

Санкт-Петербургский политехнический университет
Петра Великого Физико-механический институт Высшая
школа прикладной математики и вычислительной
физики

Отчёт
по лабораторным работам №1-4
по дисциплине
«Математическая статистика»

Выполнил студент:
Басалаев Даниил Александрович
группа:
5030102/10201
Проверил:
доцент Баженов Александр Николаевич

Санкт-Петербург

1 Лабораторные работы №1-2

1.1 Постановка задачи

1.1.1 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.

1.1.2 Точечное оценивание характеристик положения и рассеяния

Сгенерировать выборки размером 10, 50, 1000 элементов.

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

1.2 Теоретическое обоснование

1.3 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

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

1.3.1 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

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

- Среднее характеристики

$$E(z) = \bar{z} \quad (12)$$

- Оценка дисперсии

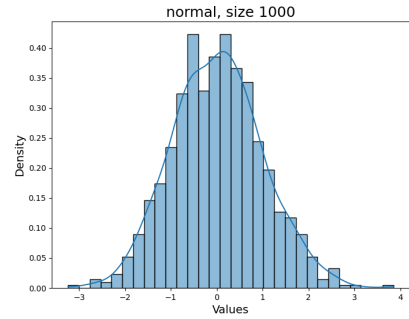
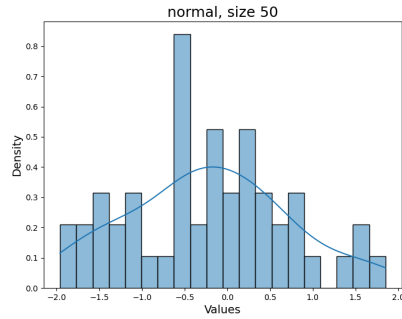
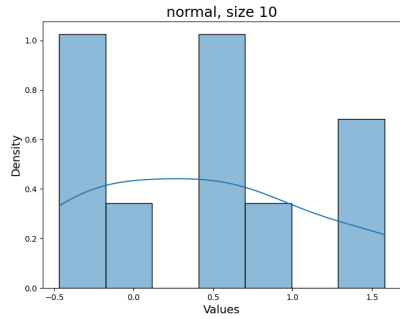
$$D(z) = \overline{z^2} - \bar{z}^2 \quad (13)$$

1.4 Ход работы

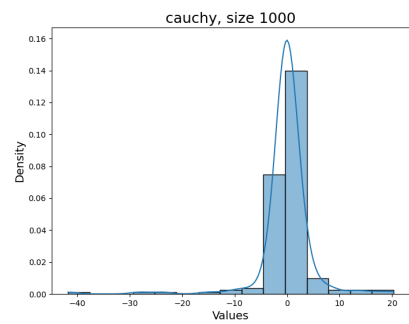
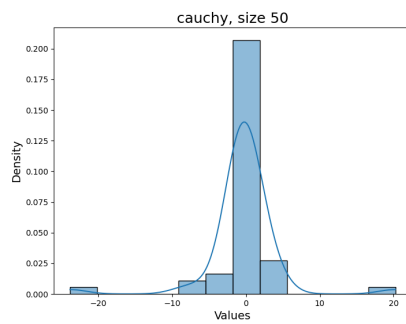
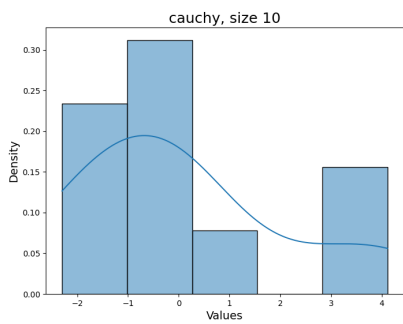
Во время выполнения лабораторной работы все расчеты значений и построения графиков были выполнены на языке программирования Python с использованием таких библиотек, как matplotlib, statistics, pandas, numpy, math.

1.5 Графики

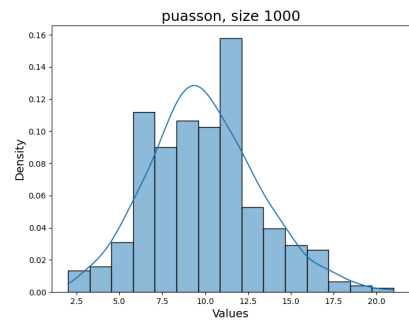
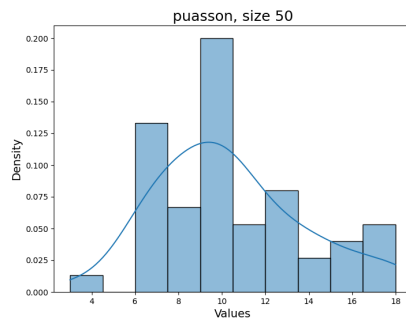
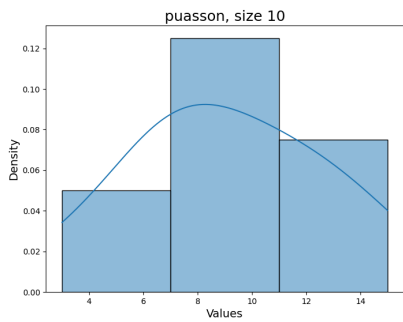
1.5.1 Нормальное распределение



