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. 23531/2 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ М.А. Косенков

(подпись)

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

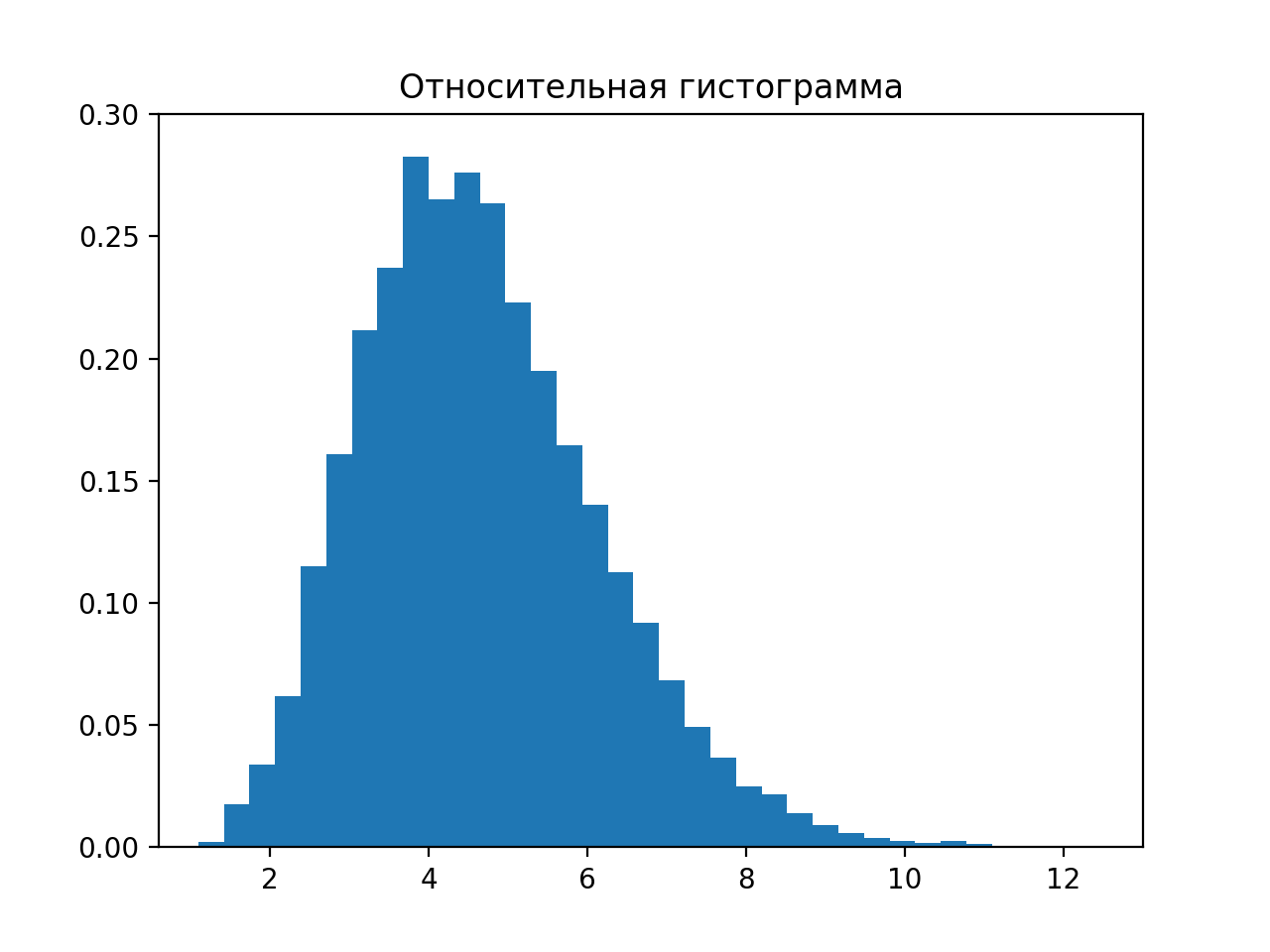
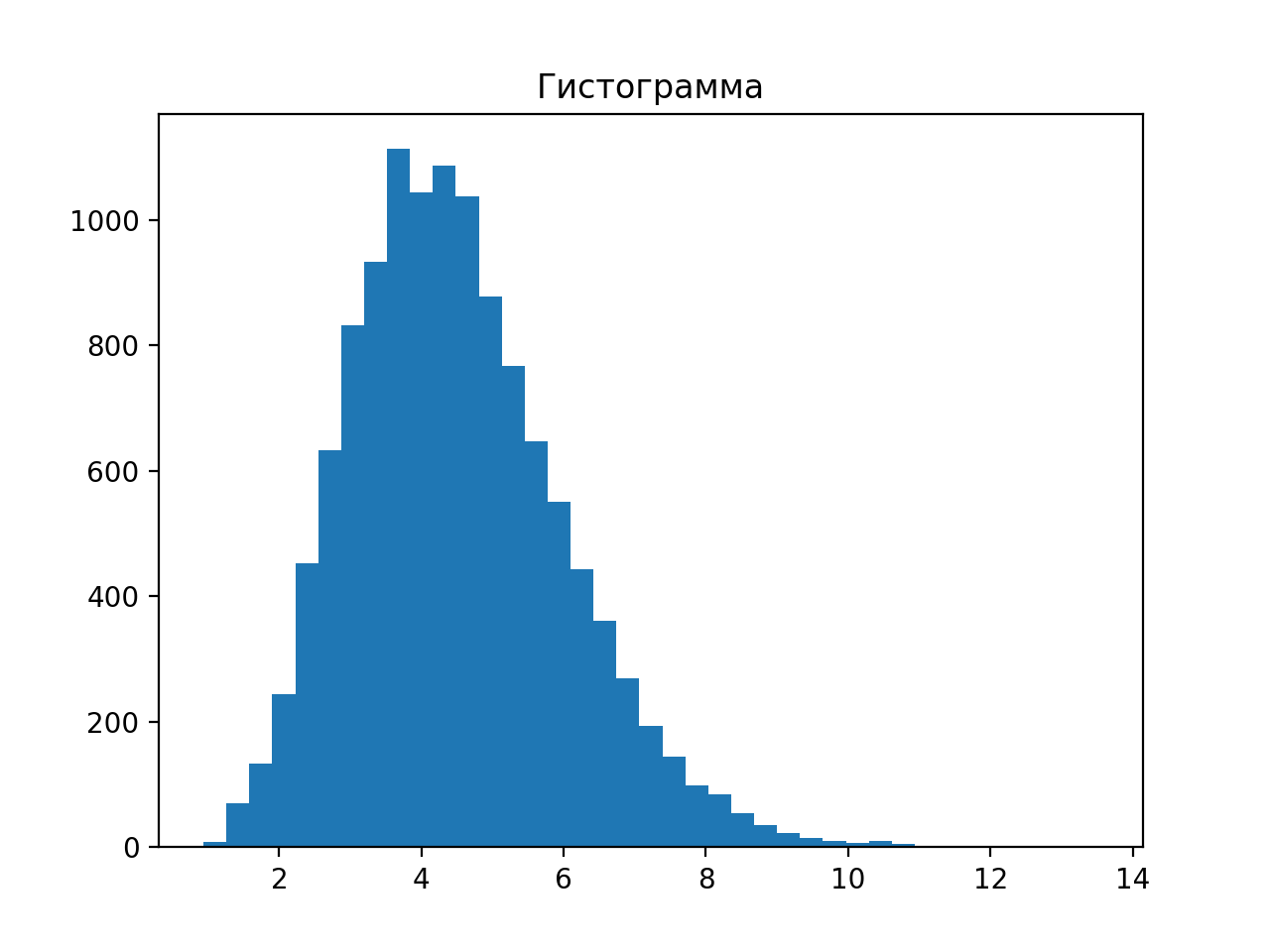
(подпись)  
“\_\_\_” \_\_\_\_\_\_\_\_\_\_\_\_ 2018 г.

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 поделены на 10 равных частей. Далее по полной выборке и по подвыборкам были посчитаны точечные оценки. Результаты представлены в таблице.

