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. Хираев

(подпись)

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

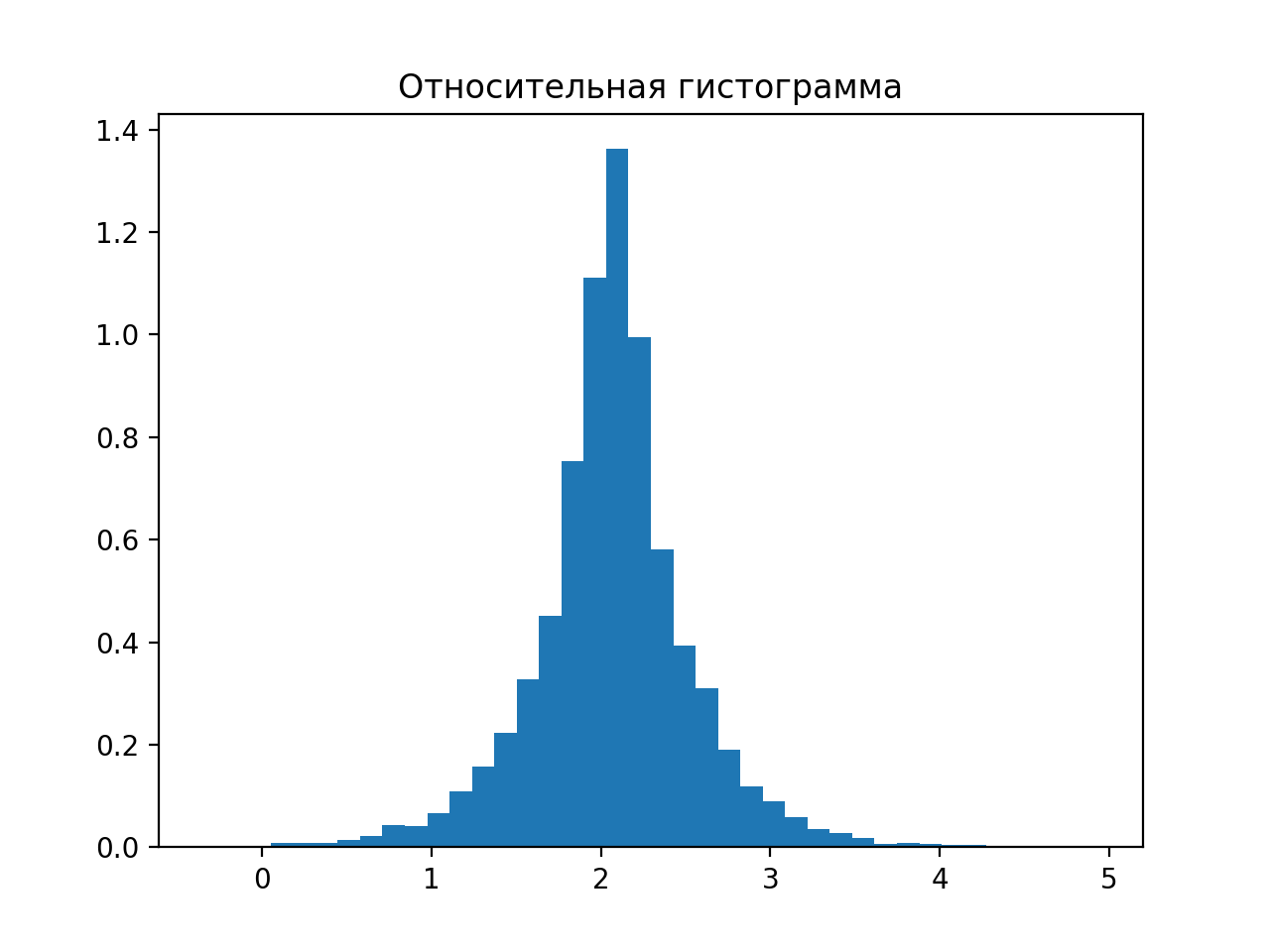
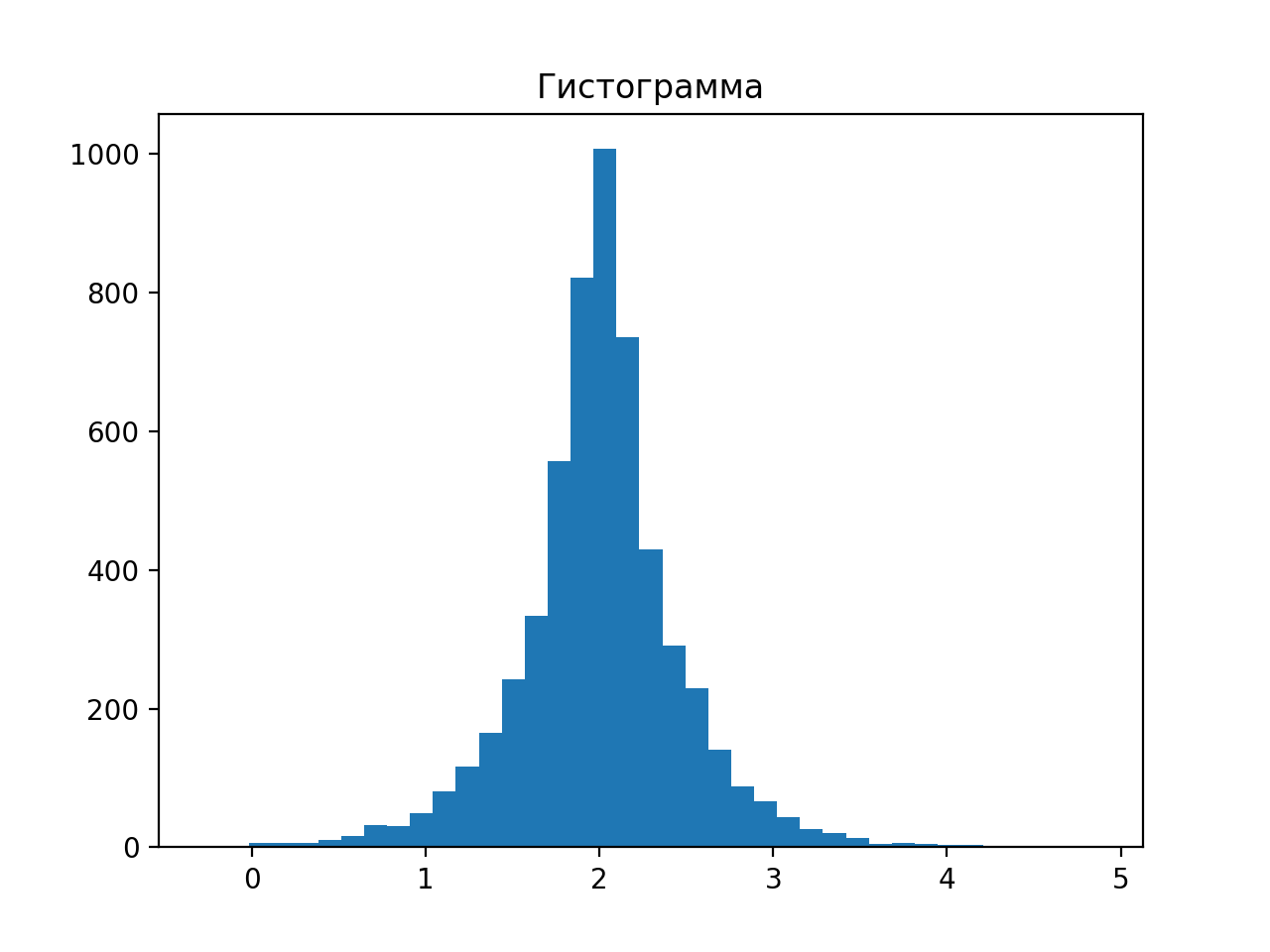
