Санкт-Петербургский политехнический университет Петра Великого

Кафедра компьютерных систем и программных технологий

**Отчёт по лабораторной работе**

**Дисциплина**: Теория вероятностей

**Тема**: Статистическая обработка случайных последовательностей. Идентификация законов распределения.

Выполнил студент гр. 23531/1 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ А.В. Медведев

(подпись)

Преподаватель \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ К.В. Никитин

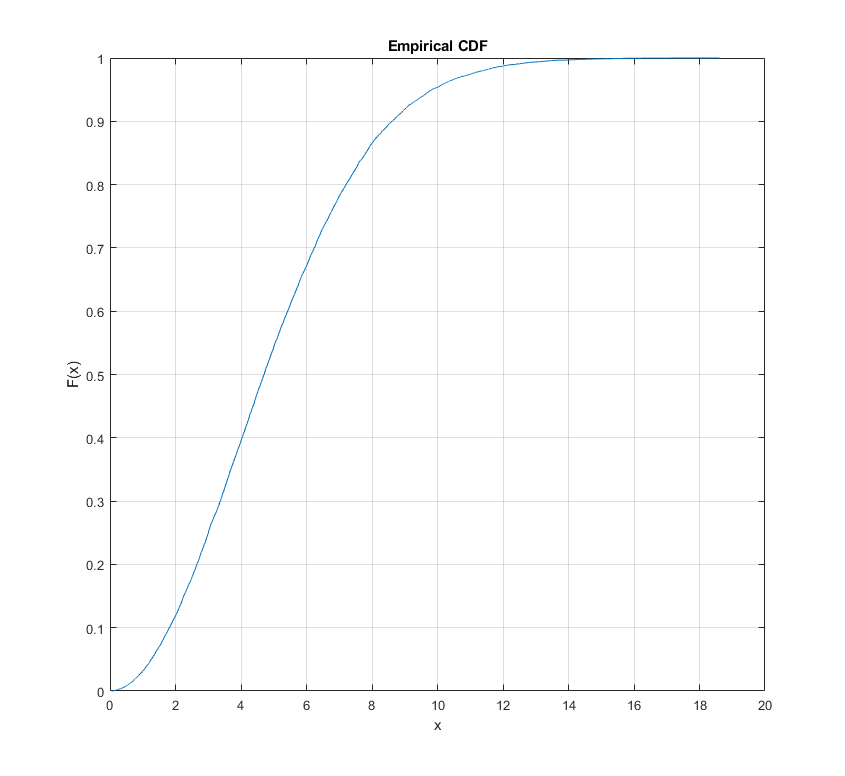
(подпись)  
“\_\_\_” \_\_\_\_\_\_\_\_\_\_\_\_ 2018 г.

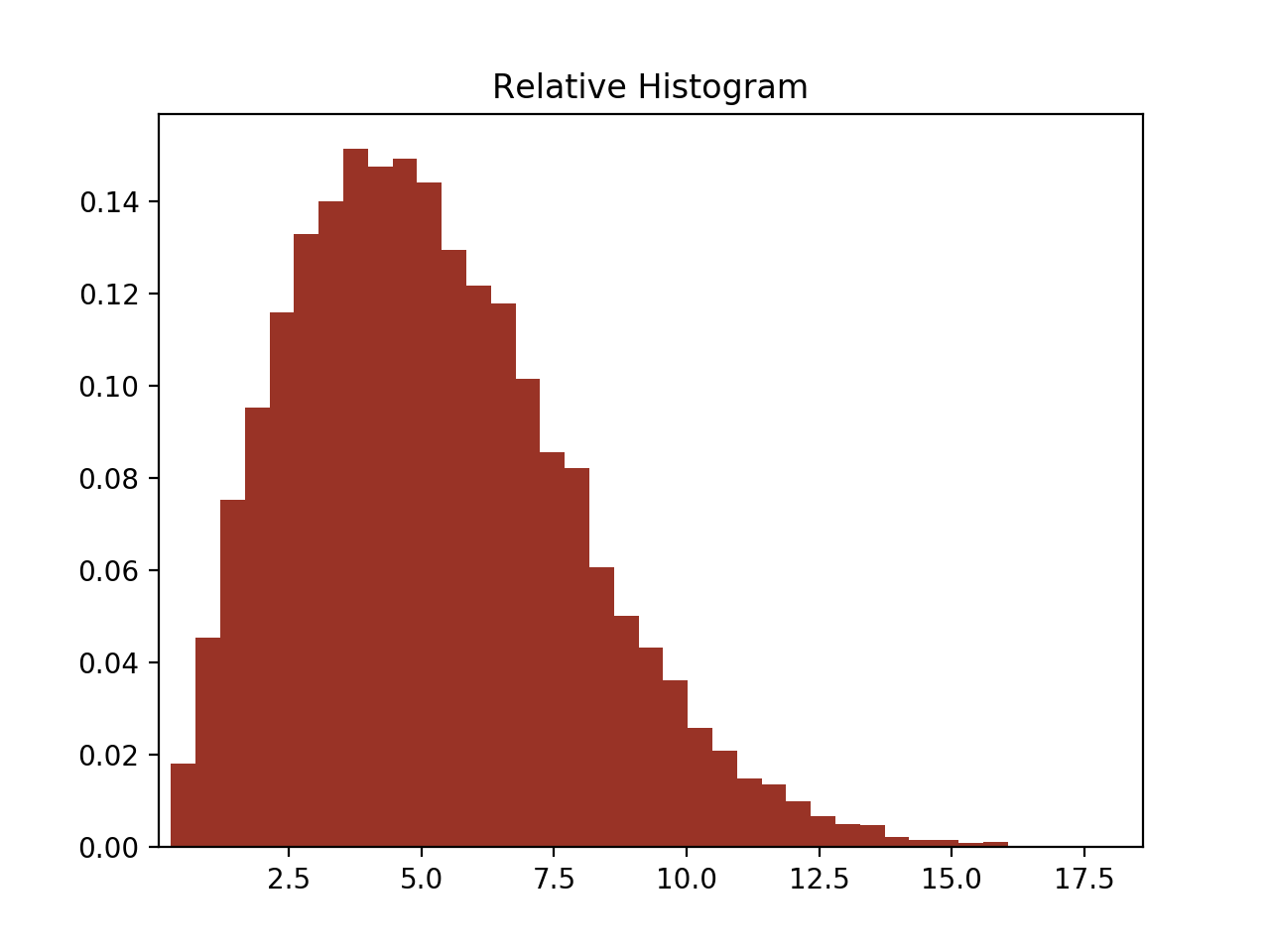
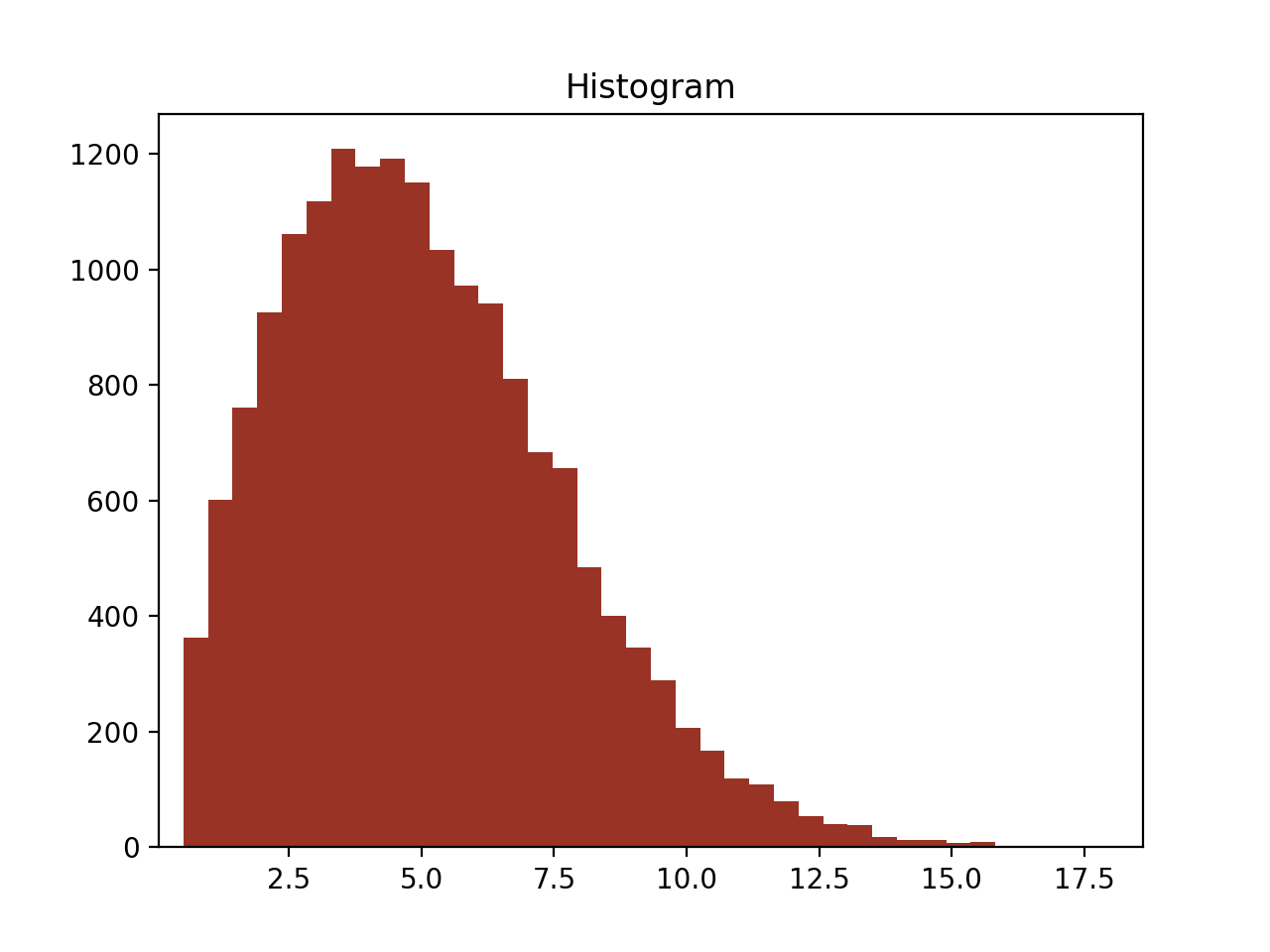
Санкт-Петербург  
2018

**1. Статистическая обработка экспериментальных данных**

**1.2 Выборочная функция распределения**

По исходным данным, находящимся в файле *Task\_2.txt* построена функция распределения и гистограмма. Количество интервалов m = 40.