**1.3 Определение точечных оценок**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
| Full | 4.5031 | 4.3462 | 6.7586 | 1.4962 | 2.2385 | 2.1326 | 18.0734 | 1.1653098867816727 | 3.606905307101759 |
| 1 | 4.5198 | 4.3617 | 6.2994 | 1.4962 | 2.2388 | 2.0202 | 17.9344 | 0.6030996364849598 | 3.578275314419131 |
| 2 | 4.5405 | 4.3037 | 5.8359 | 1.4966 | 2.2399 | 1.8812 | 17.7730 | 0.5611892278877412 | 3.5425174051826005 |
| 3 | 4.5047 | 4.3249 | 6.0428 | 1.4962 | 2.2385 | 2.1218 | 18.0597 | 0.6335449591616487 | 3.60417310534377 |
| 4 | 4.5217 | 4.3995 | 5.8956 | 1.4963 | 2.2388 | 2.0071 | 17.9187 | 0.5991651738240205 | 3.574923320491032 |
| 5 | 4.5536 | 4.4271 | 5.8209 | 1.4970 | 2.2410 | 1.7930 | 17.6765 | 0.5344645639170614 | 3.5196650075976477 |
| 6 | 4.4638 | 4.2693 | 7.0003 | 1.4967 | 2.2400 | 2.3963 | 18.4290 | 0.7147530187409794 | 3.67280929175234 |
| 7 | 4.4437 | 4.3036 | 5.8815 | 1.4973 | 2.2419 | 2.5310 | 18.6265 | 0.753958456782083 | 3.705645654096414 |
| 8 | 4.4990 | 4.3913 | 5.8849 | 1.4962 | 2.2385 | 2.1596 | 18.1079 | 0.6448177305759788 | 3.6137421598059816 |
| 9 | 4.4694 | 4.2769 | 6.0933 | 1.4965 | 2.2396 | 2.3583 | 18.3752 | 0.703628242974118 | 3.66345216166828 |
| 10 | 4.5143 | 4.4037 | 6.5264 | 1.4962 | 2.2386 | 2.0573 | 17.9794 | 0.6142307567693833 | 3.5877523620113405 |

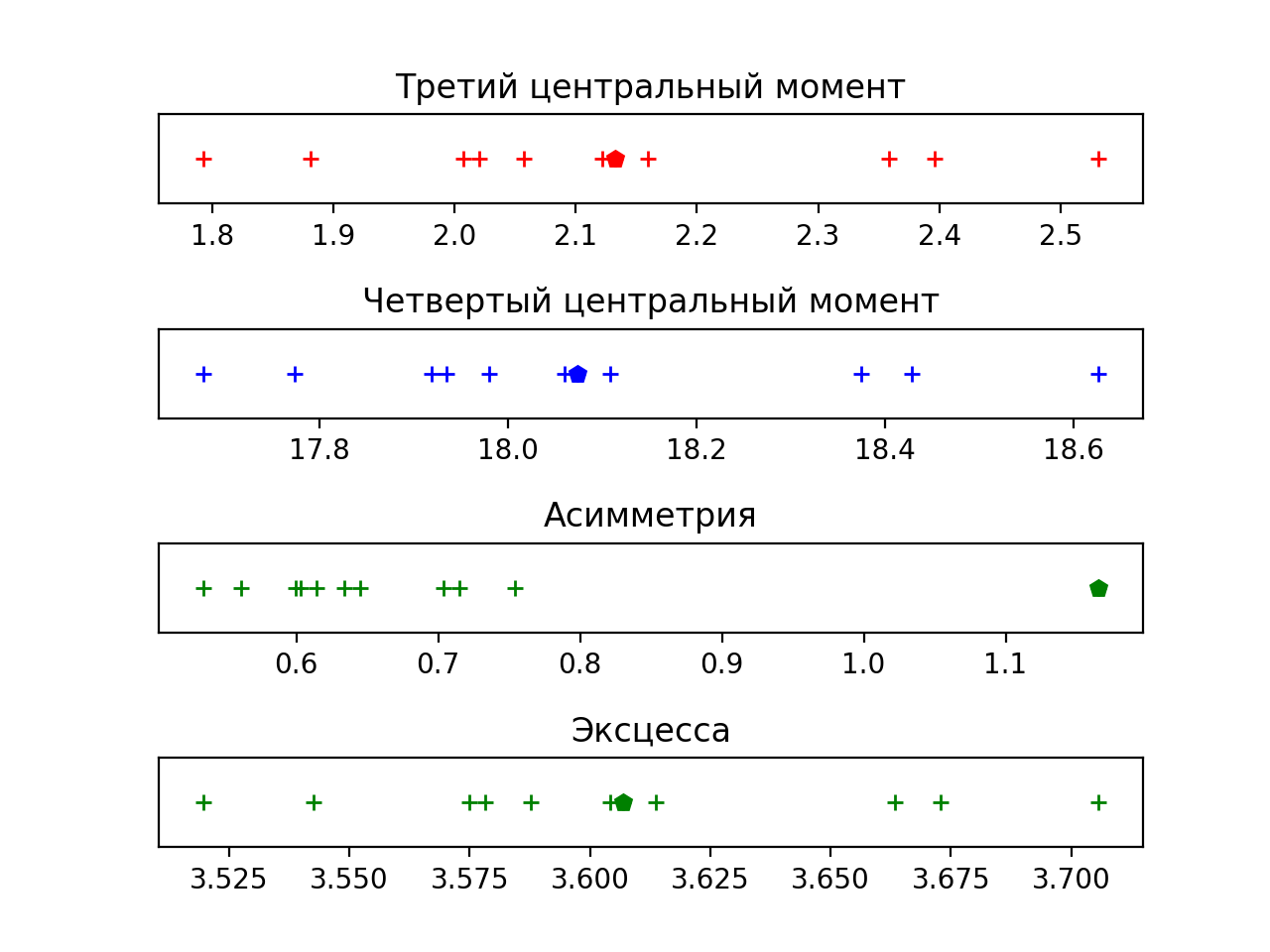
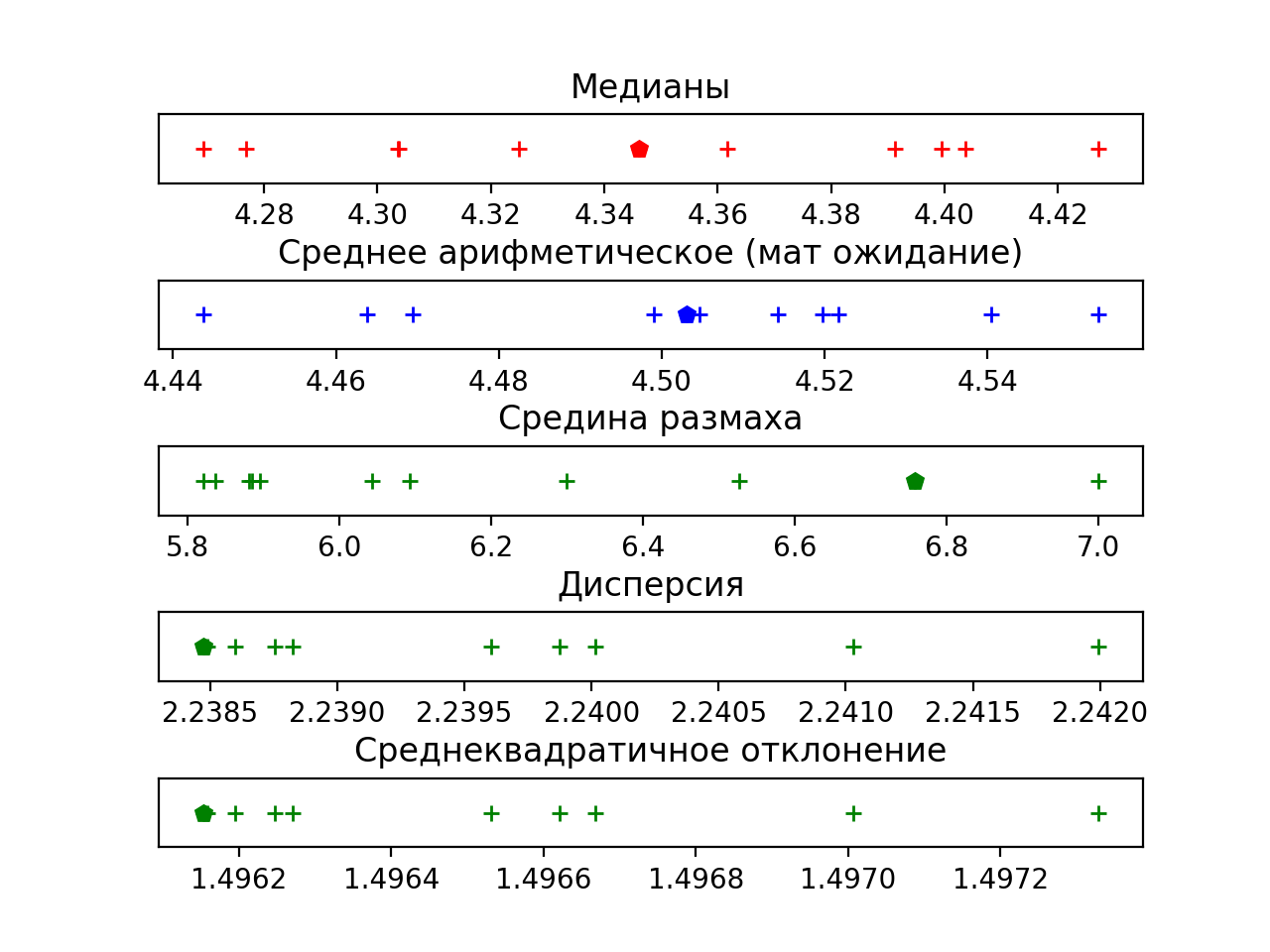
Границы интерквантильного промежутка для P = 0.95:

J = [3.42706, 5.40958]

По номерам точек

J = [3050, 9150]

**Графики точечных оценок**



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

**Интервальный оценки мат. ожидания и дисперсии**

Значение функции распределения Стьюдента посчитано в MATLAB с помощью функции:

|  |
| --- |
| tinv(0.9, 12199) = **1.2816** |

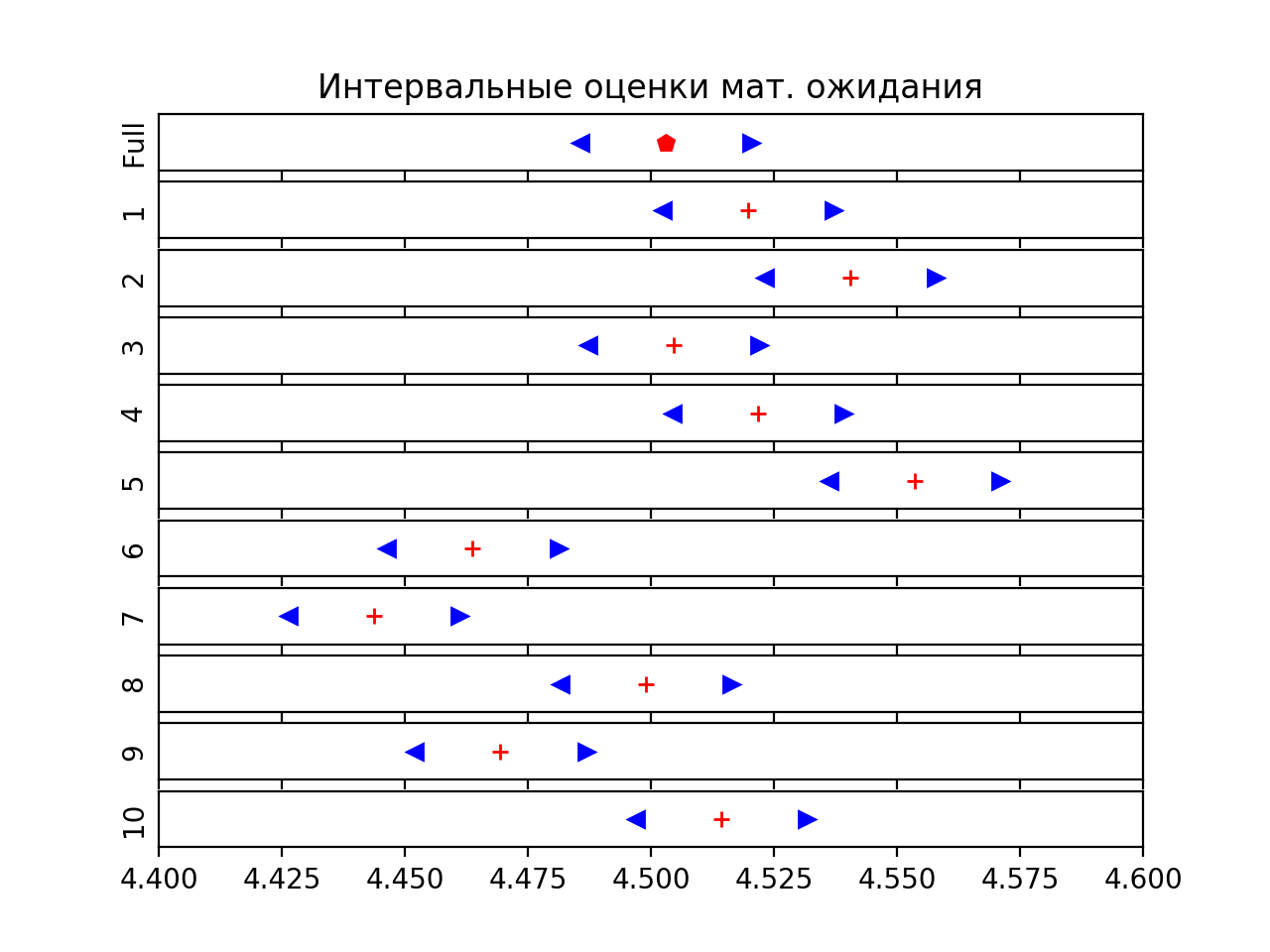
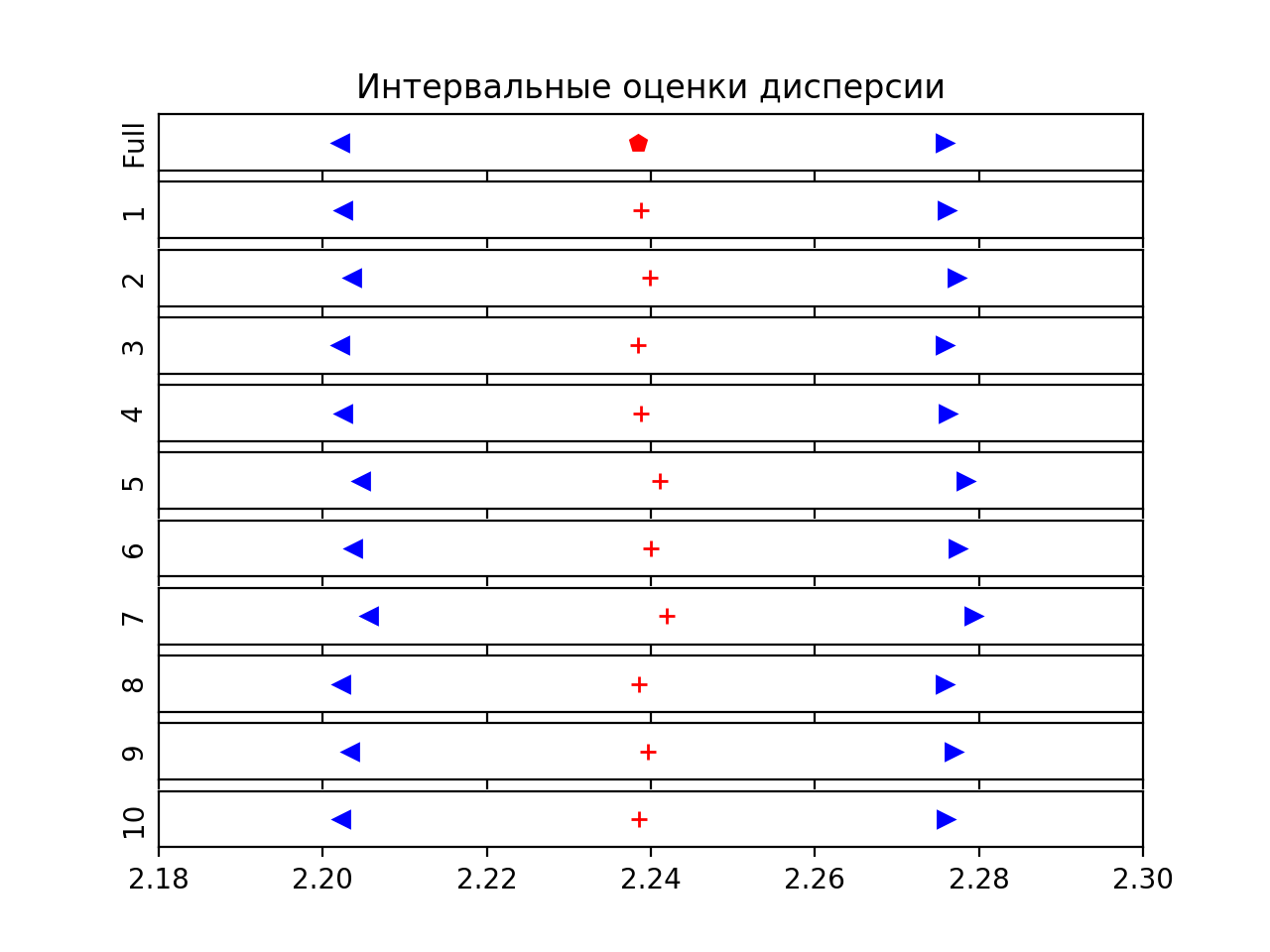
Значения функции распределения Хи-квадрат также посчитаны в MATLAB, с помощью функций:

|  |
| --- |
| chi2inv(0.9, 12199) = **1.2400е+04**  chi2inv(0.1, 12199) = **1.1999e+04** |