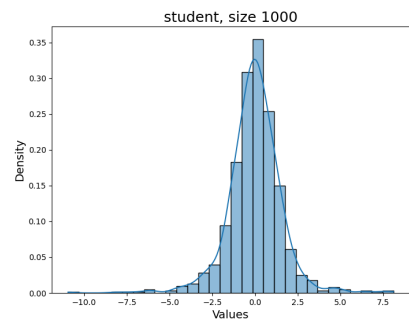
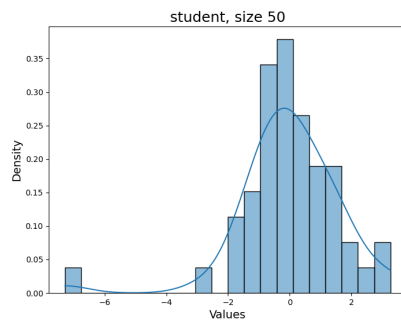
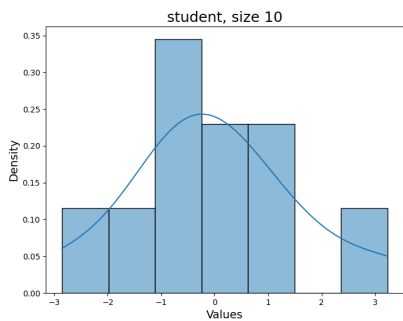
1.5.2 Коши



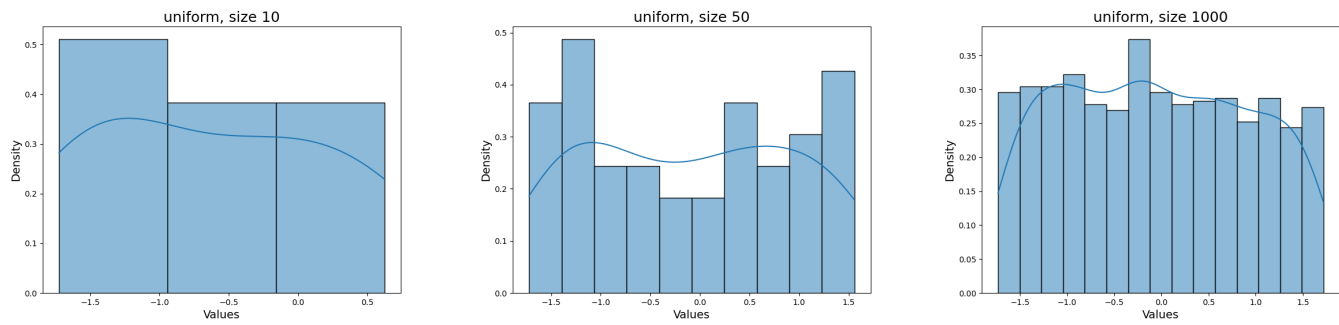
1.5.3 Пуассон



1.5.4 Стьюдент



1.5.5 Равномерное распределение



1.6 Таблицы

1.6.1 Нормальное распределение

normal n = 10	MX	med x	Z_R	Z_Q	Z_{tr}
$E(z)$	0.0029	-0.0049	0.0071	0.0066	0.0015
$D(z)$	0.101	0.1403	0.3761	0.2337	0.1132
normal n = 50	MX	med x	Z_R	Z_Q	Z_{tr}
$E(z)$	-0.0034	-0.004	-0.0075	-0.0027	-0.005
$D(z)$	0.0199	0.0277	0.2403	0.0491	0.0219
normal n = 1000	MX	med x	Z_R	Z_Q	Z_{tr}
$E(z)$	0.0005	-0.0003	0.0018	-0.0017	-0.0001
$D(z)$	0.001	0.0016	0.125	0.0025	0.0012

1.6.2 Коши

cauchy n = 10	MX	med x	Z_R	Z_Q	Z_{tr}
$E(z)$	-0.8588	0.0018	-4.0287	-0.0241	-0.0063
$D(z)$	584.999	0.3453	27944.2716	2.7613	0.5354
cauchy n = 50	MX	med x	Z_R	Z_Q	Z_{tr}
$E(z)$	-1.1602	0.0034	-27.9837	-0.0054	-0.0015
$D(z)$	1476.3684	0.0488	1825915.1848	0.2	0.053
cauchy n = 1000	MX	med x	Z_R	Z_Q	Z_{tr}
$E(z)$	0.2727	-0.0014	161.0178	-0.0064	-0.0023
$D(z)$	1258.889	0.0024	624762602.4426	0.0095	0.0024

