

Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики
Кафедра «Прикладная математика»

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №2

по дисциплине
"Математическая статистика"

Выполнил студент
Группы 3630102/80101

Шао Цзяци

Проверил
доцент, к.ф.-м.н.

Баженов Александр Николаевич

Санкт-Петербург
2021 г.

Содержание

1. Постановка задачи	4
2. Теория	4
2.1. Рассматриваемые 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. Реализация	5
4. Результаты	6
5. Обсуждение	8

Список таблиц

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

1. Постановка задачи

Сгенерировать выборки размером 10, 100, 1000 элементов.

Для каждой выборки вычислить следующие статистические характеристики положения данных: \bar{x} , $medx$, z_R , z_Q , z_{tr} . Повторить такие вычисления 1000 раз для каждой выборки и найти среднее характеристики положения и их квадратов.

$$E(z) = \bar{z} \quad (1)$$

Вычислить оценку дисперсии по формуле

$$D(z) = \overline{z^2} - \bar{z}^2 \quad (2)$$

Представить полученные данные в виде таблиц.

2. Теория

2.1. Рассматриваемые распределения

Плотности распределений

- Нормальное распределение

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

- Распределение Коши

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

- Распределение Лапласа

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

- Распределение Пуассона

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

- Равномерное распределение

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

2.2. Вариационный ряд

последовательность элементов выборки X_1, X_2, \dots, X_n , полученная в результате расположения в порядке не убывания исходной последовательности независимых одинаково распределенных случайных величин.

2.3. Выборочные числовые характеристики

2.3.1. Характеристики положения

- Выборочное среднее

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (8)$$

- Выборочная медиана

$$medx = \begin{cases} x_{l+1}, n = 2l + 1 \\ \frac{x_l + x_{l+1}}{2}, n = 2l \end{cases} \quad (9)$$

- Полусумма экстремальных выборочных элементов

$$z_R = \frac{x_1 + x_n}{2} \quad (10)$$

- Полусумма квартилей Выборочная квартиль z_p порядка p определяется формулой

$$z_p = \begin{cases} x_{[np]+1}, np - \text{дробь} \\ x_{np}, np - \text{целое.} \end{cases} \quad (11)$$

Полусумма квартилей

$$z_Q = \frac{z_{1/4} + z_{3/4}}{2} \quad (12)$$

- Усеченное среднее

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

2.3.2. Характеристики рассеяния

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

3. Реализация

Лабораторная работа выполнена с помощью встроенных средств языка программирования python в среде разработки Pycharm с дополнительными библиотеками.

- scipy
- numpy
- matplotlib
- math

Исходный код лабораторной работы размещен в Github-репозитории.

URL: <https://github.com/ShaoTs/shaoMathStatistic/tree/master/Lab2>

4. Результаты

Characteristic	Mean	Median	Z_R	Z_Q	Z_{tr}
$E(z): n=10$	-0.001818	-0.002782	-0.015130	0.319813	0.280922
$D(z): n=10$	0.104987	0.140621	0.186021	0.128988	0.118588
$E(z) \pm \sqrt{D(z)}$	[-0.325,0.322]	[-0.377,0.372]	[-0.446,0.416]	[-0.039,0.678]	[-0.063,0.625]
$\widehat{E}(z)$	0	0	0	0	0
$E(z): n=100$	0.002780	0.008094	0.005101	0.165416	0.154575
$D(z): n=100$	0.052122	0.072270	0.129649	0.088134	0.076938
$E(z) \pm \sqrt{D(z)}$	[-0.225,0.231]	[-0.260,0.276]	[-0.354,0.365]	[-0.131,0.462]	[-0.122,0.431]
$\widehat{E}(z)$	0	0	0	0	0
$E(z): n=1000$	0.001324	0.005283	0.000666	0.110445	0.103572
$D(z): n=1000$	0.035100	0.048756	0.107580	0.065216	0.056914
$E(z) \pm \sqrt{D(z)}$	[-0.186,0.188]	[-0.215,0.226]	[-0.327,0.326]	[-0.144,0.365]	[-0.134,0.342]
$\widehat{E}(z)$	0	0	0	0	0

Таблица 1. Нормальное распределение(3)

Characteristic	Mean	Median	Z_R	Z_Q	Z_{tr}
$E(z): n=10$	0.848992	-0.007079	4.446280	1.029138	0.643950
$D(z): n=10$	3112	0.303237	77478	3.171253	0.914280
$E(z) \pm \sqrt{D(z)}$	[-54.941,56.639]	[-0.558,0.544]	[-273.903,282.796]	[-0.752,2.810]	[-0.312,1.600]
$\widehat{E}(z)$	-	0	-	-	-
$E(z): n=100$	-1.077177	-0.001224	20.221538	0.617391	0.375632
$D(z): n=100$	15212	0.169123	2368134	3.762948	0.778370
$E(z) \pm \sqrt{D(z)}$	[-124,122]	[-0.412,0.410]	[-1518,1559]	[-1.322,2.557]	[-0.507,1.258]
$\widehat{E}(z)$	-	0	-	-	-
$E(z): n=1000$	-6.795911	-0.000349	-3026.927570	0.412271	0.251891
$D(z): n=1000$	120613	0.113600	27619147692	2.594488	0.550416
$E(z) \pm \sqrt{D(z)}$	[-354.091,340.499]	[-0.337,0.337]	[-169217,163163]	[-1.198,2.023]	[-0.490,0.994]
$\widehat{E}(z)$	-	0	-	-	0