Входные данные были перемешаны, после чего список был поделен на 10 равных частей. Далее по полной выборке и поп подвыборкам были посчитаны точечные оценки. Результаты представлены в таблице.

**1.3 Определение точечных оценок**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
| Full | 5.0195 | 4.7039 | 9.3254 | 2.6554 | 7.0514 | 12.3763 | 165.8085 | 2.8601 | 3.3346 |
| 1 | 4.9806 | 4.6711 | 8.2107 | 2.6557 | 7.0529 | 13.2004 | 167.8011 | 0.7047 | 3.3732 |
| 2 | 4.9375 | 4.5868 | 7.7843 | 2.6567 | 7.0581 | 14.1121 | 170.1543 | 0.7525 | 3.4155 |
| 3 | 5.0096 | 4.7056 | 9.3254 | 2.6554 | 7.0515 | 12.5874 | 166.3067 | 0.6722 | 3.3445 |
| 4 | 5.1182 | 4.8044 | 7.8826 | 2.6572 | 7.0611 | 10.2878 | 161.3351 | 0.5482 | 3.2357 |
| 5 | 5.1071 | 4.7423 | 7.8791 | 2.6568 | 7.0590 | 10.5230 | 161.7973 | 0.5610 | 3.2469 |
| 6 | 5.1386 | 4.7579 | 8.6287 | 2.6581 | 7.0656 | 9.85599 | 160.5139 | 0.5247 | 3.2152 |
| 7 | 4.9604 | 4.6771 | 7.6170 | 2.6561 | 7.0549 | 13.6279 | 168.8853 | 0.7272 | 3.3931 |
| 8 | 4.9382 | 4.6381 | 8.5915 | 2.6566 | 7.0580 | 14.0969 | 170.1138 | 0.7517 | 3.4148 |
| 9 | 4.9862 | 4.6876 | 7.9180 | 2.6556 | 7.0525 | 13.0808 | 167.5040 | 0.6984 | 3.3677 |
| 10 | 5.0189 | 4.7197 | 7.3362 | 2.6554 | 7.0514 | 12.3889 | 165.8380 | 0.6616 | 3.3352 |

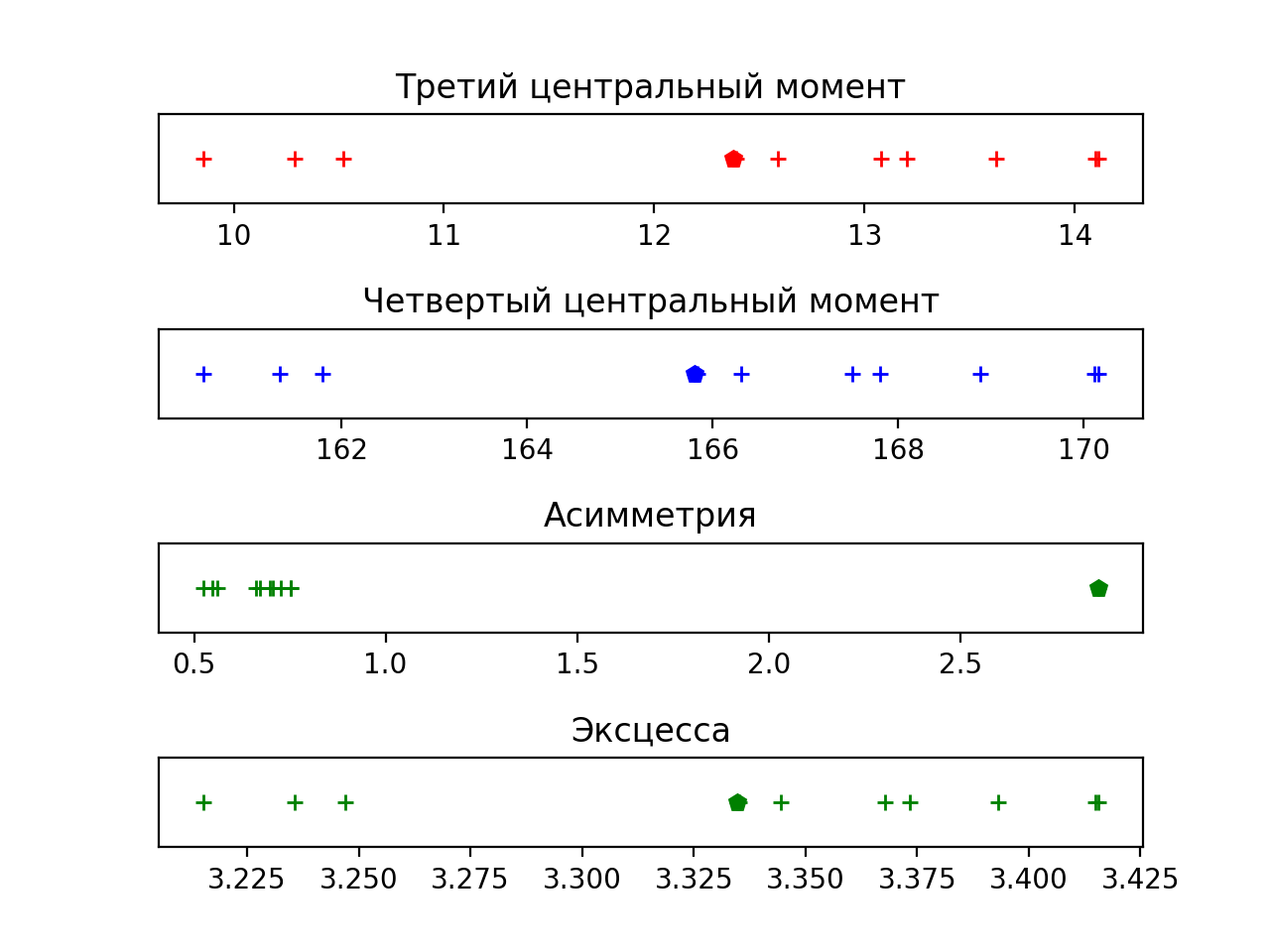
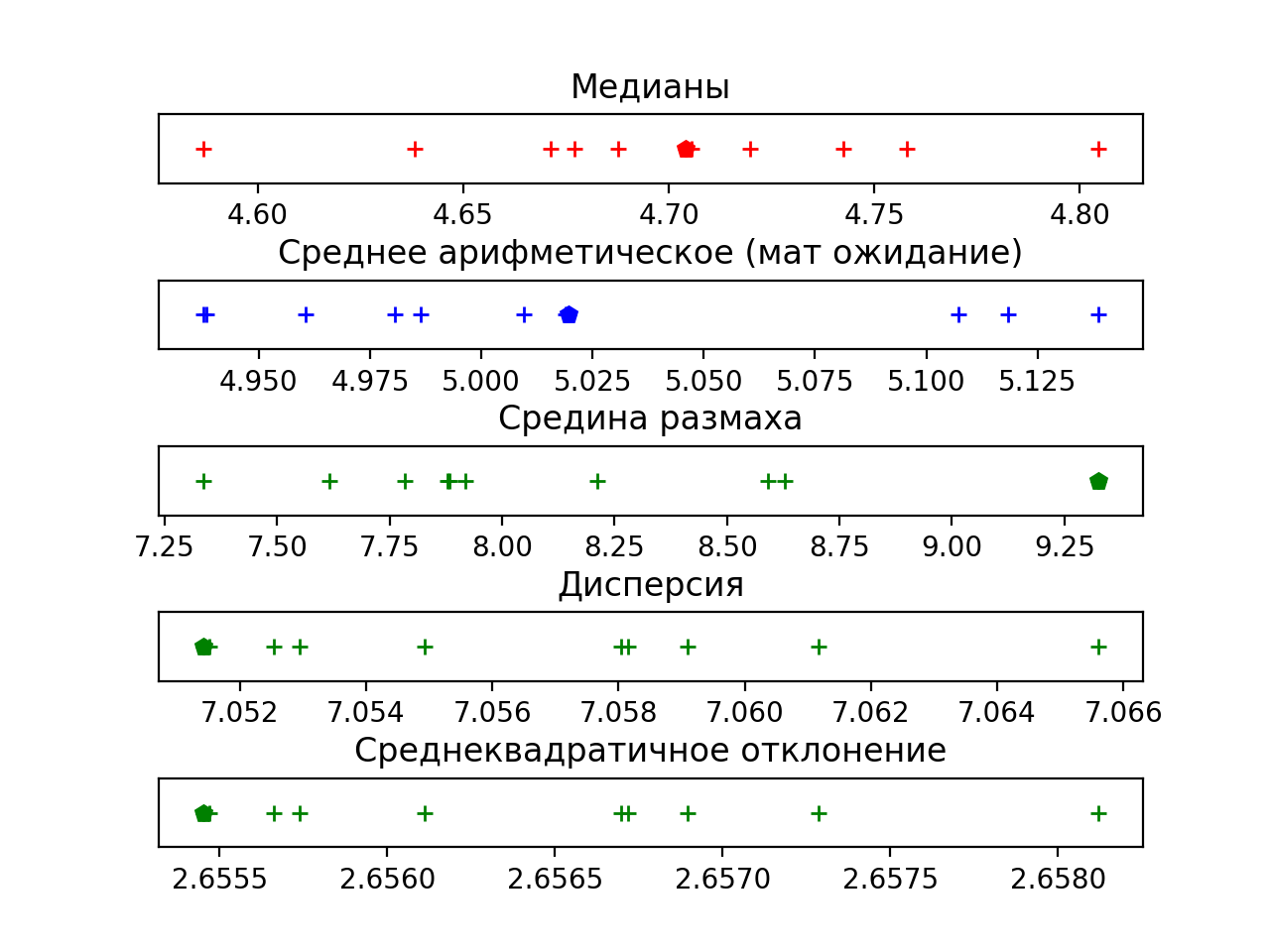
Границы интерквантильного промежутка для P = 0.95:

J = (3.00131, 6.68906)

По номерам точек

J = (4300, 12900)

**Графики точечных оценок**



**1.4 Интервальные оценки с доверительной вероятностью Q=0.8**

**Интервальный оценки мат. ожидания и дисперсии**

Значение фукции распределения Стьюдента поститано в MATLAB с помощью функции:

|  |
| --- |
| tinv(0.9, 17199) = **1.2816** |

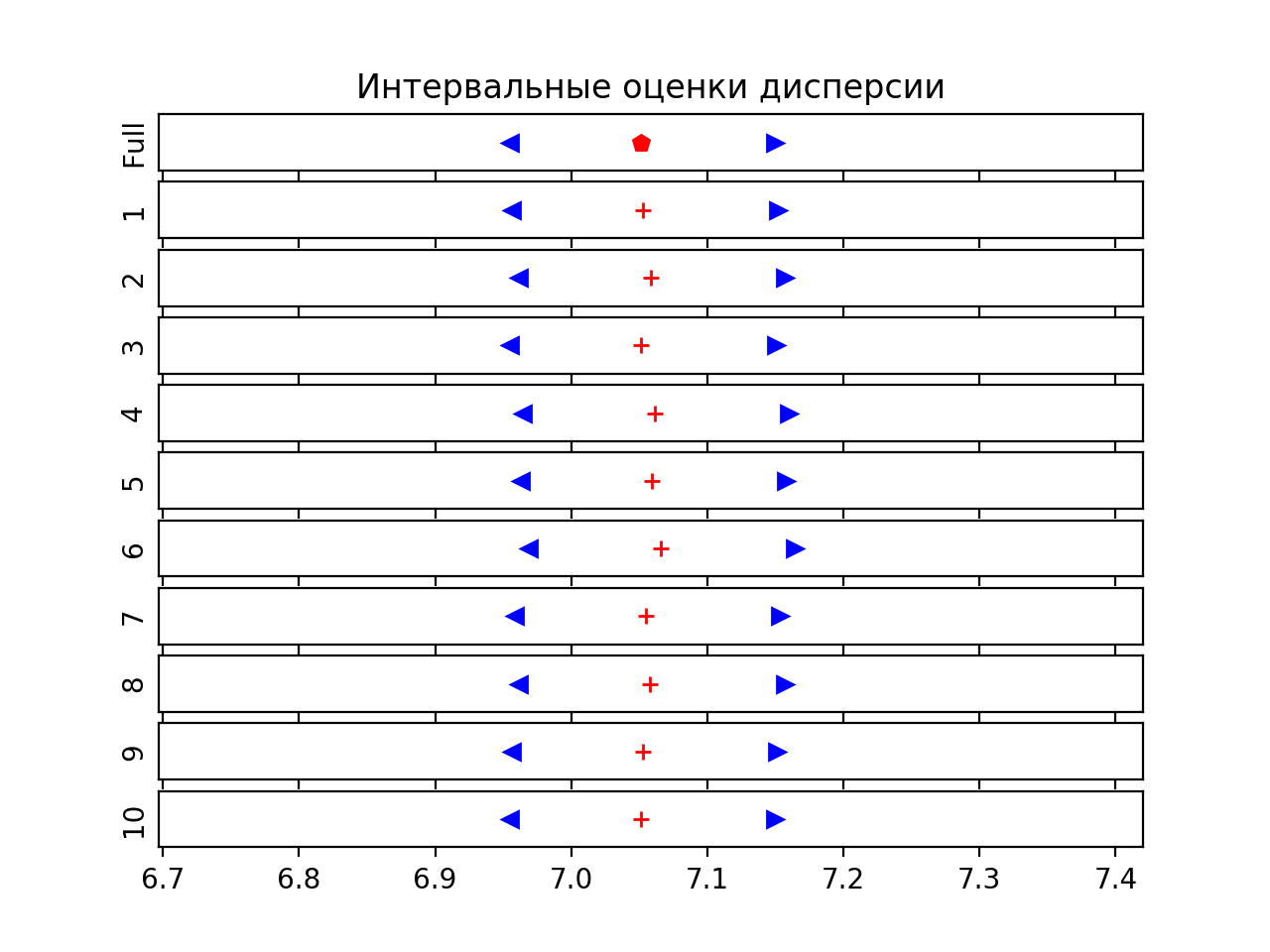
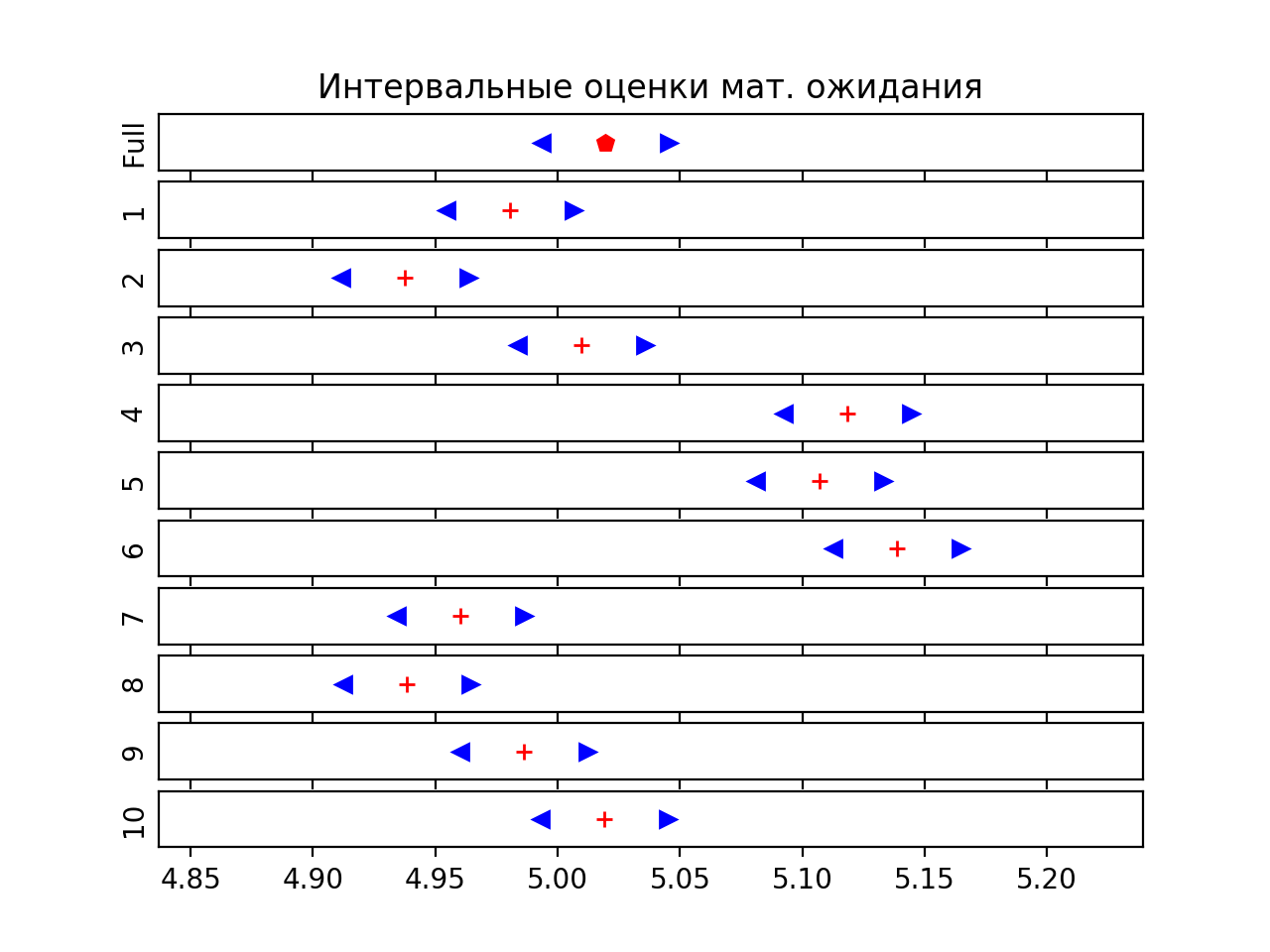
Значения функции распределения Хи-квадрат также посчитаны в MATLAB, с помощью функций:

|  |
| --- |
| chi2inv(0.9, 17199) = **1.7437e+04**  chi2inv(0.1, 17199) = **1.6962e+04** |

