

Санкт-Петербургский
Политехнический университет Петра Великого

Отчет по лабораторным работам №1-2
по дисциплине
"Математическая статистика"

**Построение гистограм различных вероятностных
распределений и получение оценок положения и
оценок рассеяния соответствующих распределений**

Студент:	Белоус Фёдор Васильевич
Преподаватель:	Баженов Александр Николаевич
Группа:	5030102/10101

Санкт-Петербург
2024

Содержание

1	Постановка задачи	2
1.1	Описательная статистика	2
1.2	Точечное оценивание характеристик положения и рассеяния	2
2	Теоретическое обоснование	2
2.1	Функции распределения	2
2.2	Характеристики положения и рассеяния	3
3	Описание работы	3
4	Результаты	4
4.1	Гистограммы и графики плотности распределения	4
4.2	Характеристики положения и рассеяния	6
5	Выводы	8

1 Постановка задачи

1.1 Описательная статистика

Для 5 распределений:

- Нормальное распределение $N(x, 0, 1)$
- распределение Коши $C(x, 0, 1)$
- Распределение Стьюдента $t(x, 0, 3)$ с тремя степенями свободы
- 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 Точечное оценивание характеристик положения и рассеяния

Сгенерировать выборки размером 10, 100, 1000 элементов.

Для каждой выборки вычислить следующие статистические характеристики положения данных: \bar{x} , $medx$, z_Q , z_R , 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 $D(z) = \overline{z^2} - \bar{z}^2$.

2 Теоретическое обоснование

2.1 Функции распределения

- Нормальное распределение

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

- Распределение Коши

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

- Распределение Стьюдента $t(x, 0, 3)$ с тремя степенями свободы

$$t(x, 0, 3) = \frac{6\sqrt{3}}{\pi(3 + t^2)^2} \quad (3)$$

- Распределение Пуассона

$$P(k, 10) = \frac{10^k}{k!} e^{-10} \quad (4)$$

- Равномерное распределение

$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}} & \text{при } |x| \leq \sqrt{3} \\ 0 & \text{при } |x| > \sqrt{3} \end{cases} \quad (5)$$

2.2 Характеристики положения и рассеяния

- Выборочное среднее

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (6)$$

- Выборочная медиана

$$\text{med } x = \begin{cases} x_{(l+1)} & \text{при } n = 2l + 1 \\ \frac{x_{(l)} + x_{(l+1)}}{2} & \text{при } n = 2l \end{cases} \quad (7)$$

- Полусумма экстремальных выборочных элементов

$$z_R = \frac{x_{(1)} + x_{(n)}}{2} \quad (8)$$

- Полусумма квартилей

Выборочная квартиль z_p порядка p определяется формулой

$$z_p = \begin{cases} x_{([np]+1)} & \text{при } np \text{ дробном} \\ x_{(np)} & \text{при } np \text{ целом} \end{cases} \quad (9)$$

Полусумма квартилей

$$z_Q = \frac{z_{1/4} + z_{3/4}}{2} \quad (10)$$

- Усечённое среднее

$$z_{tr} = \frac{1}{n-2r} \sum_{i=r+1}^{n-r} x_{(i)}, \quad r \approx \frac{n}{4} \quad (11)$$

- Оценка дисперсии

$$D(z) = \overline{z^2} - \bar{z}^2 \quad (12)$$

3 Описание работы

Лабораторные работы выполнены на языке программирования Python. С использованием сторонних библиотек `numpy`, `matplotlib`, `pandas`, `IPython` были построены гистограммы распределений и посчитаны характеристики положения. Исходный код лабораторной работы: <https://github.com/feodorrussia/Mathematical-statistics/tree/master/lab1>