Таблица 2. Распределение Коши(4)

Characteristic	Mean	Median	Z_R	Z_Q	Z_{tr}
E(z): n=10	0.017573	0.001739	0.059028	0.303693	0.237074
D(z): n=10	0.101685	0.068897	0.403718	0.126860	0.085271
$E(z) \pm \sqrt{D(z)}$	[-0.301,0.336]	[-0.261,0.264]	[-0.576,0.694]	[-0.052,0.660]	[-0.055,0.529]
$\widehat{E}(z)$	0	0	0	0	0
E(z): n=100	0.011739	0.002463	0.041044	0.160417	0.129883
D(z): n=100	0.056120	0.037498	0.403772	0.088973	0.057372
$E(z) \pm \sqrt{D(z)}$	[-0.225,0.249]	[-0.191,0.196]	[-0.594,0.676]	[-0.138,0.459]	[-0.110,0.369]
$\widehat{E}(z)$	0	0	0	0	0
E(z): n=1000	0.007945	0.001944	0.023661	0.107283	0.087262
D(z): n=1000	0.037755	0.025178	0.406219	0.065280	0.042072
$E(z) \pm \sqrt{D(z)}$	[-0.186,0.202]	[-0.157,0.161]	[-0.614,0.661]	[-0.148,0.363]	[-0.118,0.292]
$\widehat{E}(z)$	0	0	0	0	0

Таблица 3. Распределение Лапласа(5)

Characteristic	Mean	Median	Z_R	Z_Q	Z_{tr}
E(z): n=10	9.952100	9.798000	10.244500	10.881000	10.732000
D(z): n=10	0.965716	1.394696	1.840970	1.381339	1.239787
$E(z) \pm \sqrt{D(z)}$	[8.969,10.935]	[8.617,10.979]	[8.888,11.601]	[9.706,12.056]	[9.619,11.845]
$\widehat{E}(z)$	-	-	-	-	-
E(z): n=100	9.980945	9.825750	10.578750	10.421750	10.340170
D(z): n=100	0.530840	0.792262	1.493673	0.975752	0.828429
$E(z) \pm \sqrt{D(z)}$	[9.252,10.710]	[8.936,10.716]	[9.357,11.801]	[9.434,11.410]	[9.430,11.250]
$\widehat{E}(z)$	-	-	-	-	-
E(z): n=1000	9.990116	9.883667	10.948667	10.280000	10.184519
D(z): n=1000	0.357495	0.534967	1.492698	0.691767	0.604452
$E(z) \pm \sqrt{D(z)}$	[9.392,10.588]	[9.152,10.615]	[9.727,12.170]	[9.448,11.112]	[9.407,10.962]
$\widehat{E}(z)$	-	-	-	-	-

Таблица 4. Распределение Пуассона(6)

Characteristic	Mean	Median	Z_R	Z_Q	Z_{tr}
E(z): n=10	-0.009888	-0.023906	-0.005819	0.302109	0.304990
D(z): n=10	0.090412	0.208104	0.044468	0.119841	0.138929
$E(z) \pm \sqrt{D(z)}$	[-0.311,0.291]	[-0.480,0.432]	[-0.217,0.205]	[-0.044,0.648]	[-0.068,0.678]
$\widehat{E}(z)$	0	0	0	0	0
E(z): n=100	-0.005393	-0.012663	-0.002866	0.159118	0.169319
D(z): n=100	0.050531	0.119969	0.022533	0.087968	0.098547
$E(z) \pm \sqrt{D(z)}$	[-0.230,0.219]	[-0.359,0.334]	[-0.153,0.147]	[-0.137,0.456]	[-0.145,0.483]
$\widehat{E}(z)$	0	0	0	0	0
E(z): n=1000	-0.003590	-0.008291	-0.001903	0.106701	0.113998
D(z): n=1000	0.034037	0.081034	0.015026	0.064659	0.072505
$E(z) \pm \sqrt{D(z)}$	[-0.188,0.181]	[-0.293,0.276]	[-0.124,0.121]	[-0.148,0.361]	[-0.155,0.383]
$\widehat{E}(z)$	0	0	0	0	0

Таблица 5. Равномерное распределение(7)

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.

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.