

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/Lab_1

4 Результаты

4.1 Гистограммы и графики плотности распределения

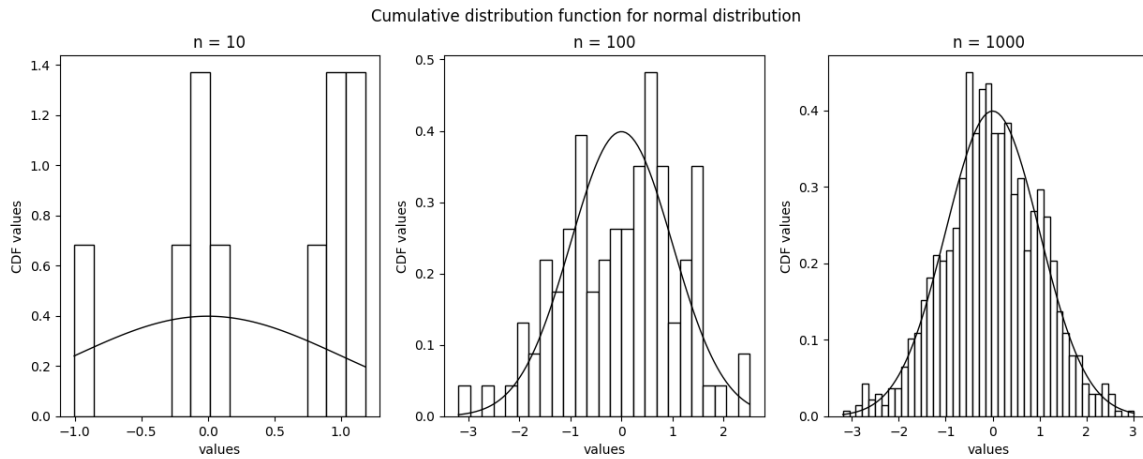


Рис. 1: Нормальное распределение (1)

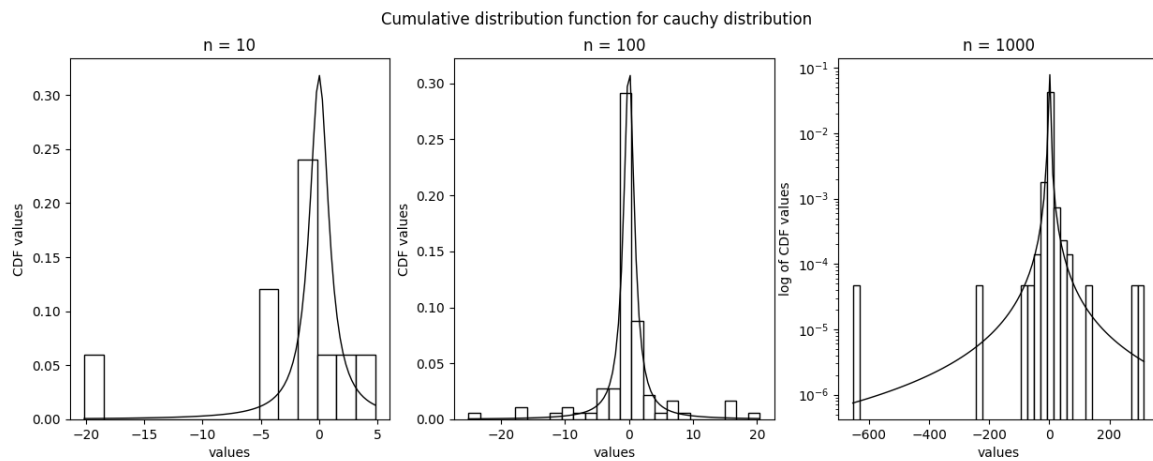


Рис. 2: Распределение Коши (2)

