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

Отчёт по лабораторной работе №2

по дисциплине
«Математическая статистика»

Выполнила студент
группы 5030102/90101 Кузин Иван Никитович

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

Санкт-Петербург
2022

СОДЕРЖАНИЕ

СПИСОК ТАБЛИЦ

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: $x_{(1)}, x_{(2)}, \ldots,$ 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 $x_{(1)}, x_{(2)}, \dots, x_{(n)}$.

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 JupyterLab. Использовались дополнительные библиотеки:

1. scipy
2. numpy

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}
Normal E(z) 10	-0.017468	-0.019852	-0.025488	0.300647	0.260236
Normal D(z) 10	0.099305	0.145161	0.182558	0.119238	0.11468
E(z) $\pm\sqrt{D(z)}$	[0.332595; 0.297659]	[0.400852; 0.361148]	[0.452756; 0.40178]	[0.044662; 0.645956]	[0.078408; 0.59888]
$\widehat{E}(z)$	0	0	0	0	0
Normal E(z) 100	-0.000866	0.002153	-0.004746	0.013516	0.027086
Normal D(z) 100	0.010386	0.01633	0.085266	0.012678	0.012396
E(z) $\pm\sqrt{D(z)}$	[0.102778; 0.101046]	[0.125636; 0.129942]	[0.296749; 0.287257]	[0.099081; 0.126113]	[0.084251; 0.138423]
$\widehat{E}(z)$	0.0	0.0	0.0	0.0	0.0
Normal E(z) 1000	0.000153	0.000089	0.007519	0.001509	0.002836
Normal D(z) 1000	0.000981	0.001642	0.059635	0.001228	0.001195
E(z) $\pm\sqrt{D(z)}$	[0.031168; 0.031474]	[0.040433; 0.040611]	[0.236684; 0.251722]	[0.033534; 0.036552]	[0.031733; 0.037405]
$\widehat{E}(z)$	0.00	0.00	0.0	0.00	0.00

Таблица 1: Нормальное распределение (3)

Characteristic	Mean	Median	z_R	z_Q	z_{tr}
Cauchy E(z) 10	0.562654	-0.006675	2.898445	1.104381	0.671502
Cauchy D(z) 10	523.363485	0.295683	12824.001857	4.591925	1.076515
E(z) $\pm\sqrt{D(z)}$	[22.314485; 23.439793]	[0.550442; 0.537092]	[110.344665; 116.141555]	[1.038497; 3.247259]	[0.36605; 1.709054]
$\widehat{E}(z)$	-	0	-	-	-
Cauchy E(z) 100	-0.7079	-0.001573	-34.396555	0.03319	0.038699
Cauchy D(z) 100	398.608006	0.027015	945555.067	0.057985	0.027952
E(z) $\pm\sqrt{D(z)}$	[20.67307; 19.25727]	[0.165935; 0.162789]	[1006.793114; 938.000004]	[0.207611; 0.273991]	[0.12849; 0.205888]
$\widehat{E}(z)$	-	0	-	0	0
Cauchy E(z) 1000	-0.510318	0.000322	-260.66918	0.004265	0.003941
Cauchy D(z) 1000	290.388668	0.002172	69909833.801229	0.00484	0.002387
E(z) $\pm\sqrt{D(z)}$	[17.551112; 16.530476]	[0.046283; 0.046927]	[8621.879248; 8100.540888]	[0.065305; 0.073835]	[0.044916; 0.052798]
$\widehat{E}(z)$	-	0.0	-	0.0	0.0

Таблица 2: Распределение Коши (4)

Characteristic	Mean	Median	z_R	z_Q	z_{tr}
Laplace E(z) 10	-0.006405	-0.006188	-0.005512	0.293618	0.229066
Laplace D(z) 10	0.010247	0.006208	0.425481	0.010082	0.006497
E(z) $\pm\sqrt{D(z)}$	[0.031154; 0.030002]	[0.022723; 0.021909]	[0.651449; 0.626013]	[0.029889; 0.032657]	[0.022585; 0.025539]
$\widehat{E}(z)$	0	0	0	0	0
Laplace E(z) 100	0.002803	0.001905	0.023725	0.017159	0.021071
Laplace D(z) 100	0.057048	0.041891	0.492694	0.492542	0.095815
E(z) $\pm\sqrt{D(z)}$	[-0.226576; 0.251118]	[-0.199293; 0.210051]	[-0.680069; 0.723773]	[-0.679985; 0.723641]	[-0.294498; 0.324580]
$\widehat{E}(z)$	0.0	0.0	0	0	0.0
Laplace E(z) 1000	-0.000576	-0.000407	-0.012718	0.001384	0.001477
Laplace D(z) 1000	0.000935	0.000498	0.407977	0.000978	0.000579
E(z) $\pm\sqrt{D(z)}$	[0.022723; 0.021909]	[0.651449; 0.626013]	[0.029889; 0.032657]	[0.022585; 0.025539]	[0.022652; 0.023563]
$\widehat{E}(z)$	0.00	0.00	0	0.00	0.00

