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»

Выполнил студент
группы 5030102/90101 Лаэтин Андрей Алексеевич

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

Санкт-Петербург
2022

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 **9**

6 Приложение **9**

Список таблиц

1 Распределение Лапласа (5) 6
2 Равномерное распределение (7) 7
3 Нормальное распределение (3) 7
4 Распределение Коши (4) 8
5 Распределение Пуассона (6) 8

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 3.9 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 $x = \hat{E}$ вариации подлежит разные цифры после точки, в зависимости от распределения. Например в случае распределения Коши(4) вариации подлежат все цифры, так что ни одна не валидна.

Characteristic	Mean	Median	z_R	z_Q	z_{tr}
Laplace E(z) 10	0.0122645	-0.0081925	0.013314	0.0485472	0.0284073
Laplace D(z) 10	0.0983683	0.0709627	0.4870477	0.5423441	0.1587122
E(z) $\pm\sqrt{D(z)}$	[-0.3013727 ; 0.3259017]	[-0.2735248 ; 0.2581958]	[-0.2745808 ; 0.711202]	[-0.684574 ; 0.7849874]	[-0.3699797 ; 0.4267943]
$\hat{E}$(z)	0.0	0.0	0.0	0.0	0.0
Laplace E(z) 100	0.0073201	0.0056240	0.0477397	0.0693713	0.0043285
Laplace D(z) 100	0.0094611	0.005750	0.4795924	0.4815146	0.0194463
E(z) $\pm\sqrt{D(z)}$	[-0.0899481 ; 0.1045883]	[-0.0702048 ; 0.0814528]	[-0.6447864 ; 0.7402658]	[-0.6245412 ; 0.7632838]	[-0.1388017 ; 0.1437785]
$\hat{E}$(z)	0.0	0.0	0.0	0.0	0.0
Laplace E(z) 1000	-0.0000240	-0.0002828	0.0070518	0.0402517	0.0012168
Laplace D(z) 1000	0.0009535	0.0004756	0.4997571	0.5332303	0.0019536
E(z) $\pm\sqrt{D(z)}$	[-0.0308548 ; 0.0309028]	[-0.6998832 ; 0.0215255]	[-0.7065385 ; 0.7139868]	[-0.6899745 ; 0.7704779]	[-0.0454163 ; 0.0429827]
$\hat{E}$(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	-0.004791	-0.00522	-0.005364	0.312004	0.307351
Uniform D(z) 10	0.095645	0.214577	0.043371	0.125257	0.150532
E(z) $\pm\sqrt{D(z)}$	[-0.314056 ; 0.304474]	[-0.468445 ; 0.458005]	[-0.213621 ; 0.202893]	[-0.041913 ; 0.665921]	[-0.080634 ; 0.695336]
$\hat{E}$(z)	0.	0.	0.	0.	0.
Uniform E(z) 100	-0.00507	-0.00721	0.001523	0.009718	0.027495
Uniform D(z) 100	0.009684	0.028672	0.000586	0.015108	0.019185
E(z) $\pm\sqrt{D(z)}$	[-0.103477 ; 0.093337]	[-0.176538 ; 0.162118]	[-0.022684 ; 0.02573]	[-0.113197 ; 0.132633]	[-0.111015 ; 0.166005]
$\hat{E}$(z)	0.	0.	0.	0.	0.
Uniform E(z) 1000	-0.001592	-0.003927	2.4e-05	0.00049	0.000673
Uniform D(z) 1000	0.001021	0.003154	7e-06	0.001535	0.002043
E(z) $\pm\sqrt{D(z)}$	[-0.033545 ; 0.030361]	[-0.060087 ; 0.052233]	[-0.002622 ; 0.00267]	[-0.038689 ; 0.039669]	[-0.044527 ; 0.045873]
$\hat{E}$(z)	0.	0.	0.	0.	0.