1.6.3 Пуассон

poisson n = 10	MX	med x	z_R	z_Q	z_{tr}
$E(z)$	9.9945	9.8645	10.279	9.937	9.9002
$D(z)$	100.9016	98.7267	215.26	199.953	99.1721
poisson n = 50	MX	med x	z_R	z_Q	z_{tr}
$E(z)$	9.9754	9.8225	10.74	9.88	9.8427
$D(z)$	99.7091	96.8537	232.936	195.799	97.1256
poisson n = 1000	MX	med x	z_R	z_Q	z_{tr}
$E(z)$	9.9981	9.996	11.6705	9.9955	9.8567
$D(z)$	99.9713	99.924	273.7785	199.8245	97.1652

1.6.4 Стьюдент

student n = 10	MX	med x	z_R	z_Q	z_{tr}
$E(z)$	0.0071	-0.0032	0.0172	0.0008	0.0018
$D(z)$	0.2774	0.1671	4.1802	0.3923	0.1526
student n = 50	MX	med x	z_R	z_Q	z_{tr}
$E(z)$	0.0011	0.0043	-0.028	0.0039	0.0025
$D(z)$	0.0583	0.0345	9.5648	0.0759	0.0293
student n = 1000	MX	med x	z_R	z_Q	z_{tr}
$E(z)$	0.0006	-0.0001	-0.0726	-0.0015	0.0006
$D(z)$	0.0031	0.0018	101.8467	0.0037	0.0016

1.6.5 Равномерное распределение

uniform n = 10	MX	med x	z_R	z_Q	z_{tr}
$E(z)$	-0.0085	-0.0142	-0.0102	-0.0016	-0.0073
$D(z)$	0.1058	0.2442	0.0979	0.2827	0.1714
uniform n = 50	MX	med x	z_R	z_Q	z_{tr}
$E(z)$	0.0024	0.0017	-0.0015	0.0038	0.0043
$D(z)$	0.0209	0.0587	0.0045	0.0619	0.0399
uniform n = 1000	MX	med x	z_R	z_Q	z_{tr}
$E(z)$	0.0003	0.0002	-0.0001	-0.001	0.0005
$D(z)$	0.001	0.0028	0.0	0.003	0.002

1.7 Вывод

В ходе выполнения лабораторной работы были построены гистограммы и графики плотности на выборках разной мощности (10, 50, 1000) для 5 распределений: Коши, нормального, Пуассона, Стьюдента и равномерного.

В ходе проведения лабораторной работы эти распределения также были исследованы. Суть исследования заключалась в нахождении мат. ожидания и дисперсии 5 величин: выборочного среднего $M[X]$, выборочной медианы $med x$, полусуммы выборочных элементов z_R , полусуммы квартилей z_p и усеченного среднего z_{tr} .

По результатам проведенных исследований можно сделать следующие выводы:

- По полученным графикам легко видеть, что с увеличением мощности выборки характер распределения (форма гистограммы и графика распределения) приближается к теоретическому.

- Для распределения Коши полученные значения исследуемых величин сильно отличаются от теоретических.
- Распределение Стьюдента при малых мощностях выборки дает отличные от теоретических значения, но с ростом мощности они становятся точнее.
- Для нормального, равномерного распределений и распределения Пуассона полученные значения характеристик оказались близки к теоретическим.
- Переходя к оценке поведения исследуемых характеристик, можно сказать, что выборочное среднее является наиболее чувствительным к экстремальным значениям по сравнению с медианой, особенно в выборках малых мощностей. Однако с увеличением мощности выборки влияние этих экстремальных значений на среднее значение уменьшается. В то же время, медиана обычно более устойчива к выбросам и мало варьирует с изменением размера выборки.
- Медиана же является чувствительной к виду распределения: в распределении Коши она дает надежные, устойчивые к выбросам оценки, в нормальном и распределении Стьюдента медиана равна среднему, в Пуассоновском приближается к среднему, и в равномерном равна половине суммы минимального и максимального значений.

2 Лабораторные работы №3-4

2.1 Постановка задачи

2.1.1 Боксплот Тьюки

Сгенерировать выборки размером 20 и 100 элементов. Построить для них боксплот Тьюки.

2.1.2 Доверительные интервалы для параметров нормального распределения

Сгенерировать выборки размером 20 и 100 элементов. Вычислить параметры положения и рассеяния:

- для нормального распределения,
- для произвольного распределения.

2.2 Теоретическое обоснование

2.2.1 Функции распределения

- Нормальное распределение

$$N(x, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (14)$$

- Распределение Коши

$$C(x, 0, 1) = \frac{1}{\pi} \frac{1}{x^2 + 1} \quad (15)$$

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

$$t(x, 0, 3) = \frac{6\sqrt{3}}{\pi(3 + t^2)^2} \quad (16)$$

- Распределение Пуассона

$$P(k, 10) = \frac{10^k}{k!} e^{-10} \quad (17)$$

- Равномерное распределение

$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}}, & |x| \leq \sqrt{3} \\ 0, & |x| > \sqrt{3} \end{cases} \quad (18)$$