Таблица 3: Распределение Лапласа (5)

Characteristic	Mean	Median	z_R	z_Q	z_{tr}
Poisson E(z) 10	10.0015	9.87	10.265	10.9455	10.786167
Poisson D(z) 10	1.101808	1.5811	1.936775	1.46278	1.362525
E(z) $\pm\sqrt{D(z)}$	[0.031154; 0.030002]	[0.022723; 0.021909]	[0.651449; 0.626013]	[0.029889; 0.032657]	[0.022585; 0.025539]
$\widehat{E}(z)$	10^{+1}_{-1}	10^{+1}_{-1}	10^{+2}_{-2}	10^{+2}_{-2}	10^{+1}_{-1}
Poisson E(z) 100	9.9932	9.844	10.9475	9.959	9.93526
Poisson D(z) 100	0.10332	0.204664	0.997494	0.159819	0.122566
E(z) $\pm\sqrt{D(z)}$	[9.671766; 10.314634]	[9.391602; 10.296398]	[9.948754; 11.946246]	[9.559226; 10.358774]	[9.585166; 10.285354]
$\widehat{E}(z)$	10^{+1}_{-1}	10^{+1}_{-1}	10^{+2}_{-2}	10^{+2}_{-2}	10^{+1}_{-1}
Poisson E(z) 1000	10.000822	9.997	11.671	9.9955	9.86806
Poisson D(z) 1000	0.00993	0.002991	0.746759	0.00223	0.01131
E(z) $\pm\sqrt{D(z)}$	[9.901173; 10.100471]	[9.94231; 10.05169]	[10.806848; 12.535152]	[9.948277; 10.042723]	[9.761712; 9.974408]
$\widehat{E}(z)$	10^{+1}_{-1}	10^{+1}_{-1}	10^{+2}_{-2}	10^{+2}_{-2}	10^{+1}_{-1}

Таблица 4: Распределение Пуассона (6)

Characteristic	Mean	Median	z_R	z_Q	z_{tr}
Uniform E(z) 10	-0.007779	-0.011541	-0.00537	0.316201	0.304339
Uniform D(z) 10	0.104723	0.233967	0.047982	0.130148	0.159694
E(z) $\pm\sqrt{D(z)}$	[0.331388; 0.31583]	[0.495242; 0.47216]	[0.224418; 0.213678]	[0.044559; 0.676961]	[0.095278; 0.703956]
$\widehat{E}(z)$	0	0	0.0	0	0
Uniform E(z) 100	-0.003973	-0.004643	0.001269	0.012524	0.02923
Uniform D(z) 100	0.009763	0.029286	0.000551	0.014414	0.019828
E(z) $\pm\sqrt{D(z)}$	[0.102781; 0.094835]	[0.175775; 0.166489]	[0.022204; 0.024742]	[0.107534; 0.132582]	[0.111582; 0.170042]
$\widehat{E}(z)$	0.00	0.0	0.00	0.0	0.0
Uniform E(z) 1000	-0.001949	-0.002912	- 0.00696	-0.000109	0.000962
Uniform D(z) 1000	0.001006	0.00297	0.00006	0.001499	0.001982
E(z) $\pm\sqrt{D(z)}$	[0.033667; 0.029769]	[0.05741; 0.051586]	[0.002545; 0.002353]	[0.038826; 0.038608]	[0.043558; 0.045482]
$\widehat{E}(z)$	0.00	0.00	0.0	0.00	0.0

Таблица 5: Равномерное распределение (7)

5 Обсуждение

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 Приложение

Код программы GitHub URL:

https://github.com/workivan/mat-ver-stat