4 Результаты

4.1 Гистограммы и графики плотности распределения

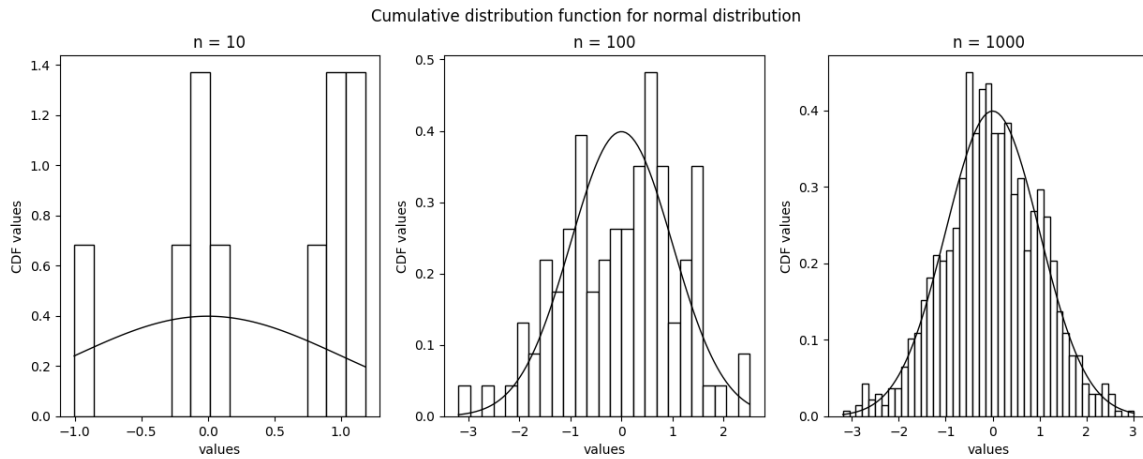


Рис. 1: Нормальное распределение

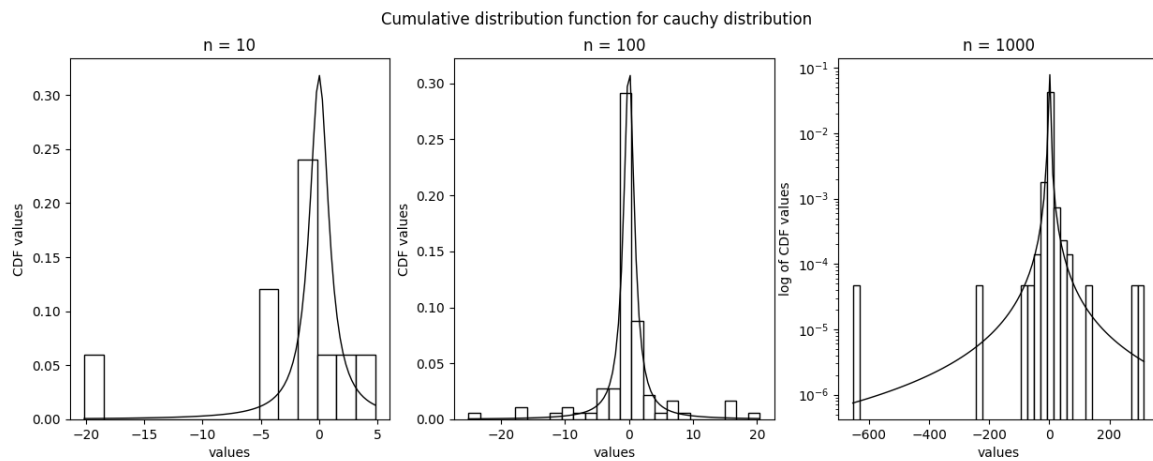


Рис. 2: Распределение Коши

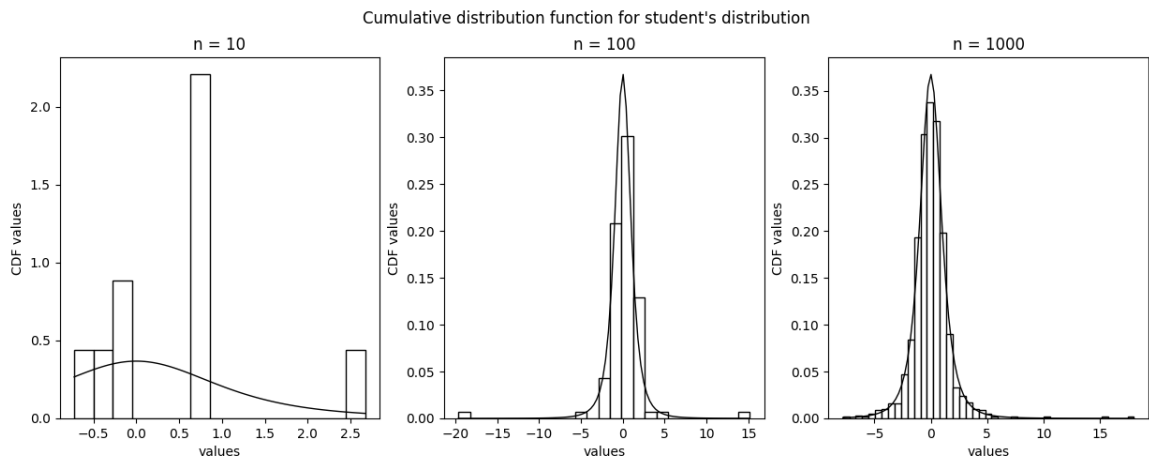


Рис. 3: Распределение Стьюдента

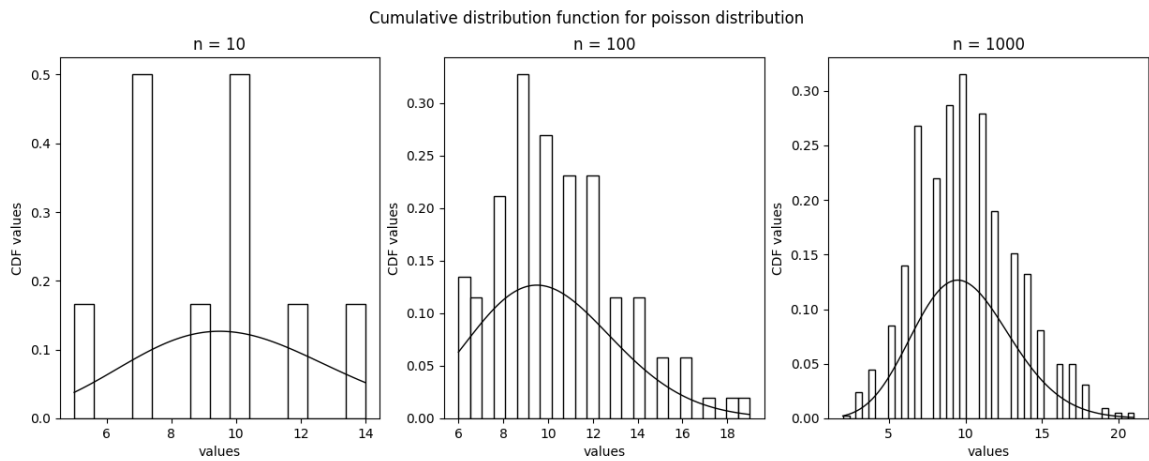


Рис. 4: Распределение Пуассона

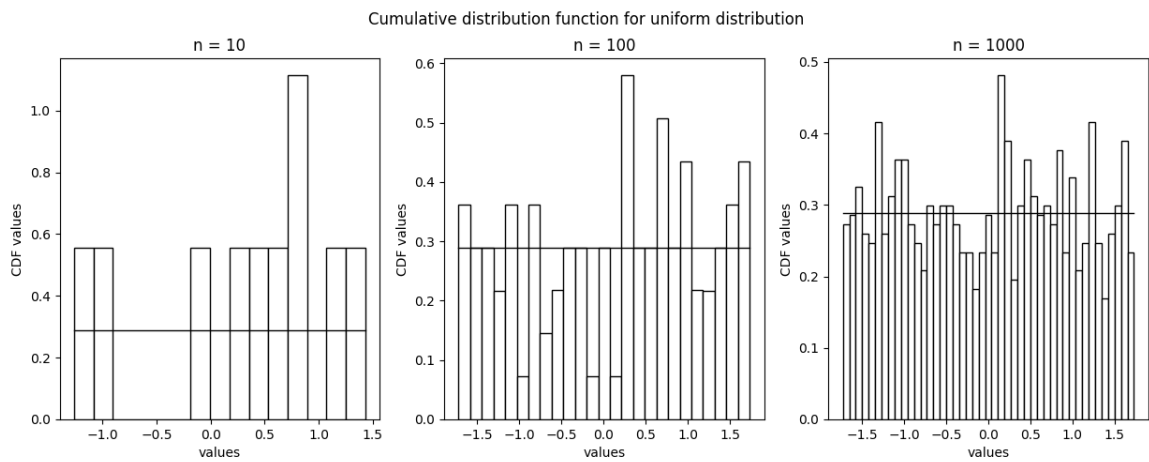


Рис. 5: Равномерное распределение

4.2 Характеристики положения и рассеяния

n = 10					
	\bar{x} (6)	$med\ x$ (7)	z_R (8)	z_Q (10)	z_{tr} (11)
$E(z)$	-0.004125	-0.010350	-0.009115	0.001618	-0.004184
$D(z)$ (12)	0.099243	0.140462	0.179263	0.114004	0.113237
n = 100					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.002628	-0.003811	-0.009304	-0.018760	-0.002022
$D(z)$	0.009988	0.015758	0.093964	0.012089	0.011859
n = 1000					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.000577	0.000381	0.003718	-0.000758	0.000752
$D(z)$	0.000924	0.001526	0.061540	0.001177	0.001160

Таблица 1: Нормальное распределение

n = 10					
	\bar{x} (6)	$med\ x$ (7)	z_R (8)	z_Q (10)	z_{tr} (11)
$E(z)$	0.000984	0.003021	0.000006	0.001482	0.001992
$D(z)$ (12)	16471.200669	0.384123	411497.578300	1.383175	0.587067
n = 100					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-2.744370	-0.000842	-1.368798e+02	-0.020361	0.003795
$D(z)$	3118.558067	0.025065	7.752087e+06	0.054164	0.026601
n = 1000					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-3.591202	0.000279	-1.767143e+03	-0.004553	-0.000679
$D(z)$	9022.756964	0.002596	2.252032e+09	0.005314	0.002774