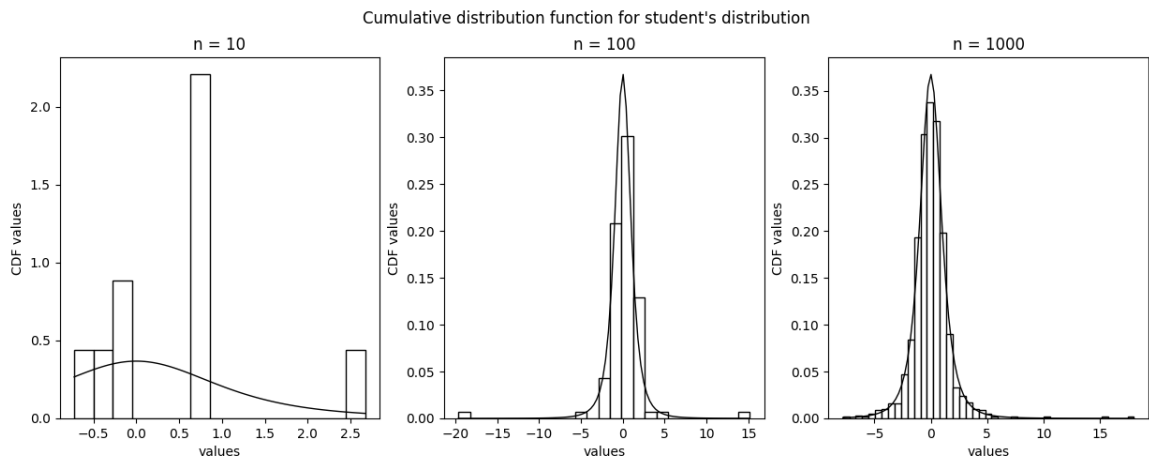


Рис. 3: Распределение Стьюдента (3)

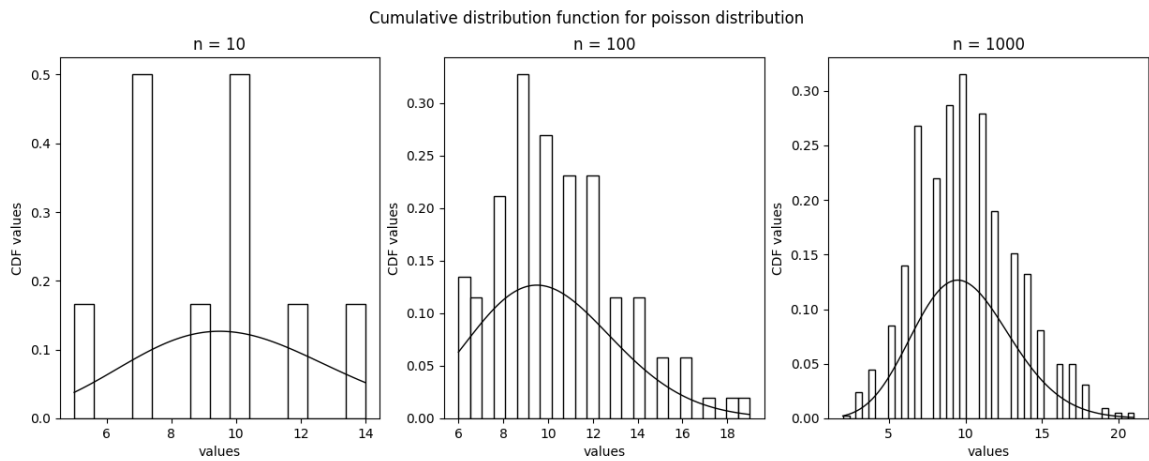


Рис. 4: Распределение Пуассона (4)

