

Санкт-Петербургский политехнический университет Петра Великого

Институт прикладной математики и механики

Кафедра «Телематика (при ЦНИИ РТК)»

Отчет по лабораторной работе

Нахождение числовых характеристик выборки

По дисциплине «Теория вероятностей и математическая статистика»

Выполнил

Студент гр.3630201/80101

В.Н. Сеннов

Руководитель

доцент к.ф.-м.н.

А.Н. Баженов

«___» _____ 202__г.

Санкт-Петербург
2020

Содержание

1	Постановка задачи	4
2	Математическое описание	5
2.1	Вариационный ряд	5
2.2	Выборочные числовые характеристики	5
2.3	Характеристики рассеяния	5
3	Особенности реализации	6
4	Результаты работы программы	7
	Заключение	9
	Список литературы	10
A	Репозиторий с исходным кодом	11

Список таблиц

1	Числовые характеристики выборки, соответствующей нормальному распределению	7
2	Числовые характеристики выборки, соответствующей распределению Коши . .	7
3	Числовые характеристики выборки, соответствующей распределению Лапласа .	7
4	Числовые характеристики выборки, соответствующей распределению Пуассона	8
5	Числовые характеристики выборки, соответствующей равномерному распределению	8

1 Постановка задачи

Для заданных распределений нужно сгенерировать выборки размером 10, 100, 1000 элементов. Для каждой выборки нужно найти следующие характеристики:

1. среднее значение;
2. медиану;
3. полусумму экстремальных значений;
4. полусумму квартилей;
5. усеченное среднее.

Это нужно повторить 1000 раз и найти среднее значение и дисперсию этих характеристик для каждого размера выборки.

Заданные распределения:

1. Нормальное (гауссово) распределение с параметрами $\mu = 0$, $\sigma = 1$;
2. Распределение Коши с параметрами $\mu = 0$, $\lambda = 1$;
3. Распределение Лапласа с параметрами $\mu = 0$, $\lambda = \frac{1}{\sqrt{2}}$;
4. Распределение Пуассона с параметром $\mu = 10$;
5. Равномерное распределение с параметрами $a = -\sqrt{3}$, $b = \sqrt{3}$.

2 Математическое описание

2.1 Вариационный ряд

Вариационным рядом называется последовательность элементов выборки, расположенных в неубывающем порядке, причем одинаковые элементы повторяются [2].

2.2 Выборочные числовые характеристики

Для вариационного ряда x_1, x_2, \dots, x_n определяют следующие числовые характеристики:

1. Выборочное среднее:

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

2. Выборочная медиана:

$$\text{med } x = \begin{cases} x_{l+1} & \text{если } n = 2l + 1 \\ \frac{x_l + x_{l+1}}{2} & \text{если } n = 2l \end{cases} \quad (2)$$

3. Полусумма экстремалей:

$$z_R = \frac{x_1 + x_n}{2} \quad (3)$$

4. Выборочный квантиль порядка p :

$$z_p = x_{[np]}$$

5. Полусумма квартилей:

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

6. Усеченное среднее:

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

[1]

2.3 Характеристики рассеяния

Для выборки можно рассчитать выборочную дисперсию по формуле:

$$D = \overline{x^2} - (\bar{x})^2, \quad (6)$$

где \bar{x} — выборочное среднее, а $\overline{x^2} = \frac{1}{n} \sum_{i=1}^n x_i^2$.

3 Особенности реализации

Программа для выполнения лабораторной была написана на языке Python 3.8.2. Для генерации выборок использовался модуль **distributions**, написанный для лабораторной №1. Для расчета численных характеристик был написан модуль **characteristics**. В нем на основе формул (1-5) рассчитываются числовые характеристики. При помощи формулы (1) рассчитывается среднее значение для каждой характеристики, при помощи формулы (6) рассчитывается дисперсия.

В приложении А приведена ссылка на репозиторий с исходным кодом.

4 Результаты работы программы

В данном разделе в таблицах 1-5 представлены результаты расчета числовых характеристик выборки.

Выборочные характеристики \bar{x} - z_{tr} были рассчитаны по формулам (1-5).

Расчет погрешности для средней величины был произведен по формуле:

$$\Delta_z = \sqrt{D(z)}$$

Нормальное распределение						
		\bar{x}	$\text{med } x$	z_R	z_Q	z_{tr}
$n = 10$	$\bar{z} \pm \Delta_z$	0.0 ± 0.4	0.0 ± 0.4	0.0 ± 0.5	0.0 ± 0.4	0.0 ± 0.4
	D_z	0.103	0.146	0.183	0.124	0.124
$n = 100$	$\bar{z} \pm \Delta_z$	0.0 ± 0.1	0.0 ± 0.2	0.0 ± 0.3	0.0 ± 0.2	0.0 ± 0.2
	D_z	0.009	0.015	0.074	0.012	0.011
$n = 1000$	$\bar{z} \pm \Delta_z$	0.00 ± 0.04	0.00 ± 0.04	0.0 ± 0.2	0.00 ± 0.04	0.00 ± 0.04
	D_z	0.001	0.002	0.042	0.001	0.001

Таблица 1: Числовые характеристики выборки, соответствующей нормальному распределению

