Математическая статистика

Домашняя работа № 3

Оценки

Попов Юрий, СКБ-172

ОГЛАВЛЕНИЕ

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

Подробности работа с данными находятся в двух других юпитерских файлах: "Database_Geom_ "Database_Expon_3_New.ipynb в которых находятся расчеты для геометрически и экспоненциально, распределенных данных

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	5	7	16
f	1	2	1	1

Выборочное среднее этой выборки равно: 6.8
Выборочная дисперсия этой выборки равна: 24.96

Вариационный ряд выборки 2 объема 5:

	0	1	4	5	14
f	1	1	1	1	1

Выборочное среднее этой выборки равно: 4.8
Выборочная дисперсия этой выборки равна: 24.56

Вариационный ряд выборки 3 объема 5:

	4	5	6	8	19
f	1	1	1	1	1

Выборочное среднее этой выборки равно: 8.4
Выборочная дисперсия этой выборки равна: 29.84

Вариационный ряд выборки 4 объема 5:

	1	2	6	26
f	2	1	1	1

Выборочное среднее этой выборки равно: 7.2
Выборочная дисперсия этой выборки равна: 91.76

Вариационный ряд выборки 5 объема 5:

	0	5	17
f	3	1	1

Выборочное среднее этой выборки равно: 4.4
Выборочная дисперсия этой выборки равна: 43.44

n = 5

Вариационный ряд выборки 1 объема 10:

	0	1	2	3	4	6
f	1	3	1	1	3	1

Выборочное среднее этой выборки равно: 2.6
Выборочная дисперсия этой выборки равна: 3.24

Вариационный ряд выборки 2 объема 10:

	0	1	2	3	4	10	14
f	3	2	1	1	1	1	1

Выборочное среднее этой выборки равно: 3.5
Выборочная дисперсия этой выборки равна: 20.45

Вариационный ряд выборки 3 объема 10:

	0	2	3	5	6	10	11	12
f	2	1	1	2	1	1	1	1

Выборочное среднее этой выборки равно: 5.4
Выборочная дисперсия этой выборки равна: 17.24

Вариационный ряд выборки 4 объема 10:

	0	1	2	8	10
f	3	4	1	1	1

Выборочное среднее этой выборки равно: 2.4
Выборочная дисперсия этой выборки равна: 11.44

Вариационный ряд выборки 5 объема 10:

	0	1	2	4	10	11	13	22
f	3	1	1	1	1	1	1	1

Выборочное среднее этой выборки равно: 6.3
Выборочная дисперсия этой выборки равна: 49.81

n = 10

```
Выборочное среднее этой выборки равно:  4.08
Выборочная дисперсия этой выборки  равна:  17.854

Выборочное среднее этой выборки равно:  4.46
Выборочная дисперсия этой выборки  равна:  20.268

Выборочное среднее этой выборки равно:  4.11
Выборочная дисперсия этой выборки  равна:  19.878

Выборочное среднее этой выборки равно:  3.23
Выборочная дисперсия этой выборки  равна:  15.217

Выборочное среднее этой выборки равно:  3.88
Выборочная дисперсия этой выборки  равна:  13.126
```

n = 100

```
Выборочное среднее этой выборки равно:  4.176
Выборочная дисперсия этой выборки  равна:  22.637

Выборочное среднее этой выборки равно:  3.931
Выборочная дисперсия этой выборки  равна:  18.76

Выборочное среднее этой выборки равно:  3.951
Выборочная дисперсия этой выборки  равна:  18.587

Выборочное среднее этой выборки равно:  4.093
Выборочная дисперсия этой выборки  равна:  21.036

Выборочное среднее этой выборки равно:  3.937
Выборочная дисперсия этой выборки  равна:  17.621
```

n = 1000

```
Выборочное среднее этой выборки равно:  4.001
Выборочная дисперсия этой выборки  равна:  19.796

Выборочное среднее этой выборки равно:  4.003
Выборочная дисперсия этой выборки  равна:  20.066

Выборочное среднее этой выборки равно:  4.01
Выборочная дисперсия этой выборки  равна:  19.955

Выборочное среднее этой выборки равно:  4.014
Выборочная дисперсия этой выборки  равна:  20.095

Выборочное среднее этой выборки равно:  4.029
Выборочная дисперсия этой выборки  равна:  20.277
```

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:

	0.45	0.453	1.318	1.724	2.874
f	1	1	1	1	1

Выборочное среднее этой выборки равно: 1.364
Выборочная дисперсия этой выборки равна: 0.815