Таблица интервальных оценок

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Выборка | Мат. ожидание | | Дисперсия | |
| Левая | Правая | Левая | Правая |
| Full | 4.48569 | 4.52041 | 2.20219 | 2.27579 |
| 1 | 4.50243 | 4.53715 | 2.20247 | 2.27607 |
| 2 | 4.52311 | 4.55784 | 2.20357 | 2.27721 |
| 3 | 4.48730 | 4.52202 | 2.20219 | 2.27579 |
| 4 | 4.50437 | 4.53909 | 2.20253 | 2.27614 |
| 5 | 4.53623 | 4.57097 | 2.20470 | 2.27838 |
| 6 | 4.44643 | 4.48116 | 2.20371 | 2.27735 |
| 7 | 4.42637 | 4.46112 | 2.20565 | 2.27936 |
| 8 | 4.48167 | 4.51639 | 2.20221 | 2.27580 |
| 9 | 4.45208 | 4.48681 | 2.20330 | 2.27693 |
| 10 | 4.49691 | 4.53163 | 2.20231 | 2.27591 |

Графики интервальных оценок мат. ожидания и дисперсии

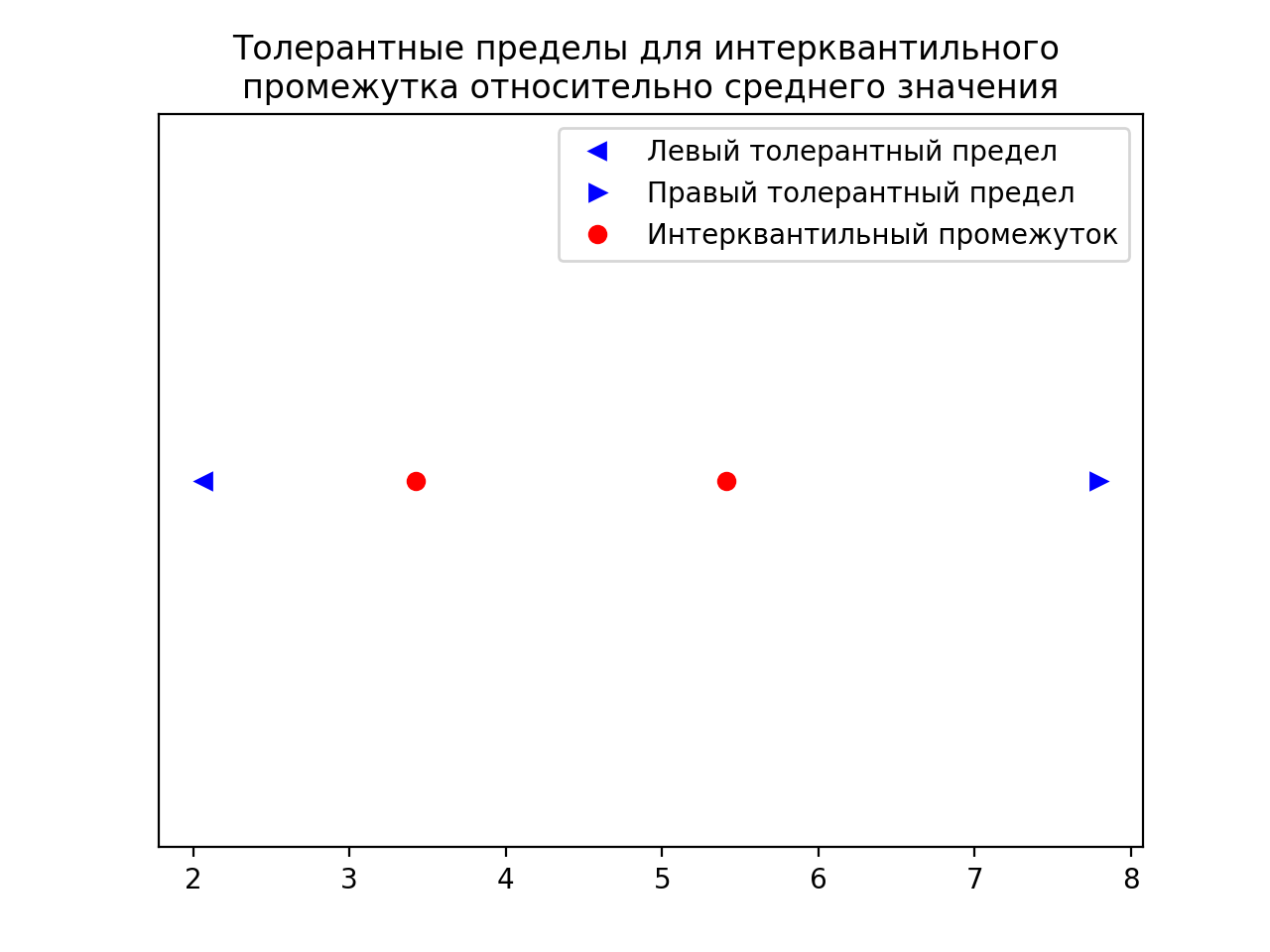


**Интервальные оценки интерквантильного промежутка для P = 0.95**

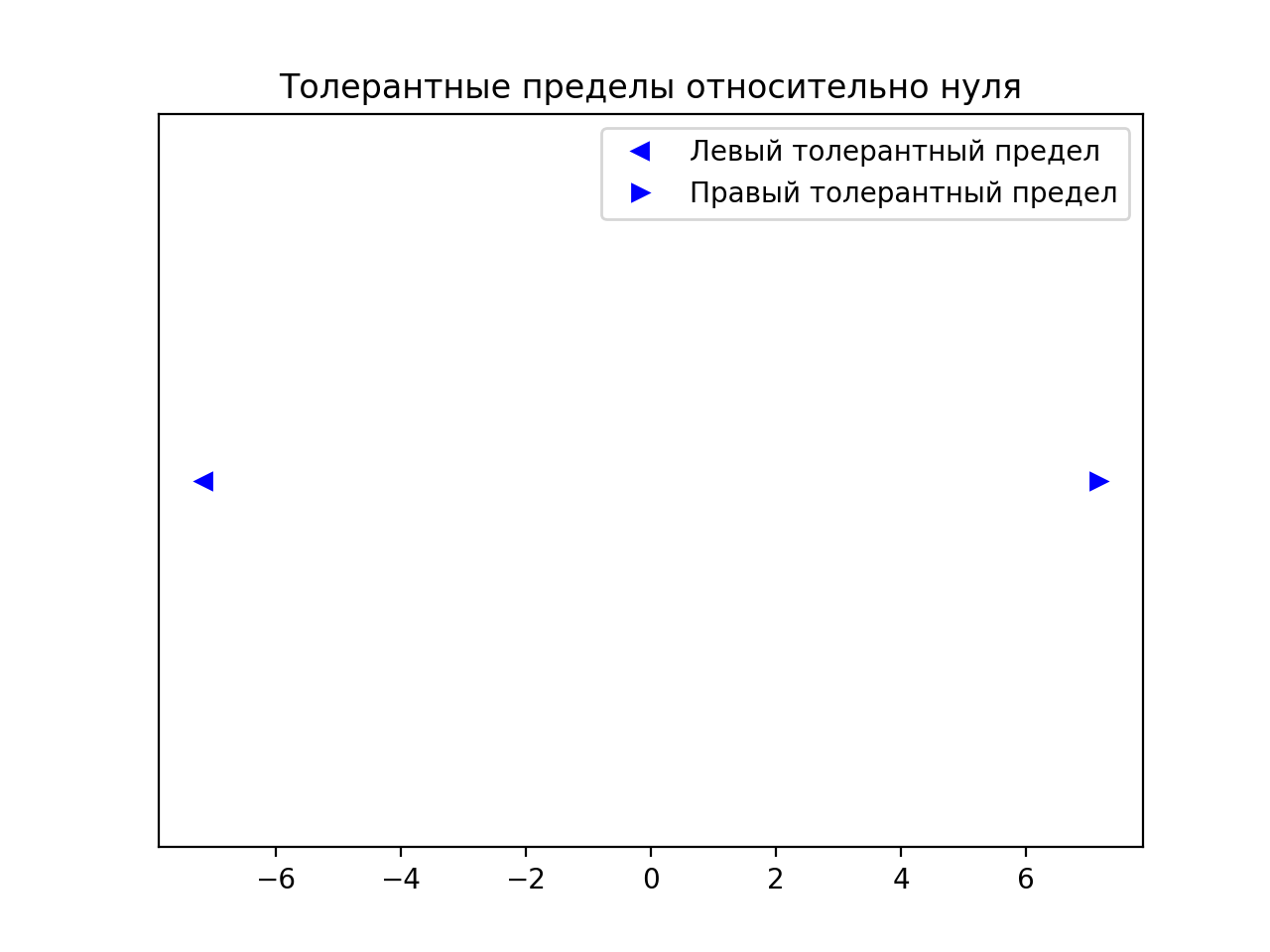
Непараметрические толерантные пределы для всей выборки симметричные относительно среднего значения.