Таблица 2: Равномерное распределение (7)

Characteristic	Mean	Median	z_R	z_Q	z_{tr}
Normal E(z) 10	0.002953	0.000333	-0.004257	0.308114	0.27533
Normal D(z) 10	0.096994	0.134541	0.194008	0.116832	0.111414
E(z) $\pm\sqrt{D(z)}$	[-0.308486 ; 0.314392]	[-0.366465 ; 0.367131]	[-0.44472 ; 0.436206]	[-0.033693 ; 0.649921]	[-0.058457 ; 0.609117]
$\hat{E}$(z)	0.0	0.0	0.0	0.0	0.0
Normal E(z) 100	-0.003213	-0.003589	-0.006834	0.012519	0.022774
Normal D(z) 100	0.010051	0.015256	0.088788	0.012047	0.011647
E(z) $\pm\sqrt{D(z)}$	[-0.103468 ; 0.097042]	[-0.127104 ; 0.119926]	[-0.304807 ; , 0.291139]	[-0.09724 ; 0.122278]	[-0.085147 ; 0.130695]
$\hat{E}$(z)	0.0	0.0	0.0	0.0	0.0
Normal E(z) 1000	1.9e-05	0.001198	0.011931	0.00227	0.003135
Normal D(z) 1000	0.000965	0.001516	0.055366	0.001155	0.001121
E(z) $\pm\sqrt{D(z)}$	[-0.031045 ; 0.031083]	[-0.037738 ; 0.040134]	[-0.223369 ; 0.247231]	[-0.031715 ; 0.036255]	[-0.030346 ; 0.036616]
$\hat{E}$(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	-1.118927	-0.03448	-5.195582	0.969381	0.611771
Cauchy D(z) 10	1112.729129	0.326576	27684.070358	2.802831	0.838361
E(z) $\pm\sqrt{D(z)}$	[-34.476522 ;	[-0.605948 ;	[-171.580889 ;	[-0.704785;	[-0.30385 ;
	32.238668]	0.536988]	161.189725]	2.643547]	1.527392]
$\hat{E}$(z)	-	0	-	-	-
Cauchy E(z) 100	-17.265172	-0.0046	-860.016538	0.029137	0.035369
Cauchy D(z) 100	287444.075602	0.026214	718425982.718221	0.05561	0.028469
E(z) $\pm\sqrt{D(z)}$	[-553.403284 ;	[-0.166507 ;	[-27663.48614 ;	[-0.206681 ;	[-0.133359 ;
	518.87294]	0.157307]	25943.453064]	0.264955]	0.204097]
$\hat{E}$(z)	-	0	-	-	-
Cauchy E(z) 1000	-2.01552	-0.000491	-1007.901814	0.000396	0.002662
Cauchy D(z)	2291.416024	0.002407	569808409.909072	0.004912	0.002556
E(z) $\pm\sqrt{D(z)}$	[-49.884257 ;	[-0.049552 ;	[-24878.561836 ;	[-0.06969 ;	[-0.047895 ;
	45.853217]	0.04857]	22862.758208]	10.070482]	0.053219]
$\hat{E}$(z)	-	0	-	-	-

Таблица 4: Распределение Коши (4)

Characteristic	Mean	Median	z_R	z_Q	z_{tr}
Poisson E(z) 10	10.0019	9.854	10.2945	10.943	10.78
Poisson D(z) 10	0.978186	1.510684	1.82902	1.371751	1.259267
E(z) $\pm\sqrt{D(z)}$	[9.012867;	[8.624901;	[8.942087;	[9.771782;	[9.657829;
	10.990933]	11.083099]	11.646913]	12.114218]	11.902171]
Poisson E(z) 100	10.00597	9.86	10.9825	9.97	9.94826
Poisson D(z) 100	0.101249	0.1969	0.906944	0.1591	0.119652
E(z) $\pm\sqrt{D(z)}$	[9.687774;	[9.416266;	[10.030164;	[9.571127;	[9.602352;
	10.324166]	10.303734]	11.934836]	10.368873]	10.294168]
$\hat{E}$(z)	10^{+0}_{-0}	10^{+0}_{-1}	10^{+2}_{-2}	10^{+2}_{-2}	10^{+0}_{-0}
Poisson E(z) 1000	10.004157	9.995	11.6795	9.993	9.871028
Poisson D(z) 1000	0.009747	0.004975	0.68603	0.003451	0.010302
E(z) $\pm\sqrt{D(z)}$	[9.90543;	[9.924466;	[10.851231;	[9.934255;	[9.769529;
	10.102884]	10.065534]	12.507769]	10.051745]	9.972527]
$\hat{E}$(z)	10^{+0}_{-0}	10^{+0}_{-0}	10^{+2}_{-2}	10^{+2}_{-2}	10^{+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/A21l63/math-prob-stat