Вариационный ряд выборки 2 объема 5:

	0.252	0.32	1.716	1.817	2.337
f	1	1	1	1	1

Выборочное среднее этой выборки равно: 1.288
Выборочная дисперсия этой выборки равна: 0.715

Вариационный ряд выборки 3 объема 5:

	0.598	0.828	1.175	1.44	1.67
f	1	1	1	1	1

Выборочное среднее этой выборки равно: 1.142
Выборочная дисперсия этой выборки равна: 0.153

Вариационный ряд выборки 4 объема 5:

	0.037	0.371	1.197	2.05	2.731
f	1	1	1	1	1

Выборочное среднее этой выборки равно: 1.277
Выборочная дисперсия этой выборки равна: 1.015

Вариационный ряд выборки 5 объема 5:

	0.66	0.807	1.007	1.356	1.451
f	1	1	1	1	1

Выборочное среднее этой выборки равно: 1.056
Выборочная дисперсия этой выборки равна: 0.093

$n = 5$

Вариационный ряд выборки 1 объема 10:

	0.178	0.294	0.387	0.475	0.524	0.589	0.634	1.077	1.316	3.224
f	1	1	1	1	1	1	1	1	1	1

Выборочное среднее этой выборки равно: 0.87
Выборочная дисперсия этой выборки равна: 0.724

Вариационный ряд выборки 2 объема 10:

	0.063	0.123	0.325	0.497	0.531	1.217	1.593	2.098	4.378	4.776
f	1	1	1	1	1	1	1	1	1	1

Выборочное среднее этой выборки равно: 1.56
Выборочная дисперсия этой выборки равна: 2.671

Вариационный ряд выборки 3 объема 10:

	0.279	0.43	0.82	1.117	1.308	1.49	1.914	2.142	2.168	2.405
f	1	1	1	1	1	1	1	1	1	1

Выборочное среднее этой выборки равно: 1.407
Выборочная дисперсия этой выборки равна: 0.504

Вариационный ряд выборки 4 объема 10:

	0.14	0.46	0.608	0.704	0.934	1.157	1.218	2.16	2.833	3.426
f	1	1	1	1	1	1	1	1	1	1

Выборочное среднее этой выборки равно: 1.364
Выборочная дисперсия этой выборки равна: 1.062

Вариационный ряд выборки 5 объема 10:

	0.261	0.376	0.407	0.563	0.845	0.901	0.992	2.212	2.224	2.494
f	1	1	1	1	1	1	1	1	1	1

Выборочное среднее этой выборки равно: 1.128
Выборочная дисперсия этой выборки равна: 0.655

$n = 10$

```
Выборочное среднее этой выборки равно:  1.114
Выборочная дисперсия этой выборки  равна:  0.807

Выборочное среднее этой выборки равно:  1.566
Выборочная дисперсия этой выборки  равна:  2.187

Выборочное среднее этой выборки равно:  1.147
Выборочная дисперсия этой выборки  равна:  1.117

Выборочное среднее этой выборки равно:  1.311
Выборочная дисперсия этой выборки  равна:  1.487

Выборочное среднее этой выборки равно:  1.005
Выборочная дисперсия этой выборки  равна:  0.851
```

n = 100

```
Выборочное среднее этой выборки равно:  1.183
Выборочная дисперсия этой выборки  равна:  1.255

Выборочное среднее этой выборки равно:  1.274
Выборочная дисперсия этой выборки  равна:  1.686

Выборочное среднее этой выборки равно:  1.238
Выборочная дисперсия этой выборки  равна:  1.623

Выборочное среднее этой выборки равно:  1.281
Выборочная дисперсия этой выборки  равна:  1.881

Выборочное среднее этой выборки равно:  1.271
Выборочная дисперсия этой выборки  равна:  1.475
```

n = 1000

```
Выборочное среднее этой выборки равно:  1.252
Выборочная дисперсия этой выборки  равна:  1.582

Выборочное среднее этой выборки равно:  1.25
Выборочная дисперсия этой выборки  равна:  1.562

Выборочное среднее этой выборки равно:  1.251
Выборочная дисперсия этой выборки  равна:  1.568

Выборочное среднее этой выборки равно:  1.248
Выборочная дисперсия этой выборки  равна:  1.549

Выборочное среднее этой выборки равно:  1.248
Выборочная дисперсия этой выборки  равна:  1.549
```