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

Толерантные пределы всей выборки симметричные относительно среднего значения  
[2.06611, 7.79174]

График  


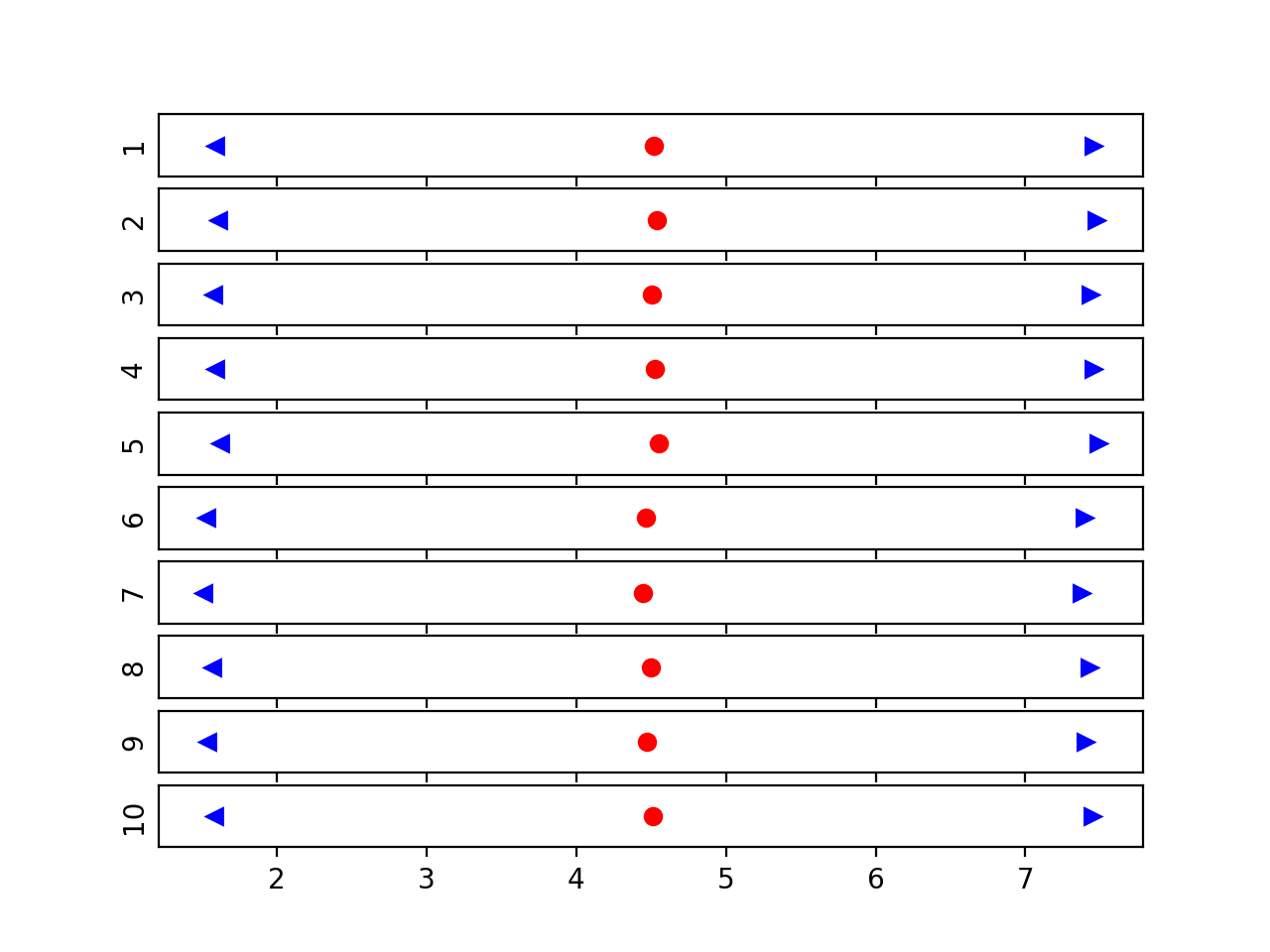
Толерантные пределы всей выборки симметричные относительно нуля  
[-7.15177, 7.15177]

График  


Параметрические толерантные пределы для подвыборок

Таблица результатов

|  |  |  |
| --- | --- | --- |
| Номер подвыборки | Левый | Правый |
| 1 | 1.5871425693856538 | 7.452430233893034 |
| 2 | 1.6070985271849008 | 7.47385384986428 |
| 3 | 1.5721960024104975 | 7.437120248409175 |
| 4 | 1.5890459895197795 | 7.45442405638186 |
| 5 | 1.6194665047268209 | 7.487733691994491 |
| 6 | 1.5303263675124903 | 7.397266042323576 |
| 7 | 1.5089819718056932 | 7.378509405243487 |
| 8 | 1.5665617732706725 | 7.431503833286705 |
| 9 | 1.536243644494105 | 7.402644306325566 |
| 10 | 1.581722998332813 | 7.446808559044236 |

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.

**2 Идентификация закона**

**2.1 Начальный выбор распределения**

В качестве распределений-кандидатов (учитывая точечные показатели) были выбраны следующие: *Рэлея,* *Лапласа, Гамма*

**2.2 Определение параметров теоретических распределений**

**Метод моментов**

*Для распределения Рэлея:*

*Для распределения Лапласа:*

*Для Гамма-распределения:*

**Метод максимального правдоподобия**

Пусть дана некоторая функция. Возьмем от нее логарифм.

*Для распределения Рэлея:*

Продифференцируем по и найдем его значение

*Для распределения Лапласа:*

Продифференцируем полученное выражение по

Данное выражение не дифференцируемо. Данная сумма принимает наименьшее свое значение значение, когда сумма расстояний от до каждой точки минимальна. Это достигается, когда равно выборочной медиане, то есть.