Таблица интервальных оценок

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Выборка | Мат. ожидание | | Дисперсия | |
| Левая | Правая | Левая | Правая |
| Full | (4.9936, 5.0455) | | (6.9551, 7.1499) | |
| 1 | (4.9546, 5.0065) | | (6.9566, 7.1514) | |
| 2 | (4.9115, 4.9635) | | (6.9618, 7.1567) | |
| 3 | (4.9836, 5.0355) | | (6.9552, 7.1500) | |
| 4 | (5.0923, 5.1442) | | (6.9647, 7.1598) | |
| 5 | (5.0812, 5.1331) | | (6.9627, 7.1577) | |
| 6 | (5.1126, 5.1646) | | (6.9691, 7.1643) | |
| 7 | (4.9344, 4.9863) | | (6.9586, 7.1534) | |
| 8 | (4.9123, 4.9642) | | (6.9617, 7.1566) | |
| 9 | (4.9603, 5.0122) | | (6.9562, 7.1510) | |
| 10 | (4.9930, 5.0449) | | (6.9551, 7.1499) | |

Графики интервальных оценок мат. ожидания и дисперсии

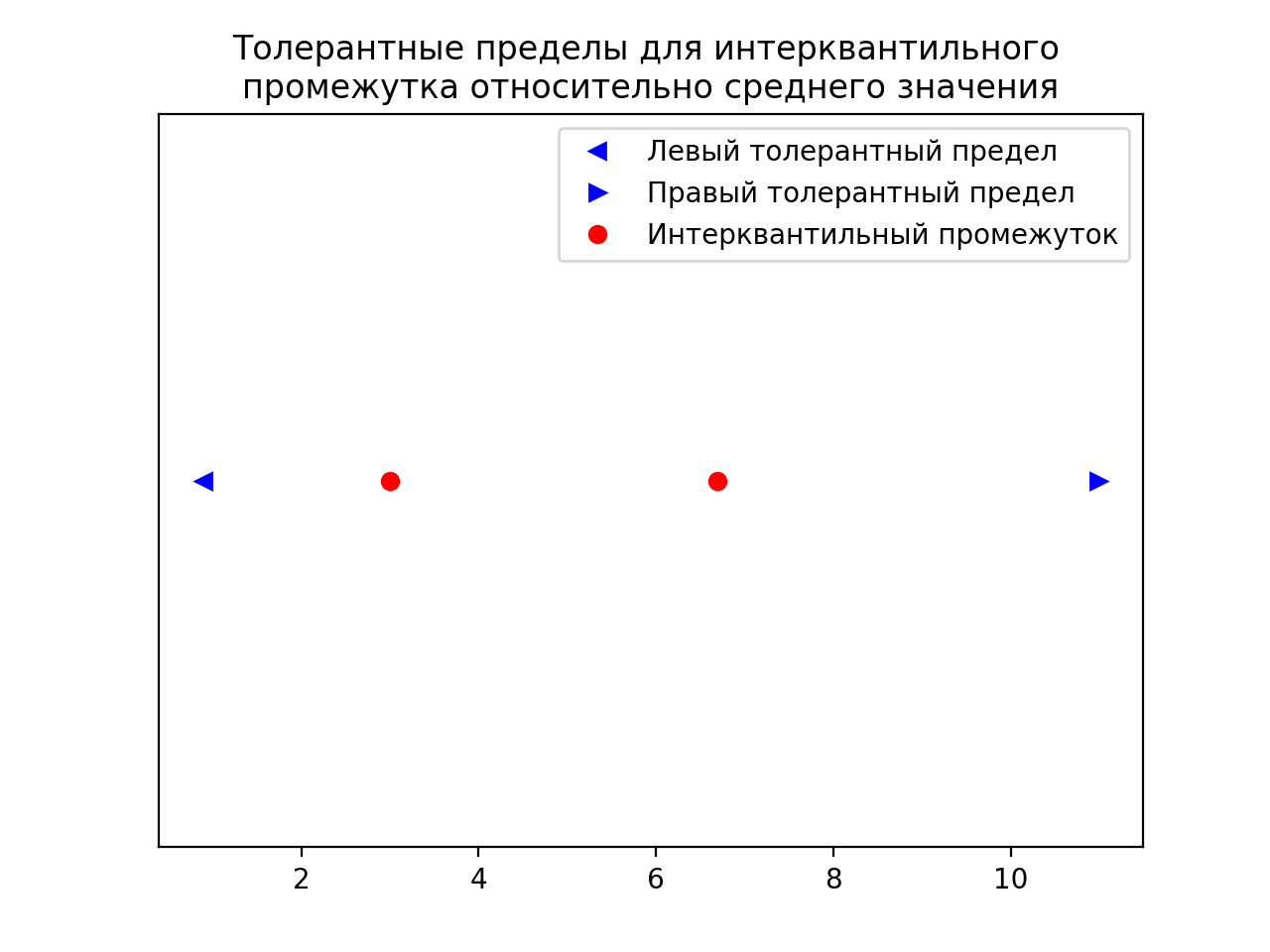


**Интервальные оценки интерквантильного промежутка для P = 0.95**

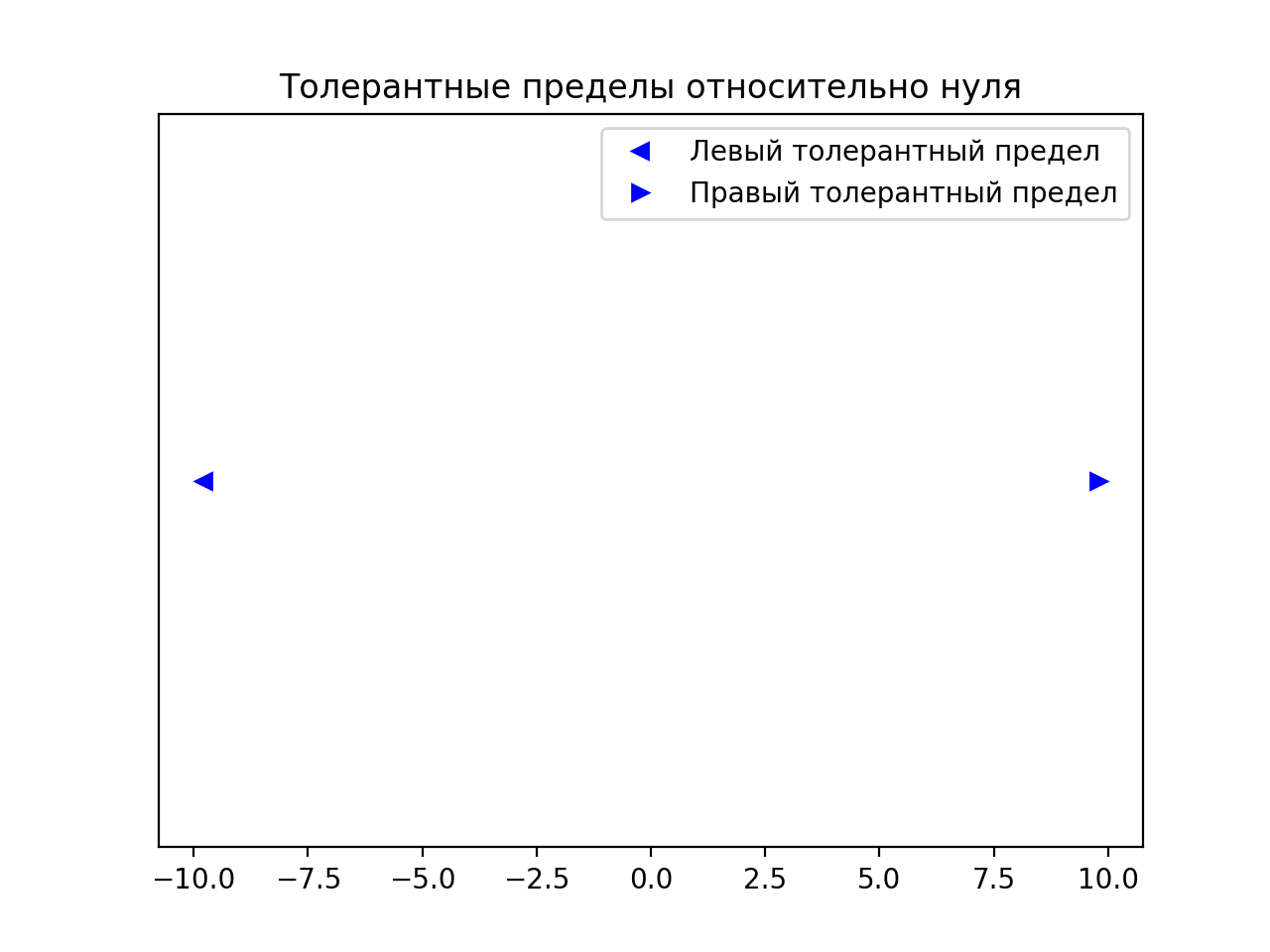
Непараметрические толеранные пределы для всей выборки симметричные относительно среднего значения.