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.

```
Для 1 реализации выборки объема 5 доверительный интервал равен: (0.442 <= a <=  7.648)
Для 2 реализации выборки объема 5 доверительный интервал равен: (-0.106 <= a <=  8.196)
Для 3 реализации выборки объема 5 доверительный интервал равен: (-0.327 <= a <=  8.417)
Для 4 реализации выборки объема 5 доверительный интервал равен: (-0.445 <= a <=  8.535)
Для 5 реализации выборки объема 5 доверительный интервал равен: (0.31 <= a <=  7.78)
```

n = 5

```
Для 1 реализации выборки объема 10 доверительный интервал равен: (1.895 <= a <=  6.195)
Для 2 реализации выборки объема 10 доверительный интервал равен: (0.713 <= a <=  7.377)
Для 3 реализации выборки объема 10 доверительный интервал равен: (-2.915 <= a <=  11.005)
Для 4 реализации выборки объема 10 доверительный интервал равен: (-0.79 <= a <=  8.88)
Для 5 реализации выборки объема 10 доверительный интервал равен: (2.036 <= a <=  6.054)
```

n = 10

```
Для 1 реализации выборки объема 100 доверительный интервал равен: (3.25 <= a <=  4.84)
Для 2 реализации выборки объема 100 доверительный интервал равен: (3.077 <= a <=  5.013)
Для 3 реализации выборки объема 100 доверительный интервал равен: (3.246 <= a <=  4.844)
Для 4 реализации выборки объема 100 доверительный интервал равен: (3.188 <= a <=  4.902)
Для 5 реализации выборки объема 100 доверительный интервал равен: (3.115 <= a <=  4.975)
```

n = 100

```
Для 1 реализации выборки объема 1000 доверительный интервал равен: (3.765 <= a <=  4.325)
Для 2 реализации выборки объема 1000 доверительный интервал равен: (3.762 <= a <=  4.328)
Для 3 реализации выборки объема 1000 доверительный интервал равен: (3.782 <= a <=  4.308)
Для 4 реализации выборки объема 1000 доверительный интервал равен: (3.76 <= a <=  4.33)
Для 5 реализации выборки объема 1000 доверительный интервал равен: (3.728 <= a <=  4.362)
```

n = 1000

```
Для 1 реализации выборки объема 100000 доверительный интервал равен: (4.017 <= a <=  4.073)
Для 2 реализации выборки объема 100000 доверительный интервал равен: (4.017 <= a <=  4.073)
Для 3 реализации выборки объема 100000 доверительный интервал равен: (4.017 <= a <=  4.073)
Для 4 реализации выборки объема 100000 доверительный интервал равен: (4.017 <= a <=  4.073)
Для 5 реализации выборки объема 100000 доверительный интервал равен: (4.017 <= a <=  4.073)
```

n = 100000

3.2.2 Экспоненциальное распределение

Для всех выборок построим доверительные интервалы.

```
Для 1 реализации выборки объема 5 доверительный интервал равен: (0.442 <= a <=  7.648)
Для 2 реализации выборки объема 5 доверительный интервал равен: (-0.106 <= a <=  8.196)
Для 3 реализации выборки объема 5 доверительный интервал равен: (-0.327 <= a <=  8.417)
Для 4 реализации выборки объема 5 доверительный интервал равен: (-0.445 <= a <=  8.535)
Для 5 реализации выборки объема 5 доверительный интервал равен: (0.31 <= a <=  7.78)
```

$n = 5$

```
Для 1 реализации выборки объема 10 доверительный интервал равен: (1.895 <= a <=  6.195)
Для 2 реализации выборки объема 10 доверительный интервал равен: (0.713 <= a <=  7.377)
Для 3 реализации выборки объема 10 доверительный интервал равен: (-2.915 <= a <=  11.005)
Для 4 реализации выборки объема 10 доверительный интервал равен: (-0.79 <= a <=  8.88)
Для 5 реализации выборки объема 10 доверительный интервал равен: (2.036 <= a <=  6.054)
```

$n = 10$

```
Для 1 реализации выборки объема 100 доверительный интервал равен: (3.25 <= a <=  4.84)
Для 2 реализации выборки объема 100 доверительный интервал равен: (3.077 <= a <=  5.013)
Для 3 реализации выборки объема 100 доверительный интервал равен: (3.246 <= a <=  4.844)
Для 4 реализации выборки объема 100 доверительный интервал равен: (3.188 <= a <=  4.902)
Для 5 реализации выборки объема 100 доверительный интервал равен: (3.115 <= a <=  4.975)
```