Продифференцруем по

*Для Гамма-распределения:*

Продифференцируем по

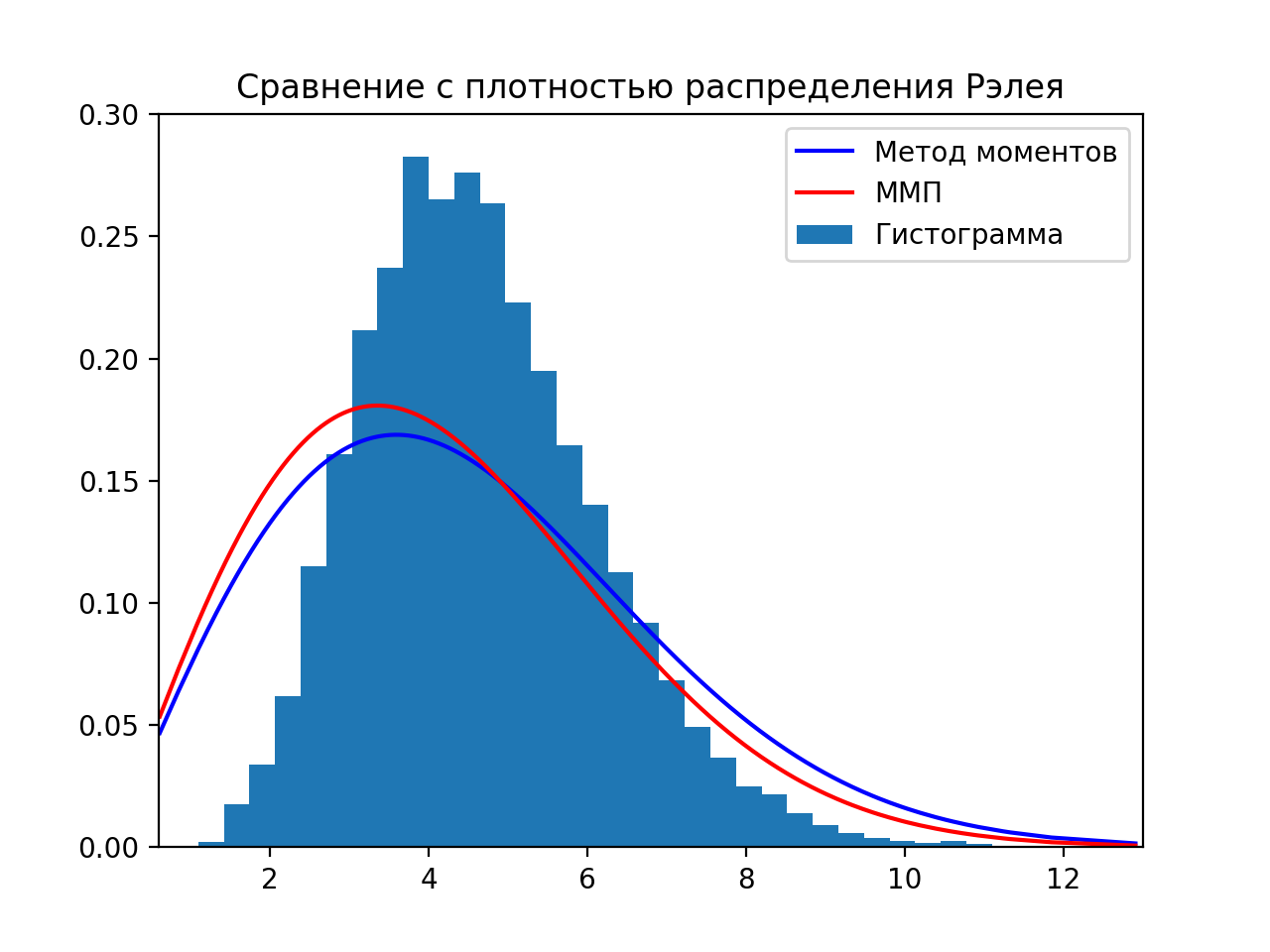
Продифференцируем по

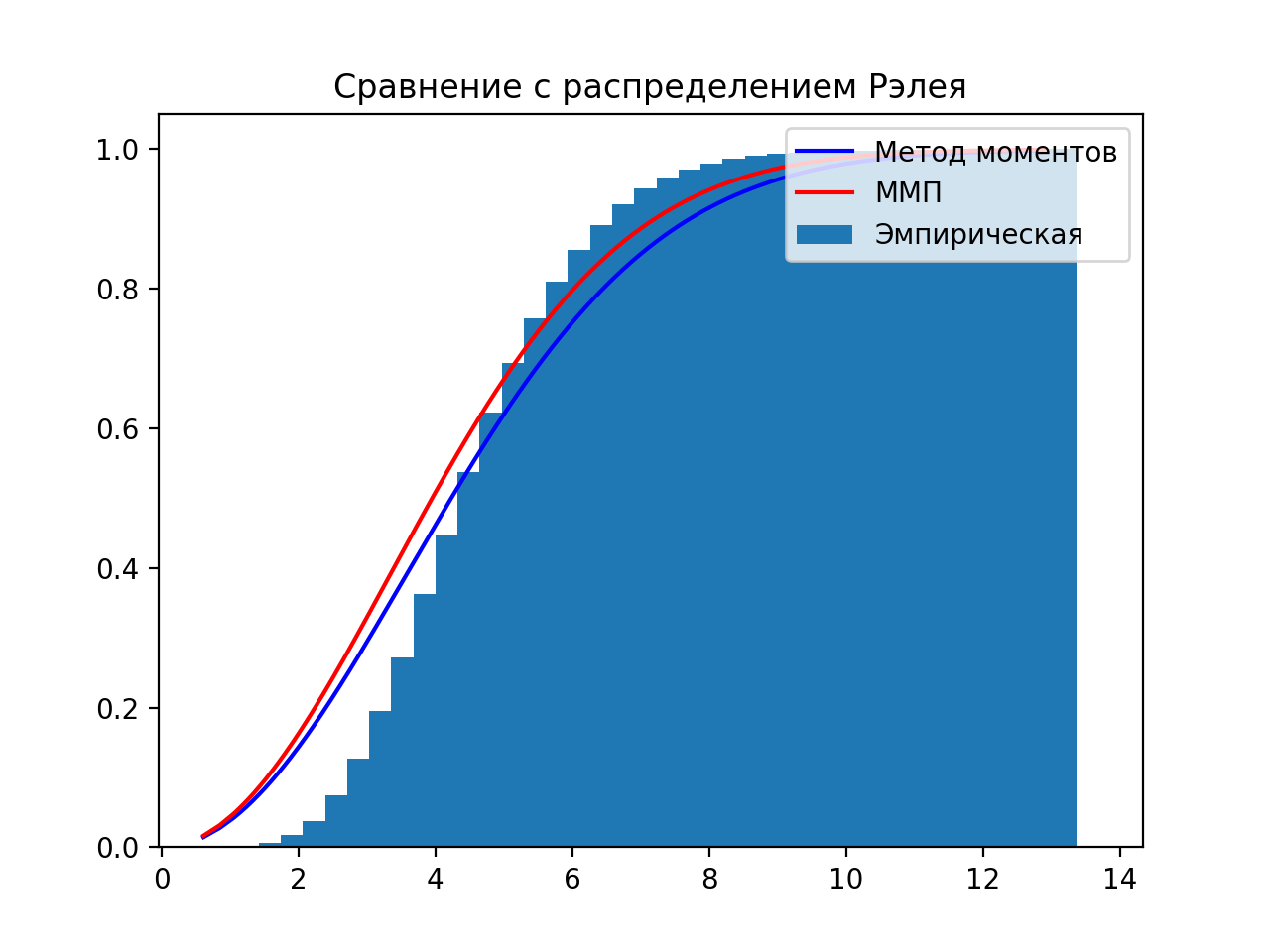
– дигамма функция Эйлера. Заменим на

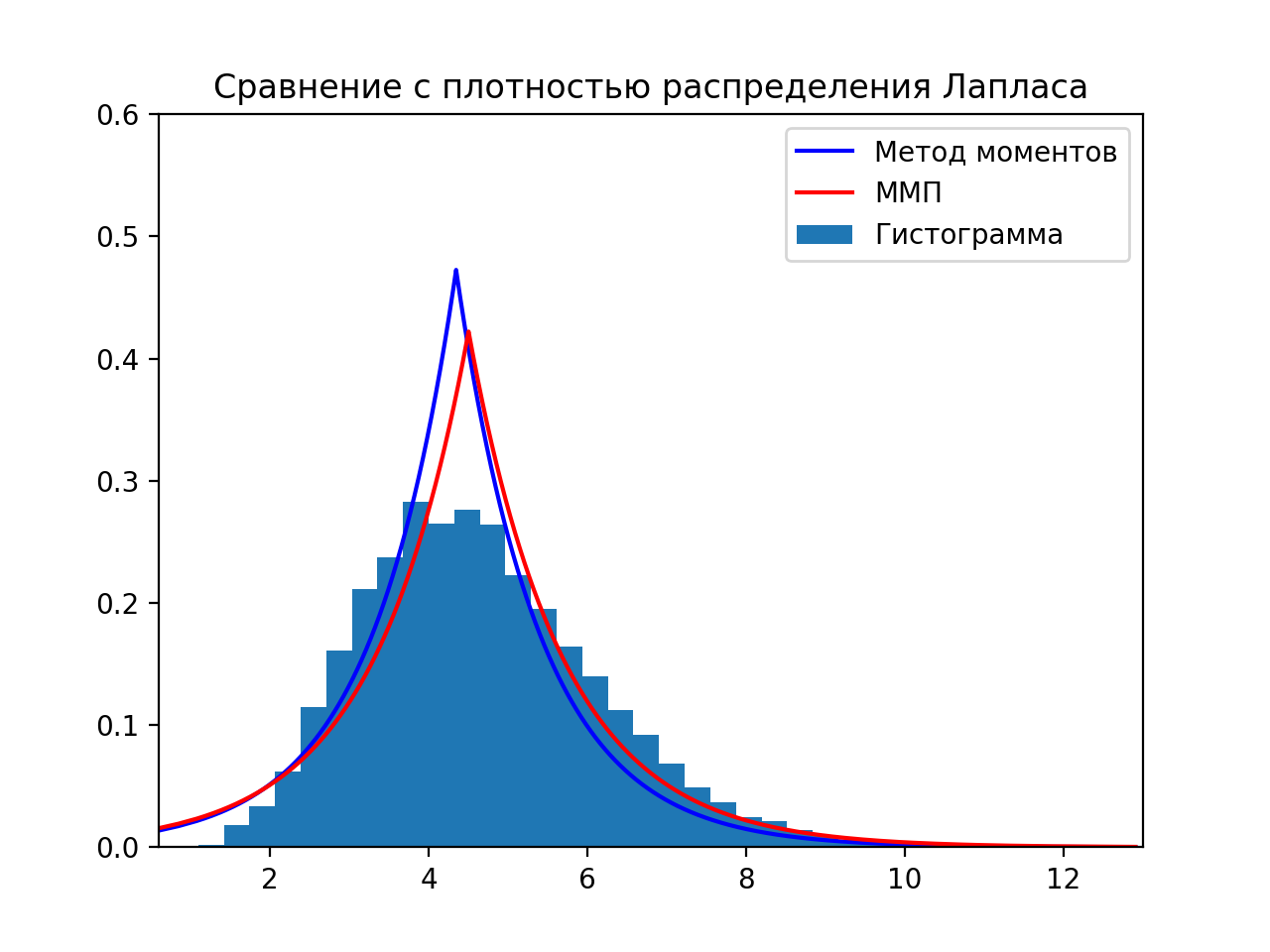
Подставим полученное ранее в последнее уравнение