2.2.2 Боксплот Тьюки

Боксплот (англ. box plot) — график, использующихся в описательной статистике, компактно изображающий одномерное распределение вероятностей. Такой вид диаграммы в удобной форме показывает медиану, нижний и верхний квартили и выбросы. Границами ящика служат первый и третий квартили, линия в середине ящика — медиана. Концы усов — края статистически значимой выборки (без выброса). Длину «усов» определяют разность первого квартиля и полутора межквартильных расстояний и сумма третьего квартиля и полутора межквартильных расстояний. Формула имеет вид

$$X_1 = Q_1 - \frac{3}{2}(Q_3 - Q_1), \quad X_2 = Q_3 + \frac{3}{2}(Q_3 - Q_1), \quad (19)$$

где X_1 — нижняя граница уса, X_2 — верхняя граница уса, Q_1 — первый квартиль, Q_3 — третий квартиль. Данные, выходящие за границы усов (выбросы), отображаются на графике в виде маленьких кружков. Выбросами считаются величины, такие что:

$$\begin{cases} x < X_1^T \\ x > X_2^T \end{cases} \quad (20)$$

2.2.3 Доверительные интервалы для параметров нормального распределения

Пусть $st_{1-\alpha/2}(n-1)$ — квантиль распределения Стьюдента с $n-1$ степенями свободы и порядка $1-\alpha/2$, где α — выбранный уровень значимости: $2F_T(x) - 1 = 1 - \alpha$. Тогда получаем

$$P\left(\bar{x} - \frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}} < m < \bar{x} + \frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}}\right) = 1 - \alpha, \quad (21)$$

что и даст доверительный интервал для m с доверительной вероятностью $\gamma = 1 - \alpha$ для нормального распределения. Случайная величина $n\frac{s^2}{\sigma^2}$ распределена по закону χ^2 с $n-1$ степенями свободы. Тогда

$$P\left(\frac{s\sqrt{n}}{\sqrt{\chi_{1-\alpha/2}^2(n-1)}} < \sigma < \frac{s\sqrt{n}}{\sqrt{\chi_{\alpha/2}^2(n-1)}}\right) = 1 - \alpha, \quad (22)$$

что и даст доверительный интервал для σ с доверительной вероятностью $\gamma = 1 - \alpha$ для нормального распределения.

2.2.4 Доверительные интервалы для параметров произвольного распределения

Пусть $u_{1-\alpha/2}$ — квантиль нормального распределения порядка $1-\alpha/2$, где α — выбранный уровень значимости: $2\Phi(x) - 1 = 1 - \alpha$. Тогда получаем

$$P\left(\bar{x} - \frac{su_{1-\alpha/2}}{\sqrt{n}} < m < \bar{x} + \frac{su_{1-\alpha/2}}{\sqrt{n}}\right) = 1 - \alpha, \quad (23)$$

что и даст доверительный интервал для m с доверительной вероятностью $\gamma = 1 - \alpha$ для произвольного распределения. Тогда для оценки σ справедлива следующая формула