$n = 100$

```
Для 1 реализации выборки объема 1000 доверительный интервал равен: (3.765 <= a <=  4.325)
Для 2 реализации выборки объема 1000 доверительный интервал равен: (3.762 <= a <=  4.328)
Для 3 реализации выборки объема 1000 доверительный интервал равен: (3.782 <= a <=  4.308)
Для 4 реализации выборки объема 1000 доверительный интервал равен: (3.76 <= a <=  4.33)
Для 5 реализации выборки объема 1000 доверительный интервал равен: (3.728 <= a <=  4.362)
```

$n = 1000$

```
Для 1 реализации выборки объема 100000 доверительный интервал равен: (4.017 <= a <=  4.073)
Для 2 реализации выборки объема 100000 доверительный интервал равен: (4.017 <= a <=  4.073)
Для 3 реализации выборки объема 100000 доверительный интервал равен: (4.017 <= a <=  4.073)
Для 4 реализации выборки объема 100000 доверительный интервал равен: (4.017 <= a <=  4.073)
Для 5 реализации выборки объема 100000 доверительный интервал равен: (4.017 <= a <=  4.073)
```

$n = 100000$

Задание 3.3 Нахождение параметров распределений событий

Определение $\mathcal{F} = F_\theta(x|\theta \in \Theta)$ называется экспоненциальной, если

$$f(x;\theta) = exp(A(\theta) \star B(x) + C(\theta) + D(x)).$$

Для того, чтобы в модели существовала эффективная оценка, необходимо и достаточно, чтобы модель принадлежала экспоненциальному семейству.

Вклад выборки для экспоненциальной модели равен:

$$V(X;\theta) = A'(\theta)\sum_{i=1}^{n} B(X_i) + nC'(\theta) = nA'(\theta)[\frac{1}{n}\sum_{i=1}^{n} B(X_i) + \frac{C'(\theta)}{A'(\theta)}]$$

Это также можно записать в виде:

$$T(X) - \tau(\theta) = \alpha(\theta)V(x;\theta)$$

При этом:

$$\tau(\theta) = -\frac{C'(\theta)}{A'(\theta)}$$

$$\alpha(\theta) = \frac{1}{nA'(\theta)}$$

$$T^* = T^*(X) = \frac{1}{n}\sum_{i=1}^{n} B(X_i)$$

По критерию Рао-Крамера заключаем, что статистика T^* является эффективной оценкой для параметрической функции $\tau(\theta)$

Оба распределения относятся к экспоненциальному семейству

3.3.1 Геометрическое распределение

Найдем оценку для геометрического распределения

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

$$E[X] = \frac{1}{p}$$

$$D(X) = \frac{1-p}{p^2}$$

Оценка методом моментов

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

Точечной оценкой параметра p является $\frac{1}{\bar{x}}$

Оценка методом 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

Для данного распределения предположим оценку $\hat{\theta} = \frac{n-1}{n}\sum_{i=1}^{n} x_i$

Достаточная статистика

Найдем достаточную статистику:

$$f_n(x;\theta)=\prod_{i=1}^{n} f(x_i;\theta)=\prod_{i=1}^{n}\theta(1-\theta)^{x_i-1}=\theta^n(1-\theta)^{-n}(1-\theta)^{\sum_{i=0}^{n} x_i}=g(T(x),\theta)h(x)$$

Получаем, что $T(x)=\sum_{i=1}^{n} x_i$ - достаточная оценка, т.к выполняется критерий факторизации. В данном случае $h(x)=1$

Полная достаточная статистика

Эта статистика является полной достаточной статистикой

3.3.2 Экспоненциальное распределение

Рассмотри "сдвинутое"экспоненциальное распределение с плотность распределения $f(x)=\theta e^{-\theta(x-\theta)}, x\geq 0, \theta\geq 0$

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

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

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

Оценка максимальным правдоподобием

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

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

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

$$\theta_{\text{омм}} = X_{(1)}$$

Несмещенность оценки

Проверим несмещенность оценки

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

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

$$
\begin{aligned}
E_\theta x_{(1)} &= \\
&= \int_0^\infty x f_{(1)}(x;\theta)dx \\
&= \int_0^\infty x\lambda n e^{-\lambda n(x-\theta} \\
&= \lambda n e^{\lambda n\theta}\left(-\frac{xe^{-\lambda xn}}{\lambda n} + \frac{1}{\lambda n}\int_0^\infty e^{-\lambda nx}dx\right) \\
&= -xe^{-\lambda n(x-\theta)}|_0^\infty - \frac{\lambda n e^{\lambda n(x-\theta)}}{\lambda^2 n^2}|_0^\infty \\
&= \theta + \frac{1}{\lambda n}
\end{aligned}
$$

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 $\theta \to$ модель не регулярна $\to$ не существует эффективной оценки

Оптимальная оценка

Согласно теореме Рао-Блекуэлла-Колмогорова, оптимальная оценка, если она существует, является функцией от полной достаточной статистики. Найдем достаточную статистику.