Кол-во отбрасываемых точек было найдено с помощью биномиального распределения.

Толерантные пределы всей выборки симметричные относительно среднего значения  
**[0.899415, 10.9851]**

График  


Толерантные пределы всей выборки симметричные относительно нуля  
**[-9.78373, 9.78373]**

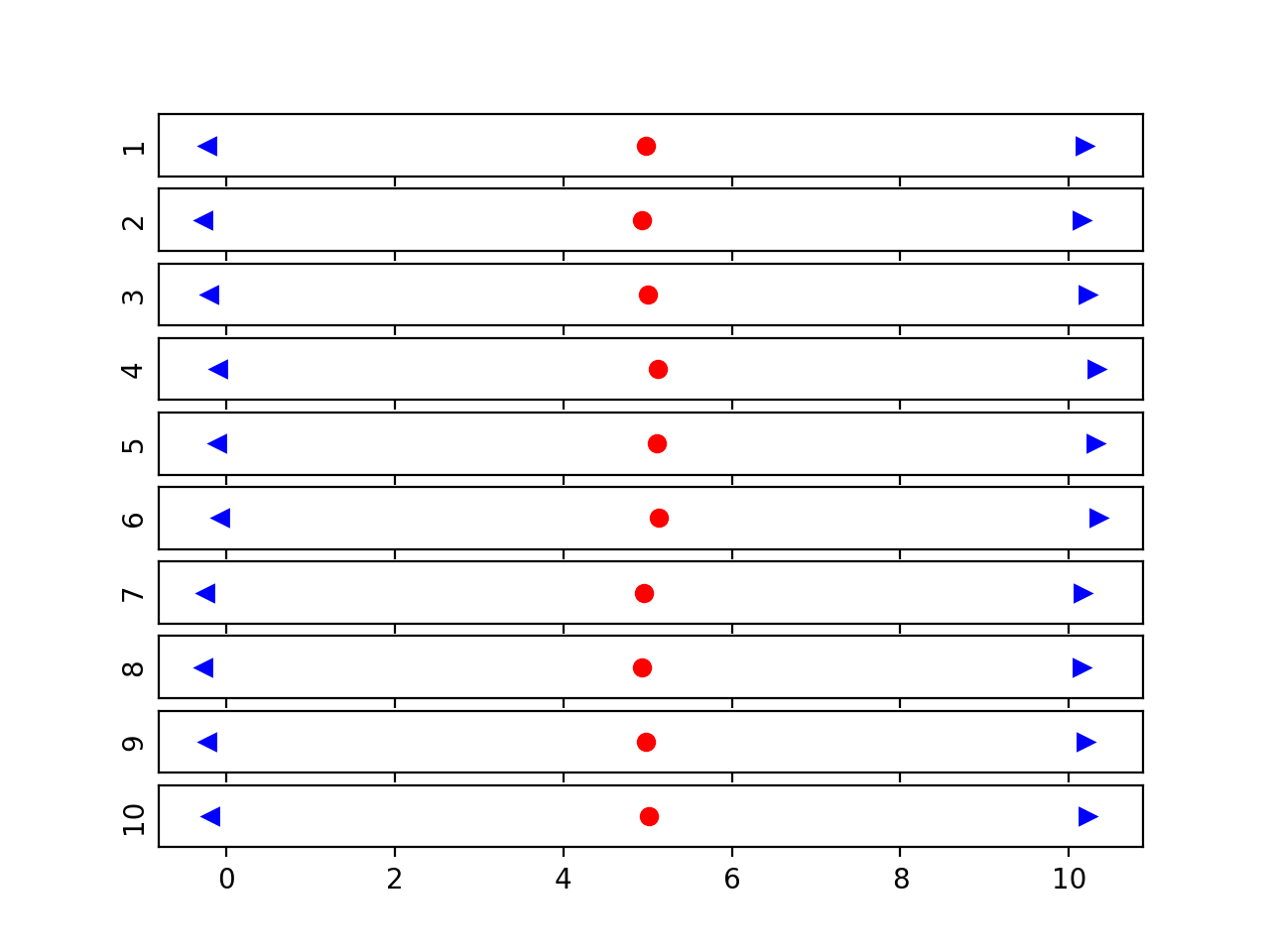
График  


Параметрические толерантные пределы для подбыборок:

Параметр

Таблица результатов

|  |  |
| --- | --- |
| Номер подвыборки | (Левый, Правый) |
| 1 | [-0.2246, 10.1858] |
| 2 | [-0.2696, 10.1447] |
| 3 | [-0.1951, 10.2143] |
| 4 | [-0.0900, 10.3265] |
| 5 | [-0.1003, 10.3146] |
| 6 | [-0.0712, 10.3485] |
| 7 | [-0.2455, 10.1664] |
| 8 | [-0.2688, 10.1454] |
| 9 | [-0.2188, 10.1913] |
| 10 | [-0.1856, 10.2236] |

График  


Красные точки на графиках – это мат. ожидания подвыборок.

Как видно из графика толерантного предела интерквантильного промежутка всей выборки, толерантные пределы шире, чем интерквантильный промежуток.

А также на всех графиках мат. ожидания лежат посередине толерантного отрезка, за исключением толерантного отрезка, симметричного относительно нуля.

**2 Идентификация закона**

**2.1 Начальный выбор распределения**

В качестве распределений-кандидатов (учитывая точечные показатели и форму гистограммы) были выбраны следующие: *Нормальное, Гамма, Рэлея*

**2.2 Определение параметров теоретических распределений**

**Метод моментов**

*Для Нормального распределения:*

*Для Гамма-распределения:*

*Для распределения Рэлея:*

**Метод максимального правдоподобия**

Пусть дана некоторая функция. Возьмем от нее логарифм.