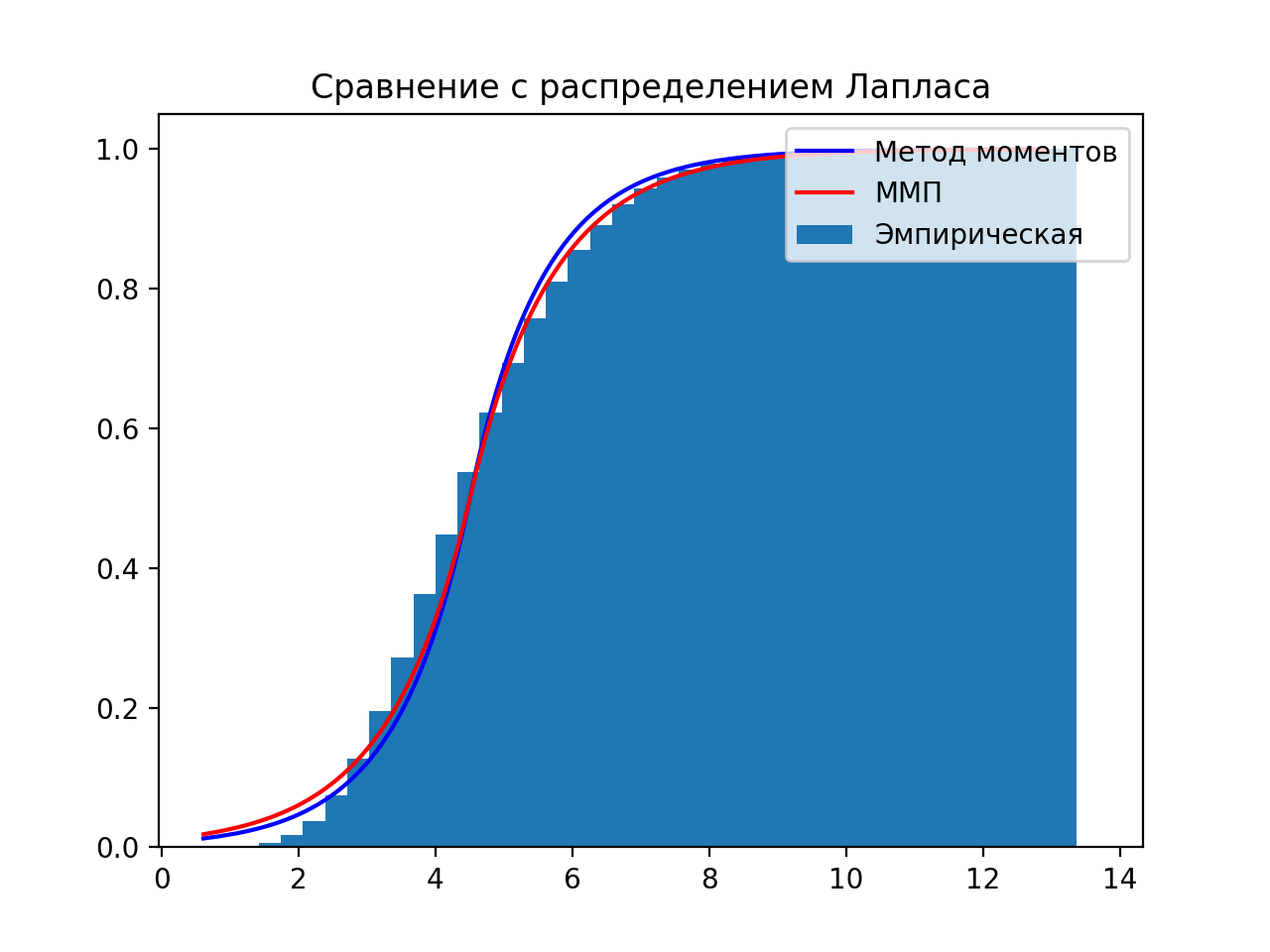
Данную функцию затруднительно считать аналитически, поэтому воспользуемся численными методами минимизации. Сначала заменим выражение справа на c, так как это значение считается единожды. Будем минимизировать функцию в следующем виде

