Санкт-Петербургский политехнический университет Петра Великого
Институт прикладной математики и механики
Кафедра «Прикладная математика»

ОТЧЁТ ПО ЛАБОРАТОРНЫМ РАБОТАМ №1-2 ПО ДИСЦИПЛИНЕ «МАТЕМАТИЧЕСКАЯ СТАТИСТИКА»

Выполнил
студент группы 3630102/70401

Кнодель Юлия Максимовна

Проверил
к. ф.-м. н., доцент

Баженов Александр Николаевич

Санкт-Петербург
2020

Содержание

Список иллюстраций

Список таблиц

1 Постановка задачи

Для 5 распределений:

1. $N(x, 0, 1)$ – нормальное распределение
2. $C(x, 0, 1)$ – распределение Коши
3. $L(x, 0, \frac{1}{\sqrt{2}})$ – распределение Лапласа
4. $P(k, 10)$ – распределение Пуассона
5. $U(x, -\sqrt{3}, \sqrt{3})$ – расномерное распределение

1.1 Задание 1

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

1.2 Задание 2

Сгенерировать выборки размером 10, 100 и 1000 элементов. Для каждой выборки вычислить следующие статистические характеристики положения данных: $\overline{x}, medx, z_R, z_Q, z_{tr}$. Повторить такие вычисления 1000 раз для каждой выборки и найти среднее характеристик положения и их квадратов:

$$E(z) = \overline{z} \tag{1}$$

Вычислить оценку дисперсии по формуле:

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

Представить полученные данные в виде таблиц.

2 Теория

2.1 Распределения

- Нормальное распределение

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

- Распределение Коши

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

- Распределение Лапласа

$$L(x, 0, \frac{1}{\sqrt{2}}) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|} \tag{3}$$

- Распределение Пуассона

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

- Равномерное распределение

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

2.2 Гистограмма

2.2.1 Определение

Гистограмма в математической статистике — это функция, приближающая плотность вероятности некоторого распределения, построенная на основе выборки из него.

2.2.2 Графическое описание

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.

2.2.3 Использование

Гистограммы применяются в основном для визуализации данных на начальном этапе статистической обработки.
Построение гистограмм используется для получения эмпирической оценки плотности распределения случайной величины. Для построения гистограммы наблюдаемый диапазон изменения случайной величины разбивается на несколько интервалов и подсчитывается доля от всех измерений, попавшая в каждый из интервалов. Величина каждой доли, отнесенная к величине интервала, принимается в качестве оценки значения плотности распределения на соответствующем интервале.

2.3 Вариационный ряд

Вариационным рядом называется последовательность элементов выборки, расположенных в неубывающем порядке. Одинаковые элементы повторяются. Запись вариационного ряда: $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$. Элементы вариационного ряда $x_{(i)}(i = 1, 2, \ldots, n)$ называются порядковыми статистиками.

2.4 Выборочные числовые характеристики

С помощью выборки образуются её числовые характеристики. Это числовые характеристики дискретной случайной величины X^*, принимающей выборочные значения $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$.

2.4.1 Характеристики положения

- Выборочное среднее

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{8}$$

- Выборочная медиана

$$medx = \begin{cases} x_{(l+1)} & n = 2l+1 \\ \frac{x_{(l)}+x_{(l+1)}}{2} & n = 2l \end{cases} \tag{9}$$

- Полусумма экстремальных выборочных элементов

$$z_R = \frac{x_{(1)} + x_{(n)}}{2} \tag{10}$$

- Полусумма квартилей
  Выборочная квартиль z_p порядка p определяется формулой

$$z_p = \begin{cases} x_{([np]+1)} & np\text{−дробное} \\ x_{(np)} & np\text{−целое} \end{cases} \tag{11}$$

  Полусумма квартилей

$$z_Q = \frac{z_{1/4} + z_{3/4}}{2} \tag{12}$$

- Усечённое среднее

$$z_{tr} = \frac{1}{n-2r}\sum_{i=r+1}^{n-r} x_{(i)}, r \approx \frac{n}{4} \tag{13}$$

2.4.2 Характеристики рассеяния

Выборочная дисперсия

$$D = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \tag{14}$$

3 Реализация

В приложении находится ссылка на репозиторий на GitHub, где находится исходный код лабораторной работы.

