Математическая статистика

Домашняя работа № 3

Оценки

Попов Юрий, СКБ-172

ОГЛАВЛЕНИЕ

Задание 3.1 Нахождение выборочного среднего и выборочной дисперсии 3

3.1.1 Геометрическое распределение 3

3.1.2 Экспоненциальное распределение 6

Задание 3.2 Построение доверительного интервала для выборочного среднего и выборочной дисперсии 9

3.2.1 Геометрическое распределение 9

3.2.2 Экспоненциальное распределение 9

Задание 3.3 Нахождение параметров распределений событий 11

3.3.1 Геометрическое распределение 11

3.3.2 Экспоненциальное распределение 12

Предисловие

Все графики, которые в дальнейшем будут вставлены в эту работу, были сконструированы с помощью различных библиотек, основные которые - это matplotlib и numpy в Jupyter Notebook

К работе приложены 2 основных файла: "Geom_Dz_3.ipynb"и "Expon_Dz_3.ipynb в которых указаны расчеты соотвественно геометрического и экспоненциального распределения

Все фотографии, использованные в работе лежат в папке *fotos*

Когда я начинал третье домашнее задание, обновленного файла с домашней работы еще не было(или я не знал о его существовании), поэтому номер 3.2 сделан из старого файла

Большая часть определений, которые представлены в этой работы взять с лекций нашего курса.

Также некоторые определения взяты из источника Г.И. Ивченко, Ю.И. Медведев "Ведение в математическую статистику"

Задание 3.1 Нахождение выборочного среднего и выборочной дисперсии

Выборочные моменты

Наиболее важными характеристиками случайной величины ξ являются ее моменты $\alpha_k = E\xi^k$, а также центральные моменты $\mu_k = E(\xi - \alpha_1)^k$(когда они существуют). Их статическими аналогами, вычисляемыми по соответствующей выборке $X = (X_1, \ldots, X_n)$, являются *выборочные моменты* соответственно *обычные*

$$\hat{\alpha}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

и центральные

$$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\alpha}_1)^k$$

Особенно важны моменты первого и второго порядков.

При $k = 1$ величину $\hat{\alpha}_1$ называют *выборочным средним* и обозначают стандартным символом $\bar{X}$:

$$\bar{X} = \hat{\alpha}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

При $k = 2$ величину $\hat{\mu}_2$ называют *выборочной дисперсией* и также обозначают стандартным символом $S^2 = S^2(X)$:

$$S^2 = \hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

3.1.1 Геометрическое распределение

Для каждой выборки из домашнего задания 2 посчитаем выборочное среднее и выборочную дисперсию. Для наглядности выведем вариационный ряд для объема 5 и 10 и посчитанные оба параметра.

Вариационный ряд выборки 1 объема 5:

	1	3	5	15
f	1	2	1	1

Выборочное среднее этой выборки равно: 4.8
Выборочная дисперсия этой выборки равна: 25.0

Вариационный ряд выборки 2 объема 5:

	1	3	5	6	12
f	1	1	1	1	1

Выборочное среднее этой выборки равно: 5.4
Выборочная дисперсия этой выборки равна: 14.2

Вариационный ряд выборки 3 объема 5:

	6	11	18	30	37
f	1	1	1	1	1

Выборочное среднее этой выборки равно: 20.4
Выборочная дисперсия этой выборки равна: 377.2

Вариационный ряд выборки 4 объема 5:

	0	1	5	7
f	2	1	1	1

Выборочное среднее этой выборки равно: 5.2
Выборочная дисперсия этой выборки равна: 13.08

Вариационный ряд выборки 5 объема 5:

	0	8	12	17
f	2	1	1	1

Выборочное среднее этой выборки равно: 14.8
Выборочная дисперсия этой выборки равна: 51.4

$n = 5$

Вариационный ряд выборки 1 объема 10:

	2	3	4	6	9	14	16	20	25
f	1	1	1	1	2	1	1	1	1

Выборочное среднее этой выборки равно: 9.9
Выборочная дисперсия этой выборки равна: 54.57

Вариационный ряд выборки 2 объема 10:

	1	2	3	5	7	17	33
f	3	1	2	1	1	1	1

Выборочное среднее этой выборки равно: 20.4
Выборочная дисперсия этой выборки равна: 101.17

Вариационный ряд выборки 3 объема 10:

	0	7	9	10	12	14	19	33
f	1	1	2	1	2	1	1	1

Выборочное среднее этой выборки равно: 10.4
Выборочная дисперсия этой выборки равна: 75.01

Вариационный ряд выборки 4 объема 10:

	0	2	4	5	9	11	15	20
f	2	2	1	1	1	1	1	1

Выборочное среднее этой выборки равно: 13.2
Выборочная дисперсия этой выборки равна: 50.97

Вариационный ряд выборки 5 объема 10:

	0	3	5	8	12	26
f	5	1	1	1	1	1

Выборочное среднее этой выборки равно: 27.0
Выборочная дисперсия этой выборки равна: 82.89

$n = 10$

```
Выборка 1 объема 100:
Выборочное среднее этой выборки равно:  48.62
Выборочная дисперсия этой выборки  равна:  1726.57

Выборка 2 объема 100:
Выборочное среднее этой выборки равно:  33.3
Выборочная дисперсия этой выборки  равна:  1566.81

Выборка 3 объема 100:
Выборочное среднее этой выборки равно:  44.52
Выборочная дисперсия этой выборки  равна:  1540.57

Выборка 4 объема 100:
Выборочное среднее этой выборки равно:  64.24
Выборочная дисперсия этой выборки  равна:  1604.11

Выборка 5 объема 100:
Выборочное среднее этой выборки равно:  42.12
Выборочная дисперсия этой выборки  равна:  1658.47
```

n = 100

```
Выборка 1 объема 1000:
Выборочное среднее этой выборки равно:  142.901
Выборочная дисперсия этой выборки  равна:  18071.18

Выборка 2 объема 1000:
Выборочное среднее этой выборки равно:  156.636
Выборочная дисперсия этой выборки  равна:  18001.53

Выборка 3 объема 1000:
Выборочное среднее этой выборки равно:  131.48
Выборочная дисперсия этой выборки  равна:  17941.4

Выборка 4 объема 1000:
Выборочное среднее этой выборки равно:  127.764
Выборочная дисперсия этой выборки  равна:  18000.1

Выборка 5 объема 1000:
Выборочное среднее этой выборки равно:  129.98
Выборочная дисперсия этой выборки  равна:  18007.92
```

n = 1000

```
# Для n =  10**5
M = make_vibor(10**5)
final_func_2(M, 10**5)
```

```
Выборка 1 объема 100000:
Выборочное среднее этой выборки равно:  397.51218
Выборочная дисперсия этой выборки  равна:  151030.27

Выборка 2 объема 100000:
Выборочное среднее этой выборки равно:  437.08329
Выборочная дисперсия этой выборки  равна:  151019.46

Выборка 3 объема 100000:
Выборочное среднее этой выборки равно:  422.30579
Выборочная дисперсия этой выборки  равна:  151048.58

Выборка 4 объема 100000:
Выборочное среднее этой выборки равно:  478.38213
Выборочная дисперсия этой выборки  равна:  151041.75

Выборка 5 объема 100000:
Выборочное среднее этой выборки равно:  413.44772
Выборочная дисперсия этой выборки  равна:  151024.1
```

n = 100000

Свойства выборочного среднего

• Выборочное среднее — несмещённая оценка теоретического среднего:

• Выборочное среднее — сильно состоятельная оценка теоретического среднего:

• Выборочное среднее — асимптотически нормальная оценка.

• Выборочное среднее из нормальной выборки — эффективная оценка её среднего.

3.1.2 Экспоненциальное распределение

Для каждой выборки из домашнего задания 2 посчитаем выборочное среднее и выборочную дисперсию. Для наглядности выведем вариационный ряд для объема 5 и 10 и посчитанные оба параметра.