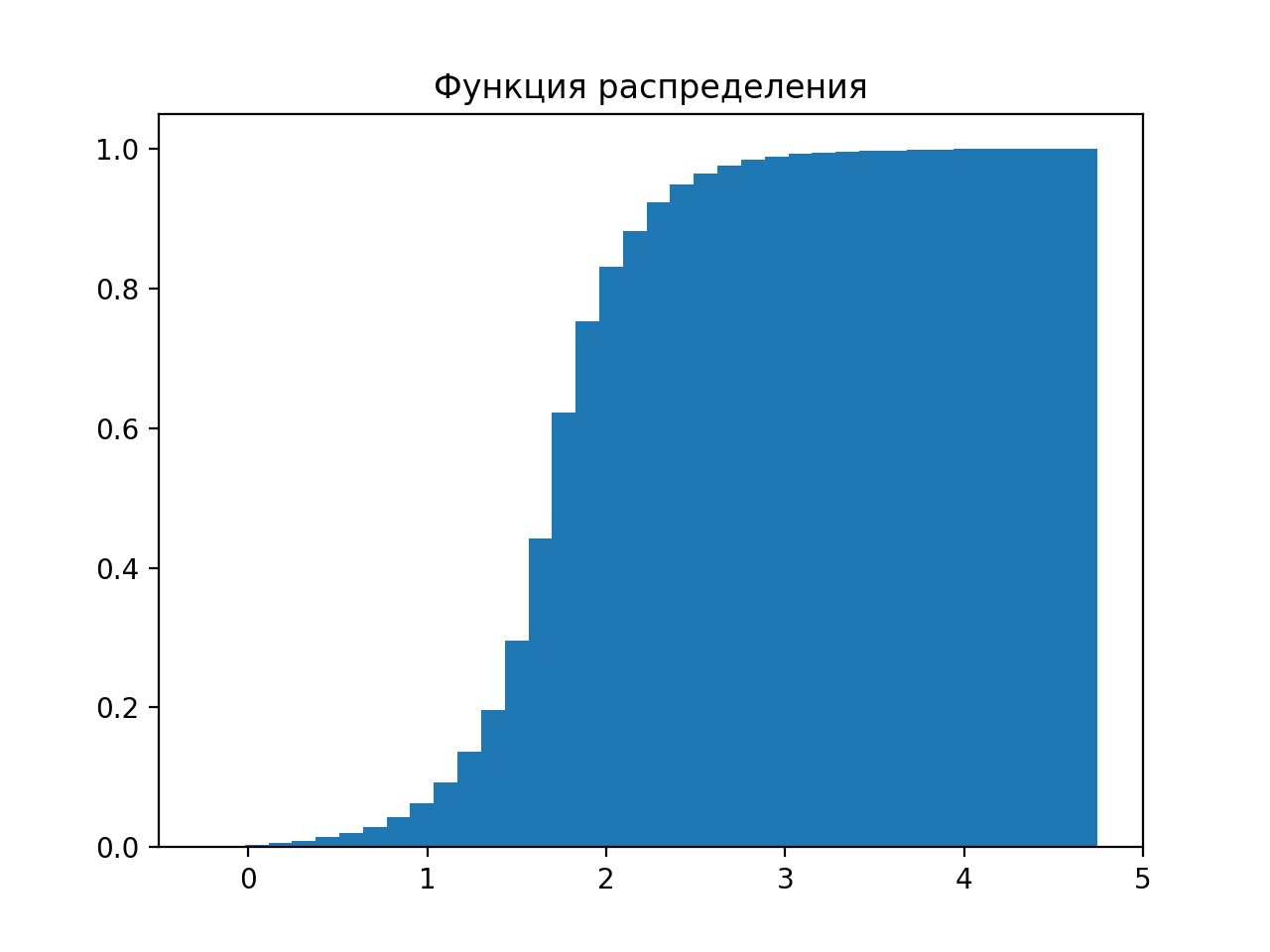
(подпись)  
“\_\_\_” \_\_\_\_\_\_\_\_\_\_\_\_ 2018 г.