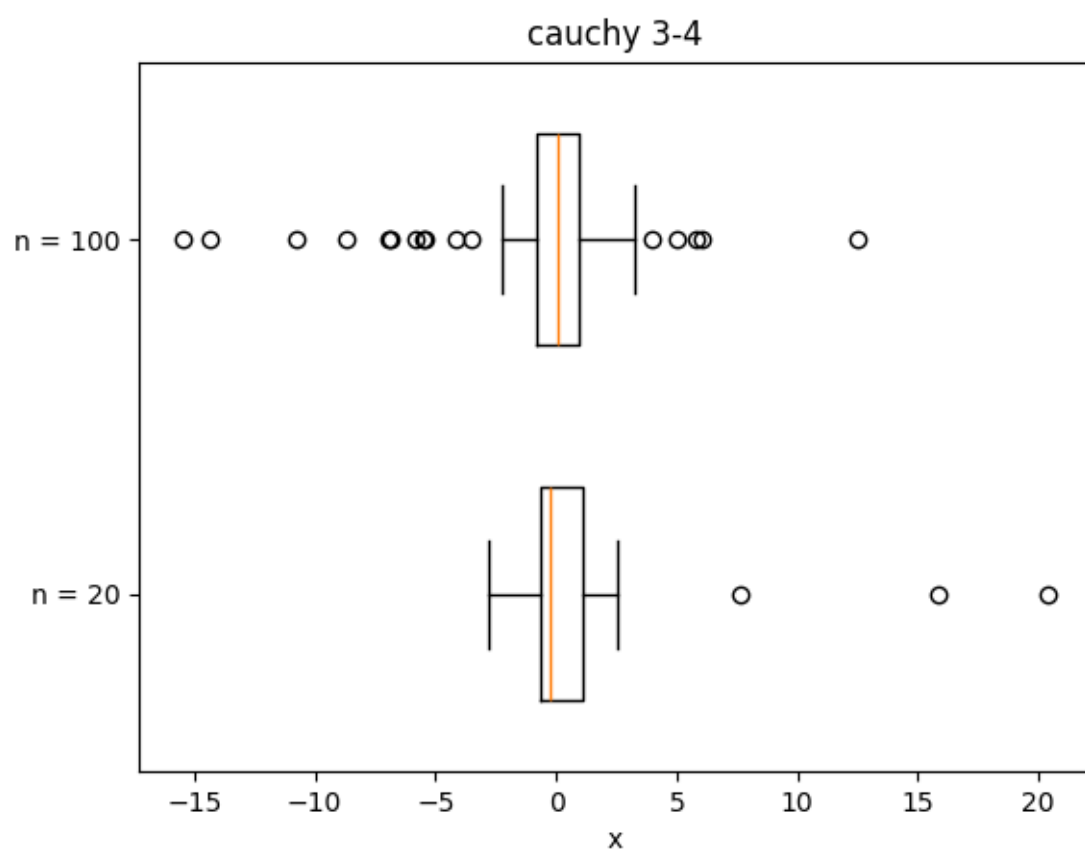
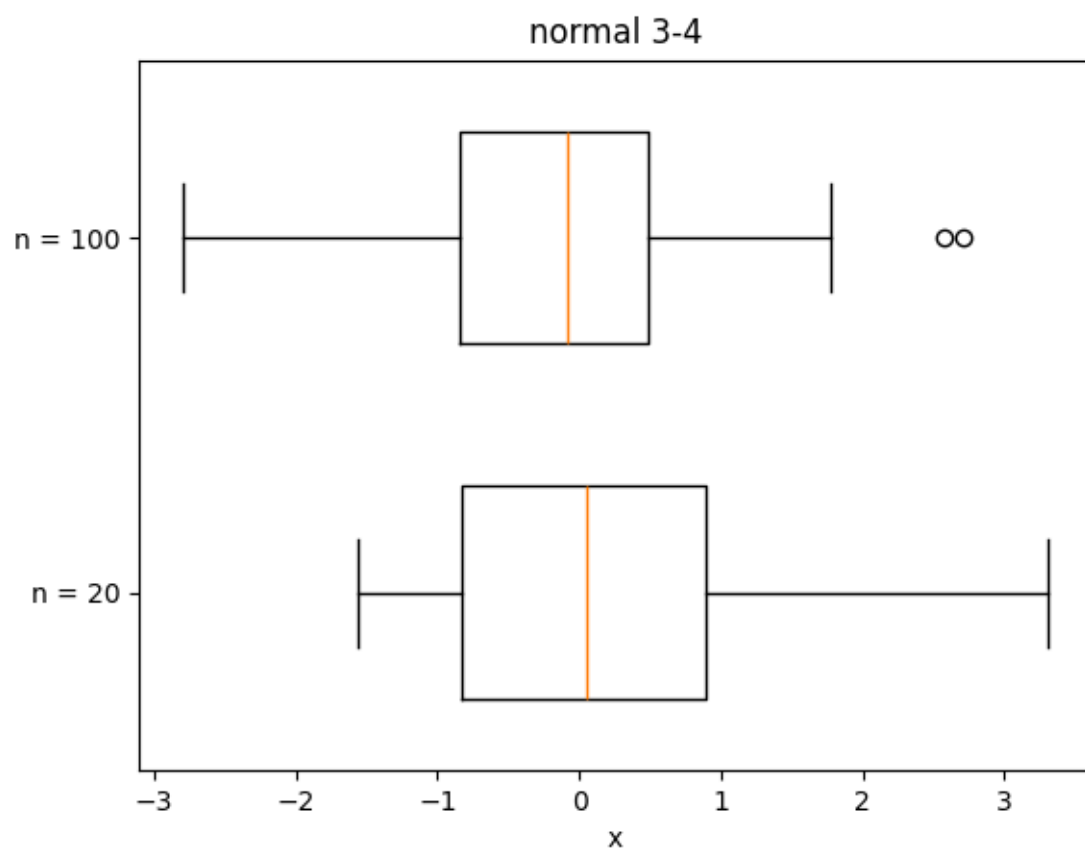
$$P\left(1 - \frac{0.5u_{1-\alpha/2}\sqrt{e+2}}{\sqrt{n}} < \sigma < 1 + \frac{0.5u_{1-\alpha/2}\sqrt{e+2}}{\sqrt{n}}\right) = 1 - \alpha, \quad (24)$$

что и даст доверительный интервал для σ с доверительной вероятностью $\gamma = 1 - \alpha$ для произвольного распределения

