Санкт-Петербургский политехнический университет Петра Великого

Институт прикладной математики и механики

Кафедра «Телематика (при ЦНИИ РТК)»

Отчет по лабораторной работе

Вычисление выборочных характеристик
По дисциплине «Теория вероятностей и Математическая статистика»

Выполнил
Студент гр. 3630201/80101 ________ М.Д. Маляренко

Руководитель
к.ф.-м.н., доцент ________ А.Н. Баженов

«___»____________ 2020г.

Санкт-Петербург
2020

Содержание

Список таблиц

1 Постановка задачи

Заданы 5 распределений случайных величин:

1. Нормальное распределение $N(x, 0, 1)$
2. Распределение Коши $C(x, 0, 1)$
3. Распределение Лапласа $L(x, 0, 1/\sqrt{2})$
4. Дискретное распределение Пуассона $P(k, 10)$
5. Равномерное распределение $U(x, -\sqrt{3}, \sqrt{3})$

Для каждого распределения необходимо сгенерировать выборки размером 10, 100 и 1000 элементов. Для каждой выборки рассчитать числовые характеристики:

1. Выборочное среднее $\overline{x}$
2. Выборочная медиана $med\ x$
3. Полусумма экстремальных выборочных элементов z_R
4. Полусумма квартилей z_Q
5. Усечённое среднее z_{tr}

Произвести генерацию каждой выборки и вычисление характеристик 1000 раз. Найти среднее характеристик $E(z)$, вычислить дисперсию $D(z)$.

2 Теория

2.1 Вариационный ряд

Вариационным рядом называется последовательность не обязательно уникальных элементов выборки, расположенных порядке неубывания. [1]

Запись элементов вариационного ряда:

$$x_{(i)},\ i = 1, \dots, n$$

2.2 Выборочные числовые характеристики

В данной лабораторной работе рассматриваются следующие числовые выборочные характеристики для вариационного ряда $x_{(1)}, \dots, x_{(n)}$. [1]

- *Выборочное среднее*

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_{(i)} \tag{1}$$

- *Выборочная медиана*

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

- *Полусумма экстремальных выборочных элементов*

$$z_R = \frac{x_{(1)} + x_{(n)}}{2} \tag{3}$$

- *Полусумма квартилей*

  Выборочная квартиль:

$$z_p = \begin{cases} x_{([np]+1)}, & np \text{ - целое} \\ x_{(np)}, & np \text{ - дробное} \end{cases}$$

  Полусумма квартилей:

$$z_Q = \frac{z_{1/4} + z_{3/4}}{2} \tag{4}$$

- *Усечённое среднее*

$$z_{tr} = \frac{1}{n - 2r} \sum_{i=r+i}^{n-r} x_{(i)}, \qquad r \approx \frac{n}{4} \tag{5}$$

2.3 Средее характеристик, дисперсия

Среднее характеристик рассчитывается как простое среднее арифметическое.

$$E(z) = \overline{z} \tag{6}$$

Рассеяние или дисперсия рассчитывается как разность среднеквадратичного значения вариационного ряда и квадрата выборочного среднего:

$$D(z) = \overline{x^2} - \overline{x}^2 \tag{7}$$

3 Реализация

Расчёты были реализованы в среде аналитических вычислений Maxima. Выборки сгенерированы встроенными функциями среды Maxima. Код скрипта представлен в репозитории на GitHub.

4 Результаты

По результатам вычисления выборочных числовых характеристик были сформированы таблицы 1-5 по количеству заданных распределений. Погрешность среднего значения характеристики выборки рассчитывалась как $\Delta_z = \sqrt{D(z)}$.

Нормальное распределение						
		$\overline{x}$ (1)	$med\ x$ (2)	z_R (3)	z_Q (4)	z_{tr} (5)
$N = 10$	$E(z)$ (6) $\pm \Delta_z$	0.0 ± 0.4	0.0 ± 0.7	0.0 ± 0.8	0.0 ± 0.8	0.0 ± 0.5
	$D(z)$ (7)	0.102	0.489	0.509	0.498	0.179
$N = 100$	$E(z) \pm \Delta_z$	0.00 ± 0.09	0.0 ± 0.7	0.0 ± 0.8	0.0 ± 0.7	0.0 ± 0.2
	$D(z)$	0.01	0.494	0.518	0.497	0.020
$N = 1000$	$E(z) \pm \Delta_z$	0.00 ± 0.04	0.0 ± 0.7	0.0 ± 0.7	0.0 ± 0.7	0.0 ± 0.05
	$D(z)$	0.001	0.528	0.501	0.520	0.002