Функция правдоподобия имеет вид:

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

Где $\nu(x) = I(x \geq 0), x \in R$ - это функция Хэвисайда

По критерию факторизации статистика $X_{(1)}$является достаточной статистикой.

Установим ее полноту

Пусть $\varphi(t), t > 0$ - произвольная функция. Найдем математическое ожидание случайной величины $\varphi(X_{(1)})$

$$E_\theta \varphi(X_{(1)}) = \int\limits_{\theta}^{\infty} \varphi(x) f_{(1)}(x;\theta) dx = \int\limits_{\theta}^{\infty} \varphi(x) \lambda n e^{-\lambda n(x-\theta)} dx = \lambda n e^{-\lambda n\theta} \int\limits_{\theta}^{\infty} \varphi(x) e^{-\lambda nx} dx$$

Из того $E_\theta \varphi(X_{(1)}) = 0$ для любого $\theta \in \Xi$ следует, что $\int\limits_{\theta}^{\infty} \varphi(x) e^{-\lambda nx} dx = 0$. Докажем, что $\varphi(\theta) = 0, t \geq 0$

Продифференцируем полученный интеграл по параметру θ. Тогда

$$\left(\int\limits_{\theta}^{\infty} \varphi(x) e^{-\lambda nx} dx \right)' = \varphi(\theta) e^{\lambda\theta n} = 0 \text{ для любого } \theta \in \Xi$$

Указанное равенство верно для любого $\theta \in \Xi$ только в случае $\varphi(t) = 0$, то есть $X_{(1)}$- полная достаточная статистика. Поскольку $E_\theta X_{(1)} = \theta + \dfrac{1}{\lambda n}$, то оптимальной оценкой для параметра θ является статистика $T = X_{(1)} - \dfrac{1}{\lambda n}$

Задание 3.4 Работа с данными

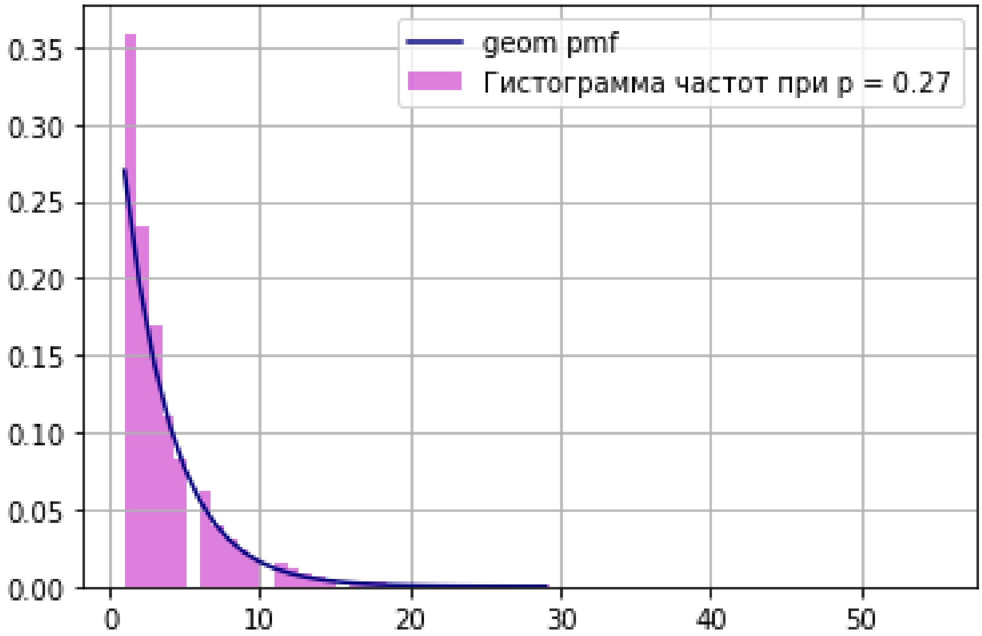
3.4.1 Геометрическое распределение

В перовой домашней работе в качестве нетипичной интерпретации геометрического распределения была выбрана модель, в которой рассматривается ДНК, то есть последовательность 4 видов нуклеотидов, и, например, перед исследователями стоит задача изучить частоту встречаемости какого-то определенного вида. Соответственно, без особых трудностей мной была найдена база данных, показывающая весь геном после применения метода дробовика.[3] Метод дробовика - метод, используемый для секвенирования длинных участков ДНК. Суть метода состоит в получении случайной массированной выборки клонированных фрагментов ДНК данного организма, на основе которых может быть восстановлена исходная последовательность ДНК.