*Для нормального распределения:*

Продифференцируем по c и найдем его значение

Теперь продифференцируем по и найдем его значение

*Для Гамма-распределения:*

Продифференцируем по

Продифференцируем по

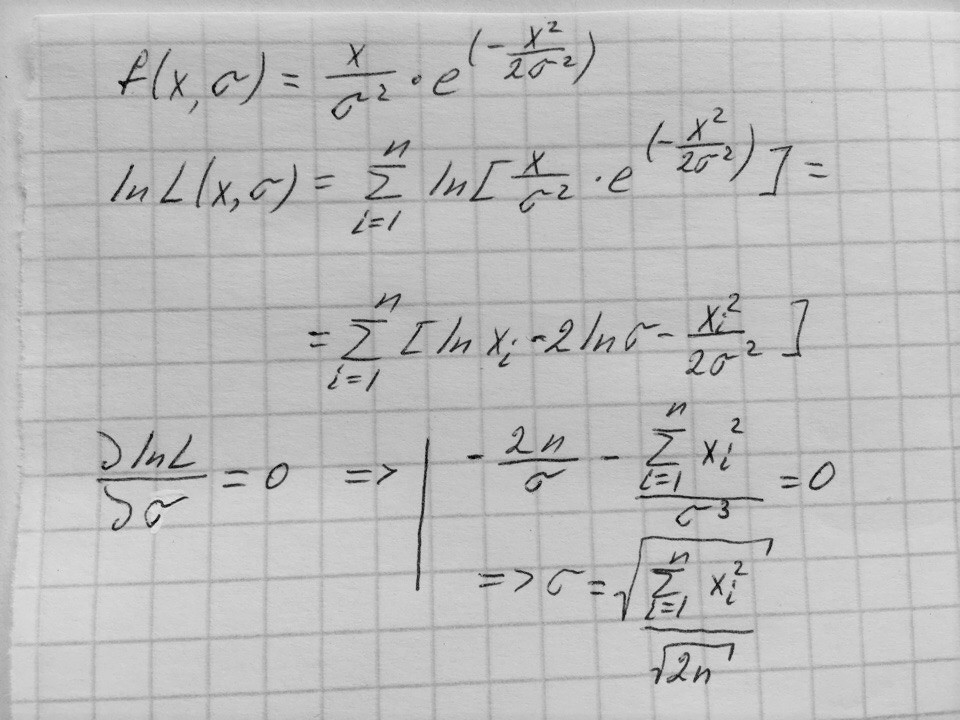
– дигамма функция Эйлера. Заменим на

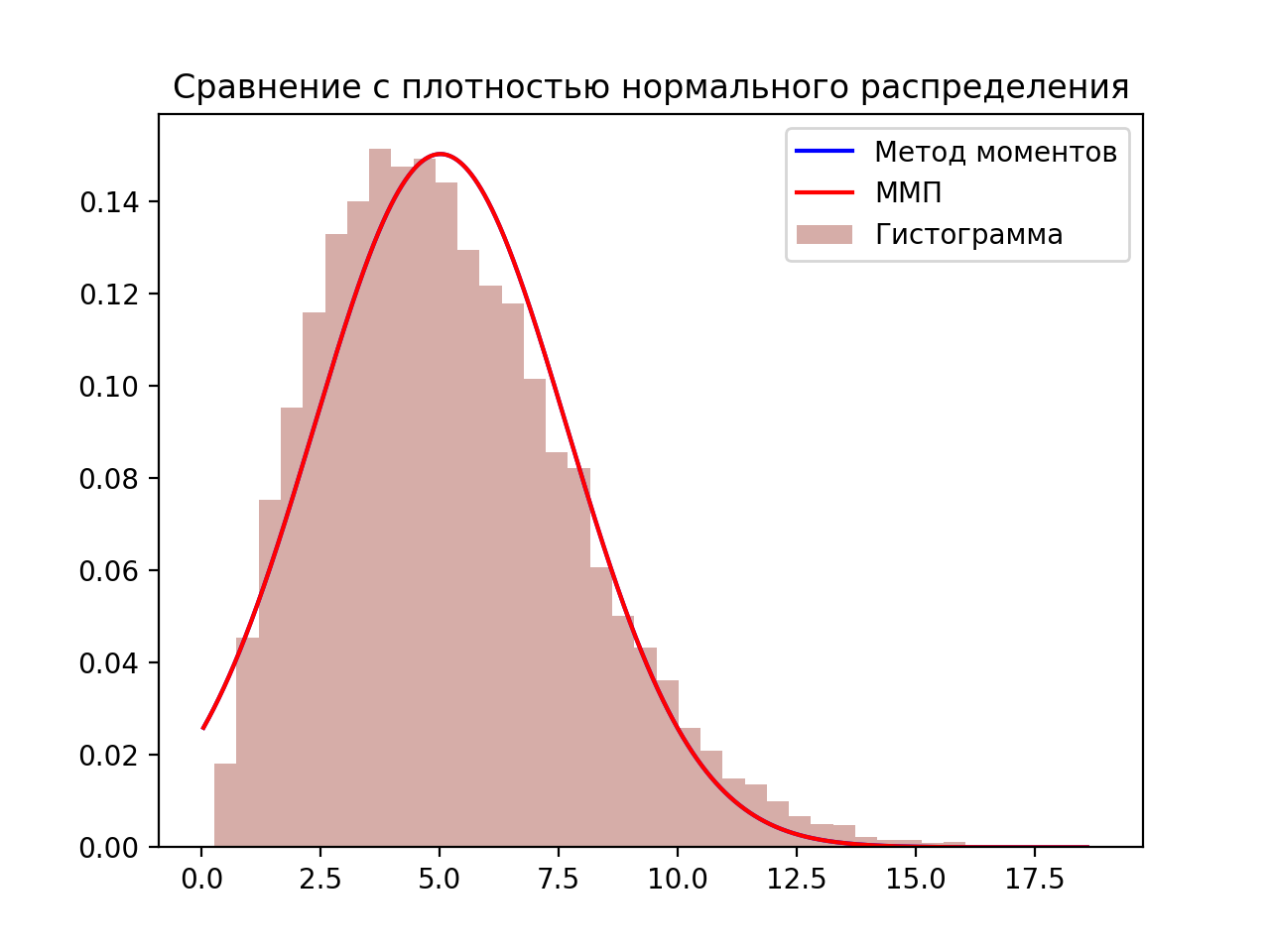
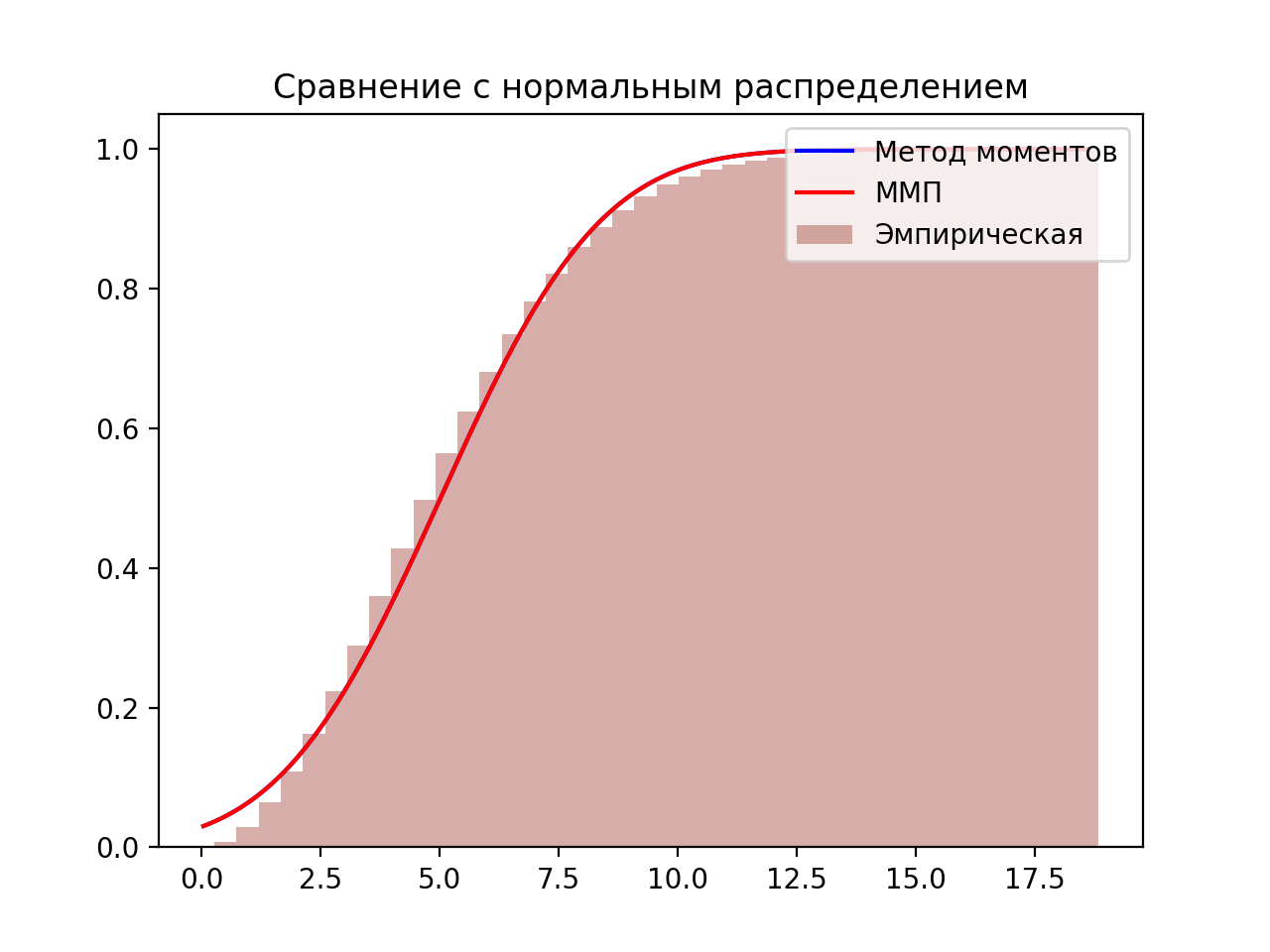
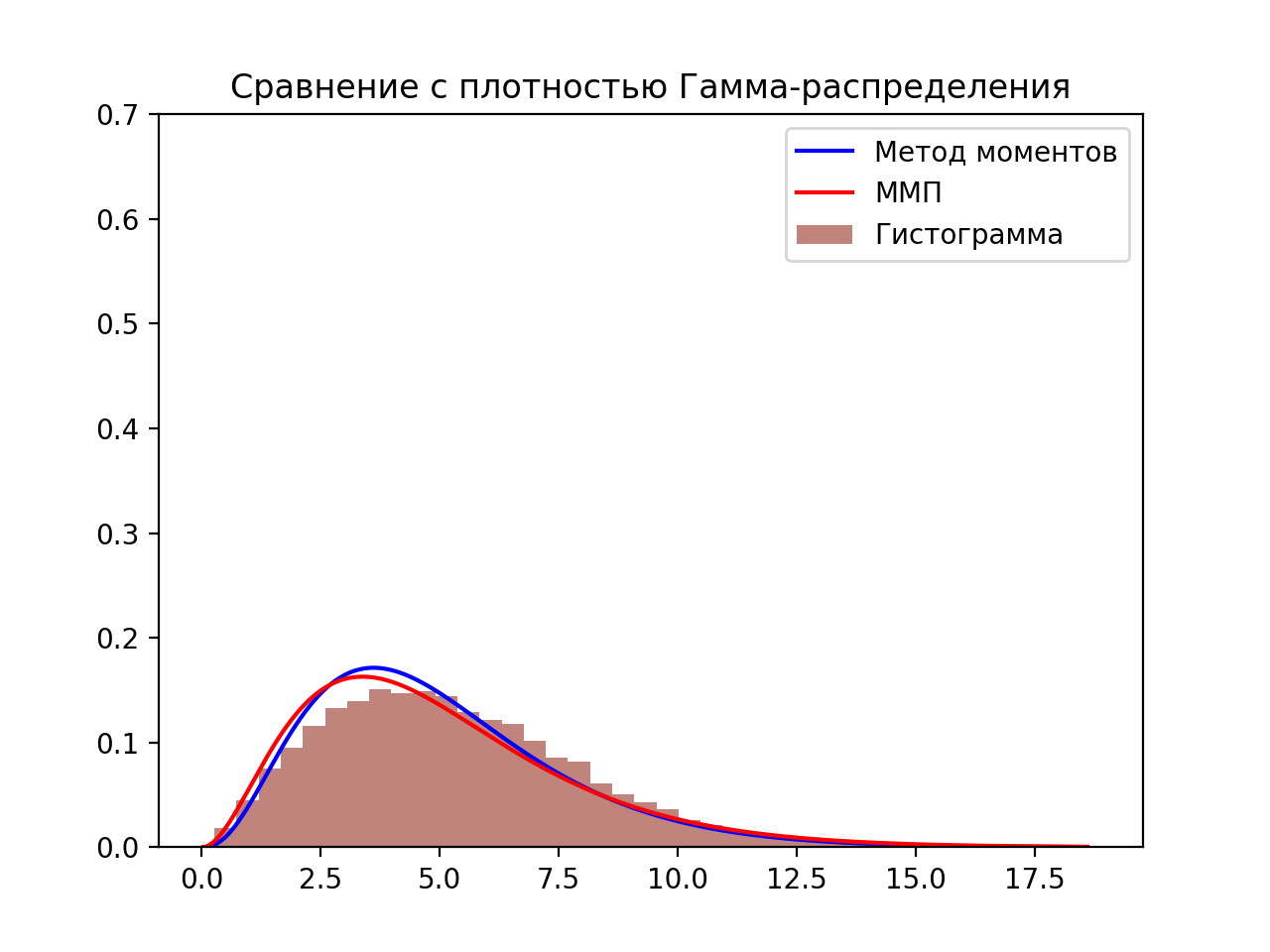
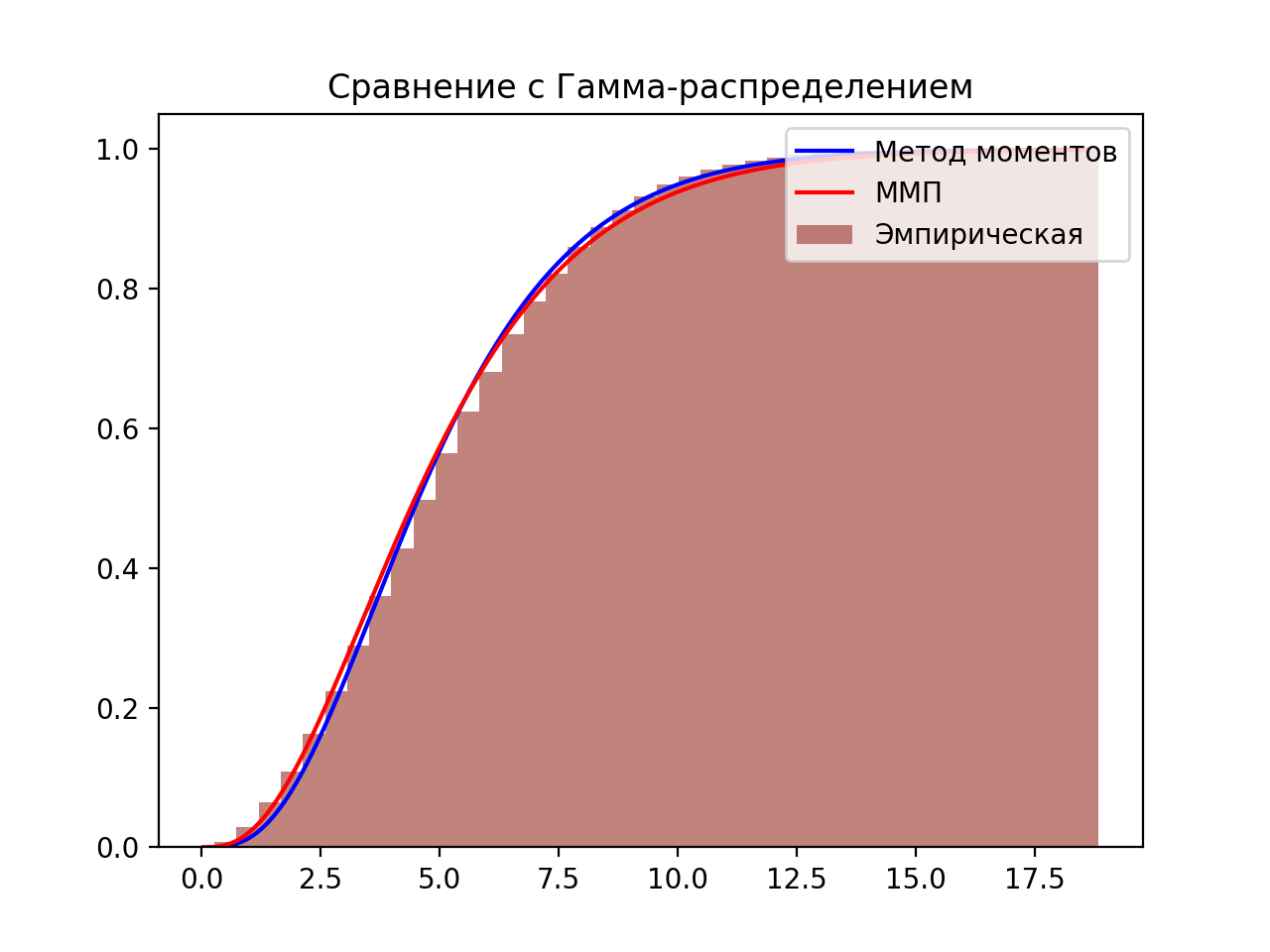
Подставим полученное ранее в последнее уравнение

