

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ
ПЕТРА ВЕЛИКОГО

ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ И МЕХАНИКИ
ВЫСШАЯ ШКОЛА ПРИКЛАДНОЙ МАТЕМАТИКИ И ФИЗИКИ

Математическая статистика
Отчёт по лабораторным работам №1-2

Выполнил:

Студент: Сачук Александр

Группа: 5030102/90201

Принял:

к. ф.-м. н., доцент

Баженов Андрей Николаевич

2022 г.

Содержание

Список иллюстраций

Список таблиц

1. Постановка задачи

Для четырех распределений:

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

Выполнить следующие задачи:

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: \bar{x} , $medx$, z_R , z_Q , z_{tr} . Повторить такие вычисления 1000 раз для каждой выборки и найти среднее характеристик положения и их квадратов:

$$E(z) = \bar{z} \quad (1)$$

Вычислить оценку дисперсии по формуле:

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

Представить полученные данные в виде таблиц.

2. Теория

2.1. Рассматриваемые распределения

Плотности распределений:

- Нормальное распределение

$$N(x, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (3)$$

- Распределение Коши

$$C(x, 0, 1) = \frac{1}{\pi [1 + x^2]} \quad (4)$$

- Распределение Пуассона

$$P(k, 10) = \frac{10^k}{k!} e^{-10} \quad (5)$$

- Равномерное распределение

$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}}, & x \in [-\sqrt{3}, \sqrt{3}] \\ 0, & x \notin [-\sqrt{3}, \sqrt{3}] \end{cases} \quad (6)$$

2.2. Гистограммы

2.2.1. Определение и описание

Гистограмма - функция, приближающая плотность вероятности некоторого распределения, построенная на основе выборки из него. Используются гистограммы для визуализации данных на начальном этапе статистической обработки. Построение гистограмм используется для получения эмпирической оценки плотности распределения случайной величины.

2.2.2. Построение гистограммы

Гистограммы строятся следующим образом: все множество значений, которые могут принимать элементы выборки, разбивается на несколько интервалов. Чаще всего, эти интервалы делают одинакового размера, но это не обязательно (в данной лабораторной работе интервалы будут одинакового размера). Если интервалы одинакового размера, то высота каждого прямоугольника гистограммы будет прямо пропорционален числу элементов выборки, попавших в этот интервал. Если же интервалы разного размера, то высота прямоугольников выбирается так, чтоб их площадь была пропорциональна числу элементов выборки, попавших в этот интервал.

2.3. Вариационный ряд

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

2.4. Выборочные числовые характеристики

2.4.1. Характеристики положения

2.4.2. Характеристики рассеяния

3. Результаты

3.1. Гистограммы

- Нормальное распределение

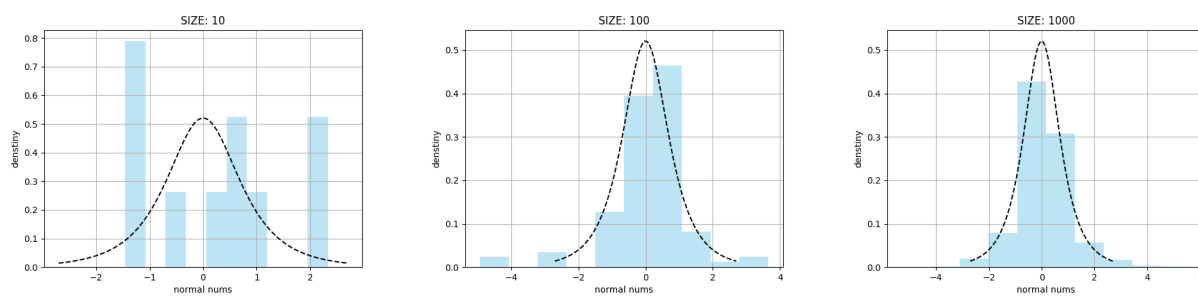


Рис. 1. Гистограмма и плотность вероятности для нормального распределения [$N = 10, 100, 1000$]