Таблица 2: Распределение Коши

n = 10					
	\bar{x} (6)	$med\ x$ (7)	z_R (8)	z_Q (10)	z_{tr} (11)
$E(z)$	0.010644	-0.001992	0.038759	0.004784	0.003353
$D(z)$ (12)	0.266455	0.202082	1.754528	0.200204	0.176812
n = 100					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.006076	-0.002772	0.293182	-0.023259	-0.001929
$D(z)$	0.029348	0.018433	7.240088	0.017931	0.015283
n = 1000					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.000733	0.002365	-0.087560	-0.000367	0.001801
$D(z)$	0.003318	0.001871	38.308169	0.001967	0.001606

Таблица 3: Распределение Стьюдента

n = 10					
	\bar{x} (6)	$med\ x$ (7)	z_R (8)	z_Q (10)	z_{tr} (11)
$E(z)$	10.006100	9.897000	10.27050	9.92050	9.914833
$D(z)$ (12)	1.045853	1.370891	1.82858	1.27043	1.143719
n = 100					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	10.010540	9.863000	10.951500	9.883000	9.869920
$D(z)$	0.101113	0.204231	0.947898	0.158311	0.121002
n = 1000					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	10.005259	9.998000	11.681500	9.994000	9.863920
$D(z)$	0.009582	0.001996	0.662808	0.003464	0.010871

Таблица 4: Распределение Пуассона

n = 10					
	\bar{x} (6)	$med\ x$ (7)	z_R (8)	z_Q (10)	z_{tr} (11)
$E(z)$	-0.007486	0.002229	-0.010002	-0.003341	-0.005203
$D(z)$ (12)	0.099234	0.227002	0.044885	0.135067	0.159930
n = 100					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.006408	-0.013009	-0.001450	-0.022577	-0.008562
$D(z)$	0.009873	0.028269	0.000633	0.015001	0.019194
n = 1000					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.009873	0.028269	0.000633	0.015001	0.019194
$D(z)$	0.000984	0.003021	0.000006	0.001482	0.001992

Таблица 5: Равномерное распределение

5 Выводы

В лабораторной работе были выполнены следующие задачи:

Построены гистограммы и графики плотности вероятности для пяти различных вероятностных распределений: нормального, равномерного, Коши, Стьюдента и Пуассона.

Рассчитаны характеристики положения для каждого распределения при различных объемах выборки (10, 100, 1000 элементов). В частности, были найдены выборочное среднее, медиана, полусумма экстремальных выборочных элементов, полусумма квартилей и усеченное среднее.

Из результатов можно сделать следующие выводы:

Увеличение объема выборки обеспечивает лучшую точность оценки параметров распределений. Это видно как по более точному совпадению гистограмм с графиками плотности вероятности, так и по уменьшению дисперсии оценок характеристик положения с увеличением объема выборки.

Распределение Коши характеризуется высокой чувствительностью к выбросам, что проявляется в больших значениях дисперсии оценок. Также несколько выбросов могут существенно влиять на среднее значение. Поэтому среднее в данном случае может быть нестабильным и неадекватным описанием центральной тенденции. В распределении Коши медиана может быть менее чувствительной к выбросам, чем среднее. Это связано с тем, что медиана не зависит от значений хвостов распределения. В случае Коши, у которого хвосты убывают медленно, выбросы влияют на среднее сильнее, чем на медиану. Поскольку распределение Коши не имеет конечного математического ожидания и дисперсии, выборочные квартили и полусумма экстремальных выборочных значений также могут быть неустойчивыми и более изменчивыми при различных выборках.

Медиана и среднее в распределении Пуассона могут быть схожими, особенно при увеличении объема выборки. Это объясняется тем, что среднее значение в распределении Пуассона равно его параметру, и при увеличении выборки закон больших чисел делает их близкими. В данном случае оба показателя, медиана и среднее, могут быть устойчивыми оценками центральной тенденции.

Нормальное распределение показывает наилучшую устойчивость к выбросам, что отражается в меньших значениях дисперсии оценок.

Распределение Стьюдента и нормальное распределение демонстрируют улучшение оценок с увеличением размера выборки.

Равномерное распределение и распределение Пуассона достаточно хорошо аппроксимируются выборочными средними и медианами, но оценки дисперсии могут быть неточными для небольших выборок.

Таким образом, выполненная лабораторная работа позволила ознакомиться с различными вероятностными распределениями, а также научиться оценивать их параметры на основе выборочных данных.