4 Результаты

4.1 Гистограммы и графики плотности распределения

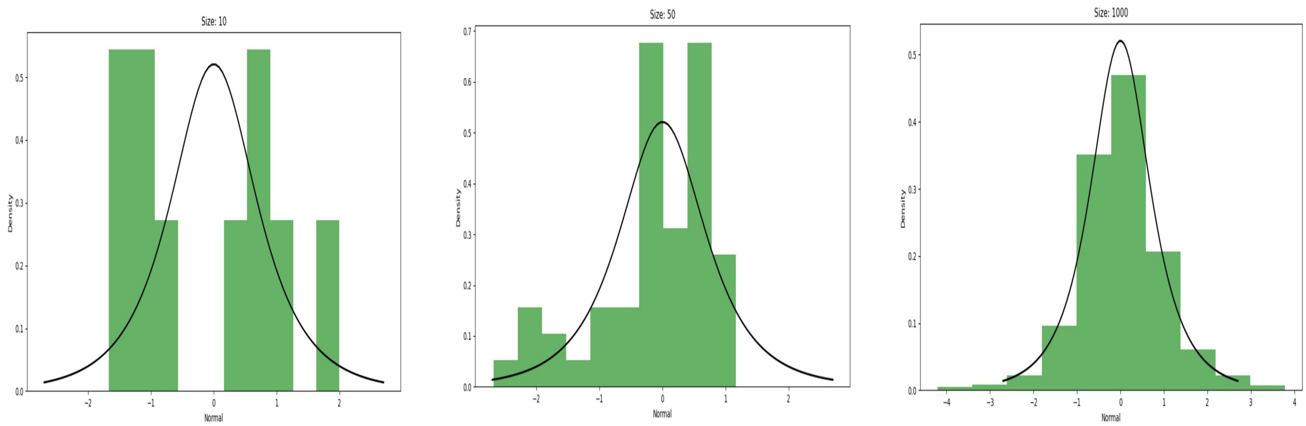

Рис. 1: Нормальное распределение

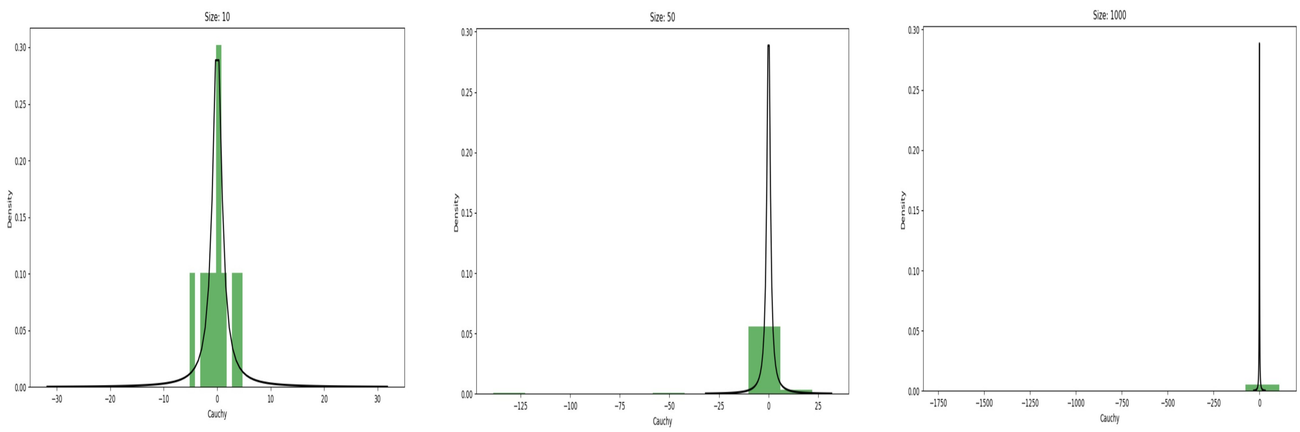

Рис. 2: Распределение Коши

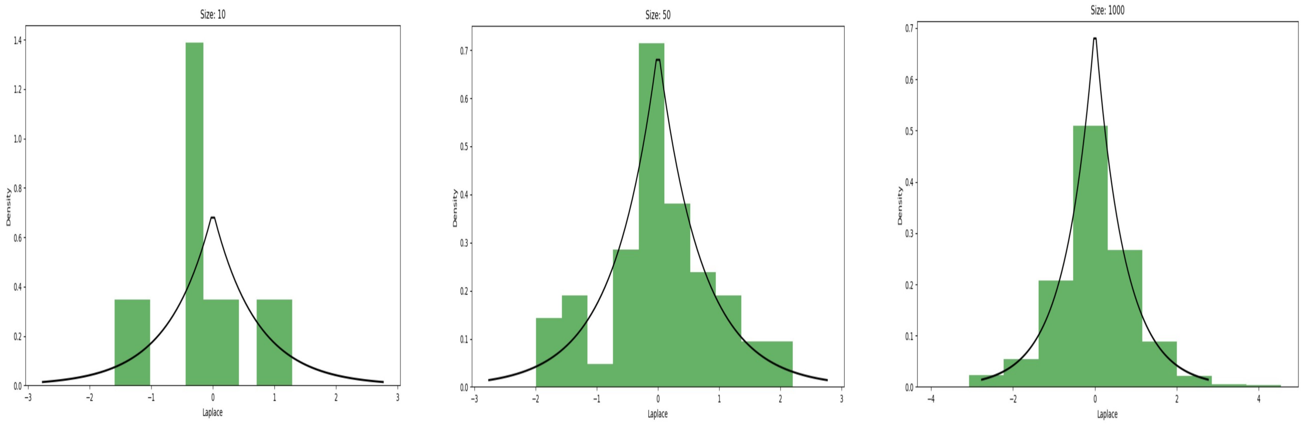

Рис. 3: Распределение Лапласа

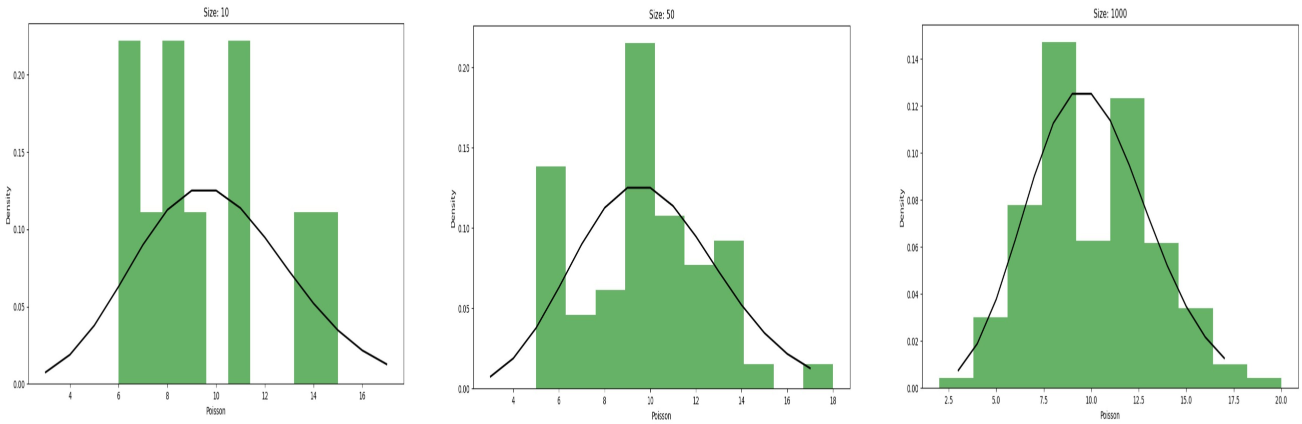

Рис. 4: Распределение Пуассона

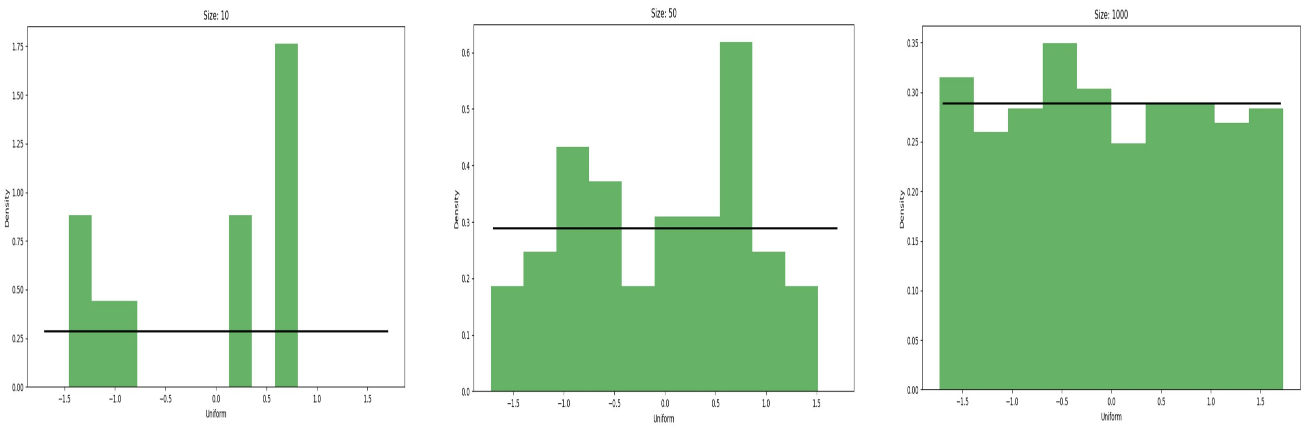

Рис. 5: Равномерное распределение

4.2 Характеристики положения и рассеяния

Characteristic	Mean	Median	z_R	z_Q	z_{tr}
Normal E(z) 10	0.025501	0.010119	0.030651	0.030570	0.025708
Normal D(z) 10	0.103765	0.092774	0.476065	0.505294	0.172139
Normal E(z) 100	0.001106	0.001414	-0.04678	0.008496	0.005806
Normal D(z) 100	0.010594	0.009886	0.461752	0.514966	0.018883
Normal E(z) 1000	-0.000306	0.00014	0.029287	0.017512	-0.001402
Normal D(z) 1000	0.0010219	0.00090	0.499531	0.498409	0.0020510

Таблица 1: Нормальное распределение

Characteristic	Mean	Median	z_R	z_Q	z_{tr}
Cauchy E(z) 10	-1.82930	-0.01278	-3.6080	-0.6565	-1.73308
Cauchy D(z) 10	987.1422	0.34462	8934.589	854.6997	1708.300
Cauchy E(z) 100	-8.04716	-0.002323	2.58638	3.76963	-17.4805