- Распределение Коши

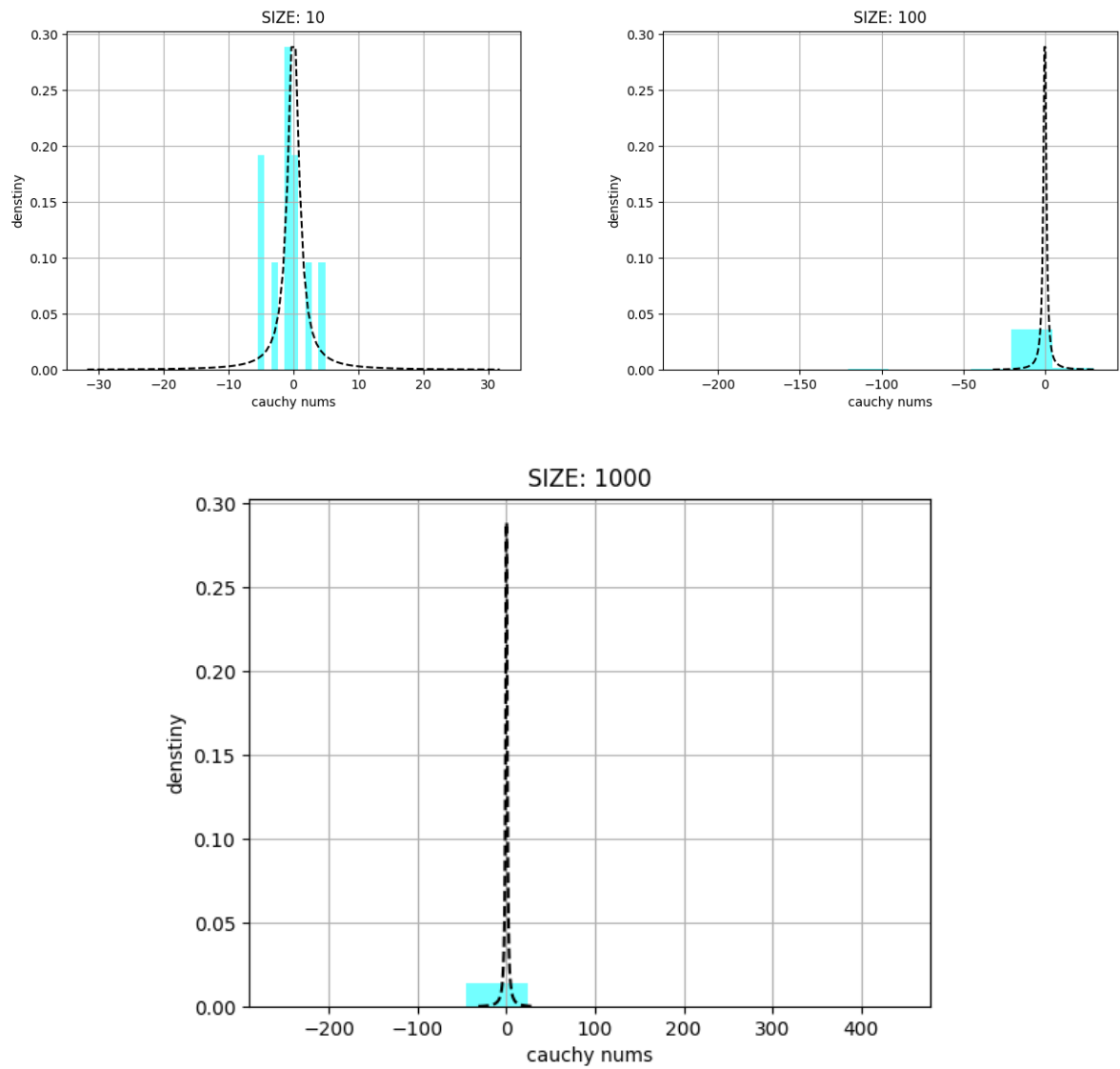


Рис. 2. Гистограмма и плотность вероятности для распределения Коши $[N = 10, 100, 1000]$

