Санкт-Петербургский политехнический университет Петра Великого
Институт прикладной математики и механики
Высшая школа прикладной математики и вычислительной физики

Отчёт по лабораторной работе №2

по дисциплине
«Математическая статистика»

Выполнил студент
группы 3630102/80401 Веденичев Дмитрий Александрович

Проверил
Доцент, к.ф.-м.н. Баженов Александр Николаевич

Санкт-Петербург
2021

Содержание

Список таблиц **3**

1 Постановка задачи **4**

2 Теория **4**

2.1 Распределения 4

2.2 Вариационный ряд 4

2.3 Выборочные числовые характеристики 5

2.3.1 Характеристики положения 5

2.3.2 Характеристики рассеяния 5

3 Программная реализация **5**

4 Результаты **6**

4.1 Характеристики положения и рассеяния 6

5 Обсуждение **8**

6 Приложение **8**

Список таблиц

1 Распределение Лапласа (5) 6
2 Равномерное распределение (7) 6
3 Нормальное распределение (3) 7
4 Распределение Коши (4) 7
5 Распределение Пуассона (6) 7

1 Постановка задачи

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

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

- Распределение Коши

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

- Распределение Лапласа

$$L(x, 0, \frac{1}{\sqrt{2}}) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|} \tag{5}$$

- Распределение Пуассона

$$P(k, 10) = \frac{10^k}{k!} e^{-10} \tag{6}$$

- Равномерное распределение

$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}} & \text{при} |x| \leq \sqrt{3} \\ 0 & \text{при} |x| > \sqrt{3} \end{cases} \tag{7}$$

2.2 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.

2.3 Выборочные числовые характеристики

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

2.3.1 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

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

2.3.2 Характеристики рассеяния

Выборочная дисперсия

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

3 Программная реализация

Лабораторная работа выполнена на языке Python вресии 3.7 в среде разработки PyCharm. Использовались дополнительные библиотеки:

1. scipy

2. numpy
3. math

В приложении находится ссылка на GitHub репозиторий с исходныи кодом.

4 Результаты

4.1 Характеристики положения и рассеяния

Как было проведено округление:
В оценке $x = E \pm D$ вариации подлежит первая цифра после точки.
В данном случае $x = 0.0 \pm 0.1k$, k - зависит от доверительной вероятности и вида распределения (рассматривается в дальнейшем цикле лабораторных работ).
Округление сделано для $k = 1$.

Characteristic	Mean	Median	z_R	z_Q	z_{tr}
Laplace E(z) 10	-0.0009839	-0.0121517	-0.0197247	-0.0076936	-0.0039645
Laplace D(z) 10	0.0978195	0.0683159	0.5134668	0.4949383	0.1625452
E(z) $\pm\sqrt{D(z)}$	0.0	0.0	0.0	0.0	0.0
Laplace E(z) 100	0.0034648	0.0033212	0.0245185	0.0027098	0.0056071
Laplace D(z) 100	0.0104284	0.0062275	0.5568239	0.4740201	0.0208539
E(z) $\pm\sqrt{D(z)}$	0.0	0.0	0.0	0.0	0.0
Laplace E(z) 1000	-0.0010085	-0.0015766	-0.0068676	-0.0020746	0.0001261
Laplace D(z) 1000	0.0011195	0.0005789	0.4895393	0.5319597	0.0020667
E(z) $\pm\sqrt{D(z)}$	0.0	0.0	0.0	0.0	0.0

Таблица 1: Распределение Лапласа (5)

Characteristic	Mean	Median	z_R	z_Q	z_{tr}
Uniform E(z) 10	0.000659	-0.0037344	0.0087508	0.0014502	-0.0009534
Uniform D(z) 10	0.1036965	0.2388555	0.4878255	0.4958455	0.1611513
E(z) $\pm\sqrt{D(z)}$	0.0	0.0	0.0	0.0	0.0
Uniform E(z) 100	0.003774	0.0071305	0.0032768	0.0214441	-0.0007186
Uniform D(z) 100	0.009722	0.0292367	0.4905364	0.5358395	0.0204128
E(z) $\pm\sqrt{D(z)}$	0.0	0.0	0.0	0.0	0.0
Uniform E(z) 1000	0.0004006	0.0011303	0.0075157	0.0083656	-0.0002115
Uniform D(z) 1000	0.0009215	0.0026702	0.5022606	0.4949788	0.0018109
E(z) $\pm\sqrt{D(z)}$	0.0	0.0	0.0	0.0	0.0

