

Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики
Кафедра «Прикладная математика»

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №2

по дисциплине
"Математическая статистика"

Выполнила студентка
группы 3630102/80201

Деркаченко Анна Олеговна

Проверил
доцент, к.ф.-м.н.

Баженов Александр Николаевич

Санкт-Петербург
2021 г.

Содержание

1. Постановка задачи	4
2. Теория	4
2.1. Распределения	4
2.2. Выборочные числовые характеристики	5
2.2.1. Характеристики положения	5
2.2.2. Характеристики рассеивания	5
3. Реализация	5
4. Результаты	6
5. Обсуждение	7

Список иллюстраций

1. Постановка задачи

Даны распределения:

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

Необходимо:

- 1) Сгенерировать выборки размером 10, 100 и 1000 элементов
- 2) Построить для них характеристики положения данных: $\bar{x}, medx, z_R, z_Q, z_{tr}$
- 3) Повторить данные вычисления 1000 раз для каждой выборки, найти среднее характеристик положения $E(z) = \bar{z}$ и вычислить оценку дисперсии $D(z) = \bar{z}^2 - \bar{z}^2$
- 4) Представить полученные результаты в виде таблиц

2. Теория

2.1. Распределения

Плотности рассматриваемых распределений:

- 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 Лапласа

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

- распределение Пуассона

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

- равномерное распределение

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

2.2. Выборочные числовые характеристики

Вариационный ряд - последовательность элементов выборки x_1, x_2, \dots, x_n , расположенных в неубывающем порядке.

Дискретная случайная величина имеет числовые характеристики, образующиеся с помощью выборки из этой величины.

2.2.1. Характеристики положения

- Выборочное среднее

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

- Выборочная медиана

$$medx = \begin{cases} x_{l+1}, n = 2l + 1 \\ \frac{x_l + x_{l+1}}{2}, n = 2l \end{cases} \quad (7)$$

- Полусумма экстремальных выборочных элементов

$$z_R = \frac{x_1 + x_n}{2} \quad (8)$$

- Полусумма квартилей

Выборочная квартиль z_p порядка p определяется формулой

$$z_p = \begin{cases} x_{[np]+1}, np - \text{дробное} \\ x_{np}, np - \text{целое} \end{cases} \quad (9)$$

Полусумма квартилей

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

- Усечённое среднее

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

2.2.2. Характеристики рассеивания

Выборочная дисперсия определяется по формуле:

$$D = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (12)$$

3. Реализация

Реализация лабораторной работы проводилась на языке Python в среде разработки PyCharm с использованием дополнительных библиотек:

- `scipy`

- numpy
- math

Исходный код лабораторной работы размещен в GitHub-репозитории.

URL: <https://github.com/derkanw/Mathstat/tree/main/lab2>

4. Результаты

	\bar{x}	$medx$	z_R	z_Q	z_{tr}
n=10					
$E(z)$	0.006371	-0.007288	0.023028	0.321795	0.279379
$D(z)$	0.099103	0.148119	0.167815	0.121352	0.121102
n=100					
$E(z)$	-0.00237	-0.004485	-0.013181	0.011586	0.023726
$D(z)$	0.010205	0.014457	0.081592	0.012794	0.011714
n=1000					
$E(z)$	-8.6e-05	-0.000496	-0.007813	0.001156	0.002173
$D(z)$	0.001049	0.001548	0.062136	0.001285	0.001217

Таблица 1. Таблица характеристик для нормального распределения

	\bar{x}	$medx$	z_R	z_Q	z_{tr}
n=10					
$E(z)$	-4.381265	-0.011736	-21.876557	1.167523	0.709626
$D(z)$	27051.784196	0.340525	676317.178264	4.667338	1.252677
n=100					
$E(z)$	-1.156414	-0.00829	-54.318545	0.022735	0.034599
$D(z)$	2465.615699	0.025737	5999973.185907	0.056302	0.028603
n=1000					
$E(z)$	0.624531	0.000828	316.572342	0.002332	0.004516
$D(z)$	391.702877	0.002398	95371869.306265	0.004996	0.002449

Таблица 2. Таблица характеристик для распределения Коши

	\bar{x}	$medx$	z_R	z_Q	z_{tr}
n=10					
$E(z)$	-0.002224	-0.002246	-0.004543	0.296567	0.230068
$D(z)$	0.101624	0.075864	0.384596	0.118392	0.084
n=100					
$E(z)$	-0.002469	-0.000214	-0.036222	0.011354	0.019459
$D(z)$	0.01011	0.005916	0.427356	0.010273	0.006263
n=1000					
$E(z)$	-0.00023	-0.000612	0.023942	0.001222	0.001819
$D(z)$	0.001008	0.000537	0.409581	0.000964	0.000605

Таблица 3. Таблица характеристик для распределения Лапласа

	\bar{x}	$medx$	z_R	z_Q	z_{tr}
n=10					
$E(z)$	9.9858	9.857	10.2595	10.901	10.756667
$D(z)$	1.109498	1.527051	1.96991	1.479199	1.343289
n=100					
$E(z)$	9.98466	9.8115	10.973	9.956	9.9234
$D(z)$	0.0984	0.216718	0.993271	0.154564	0.1187
n=1000					
$E(z)$	10.000605	9.996	11.6505	9.9985	9.866522
$D(z)$	0.00996	0.003484	0.6741	0.001248	0.011262

Таблица 4. Таблица характеристик для распределения Пуассона

	\bar{x}	$medx$	z_R	z_Q	z_{tr}
n=10					
$E(z)$	0.007634	-0.003641	0.00133	0.31847	0.322529
$D(z)$	0.100261	0.231998	0.046272	0.124854	0.150706
n=100					
$E(z)$	-0.00029	0.000115	-0.000439	0.014805	0.033771
$D(z)$	0.01091	0.029984	0.00057	0.016426	0.021133
n=1000					
$E(z)$	0.001661	0.002659	-0.000108	0.003872	0.005872
$D(z)$	0.001096	0.003206	6e-06	0.001635	0.00219

Таблица 5. Таблица характеристик для равномерного распределения

5. Обсуждение

Исходя из данных, представленных в таблицах, можно сделать вывод, что увеличение размерности выборки имеет уточняющее значение для выборочной оценки характеристик случайной величины. Также сходные значения соответствующих параметров имеют нормальное распределение, распределение Лапласа и равномерное распределение. Примечательно, что большая часть их значений близка к нулю.

Стоит отметить, что распределение Пуассона имеет значением среднего своих параметрах при 1000 опытах величину в окрестности 10, что подтверждается значением параметра задания данного распределения.

К тому же, в таблице характеристик распределения Коши можно выделить аномальные значение, явно превышающие ожидаемые. Такую ситуацию можно объяснить наличием различных выбросов, неопределенностью математического ожидания и бесконечностью дисперсии случайной величины, распределенной по данному закону.