

Санкт-Петербургский политехнический университет Петра Великого
Институт прикладной математики и механики
Кафедра прикладной математики

ОТЧЕТ

Тема: *Исследование выборочных характеристик положения*

Направление: 01.03.02 Прикладная математика и информатика

Выполнил студент гр. 33631/4

Камалетдинова Ю.

Преподаватель

Баженов А.

Санкт-Петербург
2019

Содержание

Постановка задачи	2
Реализация	3
Результат	4

Постановка задачи

В данной лабораторной работе требуется вычислить некоторые из характеристик положения $N = 1000$ раз для выборок объемами $n = 20, 50, 100$ и проанализировать полученные результаты. Также необходимо установить, в каком соотношении находятся вычисленные характеристики для каждого из распределений, приведенных ниже

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

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

$$L(x, 0, \frac{1}{\sqrt{2}}) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|} - \text{Лаплас} \quad (3)$$

$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}}, & |x| \leq \sqrt{3} \\ 0, & |x| > \sqrt{3} \end{cases} - \text{равномерное} \quad (4)$$

$$P(\lambda) = \frac{e^{-\lambda}}{k!} \lambda^k - \text{Пуассон} \quad (5)$$

Приведем формулы для вычисления характеристик положения

$$\bar{x} = \overline{x_n} = \frac{1}{n} \sum_{i=1}^n x_i - \text{выборочное среднее} \quad (6)$$

$$\text{med } x = \begin{cases} x_{(k+1)}, & n = 2k + 1 \\ \frac{x_{(k)} + x_{(k+1)}}{2}, & n = 2k \end{cases} - \text{медиана} \quad (7)$$

$$z_R = \frac{x_{(1)} + x_{(n)}}{2} - \text{полусумма экстремальных значений} \quad (8)$$

$$z_Q = \frac{Q_1 + Q_3}{2} - \text{полусумма квантилей} \quad (9)$$

$$z_{tr} = \frac{1}{n - 2r} \sum_{i=r+1}^{n-r} x_{(i)} - \text{усеченное среднее} \quad (10)$$

Замечание: r – число наблюдений, оставшихся после усечения в характеристике (10), $r = \alpha n$, где α , как правило, равняется 0.1. В таком случае мы не вовлекаем 10% наибольших и 10% наименьших значений в вычисление усеченного среднего.

Реализация метода

Для выполнения поставленной задачи будем пользоваться библиотеками для языка Python: *numpy*, *scipy* – расчеты, законы распределения вероятностей; *matplotlib*, *seaborn* – визуализация результатов. Ход работы:

- Задаем распределение с заданными параметрами
- Генерируем случайные выборки из распределений размерами $n = 20, 50, 100$
- Для каждого из распределений вычисляем характеристики положения $N = 1000$ раз
- Вычисляем математическое ожидание и дисперсию для каждой вычисленной характеристики по формулам:

$$E(z) = \bar{z} = \frac{1}{N} \sum_{i=1}^N z_i \quad (11)$$

$$D(z) = \overline{z^2} - (\bar{z})^2 \quad (12)$$

Результат

Нормальное распределение с параметрами 0, 1

$n = 20$	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.0041	0.0053	0.0005	0.0046	0.0054
$D(z)$	0.0511	0.0753	0.1405	0.0586	0.0542

$n = 50$	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.0013	-0.0004	0.0170	0.0003	-0.0010
$D(z)$	0.0211	0.0313	0.1146	0.0251	0.0222

$n = 100$	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.0028	0.0026	0.0037	0.0034	0.0023
$D(z)$	0.0100	0.0152	0.0887	0.0122	0.0106

Соотношение дисперсий при $n = 100$: $\bar{x} < z_{tr} < z_Q < med\ x < z_R$

Равномерное распределение на отрезке $[-\sqrt{3}, \sqrt{3}]$

$n = 20$	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.0068	-0.0119	0.0011	-0.0113	-0.0082
$D(z)$	0.0554	0.1404	0.0141	0.0728	0.0742