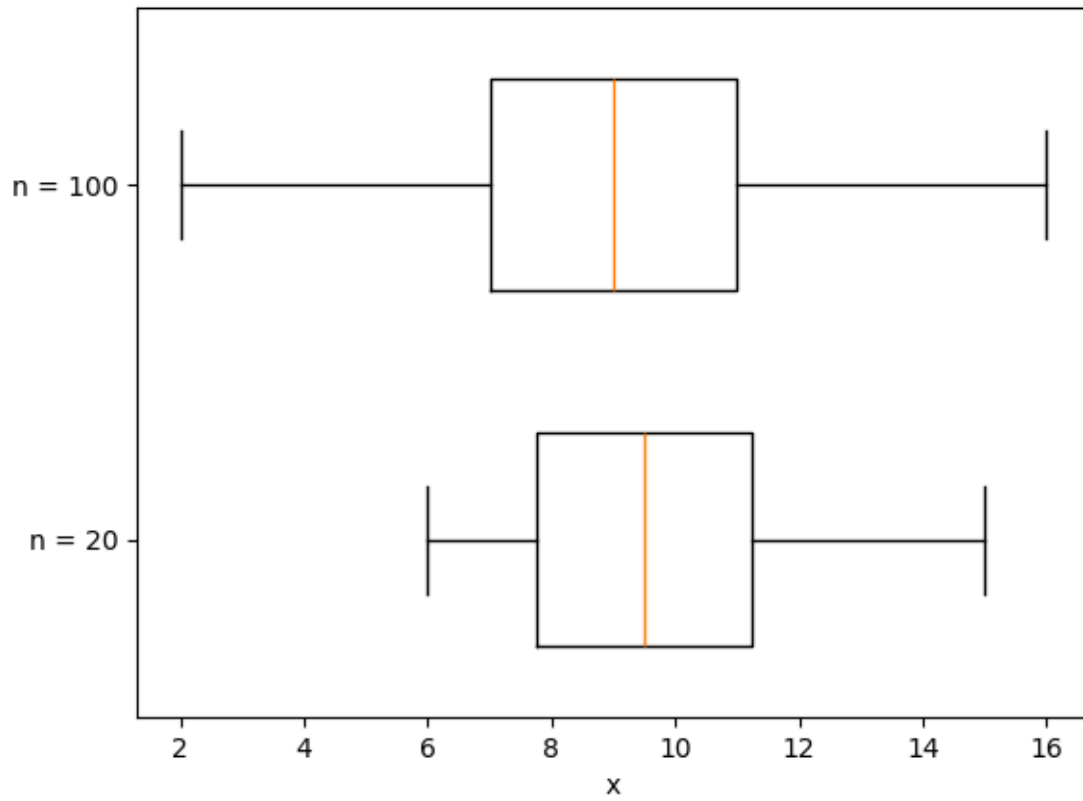
2.3 Ход работы

Во время выполнения лабораторной работы все расчеты значений и построения графиков были выполнены на языке программирования Python с использованием таких библиотек, как matplotlib, statistics, pandas, numpy, math.

