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»

Выполнила студентка
группы 3630102/80401 Мамаева Анастасия Сергеевна

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

Санкт-Петербург
2021

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

2.3.1 Характеристики положения

- Выборочное среднее

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

- Выборочная медиана

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

- 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	-0.000003	-0.008031	0.036451	-0.021004	-0.012014
Normal D(z) 10	0.101239	0.091199	0.480519	0.475769	0.172801
E(z) $\pm\sqrt{D(z)}$	[-0.318184; 0.318178]	[-0.310023; 0.293961]	[-0.656743; 0.729646]	[-0.710764; 0.668756]	[-0.427707; 0.403679]
$\widehat{E}(z)$	0.0	0.0	0.0	0.0	0.0
Normal E(z) 100	-0.002052	-0.004684	0.012205	-0.011495	-0.008917
Normal D(z) 100	0.05601	0.050246	0.478019	0.50419	0.097187
E(z) $\pm\sqrt{D(z)}$	[-0.238716; 0.234612]	[-0.228840; 0.219472]	[-0.679184; 0.703594]	[-0.721558; 0.698568]	[-0.320665; 0.302831]
$\widehat{E}(z)$	0.0	0.0	0.0	0.0	0.0
Normal E(z) 1000	-0.001219	-0.003105	0.006667	-0.010076	-0.006475
Normal D(z) 1000	0.037681	0.033809	0.475955	0.497166	0.065463
E(z) $\pm\sqrt{D(z)}$	[-0.195335; 0.192897]	[-0.186977 0.180767]	[-0.683228; 0.696562]	[-0.715176; 0.695024]	[-0.262332; 0.249382]
$\widehat{E}(z)$	0.0	0.0	0.0	0.0	0.0

Таблица 1: Нормальное распределение (3)

Characteristic	Mean	Median	z_R	z_Q	z_{tr}
Cauchy E(z) 10	-1.787828	-0.025951	-5.130831	-3.948137	-2.645886
Cauchy D(z) 10	4390.971	0.403047	71558.563	21478.02	2724.385
E(z) $\pm\sqrt{D(z)}$	[-68.0522; 64.4766]	[-0.660811; 0.608909]	[-272.635; 262.373]	[-150.502 142.606]	[-54.8415; 49.5498]
$\widehat{E}(z)$	-	0	-	-	-
Cauchy E(z) 100	-2.946291	-0.011733	-2.919987	7.274195	-5.064231
Cauchy D(z) 100	11499.485	0.214055	37287.433	194370.44	37068.767
E(z) $\pm\sqrt{D(z)}$	[-110.182; 104.289]	[-0.474394; 0.450928]	[-196.019; 190.179]	[-433.600; 448.148]	[-197.596; 187.468]
$\widehat{E}(z)$	-	0	-	-	-
Cauchy E(z) 1000	-2.272925	-0.007192	-1.740728	4.70099	-4.116761
Cauchy D(z) 1000	7794.082	0.143556	25043.95	129732.73	25183.86
E(z) $\pm\sqrt{D(z)}$	[-90.5570; 86.0111]	[-0.386079; 0.371695]	[-159.993; 156.512]	[-355.483; 364.885]	[-162.810; 154.577]
$\widehat{E}(z)$	-	0	-	-	-

Таблица 2: Распределение Коши (4)

Characteristic	Mean	Median	z_R	z_Q	z_{tr}
Laplace E(z) 10	0.022501	0.011001	0.011483	0.023003	0.027397
Laplace D(z) 10	0.103582	0.077859	0.490929	0.47279	0.169248
E(z) $\pm\sqrt{D(z)}$	[-0.299340; 0.344342]	[-0.268031; 0.290033]	[-0.689180; 0.712146]	[-0.664594; 0.710600]	[-0.384000 0.438794]
$\widehat{E}(z)$	0	0	0	0	0
Laplace E(z) 100	0.012271	0.005379	0.021852	0.021828	0.015041
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
Laplace E(z) 1000	0.008403	0.003924	0.022277	0.022505	0.010596
Laplace D(z) 1000	0.038374	0.028103	0.487406	0.505289	0.064555
E(z) $\pm\sqrt{D(z)}$	[-0.187489; 0.204295]	[-0.163715; 0.171563]	[-0.675867; 0.720421]	[-0.688331; 0.733341]	[-0.243480; 0.264672]
$\widehat{E}(z)$	0	0	0	0	0

Таблица 3: Распределение Лапласа (5)

Characteristic	Mean	Median	z_R	z_Q	z_{tr}
Poisson E(z) 10	9.979	9.8525	9.9655	10.044	9.997167
Poisson D(z) 10	0.975799	1.373994	4.68356	5.018564	1.779575
E(z) $\pm\sqrt{D(z)}$	[8.99117; 10.9668]	[8.68032; 11.0246]	[7.80134; 12.1296]	[7.80378; 12.2842]	[8.66315; 11.3311]
$\widehat{E}(z)$	10^{+1}_{-1}	10^{+1}_{-1}	10^{+2}_{-2}	10^{+2}_{-2}	10^{+1}_{-1}
Poisson E(z) 100	9.98473	9.8535	9.9615	10.0425	9.986043
Poisson D(z) 100	0.534271	0.786788	4.841018	5.073944	0.982735
E(z) $\pm\sqrt{D(z)}$	[9.25379; 10.7156]	[8.96648; 10.7405]	[7.76126; 12.1617]	[7.78995; 12.2950]	[8.99471; 10.9773]
$\widehat{E}(z)$	10^{+1}_{-1}	10^{+1}_{-1}	10^{+2}_{-2}	10^{+2}_{-2}	10^{+1}_{-1}
Poisson E(z) 1000	9.990781	9.902	9.955	10.057	9.99035
Poisson D(z) 1000	0.359408	0.529563	4.769308	5.025751	0.661672
E(z) $\pm\sqrt{D(z)}$	[9.39127; 10.5902]	[9.17428; 10.6297]	[7.77112; 12.1388]	[7.81518; 12.2988]	[9.17691; 10.8037]
$\widehat{E}(z)$	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	-0.003058	-0.008189	-0.008308	-0.058391	-0.000635
Uniform D(z) 10	0.10312	0.240584	0.509498	0.516179	0.1717448
E(z) $\pm\sqrt{D(z)}$	[-0.324181; 0.318065]	[-0.498682; 0.482304]	[-0.722099; 0.705483]	[-0.776846; 0.660064]	[-0.415056; 0.413786]
$\widehat{E}(z)$	0.0	0.0	0.0	0.0	0.0
Uniform E(z) 100	-0.001261	-0.005273	-0.013241	-0.026719	-0.001152
Uniform D(z) 100	0.056234	0.134553	0.506669	0.513444	0.095636
E(z) $\pm\sqrt{D(z)}$	[-0.238398; 0.235876]	[-0.372087; 0.361541]	[-0.725047; 0.698565]	[-0.743269; 0.689831]	[-0.310402; 0.308098]
$\widehat{E}(z)$	0.0	0.0	0.0	0.0	0.0
Uniform E(z) 1000	-0.000621	-0.003033	-0.005643	-0.017729	-0.000709
Uniform D(z) 1000	0.037829	0.090714	0.499291	0.501763	0.064392
E(z) $\pm\sqrt{D(z)}$	[-0.195117; 0.193875]	[-0.304220; 0.298154]	[-0.712248; 0.700962]	[-0.726081; 0.690623]	[-0.254464; 0.253046]
$\widehat{E}(z)$	0.0	0.0	0.0	0.0	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/Brightest-Sunshine/Math-Statistic-2021/blob/main/Lab2/Lab2.ipynb