Санкт-Петербург  
2018

**1. Статистическая обработка экспериментальных данных**

**1.2 Выборочная функция распределения**

По исходным данным, находящимся в файле *input.txt* построена функция распределения и гисограмма. Количество интервалов m = 40.



Входные данные были перемешаны и соранены в файл *shuffled.txt*. 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 | 2.00344 | 2.00169 | 2.22963 | 0.47391 | 0.22459 | 0.00390 | 0.29900 | 0.01193303003229722 | 5.927806395854247 |
| 1 | 1.98120 | 1.99316 | 2.11393 | 0.47443 | 0.22508 | 0.01888 | 0.30001 | 0.17684284101040504 | 5.921774183219552 |
| 2 | 1.99853 | 1.99569 | 2.11110 | 0.47393 | 0.22461 | 0.00720 | 0.29911 | 0.06767642059682749 | 5.928693366822748 |
| 3 | 2.01807 | 2.01444 | 2.16612 | 0.47413 | 0.22480 | -0.00597 | 0.29906 | -0.05597599491292597 | 5.917719697669919 |
| 4 | 1.98529 | 2.00218 | 1.99529 | 0.47426 | 0.22492 | 0.016134 | 0.29973 | 0.151176064412199 | 5.9248143575148875 |
| 5 | 2.03530 | 2.01583 | 2.20742 | 0.47498 | 0.22560 | -0.01760 | 0.29987 | -0.16424922169703574 | 5.891720492085266 |
| 6 | 1.98649 | 1.98956 | 2.20008 | 0.47421 | 0.22488 | 0.01532 | 0.29965 | 0.14361669206978128 | 5.925543217010234 |
| 7 | 2.02908 | 2.01617 | 1.90967 | 0.47460 | 0.22525 | -0.01340 | 0.29949 | -0.12533545879781102 | 5.902838839523527 |
| 8 | 1.98321 | 1.99118 | 1.88704 | 0.47434 | 0.22500 | 0.01753 | 0.29987 | 0.16425355020258858 | 5.923374704735412 |
| 9 | 2.00806 | 2.00265 | 2.44147 | 0.47393 | 0.22461 | 0.00078 | 0.29896 | 0.007334236035606416 | 5.925823608629138 |
| 10 | 2.00917 | 2.01271 | 2.16489 | 0.47394 | 0.22462 | 0.00003 | 0.29896 | 0.00031309811697113226 | 5.925181952119299 |