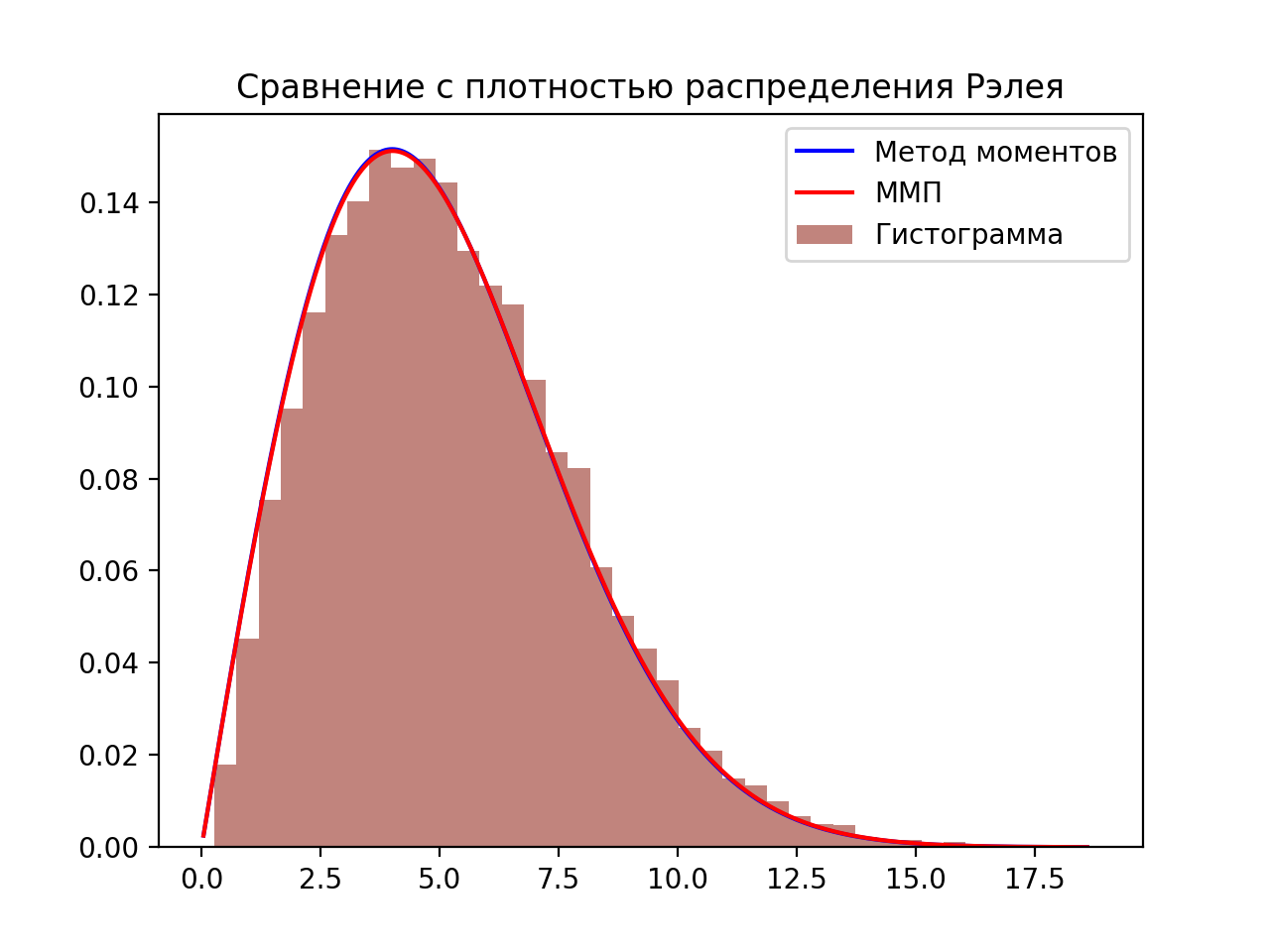
Данную функцию затруднительно считать аналитически, поэтому воспользуемся численными методами минимизации. Сначала заменим выражение справа на c, так как это значение считается единожды. Будем минимизировать функцию в следующем виде