Вариационный ряд выборки 1 объема 5:

	0.316	0.985	1.532	2.092	2.286
f	1	1	1	1	1

Выборочное среднее этой выборки равно: 1.442
Выборочная дисперсия этой выборки равна: 0.524

Вариационный ряд выборки 2 объема 5:

	0.269	1.234	2.641	2.835	3.352
f	1	1	1	1	1

Выборочное среднее этой выборки равно: 2.066
Выборочная дисперсия этой выборки равна: 1.689

Вариационный ряд выборки 3 объема 5:

	0.06	0.199	0.534	1.728	2.804
f	1	1	1	1	1

Выборочное среднее этой выборки равно: 1.065
Выборочная дисперсия этой выборки равна: 1.243

Вариационный ряд выборки 4 объема 5:

	0.035	0.18	0.302	0.524	0.883
f	1	1	1	1	1

Выборочное среднее этой выборки равно: 0.385
Выборочная дисперсия этой выборки равна: 1.206

Вариационный ряд выборки 5 объема 5:

	0.354	0.432	0.715	0.961	1.856
f	1	1	1	1	1

Выборочное среднее этой выборки равно: 0.864
Выборочная дисперсия этой выборки равна: 0.627

$n = 5$

Вариационный ряд выборки 1 объема 10:

	0.179	0.328	0.36	0.647	0.907	1.039	1.193	2.05	4.293	5.423
f	1	1	1	1	1	1	1	1	1	1

Выборочное среднее этой выборки равно: 1.642
Выборочная дисперсия этой выборки равна: 3.0

Вариационный ряд выборки 2 объема 10:

	0.069	0.313	0.341	0.43	0.768	0.825	1.756	2.278	2.303	2.769
f	1	1	1	1	1	1	1	1	1	1

Выборочное среднее этой выборки равно: 1.185
Выборочная дисперсия этой выборки равна: 1.0

Вариационный ряд выборки 3 объема 10:

	0.103	0.222	0.228	0.703	0.884	0.92	0.986	2.495	6.193
f	1	1	1	1	1	1	1	2	1

Выборочное среднее этой выборки равно: 1.273
Выборочная дисперсия этой выборки равна: 3.0

Вариационный ряд выборки 4 объема 10:

	0.063	0.452	0.492	0.553	0.773	0.816	1.391	1.686	1.863	6.171
f	1	1	1	1	1	1	1	1	1	1

Выборочное среднее этой выборки равно: 1.426
Выборочная дисперсия этой выборки равна: 3.0

Вариационный ряд выборки 5 объема 10:

	0.041	0.223	0.257	0.47	0.573	0.848	0.882	1.287	1.576	3.593
f	1	1	1	1	1	1	1	1	1	1

Выборочное среднее этой выборки равно: 0.975
Выборочная дисперсия этой выборки равна: 1.0

$n = 10$

```
Выборка 1 объема 100:
Выборочное среднее этой выборки равно:  1.161
Выборочная дисперсия этой выборки  равна:  1.236

Выборка 2 объема 100:
Выборочное среднее этой выборки равно:  1.108
Выборочная дисперсия этой выборки  равна:  1.48

Выборка 3 объема 100:
Выборочное среднее этой выборки равно:  1.113
Выборочная дисперсия этой выборки  равна:  1.596

Выборка 4 объема 100:
Выборочное среднее этой выборки равно:  1.232
Выборочная дисперсия этой выборки  равна:  1.631

Выборка 5 объема 100:
Выборочное среднее этой выборки равно:  1.228
Выборочная дисперсия этой выборки  равна:  1.092
```

$n = 100$

```
Выборка 1 объема 1000:
Выборочное среднее этой выборки равно:  1.164
Выборочная дисперсия этой выборки  равна:  1.598

Выборка 2 объема 1000:
Выборочное среднее этой выборки равно:  1.122
Выборочная дисперсия этой выборки  равна:  1.569

Выборка 3 объема 1000:
Выборочное среднее этой выборки равно:  1.213
Выборочная дисперсия этой выборки  равна:  1.853

Выборка 4 объема 1000:
Выборочное среднее этой выборки равно:  1.172
Выборочная дисперсия этой выборки  равна:  1.637

Выборка 5 объема 1000:
Выборочное среднее этой выборки равно:  1.149
Выборочная дисперсия этой выборки  равна:  1.641
```

