

Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики
Кафедра «Прикладная математика»

Отчёт
по лабораторной работе №2
по дисциплине
«Математическая статистика»

Выполнил студент:
Самутичев Евгений Романович
группа: 3630102/70201

Проверил:
к.ф.-м.н., доцент
Баженов Александр Николаевич

Санкт-Петербург
2020 г.

Содержание

1	Постановка задачи	2
2	Теория	3
2.1	Вариационный ряд	3
2.2	Выборочные характеристики	3
3	Реализация	4
4	Результаты	5
5	Обсуждение	7
5.1	Математическое ожидание и медиана	7
5.2	Полусуммы: z_R и z_Q	7
5.3	Упорядочение характеристик	7
6	Приложения	8

Список таблиц

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

1 Постановка задачи

Для каждого из 5 распределений:

1. Нормального $N(x, 0, 1)$
2. Коши $C(x, 0, 1)$
3. Лапласа $L(x, 0, \frac{1}{\sqrt{2}})$
4. Пуассона $P(k, 10)$
5. Равномерного $U(x, -\sqrt{3}, \sqrt{3})$

выборку размера: 10, 100, 1000 - сгенерировать 1000 раз, для каждой генерации произвести вычисления выборочных характеристик \bar{x} , $\text{med } x$, z_R , z_Q , z_{tr} для всех генераций в рамках одного размера выборки получить значения среднего характеристик положения:

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

и дисперсию:

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

Представить полученные данные в виде таблиц.

2 Теория

2.1 Вариационный ряд

Если элементы выборки x_1, \dots, x_n упорядочить по возрастанию на каждом элементарном исходе (рассматриваем их как случайные величины), получится новый набор случайных величин, называемый *вариационным рядом*:

$$x_{(1)} \leq \dots \leq x_{(n)}$$

Элемент $x_{(k)}$ называется *k-ой порядковой статистикой*¹.

2.2 Выборочные характеристики

При работе с выборкой нам неизвестно распределение по которому она получена, а значит и соответствующие характеристики распределения. Однако, существуют оценки - т.н. *выборочные характеристики*:

- Выборочное среднее

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

- Выборочная медиана

$$\text{med } x = \begin{cases} x_{(k+1)} & \text{при } n = 2k + 1 \\ \frac{x_{(k)} + x_{(k+1)}}{2} & \text{при } n = 2k \end{cases} \quad (4)$$

- Полусумма экстремальных выборочных элементов

$$z_R = \frac{x_{(1)} + x_{(n)}}{2} \quad (5)$$

- Выборочный квантиль уровня α

$$z_\alpha = \frac{x_{(\lfloor q \rfloor + 1)} + x_{(\lceil q \rceil + 1)}}{2}, \text{ где } q = (n - 1)\alpha \quad (6)$$

формула, используемая в **NumPy**, в этом случае $z_0 = \min_{i=1, \dots, n} x_{(i)}$, $z_1 = \max_{i=1, \dots, n} x_{(i)}$,
 $z_{0.5} = \text{med } x$

- Полусумма квантилей

$$z_Q = \frac{z_{0.25} + z_{0.75}}{2} \quad (7)$$

- Усеченное среднее

$$z_{tr} = \frac{1}{n - 2r} \sum_{i=r+1}^{n-r} x_{(i)}, \text{ где } r = \lceil \frac{n}{4} \rceil \quad (8)$$

Выборочные характеристики как борелевские функции от случайных величин (выборки) также являются случайными величинами, поэтому в работе и производится усреднение их значений для 1000 генераций и вычисление дисперсии.

¹ [1] стр. 10

3 Реализация

Работа выполнена с использованием языка **Python** в интегрированной среде разработки **PyCharm**, были задействованы библиотеки:

- **NumPy** - построение вариационного ряда и вычисления
- **SciPy** - модуль **stats** для генерации данных по распределениям

Исходный код работы приведен в приложении.

4 Результаты

	\bar{x} (3)	med x (4)	z_R (5)	z_Q (7)	z_{tr} (8)
$n = 10$					
$E(z)$	-0.0	-0.0	-0.0	-0.0	-0.1
$D(z)$	0.094467	0.130519	0.187448	0.105385	0.069492
$n = 100$					
$E(z)$	0.0	0.0	0.0	0.01	-0.01
$D(z)$	0.009651	0.015116	0.091466	0.011977	0.011309
$n = 1000$					
$E(z)$	-0.0	0.001	0.0	-0.0	-0.001
$D(z)$	0.001049	0.00153	0.062347	0.001313	0.001193

Таблица 1: Нормальное распределение

	\bar{x}	med x	z_R	z_Q	z_{tr}
$n = 10$					
$E(z)$	-0.0	-0.0	-1.0	-0.0	-0.2
$D(z)$	490.607293	0.332166	11914.643438	1.132547	0.210165
$n = 100$					
$E(z)$	2.0	0.0	79.0	0.0	-0.01
$D(z)$	3581.697241	0.026278	8922533.221739	0.04842	0.025421
$n = 1000$					
$E(z)$	1.0	0.0	642.0	0.0	-0.0
$D(z)$	2223.205146	0.00256	547218440.611133	0.004966	0.002619