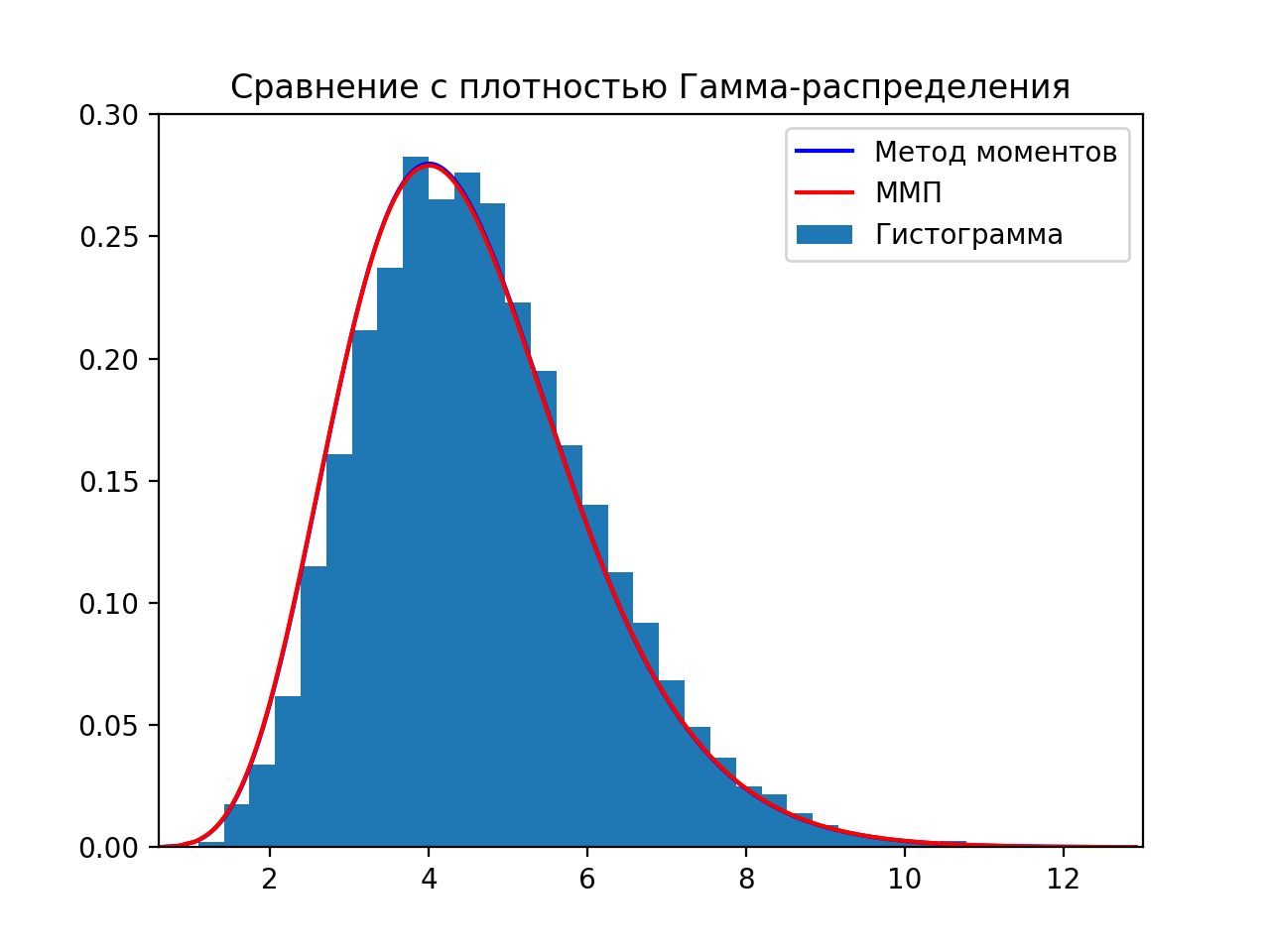
Для минимизации была написана функция, использующая метод бисекции. Точки, при которых функция принимала положительное и отрицательное значения подбирались вручную.

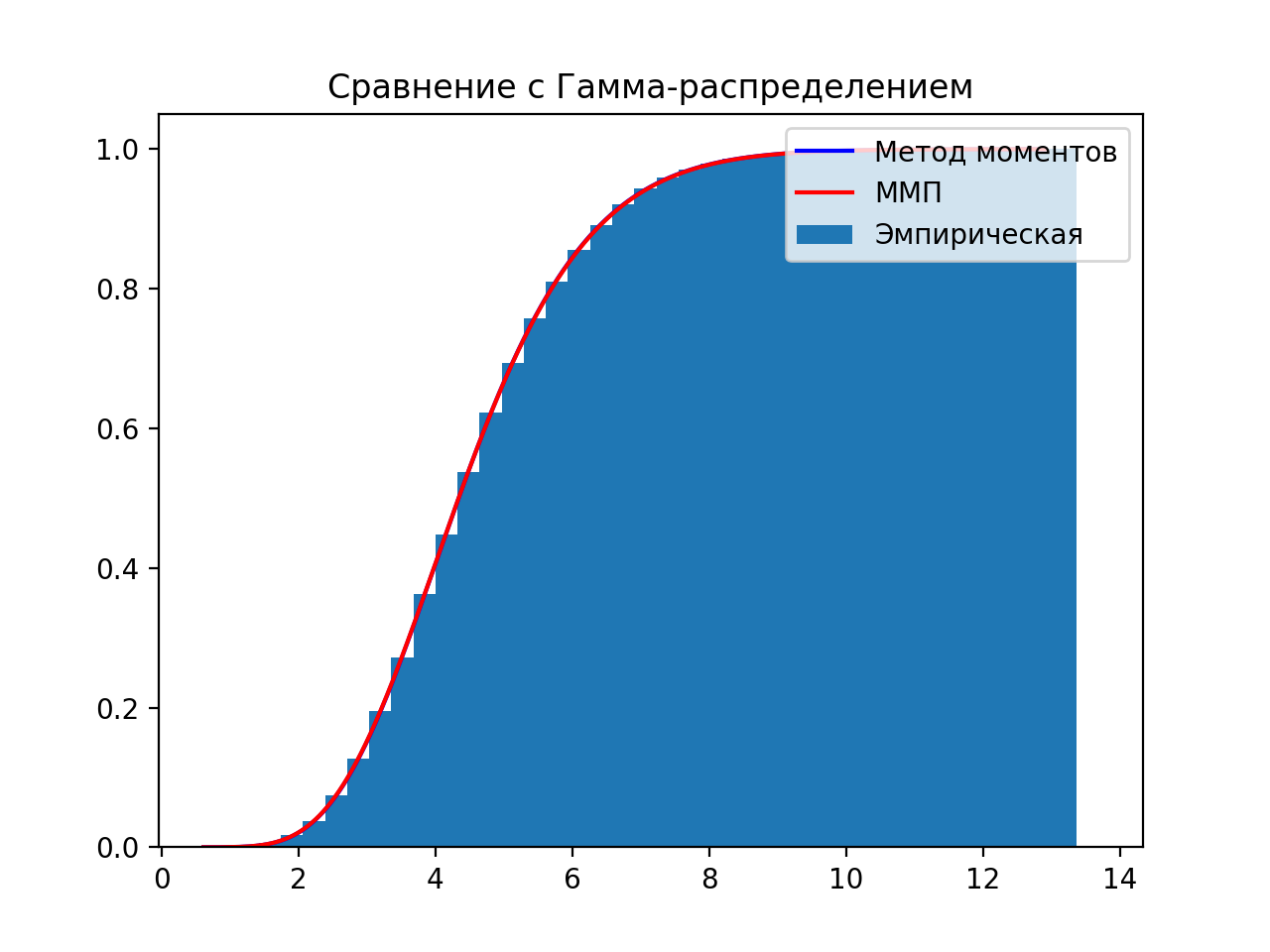
Графики сравнения результатов











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.

Таблица оценок

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Параметры | Распределение Рэлея | Распределение Лапласа | Гамма-распределение |
| Метод моментов |  | 3.592917 | 4.346195 | 9.058622 |
|  | 0.945233 | 0.497101 |
| ММП |  | 3.355278 | 4.503054 | 9.001268 |
|  | 0.844467 | 0.500269 |

**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. Для распределения Рэлея r = 1, а во всех остальных распределениях будет равно 2;

– вероятность попадания в k-ый интервал.

Найдем с помощью функции chi2inv(0.9, 39) и chi2inv(0.9, 38) в *MATLAB*.

(Для распределения Рэлея)

(Для распределения Лапласа и Гамма-распределения)

Таблица результатов

|  |  |  |  |
| --- | --- | --- | --- |
|  | Распределение Рэлея | Распределение Лапласа | Гамма-распределение |
| Метод моментов | 36498001.08173679 | 398209470.810199 | 1206305891.6884854 |
| ММП | 87530372.76508234 | 157628386.89953214 | 1150846781.1913738 |

По критерию «Хи-квадрат» все гипотезы отклонены

**Критерий Колмогорова-Смирнова**

Критическое значение при больших вычисляется по формуле

При получаем

Статистика Комогорова-Смирнова вычисляется по формуле

Таблица результатов

|  |  |  |  |
| --- | --- | --- | --- |
|  | Распределение Рэлея | Распределение Лапласа | Гамма-распределение |
| Метод моментов | 0.14935333681507945 | 0.06919159808271191 | 0.005001224522911185 |
| ММП | 0.18125083360545002 | 0.07896934650605125 | 0.004467861365912484 |

По данному критерию подтверждается гипотеза о Гамма-распределении. Остальные гипотезы были отвергнуты.

**Критерий Мизеса**

Критическое значение при равно

Статистика критерия вычисляется по следующей формуле

Таблица результатов

|  |  |  |  |
| --- | --- | --- | --- |
|  | Распределение Рэлея | Распределение Лапласа | Гамма-распределение |
| Метод моментов | 98.48051473815396 | 21.91475208122736 | 0.03647154427932453 |
| ММП | 2459.3295123091957 | 20.169375369649313 | 0.03606615484079162 |

По критерию Мизеса опять прошла проверку гипотезы о том, что распределение является Гамма-распределением.

**Вывод**

В ходе данной лабораторной работе было выяснено, что изучаемый набор случайных величин имеет Гамма-распределение с параметрами