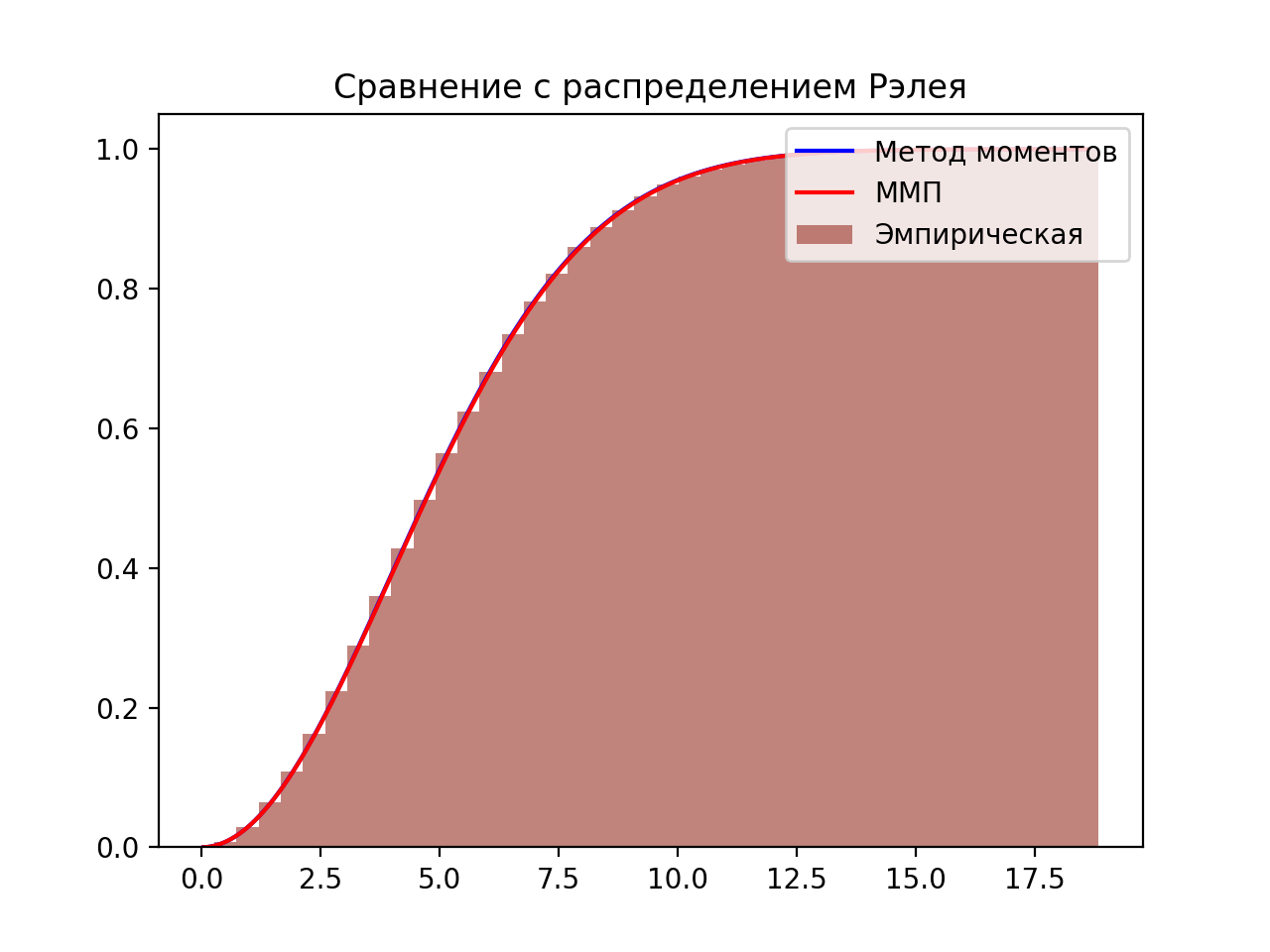
Для минимизации была написана функция, использующая метод бисекции. Точки, при которых функция принимала положительное и отрицательное значения подбирались вручную.

*Для распределения Рэлея:*



Графики сравнения результатов    





На некоторых графиках не видно кривой полученной методом моментов. Это связано с тем, что метод моментов и ММП дали очень близкие значения параметров, и одна кривая находится под другой.

График для распределения Рэлея почти совпадает с полученной ранее гистограммой.

Таблица оценок

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Параметры | Нормальное распределение | Гамма-распределение | Распределение Рэлея |
| Метод моментов |  | 5.0195 | 3.5732 | 4.0050 |
|  | 2.6554 | 1.4047 |  |
| ММП |  | 5.0195 | 3.0680 | 4.0154 |
|  | 2.6553 | 1.6360 |  |

**2.3 Проверка гипотез**

Уровень значимости

В качестве параметров будем использовать параметры, полученные с помощью ММП, так как он призван быть точнее метода моментов.

**Критерий «Хи-квадрат»**

Критическое значение

Статистика

Где

– объем выборки;

– кол-во выборочных значений, попавших в интервал ;

– общее кол-во интервалов, было выбрано 40 интервалов;

– кол-во оцениваемых параметров.

– вероятность попадания в k-ый интервал.

. Найдем его с помощью функции chi2inv(0.8, 38) в *MATLAB*.

Аналогично для r = 1,

Таблица результатов

|  |  |  |  |
| --- | --- | --- | --- |
|  | Нормальное распределение | Гамма-распределение | Распределение Рэлея |
| Метод моментов | 16991.36 | 1631.39 | 59,62 |
| ММП | 16994.63 | 1588,40 | 57,522 |

**Критерий Колмогорова-Смирнова**

Критическое значение при больших вычисляется по формуле

При получаем

Статистика Комогорова-Смирнова вычисляется по формуле

Таблица результатов

|  |  |  |  |
| --- | --- | --- | --- |
|  | Нормальное распределение | Гамма-распределение | Распределение Рэлея |
| Метод моментов | 0.049145 | 0.02855 | 0.00729 |
| ММП | 0.049147 | 0.03306 | 0.00844 |

**Критерий Мизеса**

Критическое значение при равно

Статистика критерия вычисляется по следующей формуле

Таблица результатов

|  |  |  |  |
| --- | --- | --- | --- |
|  | Нормальное распределение | Гамма-распределение | Распределение Рэлея |
| Метод моментов | 15.3824 | 5.9019 | 0.1013 |
| ММП | 15.3832 | 6.7911 | 0.1706 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Название | Нормальное распределение | | | Гамма-распределение | | | Распределение Рэлея | | | |
|  | Метод моментов | ММП - аналитич | ММП – числ. | Метод моментов | ММП - аналитич | ММП – числ. | Метод моментов | ММП - аналитич | ММП – числ. |
|  | 5.0195 | 5.0195 | - | 3.5732 | - | 3.0680 | 4.0050 | 4.0154 | - |
|  | 2.6554 | 2.6553 | - | 1.4047 | - | 1.6360 | - | - | - |
| Хи-квадрат статистика | 16991.36 | 16994.63 | - | 1631.39 | - | 1588,40 | 59,62 | 57,522 | - |
| Хи-квадрат критич. знач. |  | | | | | | | | | |
| Хи-квадрат вывод | *нет* | *нет* | *-* | *нет* | *-* | *нет* | *близко* | *близко* | - |
| Колм. - Смирн. статистика | 0.049145 | 0.049147 | - | 0.02855 | - | 0.03306 | 0.00729 | 0.00844 | - |
| Колм. - Смирн. критич. знач. |  | | | | | | | | | |
| Колм. - Смирн. вывод | *нет* | *нет* | *-* | *нет* | *-* | *нет* | *Да!* | *Да!* | - |
| Мизеса - статистика | 15.3824 | 15.3832 | - | 5.9019 | - | 6.7911 | 0.1013 | 0.1706 | - |
| Мизеса критич. знач. |  | | | | | | | | | |
| Мизеса вывод | *нет* | *нет* | *-* | *нет* | *-* | *нет* | *Да!* | *Да!* | - |

**Вывод**

В данной лабораторной работе была проанализирована некоторая выборка случайных величин. Были получены точечные и интервальные оценки. С помощью точечных оценок и гистограммы были сделаны гипотезы и различных распределениях. Далее гипотезы были проверены с помощью различных критериев.

Исходя из того, что гипотеза о распределении Рэлея единственная прошла два критерия из трёх, в то время, как другие гипотезы не прошли ни одной, можно уверенно сказать, что исходная выборка имеет **распределение Рэлея** с параметром:

(Приблизительно)