Таблица 2: Равномерное распределение (7)

Characteristic	Mean	Median	z_R	z_Q	z_{tr}
Normal E(z) 10	0.0048043	0.0035738	-0.0158131	0.029832	0.0076417
Normal D(z) 10	0.093454	0.0806334	0.4759391	0.5494387	0.1552688
E(z) $\pm\sqrt{D(z)}$	0.0	0.0	0.0	0.0	0.0
Normal E(z) 100	0.0014876	-0.004149	-0.0337122	0.0379003	0.0061809
Normal D(z) 100	0.0095173	0.0093478	0.512915	0.4902777	0.0180784
E(z) $\pm\sqrt{D(z)}$	0.0	0.0	0.0	0.0	0.0
Normal E(z) 1000	7.13e-05	-0.0002033	-0.002519	-0.0178714	-0.0012642
Normal D(z) 1000	0.0009921	0.0009216	0.4910093	0.5370306	0.0020388
E(z) $\pm\sqrt{D(z)}$	0.0	0.0	0.0	0.0	0.0

Таблица 3: Нормальное распределение (3)

Characteristic	Mean	Median	z_R	z_Q	z_{tr}
Cauchy E(z) 10	-0.9015205	-0.0102798	1.5506815	2.0303134	-1.3998546
Cauchy D(z) 10	887.5172201	0.3256897	7785.2477732	1676.7961675	1337.2031198
E(z) $\pm\sqrt{D(z)}$	0.0	0.0	0.0	0.0	0.0
Cauchy E(z) 100	-0.4475114	-0.0059119	-0.4265913	2.0329197	0.5847918
Cauchy D(z) 100	151.9747356	0.025389	550.5545438	1026.1692437	423.7994764
E(z) $\pm\sqrt{D(z)}$	0.0	0.0	0.0	0.0	0.0
Cauchy E(z) 1000	0.0604925	0.0014351	0.1957455	-0.3087999	-0.5480516
Cauchy D(z) 1000	1063.4570552	0.0023946	384.7296859	139.1980079	3363.2160612
E(z) $\pm\sqrt{D(z)}$	0.0	0.0	0.0	0.0	0.0

Таблица 4: Распределение Коши (4)

Characteristic	Mean	Median	z_R	z_Q	z_{tr}
Poisson E(z) 10	10.0151	9.8985	9.934	10.0265	10.0538333
Poisson D(z) 10	0.996602	1.3839478	5.290144	4.8850478	1.5409631
E(z) $\pm\sqrt{D(z)}$	0.0	0.0	0.0	0.0	0.0
Poisson E(z) 100	9.98215	9.82	10.113	10.094	9.9692
Poisson D(z) 100	0.1006763	0.1961	5.311731	4.826664	0.1995538
E(z) $\pm\sqrt{D(z)}$	0.0	0.0	0.0	0.0	0.0
Poisson E(z) 1000	9.995908	9.995	9.952	10.0135	9.99356
Poisson D(z) 1000	0.009907	0.004475	4.938196	4.3005678	0.0205054
E(z) $\pm\sqrt{D(z)}$	0.0	0.0	0.0	0.0	0.0

Таблица 5: Распределение Пуассона (6)

5 Обсуждение

Из полученных нами данных сильно выделяется распределение Коши. Так, даже для больших выборок, дисперсия принимает огромные значения. Кроме того, нет какой-то очевидной закономерности между увеличением выборки и изменением значения дисперсии: у mean дисперсия от выборки из 10 к 100 падает, от 100 к 1000 растет, у z_R все время убывает. Данные аномалии являются результатами выбросов, которые наблюдались в распределении Коши еще в первой лабораторной.

6 Приложение

Код программы GitHub URL:

https://github.com/PopeyeTheSailorsCat/math_stat_2021/blob/main/lab2/src/lab2.py