$n = 1000$

```
Выборка 1 объема 100000:
Выборочное среднее этой выборки равно:  8.929
Выборочная дисперсия этой выборки  равна:  60.515

Выборка 2 объема 100000:
Выборочное среднее этой выборки равно:  8.516
Выборочная дисперсия этой выборки  равна:  60.588

Выборка 3 объема 100000:
Выборочное среднее этой выборки равно:  8.336
Выборочная дисперсия этой выборки  равна:  60.631

Выборка 4 объема 100000:
Выборочное среднее этой выборки равно:  7.213
Выборочная дисперсия этой выборки  равна:  60.599

Выборка 5 объема 100000:
Выборочное среднее этой выборки равно:  5.339
Выборочная дисперсия этой выборки  равна:  60.519
```

$n = 100000$

Задание 3.2 Построение доверительного интервала для выборочного среднего и выборочной дисперсии

3.2.1 Геометрическое распределение

Для всех выборок построим доверительные интервалы.

```
func_for_2_ex(5)
```

```
Для 1 реализации выборки объема 5 доверительный интервал равен: (2.843 <= a <=  13.557)
Для 2 реализации выборки объема 5 доверительный интервал равен: (-8.925 <= a <=  25.325)
Для 3 реализации выборки объема 5 доверительный интервал равен: (0.706 <= a <=  15.694)
Для 4 реализации выборки объема 5 доверительный интервал равен: (0.882 <= a <=  15.518)
Для 5 реализации выборки объема 5 доверительный интервал равен: (-0.012 <= a <=  16.412)
```

n = 5

```
func_for_2_ex(10)
```

```
Для 1 реализации выборки объема 10 доверительный интервал равен: (1.133 <= a <=  20.067)
Для 2 реализации выборки объема 10 доверительный интервал равен: (3.706 <= a <=  17.494)
Для 3 реализации выборки объема 10 доверительный интервал равен: (4.917 <= a <=  16.283)
Для 4 реализации выборки объема 10 доверительный интервал равен: (5.407 <= a <=  15.793)
Для 5 реализации выборки объема 10 доверительный интервал равен: (5.74 <= a <=  15.46)
```

n = 10

```
func_for_2_ex(100)
```

```
Для 1 реализации выборки объема 100 доверительный интервал равен: (50.101 <= a <=  70.859)
Для 2 реализации выборки объема 100 доверительный интервал равен: (50.093 <= a <=  70.867)
Для 3 реализации выборки объема 100 доверительный интервал равен: (49.781 <= a <=  71.179)
Для 4 реализации выборки объема 100 доверительный интервал равен: (50.297 <= a <=  70.663)
Для 5 реализации выборки объема 100 доверительный интервал равен: (50.006 <= a <=  70.954)
```

n = 100

```
func_for_2_ex(1000)
```

```
Для 1 реализации выборки объема 1000 доверительный интервал равен: (115.706 <= a <=  129.91)
Для 2 реализации выборки объема 1000 доверительный интервал равен: (115.727 <= a <=  129.889)
Для 3 реализации выборки объема 1000 доверительный интервал равен: (115.696 <= a <=  129.92)
Для 4 реализации выборки объема 1000 доверительный интервал равен: (115.698 <= a <=  129.918)
Для 5 реализации выборки объема 1000 доверительный интервал равен: (115.724 <= a <=  129.892)
```

n = 1000

```
func_for_2_ex(10**5)
```

```
Для 1 реализации выборки объема 100000 доверительный интервал равен: (419.841 <= a <=  424.967)
Для 2 реализации выборки объема 100000 доверительный интервал равен: (419.841 <= a <=  424.968)
Для 3 реализации выборки объема 100000 доверительный интервал равен: (419.841 <= a <=  424.967)
Для 4 реализации выборки объема 100000 доверительный интервал равен: (419.841 <= a <=  424.967)
Для 5 реализации выборки объема 100000 доверительный интервал равен: (419.841 <= a <=  424.967)
```

n = 100000

3.2.2 Экспоненциальное распределение

Для всех выборок построим доверительные интервалы.

```
func_for_2_ex(5)
```

```
Для 1 реализации выборки объема 5 доверительный интервал равен: (2.843 <= a <=  13.557)
Для 2 реализации выборки объема 5 доверительный интервал равен: (-8.925 <= a <=  25.325)
Для 3 реализации выборки объема 5 доверительный интервал равен: (0.706 <= a <=  15.694)
Для 4 реализации выборки объема 5 доверительный интервал равен: (0.882 <= a <=  15.518)
Для 5 реализации выборки объема 5 доверительный интервал равен: (-0.012 <= a <=  16.412)
```

