить полученные данные в виде таблиц

2.2 Теоретическая справка

Характеристики положения:

- Выборочное среднее: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
- Выборочная медиана: $\text{med } x = \begin{cases} x_{(l+1)} & , n = 2l + 1 \\ \frac{x_{(l)} + x_{(l+1)}}{2} & , n = 2l \end{cases}$
- Полусумма экстремальных выборочных элементов $z_R = \frac{x_{(1)} + x_{(n)}}{2}$
- Полусумма квартилей: $z_Q = \frac{z_{1/4} + z_{3/4}}{2} \left(z_p = \begin{cases} x_{[np]+1} & , np \notin Z \\ x_{(np)} & , np \in Z \end{cases} \right)$
- Усеченное среднее: $z_{tr} = \frac{1}{n-2r} \sum_{i=r+1}^{n-r} x_i, r \approx \frac{n}{4}$
- Среднее характеристик положения: $E(z) = \bar{z}$
- Оценка дисперсии $D = \overline{z^2} - \bar{z}^2$

2.3 Реализация

Для вычисления статических характеристик положения данных будем использовать язык программирования Python3 и следующие библиотеки: `numpy`, `scipy`

Ссылка на Github-репозиторий: <https://github.com/Highlandet/MathStatistics-Labs>

2.4 Результаты и выводы

Normal: n=10	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	5.79	3041.21	-6.31	12.23	15.22
$D(z)$	0.09	0.25	0.50	0.12	0.41
Normal: n=10	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	-2.50	25478.24	-27.22	-5.36	-62.72
$D(z)$	0.01	0.24	0.51	0.14	2.83
Normal: n=1000	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	-0.16	250486.57	-12.00	-10.39	-118.51
$D(z)$	0.00	0.26	0.50	0.12	31.75

Таблица 1: Нормальное распределение $N(x, 0, 1)$

Проанализировав полученные результаты, можно заметить, что для нормального распределения, распределения Стюдента и равномерного распределения $E(z)$ и $D(z)$ для всех характеристик уменьшается с ростом выборки.

Стоит отметить, что в распределении Пуассона $E(z)$ для всех характеристик колеблется в окрестности десяти. Для $D(z)$ наблюдается уменьшений значений при росте выборки.

В таблице характеристик распределения Коши можно выделить аномальные значения, явно превышающие ожидаемые. Такой результат можно объяснить наличием различных выбросов, неопределенностью матожидания и бесконечностью дисперсии СВ, распределенной по данному закону.

Cauchy: n=10	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	1316.54	2405.20	1323.37	-368.03	2962.64
$D(z)$	652.51	52.08	3184.72	156.85	2993.08
Cauchy: n=100	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	1034.36	24548.25	-994.41	332.99	16926.15
$D(z)$	449.01	515.71	3128.20	253.32	87613.01
Cauchy: n=1000	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	8865.32	250288.69	528.44	-327.32	-369356.70
$D(z)$	61074.19	68.73	524.80	506.45	107199840.92

Таблица 2: Распределение Коши $C(x, 0, 1)$

Student: n=10	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	-7.75	2992.32	10.74	-43.07	-32.11
$D(z)$	0.28	0.71	1.42	0.32	1.27
Student: n=100	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	1.86	25521.35	-15.68	-14.60	84.54
$D(z)$	0.03	0.81	1.51	0.37	9.58
Student: n=1000	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	2.82	250506.41	-87.77	3.15	310.55
$D(z)$	0.00	0.75	1.35	0.37	92.46

Таблица 3: Распределение Стьюдента $t(x, 0, 3)$

Poisson: n=10	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	-7.75	2992.32	10.74	-43.07	-32.11
$D(z)$	0.28	0.71	1.42	0.32	1.27
Poisson: n=100	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	1.86	25521.35	-15.68	-14.60	84.54
$D(z)$	0.03	0.81	1.51	0.37	9.58
Poisson: n=1000	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	2.82	250506.41	-87.77	3.15	310.55
$D(z)$	0.00	0.75	1.35	0.37	92.46

Таблица 4: Распределение Пуассона $P(10, k)$

Uniform: n=10	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	-6.74	2977.22	-30.38	-3.45	-1.72
$D(z)$	0.09	0.25	0.53	0.12	0.41
Uniform: n=100	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	-2.60	25497.47	6.61	11.79	-88.65
$D(z)$	0.01	0.25	0.50	0.12	3.20
Uniform: n=1000	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	0.56	250474.54	19.18	-4.99	132.16
$D(z)$	0.00	0.24	0.49	0.13	33.66

Таблица 5: Равномерное распределение $U(x, -\sqrt{3}, \sqrt{3})$