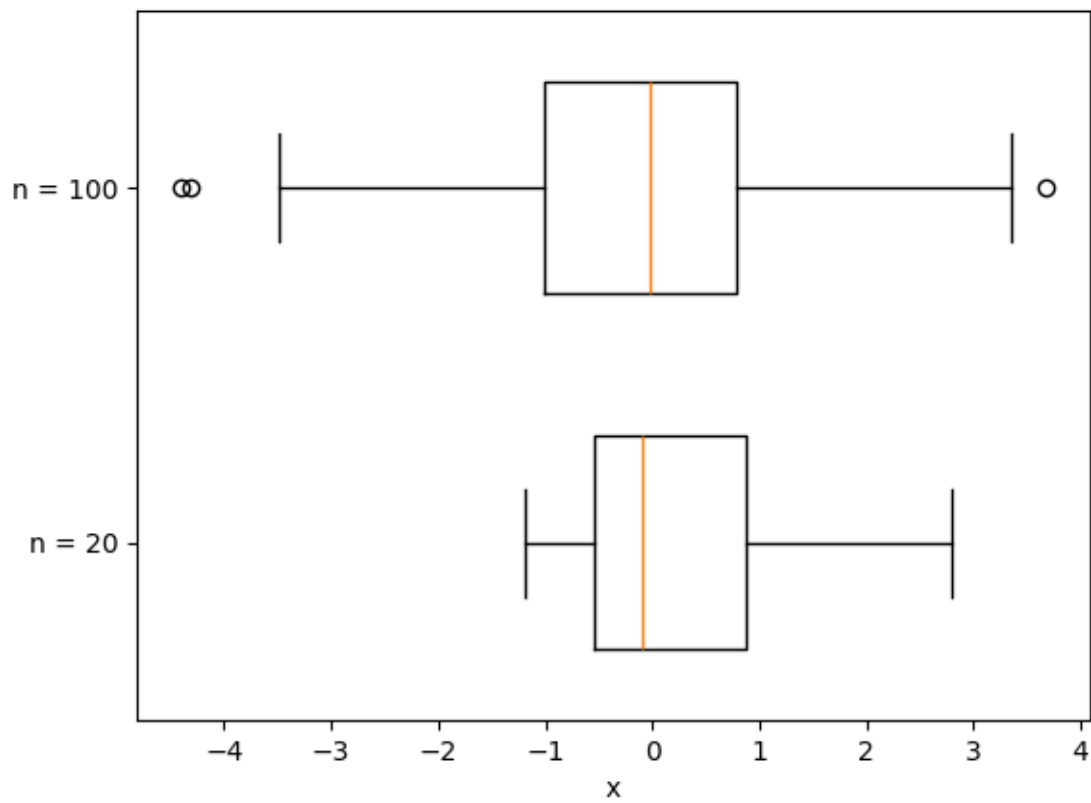
2.4 Бокс-плот Тьюки

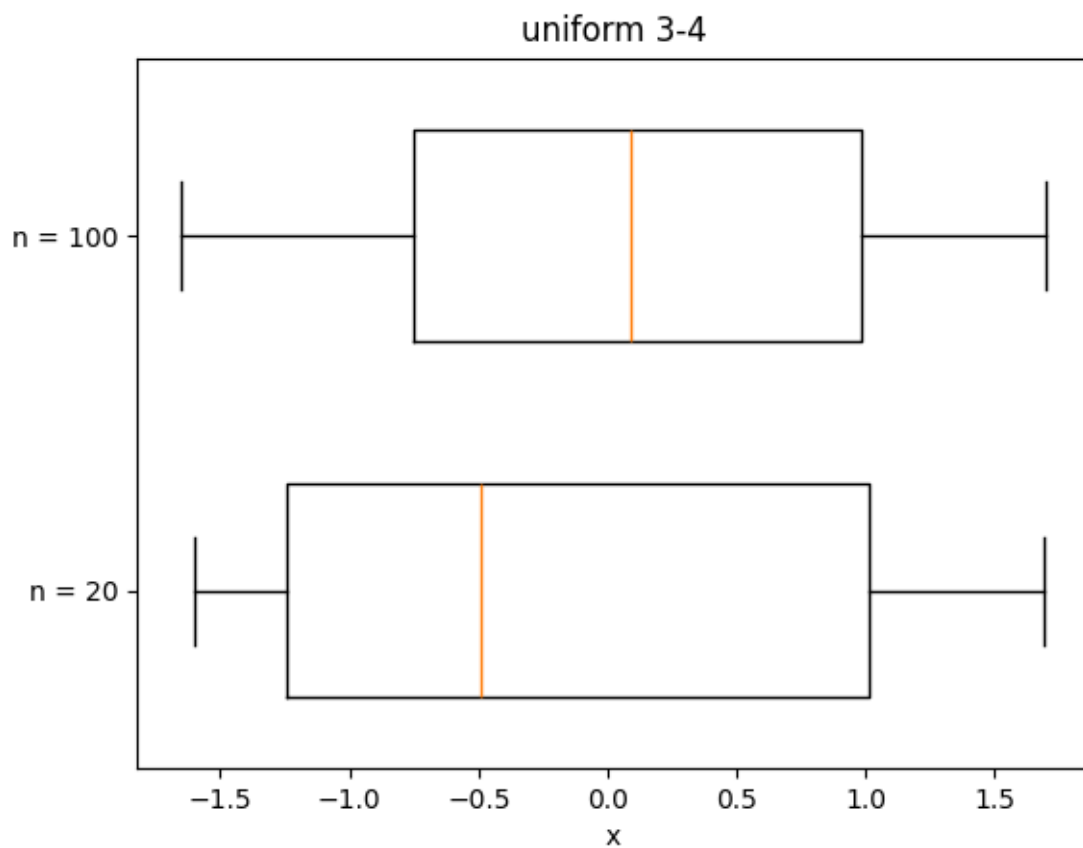


puasson 3-4



student 3-4





2.5 Доверительные интервалы для параметров распределений

n	m	σ
$n = 20$	$-0.479 < m < 0.262$	$0.604 < \sigma < 1.159$
$n = 100$	$-0.178 < m < 0.209$	$0.856 < \sigma < 1.133$

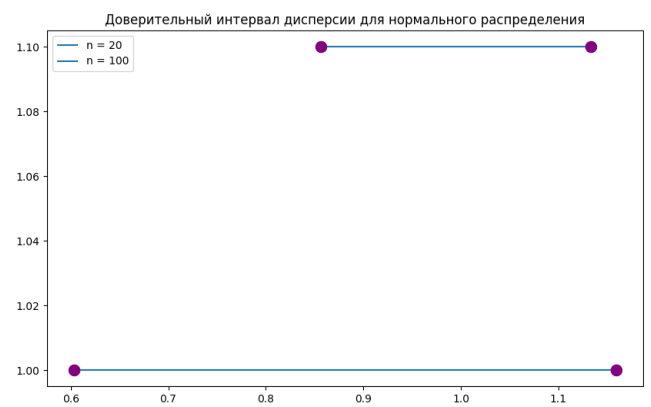
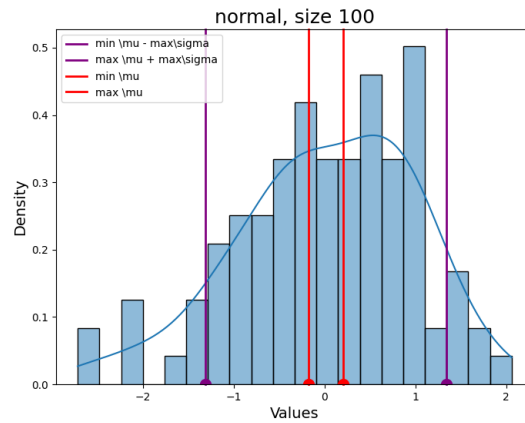
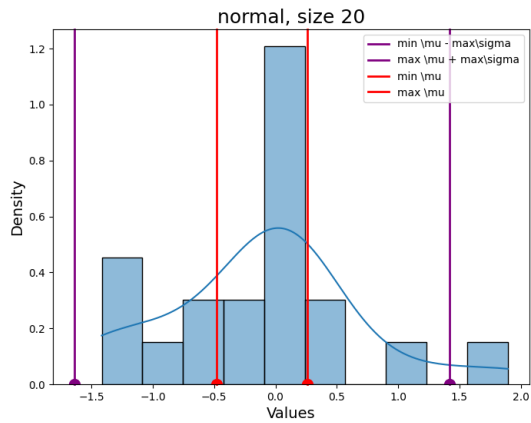
Таблица 1: Доверительные интервалы для параметров нормального распределения (14)