$n = 5$

```
func_for_2_ex(10)
```

```
Для 1 реализации выборки объема 10 доверительный интервал равен: (1.133 <= a <=  20.067)
Для 2 реализации выборки объема 10 доверительный интервал равен: (3.706 <= a <=  17.494)
Для 3 реализации выборки объема 10 доверительный интервал равен: (4.917 <= a <=  16.283)
Для 4 реализации выборки объема 10 доверительный интервал равен: (5.407 <= a <=  15.793)
Для 5 реализации выборки объема 10 доверительный интервал равен: (5.74 <= a <=  15.46)
```

$n = 10$

```
func_for_2_ex(100)
```

```
Для 1 реализации выборки объема 100 доверительный интервал равен: (50.101 <= a <=  70.859)
Для 2 реализации выборки объема 100 доверительный интервал равен: (50.093 <= a <=  70.867)
Для 3 реализации выборки объема 100 доверительный интервал равен: (49.781 <= a <=  71.179)
Для 4 реализации выборки объема 100 доверительный интервал равен: (50.297 <= a <=  70.663)
Для 5 реализации выборки объема 100 доверительный интервал равен: (50.006 <= a <=  70.954)
```

$n = 100$

```
func_for_2_ex(1000)
```

```
Для 1 реализации выборки объема 1000 доверительный интервал равен: (115.706 <= a <=  129.91)
Для 2 реализации выборки объема 1000 доверительный интервал равен: (115.727 <= a <=  129.889)
Для 3 реализации выборки объема 1000 доверительный интервал равен: (115.696 <= a <=  129.92)
Для 4 реализации выборки объема 1000 доверительный интервал равен: (115.698 <= a <=  129.918)
Для 5 реализации выборки объема 1000 доверительный интервал равен: (115.724 <= a <=  129.892)
```

$n = 1000$

```
func_for_2_ex(10**5)
```

```
Для 1 реализации выборки объема 100000 доверительный интервал равен: (419.841 <= a <=  424.967)
Для 2 реализации выборки объема 100000 доверительный интервал равен: (419.841 <= a <=  424.968)
Для 3 реализации выборки объема 100000 доверительный интервал равен: (419.841 <= a <=  424.967)
Для 4 реализации выборки объема 100000 доверительный интервал равен: (419.841 <= a <=  424.967)
Для 5 реализации выборки объема 100000 доверительный интервал равен: (419.841 <= a <=  424.967)
```

$n = 100000$

Задание 3.3 Нахождение параметров распределений событий

3.3.1 Геометрическое распределение

Найдем оценку для геометрического распределения

$$P(X = k) = q^{k-1}p$$

Найдём истинные первый и центральный второй моменты геометрического распределения:

$$\overline{x} = M_1 = \sum_{k=1}^{\infty} pq^{k-1}k = p\sum_{k=1}^{\infty} q^{k-1}k = \frac{1}{p}$$

$$D = M_c^2 = E(x - \overline{x})^2 = \sum_{k=1}^{\infty} q^{k-1}k^2 = \frac{q}{p^3} = M_1\frac{q}{p^2}$$

Метод моментов

По методу моментов находим оценку параметра p: $p = \dfrac{1}{x}$

Точечной оценкой параметра p является $\dfrac{1}{\overline{x}}$

Метод максимального правдоподобия

$$L(x_1, x_2, \cdots, x_n; \theta) = p^n q^{n(\theta-1)}$$

$$\ln(L) = n\ln(p) + n(\theta - 1)\ln(q)$$

$$\frac{\partial \ln(L)}{\partial \theta} = n\ln q = 0$$

Минимума или максимума нет. Т.е. метод максимального правдоподобия не работает в случае геометрического распределения