$n = 50$	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.0032	0.0049	-0.0004	0.0035	0.0040
$D(z)$	0.0209	0.0582	0.0022	0.0304	0.0288

$n = 100$	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.0015	0.0015	0.0011	0.0017	0.0019
$D(z)$	0.0098	0.0292	0.0007	0.0156	0.0142

Соотношение дисперсий при $n = 100$: $z_R < \bar{x} < z_{tr} < z_Q < med\ x$

Распределение Коши с параметрами 0, 1

$n = 20$	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-4.1847	-0.0185	-41.9652	-0.0013	0.0043
$D(z)$	9829.5729	0.1450	982309.3260	0.3753	0.4334

$n = 50$	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	3.7413	0.0051	95.1768	-0.0011	-0.0022
$D(z)$	14981.2878	0.0483	9324487.2831	0.1102	0.1094

$n = 100$	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	2.6928	-0.0082	133.6930	-0.0080	-0.0116
$D(z)$	4451.3761	0.0247	11101512.6370	0.0522	0.0513

Соотношение дисперсий при $n = 100$: $med\ x < z_{tr} < z_Q < \bar{x} < z_R$

Распределение Лапласа с параметрами 0, $\frac{1}{\sqrt{2}}$

$n = 20$	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.0113	-0.0035	-0.0344	-0.0087	-0.0070
$D(z)$	0.0550	0.0348	0.4406	0.0519	0.0427

$n = 50$	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.0028	-0.0010	0.0085	0.0035	0.0019
$D(z)$	0.0193	0.0130	0.3839	0.0206	0.0156

$n = 100$	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.0059	-0.0029	-0.0104	-0.0047	-0.0051
$D(z)$	0.0097	0.0056	0.3931	0.0093	0.0071

Соотношение дисперсий при $n = 100$: $med\ x < z_{tr} < z_Q < \bar{x} < z_R$

Распределение Пуассона с параметром $\lambda = 7$

$n = 20$	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	6.9844	6.8310	7.4820	6.9067	6.8978
$D(z)$	0.3285	0.5259	1.0587	0.4049	0.3441

$n = 50$	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	7.0165	6.8640	7.7515	6.9400	6.9287
$D(z)$	0.1561	0.2665	0.8315	0.2299	0.1639

$n = 100$	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	7.0063	6.8755	7.9230	6.9088	6.9181
$D(z)$	0.0714	0.1482	0.7266	0.1074	0.0746

Соотношение дисперсий при $n = 100$: $\bar{x} < z_{tr} < z_Q < med\ x < z_R$

Обратившись к полученным соотношениям дисперсий выборок из распределений, можно сделать вывод о том, что полусумма экстремальных значений z_R имеет наибольший разброс относительно математического ожидания. Данное суждение не сходится с

результатами для случая равномерного распределения на отрезке. Медиана равномерного распределения на $[a, b]$ есть $med = \frac{a+b}{2}$, что и является полусуммой значений в крайних точках отрезка.

Также полученные результаты говорят о том, что в случае симметричного распределения, исходя из дисперсий, наиболее выгодно использовать выборочное среднее, чем медиану, хотя они и оценивают одну и ту же величину в данном случае. Но для распределения Лапласа медиана становится более эффективной.

Дисперсии таких характеристик, как усеченное среднее или полусумма квартилей, показывают среднее отклонение относительно остальных. Усеченное среднее представляет собой некий баланс между медианой и выборочным средним (является ими в частных случаях параметра α)

Список литературы

- [1] *Кадырова Н. О.* Теория вероятностей и математическая статистика. Статистический анализ данных: учеб. пособие / *Н. О. Кадырова, Л. В. Павлова, И. Е. Ануфриев.* - СПб.: Изд-во Политехн. ун-та, 2010. -54с.