Санкт-Петербургский Политехнический Университет им. Петра Великого

Институт прикладной математики и механики

Кафедра прикладной математики

Лабораторная работа №2 Характеристики положения выборки

Студент группы 3630102/70301 Камянский Д.В.

Преподаватель Баженов А. Н.

Санкт-Петербург
2020 г.

Содержание

1 Список таблиц

2 Постановка задачи

Любыми средствами сгенерировать выборки размеров 10, 100, 1000 элементов для 5ти распределений. Для каждой выборки вычислить $\overline{x}$, $med\, x$, Z_R, Z_Q, Z_{tr}, при $r = \frac{n}{4}$.

Распределения:

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

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

$$L\left(x, 0, \frac{1}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|} \tag{3}$$

$$P(5, k) = \frac{5^k}{k!} e^{-5} \tag{4}$$

$$M(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}} & |x| \leqslant \sqrt{3} \\ 0 & |x| > \sqrt{3} \end{cases} \tag{5}$$

3 Теория

1. Выборочное среднее:

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{6}$$

2. Выборочная медиана:

$$med\, x = \begin{cases} x_{k+1}, & n = 2k + 1 \\ \frac{1}{2}\left(x_k + x_{k+1}\right), & n = 2k \end{cases} \tag{7}$$

3. Полусумма экстремальных значений:

$$Z_R = \frac{1}{2}\left(x_1 + x_n\right) \tag{8}$$

4. Полусумма квартилей:

$$Z_Q = \frac{1}{2}\left(Z_{\frac{1}{4}} + Z_{\frac{3}{4}}\right) \tag{9}$$

5. Усечённое среднее:

$$Z_{tr} = \frac{1}{n - 2r} \sum_{i=r+1}^{n-r} x_i \tag{10}$$

После вычисления характеристик положения 1000 раз находится среднее значение и дисперсия:

$$E(z) = \frac{1}{n} \sum_{i=1}^{n} z_i \tag{11}$$

$$D(z) = E\left(z^2\right) - E^2(z) \tag{12}$$

4 Реализация

Для генерации выборки был использован *Python* 3.8.2: модуль *stats* библиотеки *scipy* для генерации выборок различных распределений.Характеристики положения были вычислены с помощью библиотеки *numpy*.

5 Результаты

Таблица 1: Стандартное нормальное распределение.

$n = 10$	average	med	Z_R	Z_Q	Z_{tr}
$E =$	-0.0	-0.02	0.0	-0.0	-0.0
$D =$	0.106931	0.145363	0.186191	0.111027	0.165807
$n = 100$	average	med	Z_R	Z_Q	Z_{tr}
$E =$	-0.0	0.01	0.02	0.0	0.0
$D =$	0.010365	0.015172	0.092881	0.011974	0.017426
$n = 1000$	average	med	Z_R	Z_Q	Z_{tr}
$E =$	0.0	0.0	-0.0	0.0	0.0
$D =$	0.00096	0.001594	0.060543	0.001256	0.001888

Таблица 2: Стандартное распределение Лапласа.

$n = 10$	average	med	Z_R	Z_Q	Z_{tr}
$E =$	-0.01	0.01	-0.0	-0.0	-0.01
$D =$	0.097868	0.069859	0.390201	0.093469	0.167235
$n = 100$	average	med	Z_R	Z_Q	Z_{tr}
$E =$	-0.0	0.01	0.01	0.0	-0.0
$D =$	0.009778	0.005878	0.429639	0.009072	0.020247
$n = 1000$	average	med	Z_R	Z_Q	Z_{tr}
$E =$	0.0	0.0	-0.02	-0.0	0.0
$D =$	0.001009	0.000531	0.428975	0.000997	0.001992

Таблица 3: Распределение Коши.

$n = 10$	average	med	Z_R	Z_Q	Z_{tr}
$E =$	0.31	-0.01	-5.87	0.03	3.5
$D =$	260.978036	0.347366	26462.674037	0.774488	6529.2454
$n = 100$	average	med	Z_R	Z_Q	Z_{tr}
$E =$	2.26	0.01	22.79	-0.0	1.71
$D =$	2122.106445	0.023531	1741103.211025	0.05289	1778.363557
$n = 1000$	average	med	Z_R	Z_Q	Z_{tr}
$E =$	1.14	0.0	-1845.52	-0.0	-1.64
$D =$	706.431432	0.002497	21117320297.542297	0.004864	1675.403679

Таблица 4: Равномерное распределение.

$n = 10$	average	med	Z_R	Z_Q	Z_{tr}
$E =$	−0.02	0.0	0.0	−0.0	−0.02
$D =$	0.095612	0.227685	0.043735	0.135551	0.171826
$n = 100$	average	med	Z_R	Z_Q	Z_{tr}
$E =$	0.0	−0.0	−0.0	0.0	0.01
$D =$	0.009971	0.029194	0.00058	0.014725	0.019723
$n = 1000$	average	med	Z_R	Z_Q	Z_{tr}
$E =$	−0.0	0.0	−0.0	0.0	−0.0
$D =$	0.000959	0.002864	$6e-06$	0.001502	0.00205

Таблица 5: Распределение Пуассона.

$n = 10$	average	med	Z_R	Z_Q	Z_{tr}
$E =$	5.01	4.84	5.3	4.88	5.02
$D =$	0.515302	0.739579	0.942244	0.586539	0.832244
$n = 100$	average	med	Z_R	Z_Q	Z_{tr}
$E =$	5.0	4.9	5.97	4.89	4.99
$D =$	0.049874	0.102596	0.499128	0.143341	0.101492
$n = 1000$	average	med	Z_R	Z_Q	Z_{tr}
$E =$	5.0	5.0	6.83	4.67	5.0
$D =$	0.005329	0.0	0.341759	0.076798	0.010249

6 Обсуждение

При вычислении средних значений приходится отбрасывать некоторое число знаков после запятой, так как получаемая дисперсия не может гарантировать получаемое точное значение.

Другими словами дисперсия может гарантировать порядок точности среднего значения только до первого значащего знака после запятой в дисперсии включительно.

Исключением является стандартное распределение Коши, так как оно имеет бесконечную дисперсию, а значит не может гарантировать никакой точности, а в общем мат. ожидание для данного распределения не определено.

7 Выводы

В процессе работы вычислены значения характеристик положения для определённых распределений на выборках фиксированной мощности и получено следующее ранжирование характеристик положения:

1. Стандартное нормальное распределение

$$\overline{x} < Z_{tr} < Z_Q < med\ x < Z_R$$

2. Распределение Лапласа (коэффициент масштаба $\sqrt{2}$ коэффициент сдвига равен нулю)

$$Z_Q < Z_{tr} < \overline{x} < Z_R < med\ x$$

3. Стандартное распределение Коши

$$med\, x < Z_Q < \overline{x} < Z_{tr} < Z_R$$

4. Равномерное распределение на отрезке $\left[-\sqrt{3}, \sqrt{3}\right]$

$$Z_R < med\, x < Z_Q < Z_{tr} < \overline{x}$$

5. Распределение Пуассона (значение мат ожидания равно 5)

$$Z_Q < med\, x < \overline{x} < Z_{tr} < Z_R$$

8 Список литературы

Модуль scipy.stats Модуль numpy для статистики

9 Приложения

Код лабораторной: https://github.com/dkamianskii/MatStatLabs/tree/master/Lab2