- Распределение Пуассона

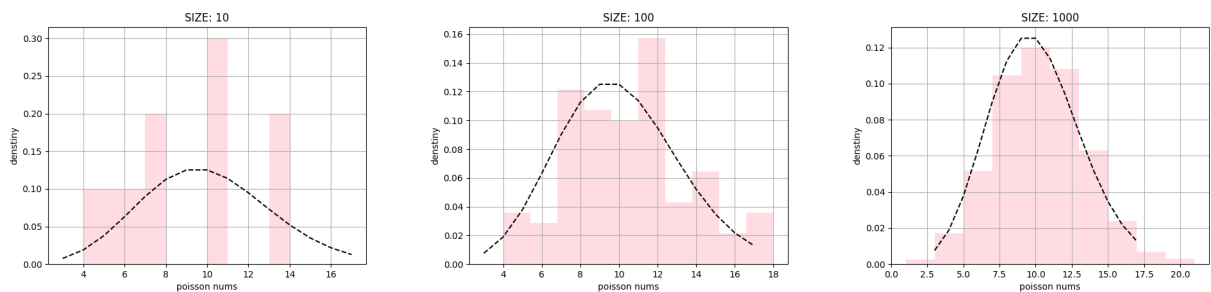


Рис. 3. Гистограмма и плотность вероятности для распределения Пуассона $[N = 10, 100, 1000]$

- Равномерное распределение

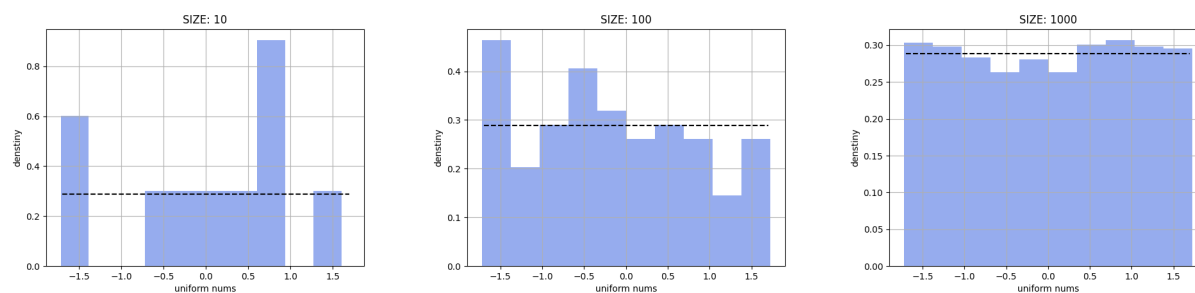


Рис. 4. Гистограмма и плотность вероятности для равномерного распределения [$N = 10, 100, 1000$]

3.2. Характеристики положения и рассеяния

4. Реализация

Данная лабораторная работа была выполнена с использованием языка программирования Python 3.10 в среде разработки Visual Studio Code с использованием следующих библиотек:

- scipy версии 1.8.0
- numpy версии 1.22.0
- matplotlib версии 3.5.1

5. Обсуждение

5.1. Гистограммы

Полученные результаты работы говорят о том, что при увеличении размеров выборок, гистограммы все ближе к графику плотности вероятности того закона, по которому были сгенерированы элементы выборок. Верно и обратное: чем меньше выборка, тем хуже по ней можно определить закон, по которой эта выборка генерировалась.

Также одним из ключевых выводов является тот факт, что по маленькому размеру выборки ($n = 10$) очень трудно отличить гистограммы, а, следовательно, и определить закон, по которой генерировалась выборка. Действительно, гистограмма выборки, построенной по распределению Пуассона при $n = 10$, могла бы с тем же успехом описывать график равномерного распределения (если не учитывать один единственный всплеск гистограммы, который вообще мог остаться незамеченным при более широких интервалах боксов гистограммы).

При выборках $n = 1000$ видно, что гистограммы уже достаточно неплохо приближаются к графикам плотностей соответствующих законов распределения: в равномерном распределении отклонения гистограммы от графика незна-

чительны, а в нормальном распределении уже наблюдаются «хвосты», которые позволяют отличить треугольное распределение от нормального.

5.2. Характеристики положения и рассеяния

6. Ссылки на библиотеки

1. <https://scipy.org/> - SciPy
2. <https://numpy.org/> - NumPy
3. <https://numpy.org/> - Matplotlib

7. Ссылки на репозиторий

<https://github.com/AS2/Mathematical-statistics> - GitHub репозиторий