Итак, моя выборка состоит из последовательности нуклеотидов 'A', 'C', 'G','T'. Анализируя эту выборку, я считал расстояние между двумя соседними нуклеотидами 'T'. Я начал двигаться по ДНК, установив при этом счетчик на значение 1. Соответственно, при каждой встрече нужного нуклеотида, значение счетчика заносилось в отдельный массив, затем счетчик устанавливался на значение 1 и движение продолжалось.

В результате анализа, я построил функцию вероятности моей выборки. Синим цветом - теоретически построена функция вероятности с подобранным коэффициентом.

Затем были посчитаны выборочное среднее и выборочная дисперсия. И согласно, полученной мною ранее оптимальной оценки, получена оценка параметра. Они приблизительно оказались равны!!!




```
Выборчоное среднее равно:  3.796
Выборочная дисперсия равна:  14.77
```

```
Оценка параметра равна: 0.263
```

3.4.2 Экспоненциальное распределение

В первой домашней работе в качетсве нетипичной интерпретации для экспоненциального распределения была выбрана теория надежности. Эта интерпретация позволяет исследовать срок службы прибора или механизма, и в нужный момент принять соответствующие меры по ремонту или замене деталей.

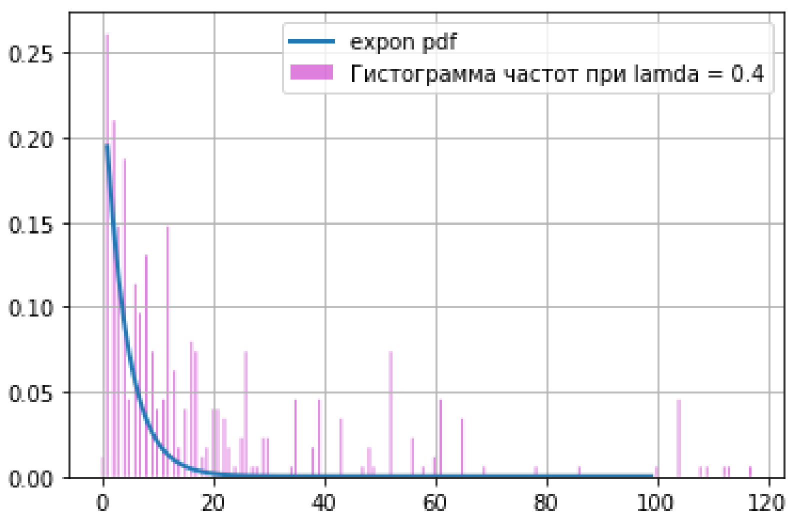
Я нашел базу данных, касающуюся безработицы. В этой базе данных главным фактором была приведена продолжить безработицы, то есть количество дней, которое она продолжалась

Почему, я считаю, что данная интерпретация соответствует экспоненциальному распределению? Работа, или другой похожий способ заработка, определяет доход человека, и соотвественно, его возможности. Если человек, по какой либо причине потеряет работу, то ему будет нечем себя кормить и не на что будет существовать. Но так, как, возможно, что у него(нее) есть накопления, то он сможет спокойно жить без работы, но так как иной прибыли денег нет, то в какой-то момент накопления будут заканчиваться, и жизнь заставит искать работу. Так что со временем, человек перестанет быть безработным

Этот процесс очень схож с жизнью приборов. Чем дольше прибор находится в эксплотации, тем выше вероятность, что он выйдет из строя, а в случае человека - найдет работу.

Для этих данных я проделал тоже самое, что и для предыдущих.

Получил вот такие результаты:



Они приблизительно оказались равны!!!

```
Выборчоное среднее равно:  18.511
Выборочная дисперсия равна:  531.732
```

```
Оценка параметра равна: 0.388
```

Литература

[1]

[2] Ссылка на базу данных для экспоненциального распределения,Unemployment, Unemployment Duration

[3] Ссылка на базу данных для геометрического распределения, NCBI - National Center of Biotechnology Information

[4] // ссылка3

[5] // ссылка4