Таблица 2: Распределение Коши

	\bar{x}	med x	z_R	z_Q	z_{tr}
$n = 10$					
$E(z)$	0.0	0.0	0.0	0.0	-0.1
$D(z)$	0.098335	0.072722	0.382323	0.089897	0.041919
$n = 100$					
$E(z)$	-0.0	0.0	-0.0	-0.0	-0.01
$D(z)$	0.009719	0.00561	0.433362	0.009396	0.005806
$n = 1000$					
$E(z)$	0.0	0.001	-0.0	0.001	-0.0
$D(z)$	0.000918	0.000483	0.441511	0.000944	0.000561

Таблица 3: Распределение Лапласа

	\bar{x}	med x	z_R	z_Q	z_{tr}
$n = 10$					
$E(z)$	10.0	10.0	10.0	10.0	7.0
$D(z)$	0.955624	1.3806	1.744716	1.128935	0.704541
$n = 100$					
$E(z)$	10.0	9.9	11.0	9.9	9.6
$D(z)$	0.097956	0.194391	0.997104	0.14328	0.110723
$n = 1000$					
$E(z)$	10.0	10.0	12.0	9.994	9.84
$D(z)$	0.010385	0.003484	0.6581	0.002748	0.011585

Таблица 4: Распределение Пуассона

	\bar{x}	med x	z_R	z_Q	z_{tr}
$n = 10$					
$E(z)$	0.0	0.0	-0.0	0.0	-0.1
$D(z)$	0.10033	0.234165	0.043909	0.136123	0.119729
$n = 100$					
$E(z)$	0.0	-0.0	0.001	0.0	-0.02
$D(z)$	0.009457	0.028559	0.00059	0.014028	0.018067
$n = 1000$					
$E(z)$	-0.001	-0.002	4e-05	-0.001	-0.003
$D(z)$	0.00102	0.003073	6e-06	0.001465	0.002005

Таблица 5: Равномерное распределение

5 Обсуждение

5.1 Математическое ожидание и медиана

Для каждого из указанных в постановке задачи распределений, приведем теоретические значения математического ожидания и медианы:

- $N(x, 0, 1) : \mathbf{E} = 0, \text{med} = 0$
- $C(x, 0, 1) : \mathbf{E} - \text{не определено}, \text{med} = 0$
- $L(x, 0, \frac{1}{\sqrt{2}}) : \mathbf{E} = 0, \text{med} = 0$
- $P(k, 10) : \mathbf{E} = 10, \text{med} = 10$
- $U(x, -\sqrt{3}, \sqrt{3}) : \mathbf{E} = 0, \text{med} = 0$

Как известно, *выборочное среднее является несмещенной и состоятельной оценкой для математического ожидания*² Это объясняет то что для всех распределений кроме распределения Коши - выборочное среднее при росте n стремится к математическому ожиданию, для распределения Коши последовательность вычислений не демонстрирует никакой сходимости (см. таблицу 2), поскольку у него отсутствует математическое ожидание. В тоже время медиана имеется у всех распределений и к ней сходится выборочная медиана.

5.2 Полусуммы: z_R и z_Q

Полусумма квартилей z_Q и экстремальных выборочных элементов z_R оценивают центр симметрии распределения, из таблиц наблюдается что z_Q ближе к медиане и последовательность вычислений $E(z)$ для z_Q при увеличении n сходится, в тоже время последовательность значений $E(z)$ для z_R расходится при распределении Коши. Таким образом оценка через полусумму квартилей лучше, хотя и требует больше вычислений.

5.3 Упорядочение характеристик

Для $n = 1000$ приведем упорядочение характеристик положения по каждому распределению:

- $N(x, 0, 1) : z_{tr} < z_Q \leq \bar{x} \leq z_R < \text{med } x$
- $C(x, 0, 1) : z_{tr} \leq z_Q \leq \text{med } x < \bar{x} < z_R$
- $L(x, 0, \frac{1}{\sqrt{2}}) : z_{tr} \leq z_R \leq \bar{x} < \text{med } x \leq z_Q$
- $P(k, 10) : z_{tr} < z_Q < \text{med } x \leq \bar{x} < z_R$
- $U(x, -\sqrt{3}, \sqrt{3}) : z_{tr} < \text{med } x < z_Q \leq \bar{x} < z_R$

² [1] стр. 17

6 Приложения

1. Исходный код лабораторной <https://github.com/zhenyatos/statlabs/tree/master/Lab2>

Список литературы

- [1] Н. И. Чернова, *Математическая статистика: Учеб. пособие*. Новосиб. гос. ун-т. Новосибирск, 2007. 148 стр.