Cauchy D(z) 100	90333.65	0.024408	3996.35	5978.85	359703.9
Cauchy E(z) 1000	0.0993672	0.000122	-2.16597	-0.68649	-0.033700
Cauchy D(z) 1000	1228.695	0.0024604	3185.117	1023.5756	4370.5033

Таблица 2: Распределение Коши

Characteristic	Mean	Median	z_R	z_Q	z_{tr}
Laplace E(z) 10	-0.006466	0.000874	-0.017804	-0.029233	-0.010286
Laplace D(z) 10	0.0967693	0.068329	0.529119	0.4894004	0.159409
Laplace E(z) 100	-0.002621	-0.002186	0.0325949	0.000488	-0.007330
Laplace D(z) 100	0.010093	0.006105	0.5098014	0.528626	0.020044
Laplace E(z) 1000	-0.000461	4.5834-05	0.005827	0.008741	-0.0003342
Laplace D(z) 1000	0.000950	0.000526	0.503384	0.483016	0.001967

Таблица 3: Распределение Лапласа

Characteristic	Mean	Median	z_R	z_Q	z_{tr}
Poisson E(z) 10	10.0305	9.887	9.9375	10.1255	10.0543
Poisson D(z) 10	1.05715	1.45523	5.1613	5.0939	1.70460
Poisson E(z) 100	9.978119	9.8255	9.8905	9.777	9.98496
Poisson D(z) 100	0.097413	0.1972	5.11775	5.105771	0.198463
Poisson E(z) 1000	10.00179	9.997	9.905	9.988	10.0054
Poisson D(z) 1000	0.009803	0.001991	5.31547	5.15535	0.018978

Таблица 4: Распределение Пуассона

Characteristic	Mean	Median	z_R	z_Q	z_{tr}
Uniform E(z) 10	0.017459	0.016779	0.043340	-0.011148	0.001145
Uniform D(z) 10	0.098445	0.219330	0.487425	0.5378816	0.166308
Uniform E(z) 100	-0.002799	-0.00437	0.017764	0.001040	0.002232
Uniform D(z) 100	0.010377	0.030492	0.443175	0.514648	0.020929
Uniform E(z) 1000	0.001501	0.002397	-0.033368	0.026533	0.001575
Uniform D(z) 1000	0.001005	0.002988	0.5136426	0.495917	0.002077

Таблица 5: Равномерное распределение

5 Обсуждение

5.1 Гистограмма и график плотности распределения

По полученным графикам можно сделать вывод о том, что чем больше размер выборки, тем точнее гистограмма будет повторять график плотности вероятности того закона распределения, для которого была сгенерирована выборка. На маленьких выборках определить закон распределения почти невозможно.

5.2 Характеристики положения и рассеяния

Исходя из данных, приведенных в таблицах, можно судить о том, что дисперсия характеристик рассеяния для распределения Коши является некой аномалией: значения слишком большие даже при увеличении размера выборки - понятно, что это результат выбросов, которые мы могли наблюдать в результатах предыдущего задания.

6 Приложения

Код программы - GitHub URL: https://github.com/juliaknodel/MS-1-2