Распределение Коши						
		\bar{x}	$\text{med } x$	z_R	z_Q	z_{tr}
$n = 10$	$\bar{z} \pm \Delta_z$	0 ± 4	0.0 ± 0.6	0 ± 20	0.0 ± 1.1	0.0 ± 0.8
	D_z	15.4	0.35	296	1.16	0.519
$n = 100$	$\bar{z} \pm \Delta_z$	0.0 ± 1.2	0.0 ± 0.2	0.0 ± 40	0.0 ± 0.3	0.0 ± 0.2
	D_z	1.521	0.024	1190	0.050	0.024
$n = 1000$	$\bar{z} \pm \Delta_z$	0.00 ± 0.5	0.00 ± 0.06	0 ± 20	0.00 ± 0.08	0.00 ± 0.06
	D_z	0.172	0.002	400	0.005	0.003

Таблица 2: Числовые характеристики выборки, соответствующей распределению Коши

Распределение Лапласа						
		\bar{x}	$\text{med } x$	z_R	z_Q	z_{tr}
$n = 10$	$\bar{z} \pm \Delta_z$	0.0 ± 0.6	0.0 ± 0.6	0.0 ± 1.2	0.0 ± 0.6	0.0 ± 0.6
	D_z	0.361	0.284	1.379	0.369	0.278
$n = 100$	$\bar{z} \pm \Delta_z$	0.0 ± 0.2	0.0 ± 0.2	0.0 ± 1.0	0.0 ± 0.2	0.0 ± 0.2
	D_z	0.041	0.027	1.025	0.040	0.028
$n = 1000$	$\bar{z} \pm \Delta_z$	0.00 ± 0.07	0.00 ± 0.05	0.0 ± 0.6	0.00 ± 0.06	0.00 ± 0.05
	D_z	0.0036	0.0019	0.3298	0.0034	0.0021

Таблица 3: Числовые характеристики выборки, соответствующей распределению Лапласа

Распределение Пуассона						
		\bar{x}	med x	z_R	z_Q	z_{tr}
$n = 10$	$\bar{z} \pm \Delta_z$ D_z	10.0 ± 1.0 1.019	9.8 ± 1.2 1.442	10 ± 2 1.762	9.9 ± 1.2 1.245	9.9 ± 1.1 1.156
$n = 100$	$\bar{z} \pm \Delta_z$ D_z	10.0 ± 0.3 0.097	9.8 ± 0.5 0.211	10.9 ± 0.8 0.689	9.9 ± 0.4 0.152	9.8 ± 0.4 0.121
$n = 1000$	$\bar{z} \pm \Delta_z$ D_z	10.0 ± 0.1 0.0096	10.00 ± 0.04 0.0019	11.3 ± 0.4 0.1105	10.00 ± 0.06 0.0027	9.86 ± 0.10 0.0109

Таблица 4: Числовые характеристики выборки, соответствующей распределению Пуассона

Равномерное распределение						
		\bar{x}	med x	z_R	z_Q	z_{tr}
$n = 10$	$\bar{z} \pm \Delta_z$ D_z	0.0 ± 0.4 0.106	0.0 ± 0.5 0.232	0.0 ± 0.3 0.046	0.0 ± 0.4 0.141	0.0 ± 0.5 0.166
$n = 100$	$\bar{z} \pm \Delta_z$ D_z	0.00 ± 0.11 0.0103	0.0 ± 0.2 0.232	0.00 ± 0.03 0.0462	-0.02 ± 0.12 0.141	0.0 ± 0.2 0.166
$n = 1000$	$\bar{z} \pm \Delta_z$ D_z	0.00 ± 0.04 0.0009	0.00 ± 0.06 0.0028	0.000 ± 0.003 0.00005	0.00 ± 0.04 0.0014	0.00 ± 0.05 0.0019

Таблица 5: Числовые характеристики выборки, соответствующей равномерному распределению

Заключение

В рамках лабораторной работы были рассчитаны средние значения числовых характеристик выборок разного размера для заданных распределений.

Заметно, что при большем количестве элементов выборки значения среднего и медианы точнее соответствует математическому ожиданию и теоретическому значению медианы.

Стоит отметить, что при реализации генерации выборки, соответствующей распределению Коши, был намеренно ограничен диапазон значений, чтобы избежать чересчур больших значений случайной величины. Несмотря на это, заметно, что дисперсия при подсчете среднего больше, чем для других распределений. Заметно также, что для медианы и усеченного среднего этого эффекта не наблюдается.

Программа для лабораторной была написана языке Python 3.8.2.

Список литературы

- [1] Максимов Ю. Д. Математика. Теория вероятностей и случайных процессов; Учебное пособие / Под ред. В. И. Антонова. — СПб.: Изд-во Политехн. ун-та, 2008. — 384 с.
- [2] Теоритическое приложение к лабораторным работам №1-4 по дисциплине «Математическая статистика». — СПб.: Сантк-Петербургский политехнический университет, 2020. — 12 с.

А Репозиторий с исходным кодом

Исходный код программы для данной лабораторной размещен на сервисе GitHub.

Ссылка на репозиторий: <https://github.com/Vovan-S/TV-Lab1>.