n	m	σ
$n = 20$	$7.674 < m < 9.726$	$1.687 < \sigma < 2.878$
$n = 100$	$8.744 < m < 9.996$	$2.781 < \sigma < 3.573$

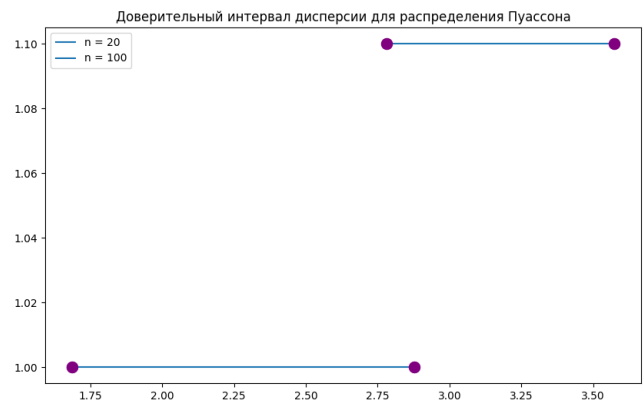
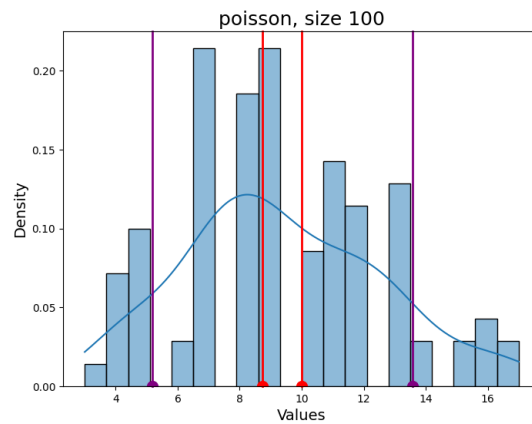
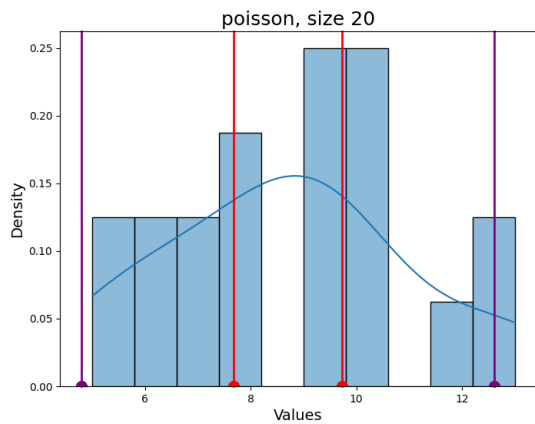
Таблица 2: Доверительные интервалы для параметров произвольного распределения. Асимптотический подход

2.6 Гистограммы

2.6.1 Гистограммы и оценки для параметров нормального распределения



2.6.2 Гистограммы и оценки для параметров произвольного распределения



2.7 Выводы

По результатам выполнения лабораторной работы были сгенерированы выборки размером 20 и 100 элементов и построены для них боксплоты Тьюки.

Боксплот позволяет наглядно представить основные характеристики выборки - медиану, квартили, межквартильный размах и выбросы. На основе построенных графиков можно увидеть разницу в распределении данных для двух выборок. Для выборки размером в 100 элементов представленные метрики имеют более проработанный вид, ведь с увеличением размера выборки улучшается точность оценок параметров распределения.

Также в ходе выполнения лабораторной работы были сгенерированы две выборки размерами 20 и 100 элементов для нормального и произвольного распределения. Затем для каждой из них были вычислены параметры распределения: среднее значение и дисперсия.

Результаты, представленные графически, демонстрируют, что количество элементов в выборке влияет на точность оценок параметров. Более большое количество наблюдений (т.е. 100 элементов) приводит к более точным и стабильным оценкам среднего и дисперсии, как для нормального, так и для произвольного распределения. Для выборки с меньшим количеством элементов (20 элементов) оценки могут сильно варьироваться в зависимости от конкретной выборки, что также наглядно отображено на графиках.

Лабораторная работа иллюстрирует важнейший статистический принцип: точность статистической оценки увеличивается с ростом объема выборки. Результаты этого исследования подчеркивают значимость использования достаточно больших выборок для надежного анализа данных.

3 Ссылка на репозиторий GitHub

<https://github.com/11AgReS1SoR11/MatStat/tree/main>