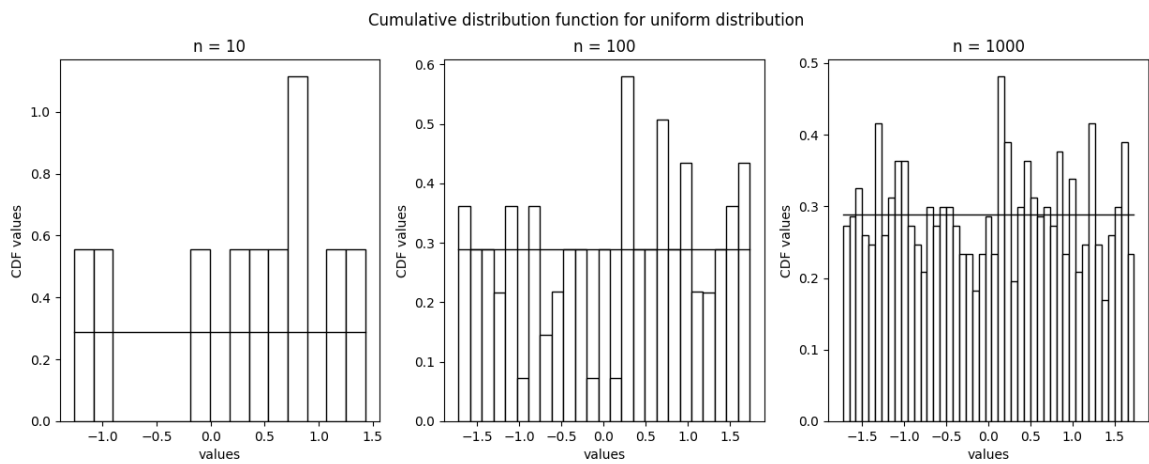


Рис. 5: Равномерное распределение (5)

4.2 Характеристики положения и рассеяния

n = 10					
	\bar{x} (6)	$med\ x$ (7)	z_R (8)	z_Q (10)	z_{tr} (11)
$E(z)$	-0.0041	-0.0104	-0.0091	0.0016	-0.0042
$D(z)$ (12)	0.099	0.14	0.18	0.11	0.11
n = 100					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.0026	-0.0038	-0.0093	-0.019	-0.00202
$D(z)$	0.010	0.016	0.094	0.012	0.012
n = 1000					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.00058	0.00038	0.0037	-0.00076	0.00075
$D(z)$	0.00092	0.0015	0.062	0.0012	0.001160

Таблица 1: Нормальное распределение

n = 10					
	\bar{x} (6)	$med\ x$ (7)	z_R (8)	z_Q (10)	z_{tr} (11)
$E(z)$	0.00098	0.00302	$0.62 \cdot 10^{-5}$	0.0018	0.0020
$D(z)$ (12)	$16.47 \cdot 10^3$	0.38	$0.41 \cdot 10^6$	1.38	0.59
n = 100					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-2.74	-0.000842	$-1.37 \cdot 10^2$	-0.0203	0.0038
$D(z)$	$3.12 \cdot 10^3$	0.025	$7.75 \cdot 10^6$	0.054	0.027
n = 1000					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-3.60	0.00028	$-1.77 \cdot 10^3$	-0.0046	-0.00068
$D(z)$	$9.02 \cdot 10^3$	0.0026	$2.25 \cdot 10^9$	0.0053	0.0028

Таблица 2: Распределение Коши

n = 10					
	\bar{x} (6)	$med\ x$ (7)	z_R (8)	z_Q (10)	z_{tr} (11)
$E(z)$	0.011	-0.0020	0.039	0.0048	0.0034
$D(z)$ (12)	0.27	0.202	1.75	0.20	0.18
n = 100					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.0061	-0.0028	0.29	-0.023	-0.0020
$D(z)$	0.030	0.018	7.24	0.018	0.015
n = 1000					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.00073	0.0024	-0.088	-0.00037	0.0018
$D(z)$	0.0033	0.0019	38.31	0.0020	0.0016

Таблица 3: Распределение Стьюдента

n = 10					
	\bar{x} (6)	$med\ x$ (7)	z_R (8)	z_Q (10)	z_{tr} (11)
$E(z)$	10.0061	9.90	10.27	9.92	9.91
$D(z)$ (12)	1.046	1.37	1.83	1.27	1.14
n = 100					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	10.011	9.86	10.95	9.88	9.87
$D(z)$	0.101	0.204	0.95	0.16	0.12
n = 1000					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	10.0053	10.00	11.68	9.99	9.86
$D(z)$	0.0096	0.0020	0.66	0.0035	0.0109

Таблица 4: Распределение Пуассона

n = 10					
	\bar{x} (6)	$med\ x$ (7)	z_R (8)	z_Q (10)	z_{tr} (11)
$E(z)$	-0.0075	0.0022	-0.010	-0.0033	-0.0052
$D(z)$ (12)	0.099	0.23	0.045	0.14	0.16
n = 100					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.0064	-0.013	-0.0015	-0.023	-0.0086
$D(z)$	0.0099	0.028	0.00063	0.015	0.019
n = 1000					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.0099	0.028	0.00063	0.015	0.019
$D(z)$	0.00098	0.00302	$0.63 \cdot 10^{-5}$	0.0015	0.0020

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.