3.3.2 Экспоненциальное распределение

Оценка методом моментов

Математическое ожидание равно $Ex = \theta + \dfrac{1}{\lambda}$

Дисперсия равна $Dx = \dfrac{1}{\lambda^2}$

$$\bar{x} = \theta + \frac{1}{\lambda}$$

Оценка:

$$\theta_{\text{омм}} = \bar{x} - \frac{1}{\lambda}$$

Проверим, является ли данная оценка несмещенной

$$E_\theta T = E\left(\bar{X} - \frac{1}{\lambda}\right) = E\bar{X} - \frac{1}{\lambda} = \theta + \frac{1}{\lambda} - \frac{1}{\lambda} = \theta$$

Следовательно, оценка $\theta_{\text{омм}}$ несмещенная оценка параметра θ

$$D\theta_{\text{омм}} = D\left(\bar{X} - \frac{1}{\lambda}\right) = D(\bar{X}) + 0 = D(\bar{X}) = \frac{1}{n^2} D\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\frac{1}{\lambda^2} = \frac{1}{n\lambda^2} \to 0ghbn$$

Следовательно, $\theta_{\text{омм}}$ состоятельная оценка параметра θ

Оценка максимальным правдоподобием

$$\theta_{\text{омм}} = argsupL(x;\theta)$$

$$L(x;\theta) = \lambda^n e^{-\lambda \sum_{i=1}^{n}(x_i - \theta)}$$

Воспользуемся тем, что экстремумы $L(x;\theta)$ и $\ln L(x;\theta)$ совпадают

$$\ln L(x;\theta) = \ln \lambda^n e^{-\lambda \sum_{i=1}^{n}(x_i - \theta)} = n \ln \lambda - n\lambda\bar{X} + n\lambda\theta$$

$$\frac{\partial \ln L(x;\theta)}{\partial \theta} = n\lambda > 0$$

Чтобы максимизировать функцию правдоподобия, возьмем максимальное θ

$$\theta_{\text{омм}} = X_{(1)}$$

Проверим несмещенность оценки

$$F_{(1)}(x) = 1 - (1 - F(x))^n = 1 - (1 - 1 + e^{-\lambda(x-\theta)})^n = 1 - \left(\frac{e^{\lambda n}}{e^{\lambda\theta}}\right)^n = 1 - \frac{e^{\lambda\theta n}}{e^{\lambda x n}}$$

$$f_{(1)}(x) = F'_(1)(x) = \frac{e^{\lambda n}\lambda n}{e^{\lambda x n}} = \lambda n e^{-\lambda n(x-\theta)}$$

$$\begin{aligned}
E_\theta x_{(1)} &= \\
\int_0^\infty x f_{(1)}(x;\theta)dx &= \\
\int_0^\infty x\lambda n e^{-\lambda n(x-\theta} &= \\
\lambda n e^{\lambda n\theta}\left(-\frac{xe^{-\lambda xn}}{\lambda n} + \frac{1}{\lambda n}\int_0^\infty e^{-\lambda nx}dx\right) &= \\
-xe^{-\lambda n(x-\theta)} \mid_0^\infty - \frac{\lambda n e^{\lambda n(x-\theta)}}{\lambda^2 n^2} \mid_0^\infty &= \\
\theta + \frac{1}{\lambda n}
\end{aligned}$$

Таким образом, оценка максимальным правдоподобием смещена со смещением $\frac{1}{\lambda n}$

$$\theta = X_{(1)} - \frac{1}{\lambda n}$$

Найдем дисперсию несмещенной оценки:

$$D\theta_{\text{омп}} = D\left(X_{(1)-\frac{1}{\lambda n}}\right) = EX_{(1)}^2 - (EX_1)^2 = \theta^2 - \frac{2\theta}{\lambda n} - \frac{2}{\lambda^2 n^2} - \left(\theta + \frac{1}{\lambda n}\right)^2 = \frac{1}{(\lambda n)^2}$$

$D\theta_{\text{омп}} \to 0$ при $n \to \infty \to$ оценка состоятельна

Эффективная оценка

x зависит от параметра θ $\rightarrow$ модель не регулярна $\rightarrow$ не существует эффективной оценки

Литература

[1]

[2] ссылка1

[3] ссылка2

[4] // ссылка3

[5] // ссылка4