Таблица 1: Характеристики выборок нормального распределения

Распределение Лапласа						
		$\overline{x}$	$med\ x$	z_R	z_Q	z_{tr}
$N = 10$	$E(z) \pm \Delta_z$	0.0 ± 0.3	0.0 ± 0.5	0.0 ± 0.4	0.0 ± 0.4	0.0 ± 0.1
	$D(z)$	0.093	0.512	0.428	0.479	0.116
$N = 100$	$E(z) \pm \Delta_z$	0.0 ± 0.1	0.0 ± 0.8	0.0 ± 0.8	0.0 ± 0.8	0.0 ± 0.2
	$D(z)$	0.01	0.494	0.518	0.497	0.020
$N = 1000$	$E(z) \pm \Delta_z$	0.00 ± 0.04	0.0 ± 0.7	0.0 ± 0.8	0.0 ± 0.7	0.00 ± 0.05
	$D(z)$	0.001	0.492	0.521	0.485	0.002

Таблица 2: Характеристики выборок распределения Лапласа

Распределение Коши						
		$\overline{x}$	$med\ x$	z_R	z_Q	z_{tr}
$N = 10$	$E(z) \pm \Delta_z$	0 ± 16	0 ± 48	0 ± 33	0 ± 13	0 ± 23
	$D(z)$	260	2342	1063	194	558
$N = 100$	$E(z) \pm \Delta_z$	0 ± 12	-1 ± 22	1 ± 12	-2 ± 105	1 ± 23
	$D(z)$	142	499	137	11084	492
$N = 1000$	$E(z) \pm \Delta_z$	0 ± 16	0 ± 34	0 ± 21	2 ± 64	0 ± 26
	$D(z)$	259	1170	461	4212	694

Таблица 3: Характеристики выборок распределения Коши

Распределение Пуассона						
		$\overline{x}$	$med\ x$	z_R	z_Q	z_{tr}
$N = 10$	$E(z) \pm \Delta_z$	10.0 ± 1.0	9.9 ± 2.2	9.9 ± 2.2	10.0 ± 2.1	10.0 ± 1.3
	$D(z)$	1.028	4.832	5.127	4.685	1.649
$N = 100$	$E(z) \pm \Delta_z$	10.0 ± 0.3	9.9 ± 2.3	10.0 ± 2.4	10.0 ± 2.2	10.0 ± 0.5
	$D(z)$	0.092	5.151	4.989	4.910	0.199
$N = 1000$	$E(z) \pm \Delta_z$	10.00 ± 0.09	9.8 ± 2.3	10.0 ± 2.2	10.0 ± 2.2	10.00 ± 0.15
	$D(z)$	0.0098	5.1100	4.8222	4.7592	0.0211

Таблица 4: Характеристики выборок распределения Пуассона

Равномерное распределение						
		$\overline{x}$	$med\ x$	z_R	z_Q	z_{tr}
$N = 10$	$E(z) \pm \Delta_z$	0.0 ± 0.3	0.0 ± 0.7	0.0 ± 0.8	0.0 ± 0.8	0.0 ± 0.4
	$D(z)$	0.091	0.477	0.509	0.506	0.150
$N = 100$	$E(z) \pm \Delta_z$	0.0 ± 0.1	0.0 ± 0.8	0.0 ± 0.8	0.0 ± 0.7	0.0 ± 0.2
	$D(z)$	0.103	0.727	0.713	0.701	0.146
$N = 1000$	$E(z) \pm \Delta_z$	0.00 ± 0.04	0.0 ± 0.7	0.0 ± 0.7	0.0 ± 0.7	0.0 ± 0.05
	$D(z)$	0.001	0.500	0.496	0.489	0.002

Таблица 5: Характеристики выборок равномерного распределения

Заключение

В результате лабораторной работы были сгенерированы выборки размером 10, 100, 1000 элементов по заданным распределений и оценены их числовые характеристики. Работа велась в среде аналитических вычислений Maxima.

Из всех рассмотренных распределений самую большую дисперсию (на 3-4 порядка относительно других) имеет распределение Коши в связи с мощными выбросами.

Можно сделать вывод, что для данных распределений, чем больше мощность выборки, тем ближе значение медианы к своему теоретическому значению и значение выборочного среднего ближе к теоретическому матожиданию.

Список литературы

[1] Теоретическое приложение к лабораторным работам №1-4 по дисциплине «Математическая статистика». – СПб.: СПбПУ, 2020. – 12 с

Приложение А. Репозиторий с исходным кодом

Исходный код скрипта для среды аналитических вычислений Maxima находится в репозитории GitHub – URL `https://github.com/malyarenko-md/TeorVer`