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

Список таблиц

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 $\bar{x}, \mathrm{med}\ x, z_R, z_Q, z_{tr}$ для всех генераций в рамках одного размера выборки получить значения среднего характеристик положения:

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

и дисперсию:

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

Представить полученные данные в виде таблиц.

2 Теория

2.1 Вариационный ряд

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

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

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
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3}$$

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

- Полусумма экстремальных выборочных элементов
$$z_R = \frac{x_{(1)} + x_{(n)}}{2} \tag{5}$$

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

  формула, используемая в **NumPy**, в этом случае $z_0 = \min\limits_{i=1,...,n} x_{(i)}, z_1 = \max\limits_{i=1,...,n} x_{(i)}$, $z_{0.5} = \operatorname{med} x$

- Полусумма квантилей
$$z_Q = \frac{z_{0.25} + z_{0.75}}{2} \tag{7}$$

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

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 (6)	z_{tr} (8)
$n = 10$					
$E(z)$ (1)	0.012715	0.014862	0.015454	0.012652	-0.083313
$D(z)$ (2)	0.097193	0.125152	0.178521	0.111685	0.068423
$n = 100$					
$E(z)$	-0.000946	-0.006489	0.004489	-0.002539	-0.016852
$D(z)$	0.00996	0.015447	0.094226	0.012023	0.011609
$n = 1000$					
$E(z)$	-5.2e-05	-0.000798	0.008211	-9.9e-05	-0.001863
$D(z)$	0.001047	0.001693	0.059478	0.001248	0.001292

Таблица 1: Нормальное распределение

	$\bar{x}$	med x	z_R	z_Q	z_{tr}
$n = 10$					
$E(z)$	0.269175	0.003662	1.301074	0.035507	-0.148399
$D(z)$	420.187861	0.35542	10321.906112	1.007555	0.211997
$n = 100$					
$E(z)$	2.907996	0.003994	148.430856	-0.006257	-0.021858
$D(z)$	6161.9537	0.024654	15382087.074375	0.053289	0.026204
$n = 1000$					
$E(z)$	-0.388287	0.002631	-198.918484	0.00337	0.001214
$D(z)$	116.377519	0.0025	26681717.853264	0.004778	0.00254

Таблица 2: Распределение Коши

	$\bar{x}$	med x	z_R	z_Q	z_{tr}
$n = 10$					
$E(z)$	0.006704	0.009902	0.009228	0.004885	-0.063547
$D(z)$	0.097983	0.077904	0.404708	0.087752	0.042383
$n = 100$					
$E(z)$	-0.001985	-0.001411	-0.016642	-0.000722	-0.011451
$D(z)$	0.009994	0.005558	0.383806	0.009967	0.005637
$n = 1000$					
$E(z)$	-0.000221	0.000461	-0.007886	0.000215	-0.000866
$D(z)$	0.001024	0.000529	0.425748	0.000971	0.000617

Таблица 3: Распределение Лапласа

	$\bar{x}$	med x	z_R	z_Q	z_{tr}
$n = 10$					
$E(z)$	10.0066	9.8445	10.2815	9.93025	7.0925
$D(z)$	0.979396	1.41507	1.820008	1.116135	0.736944
$n = 100$					
$E(z)$	10.00372	9.867	10.924	9.9125	9.62518
$D(z)$	0.09631	0.197311	1.028224	0.14825	0.110399
$n = 1000$					
$E(z)$	9.997868	9.9985	11.654	9.994375	9.832432
$D(z)$	0.009556	0.001248	0.666284	0.003234	0.010793

Таблица 4: Распределение Пуассона

	$\bar{x}$	med x	z_R	z_Q	z_{tr}
$n = 10$					
$E(z)$	0.01133	0.017503	0.008043	0.011588	-0.105403
$D(z)$	0.097846	0.228939	0.042428	0.139057	0.118902
$n = 100$					
$E(z)$	-0.002344	-0.002894	0.001376	-0.002685	-0.020092
$D(z)$	0.009287	0.026082	0.000668	0.014104	0.017473
$n = 1000$					
$E(z)$	-0.001374	-0.003283	4.3e-05	-0.00104	-0.004076
$D(z)$	0.000993	0.00292	6e-06	0.001502	0.001978

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:

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

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.

[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.