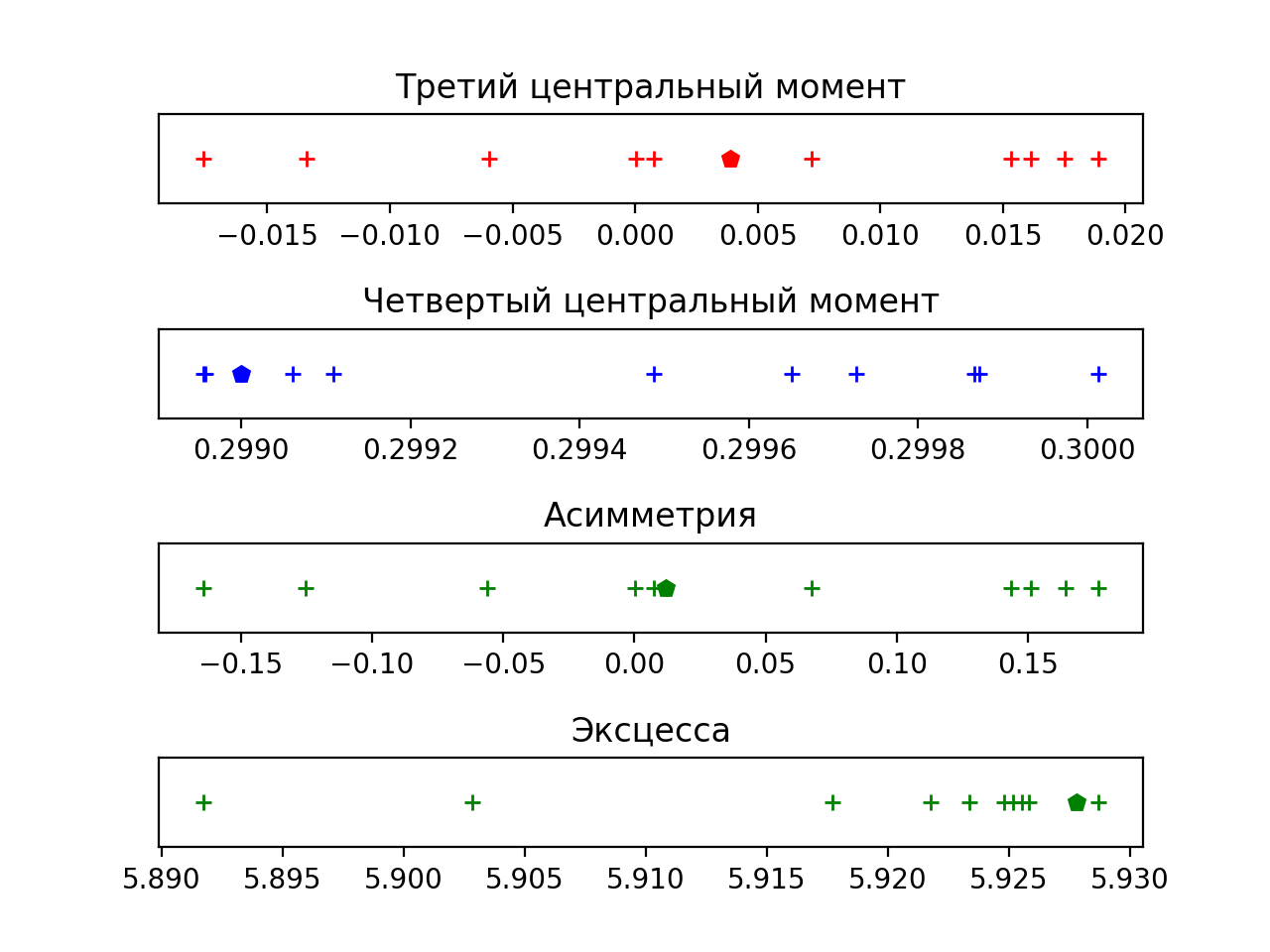
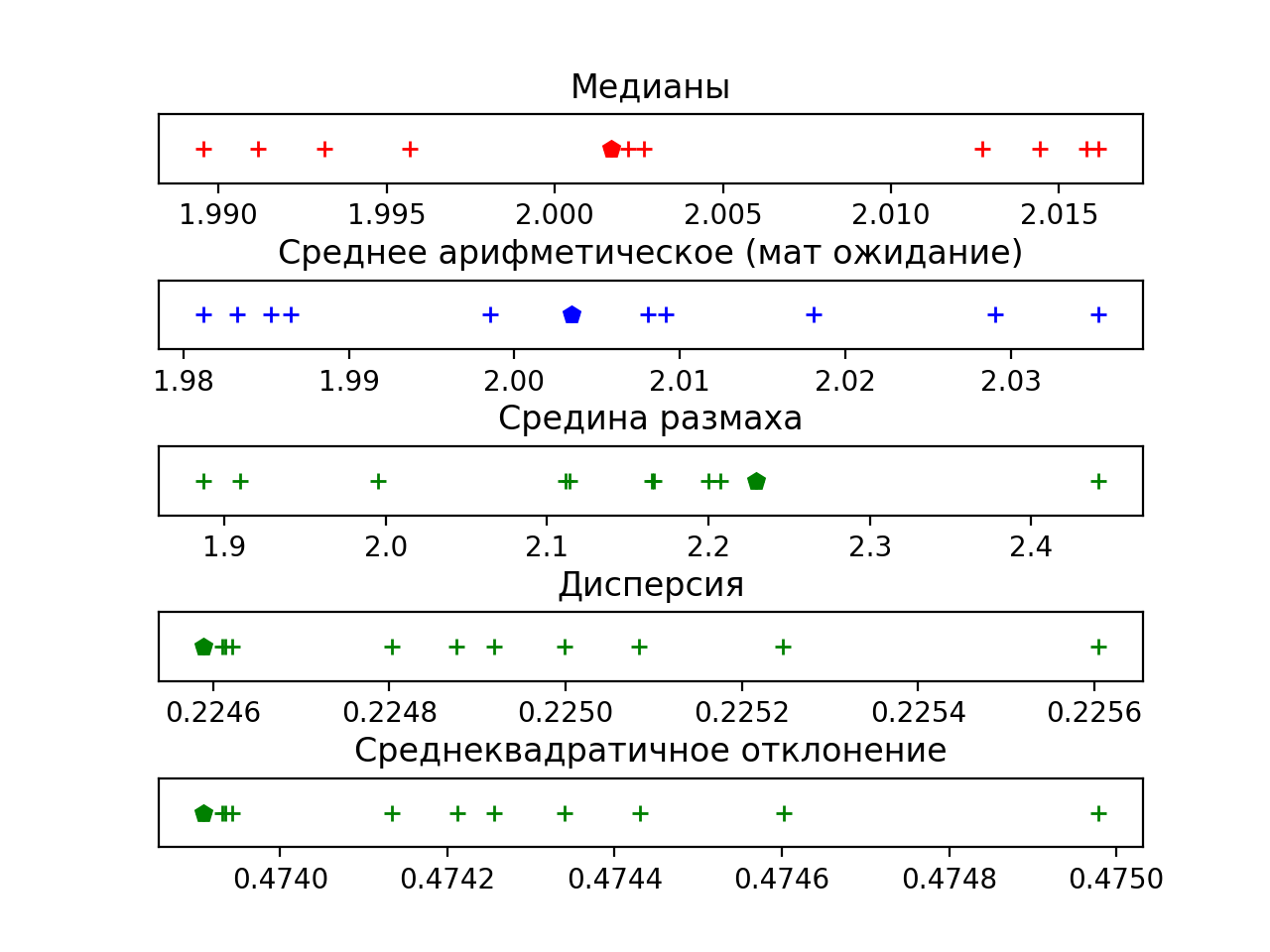
Границы интерквантильного промежутка для P = 0.95:

J = [1.77959, 2.22534]

По номерам точек

J = [1400, 4200]

**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:

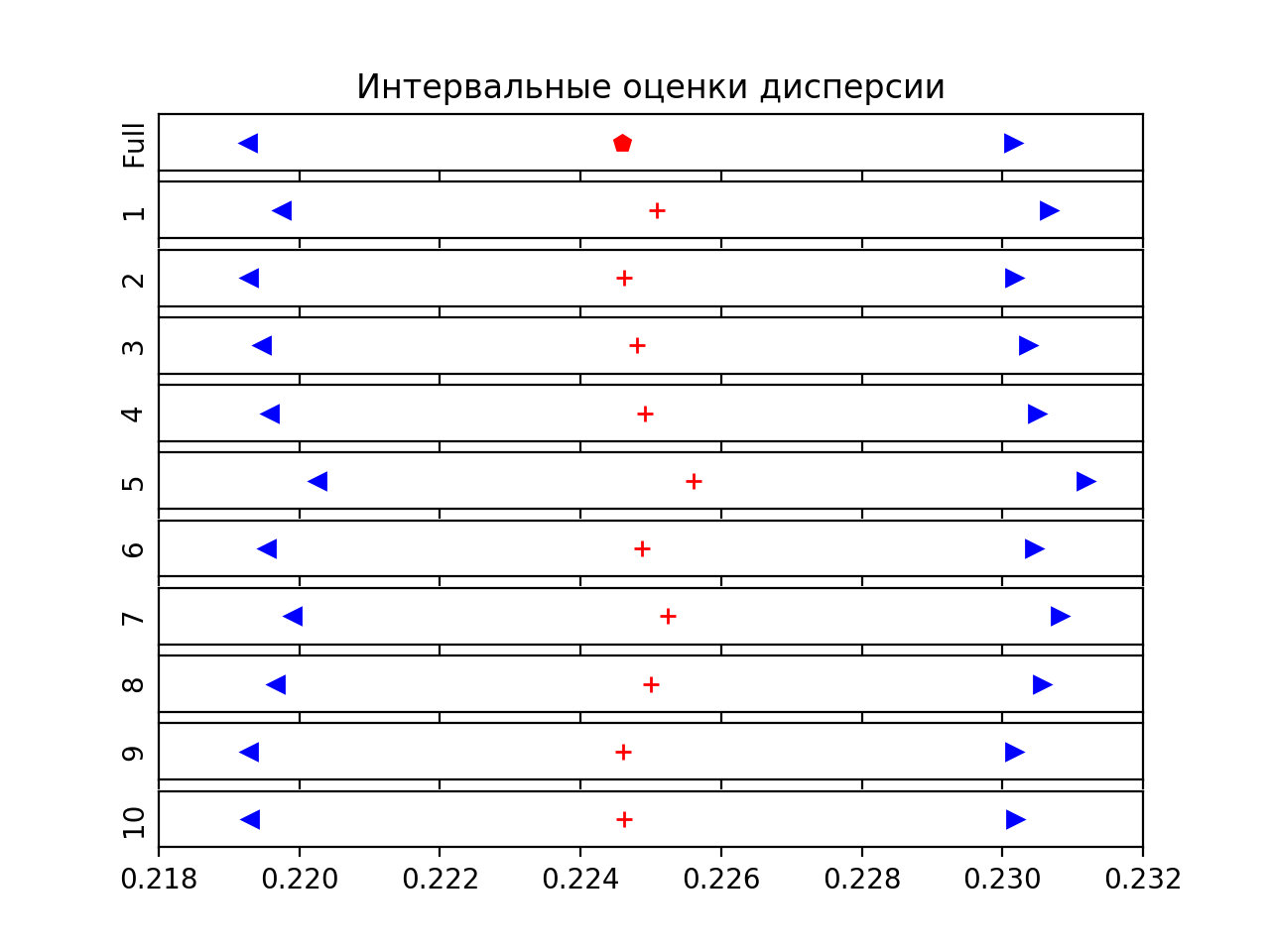
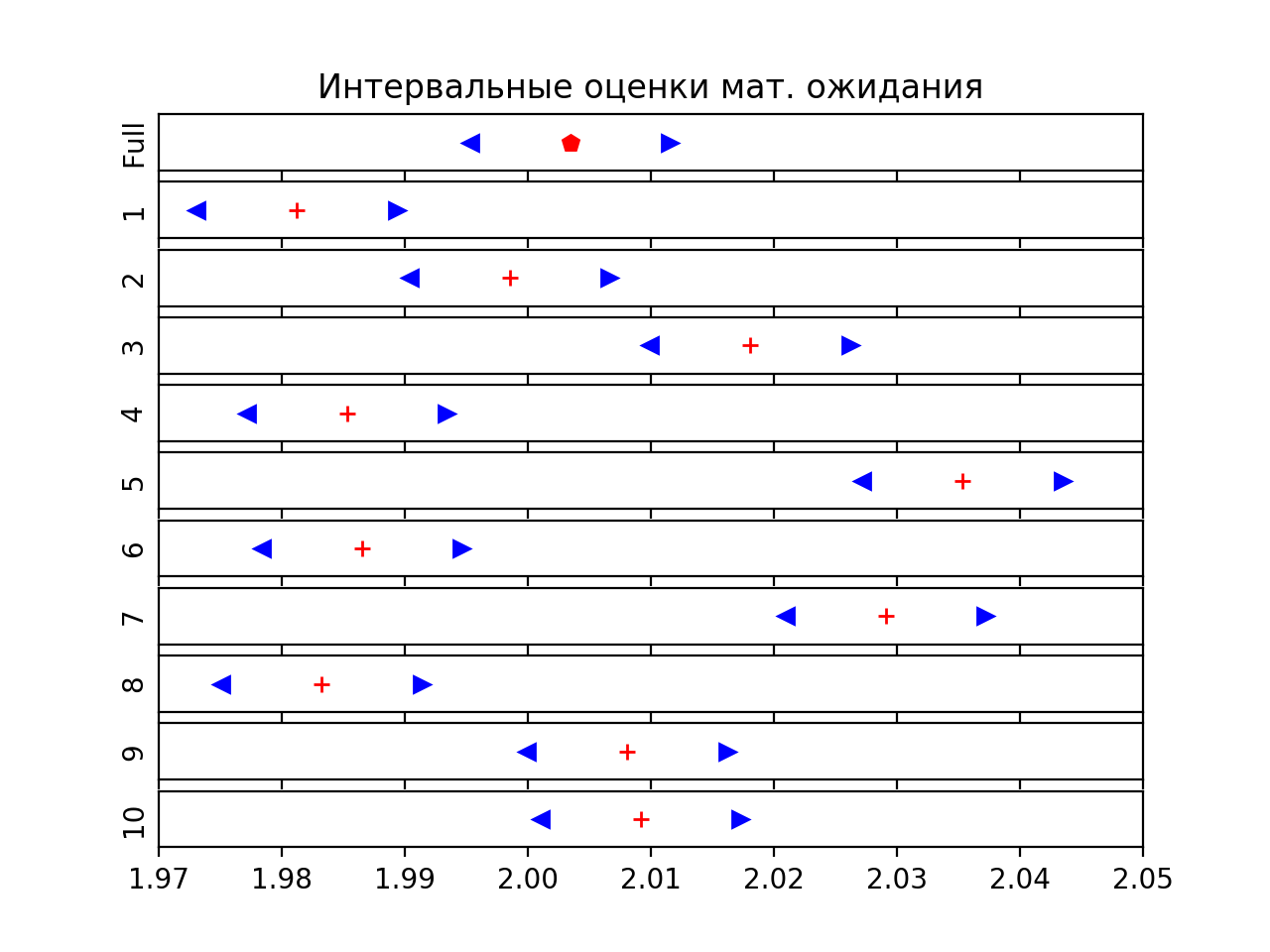
|  |
| --- |
| tinv(0.9, 5599) = **1.2817** |

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

|  |
| --- |
| chi2inv(0.9, 5599) = **5.7350e+03**  chi2inv(0.1, 5599) = **5.4638e+03** |

Таблица интервальных оценок

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Выборка | Мат. ожидание | | Дисперсия | |
| Левая | Правая | Левая | Правая |
| Full | 1.99532 | 2.01156 | 0.21926 | 0.23015 |
| 1 | 1.97308 | 1.98933 | 0.21975 | 0.23065 |
| 2 | 1.99041 | 2.00664 | 0.21929 | 0.23017 |
| 3 | 2.00995 | 2.02619 | 0.21947 | 0.23037 |
| 4 | 1.97717 | 1.99341 | 0.21958 | 0.23048 |
| 5 | 2.02716 | 2.04343 | 0.22025 | 0.23119 |
| 6 | 1.97837 | 1.99461 | 0.21954 | 0.23044 |
| 7 | 2.02096 | 2.03721 | 0.21991 | 0.23082 |
| 8 | 1.97509 | 1.99133 | 0.21966 | 0.23057 |
| 9 | 1.99994 | 2.01618 | 0.21928 | 0.23017 |
| 10 | 2.00105 | 2.01729 | 0.21930 | 0.23018 |

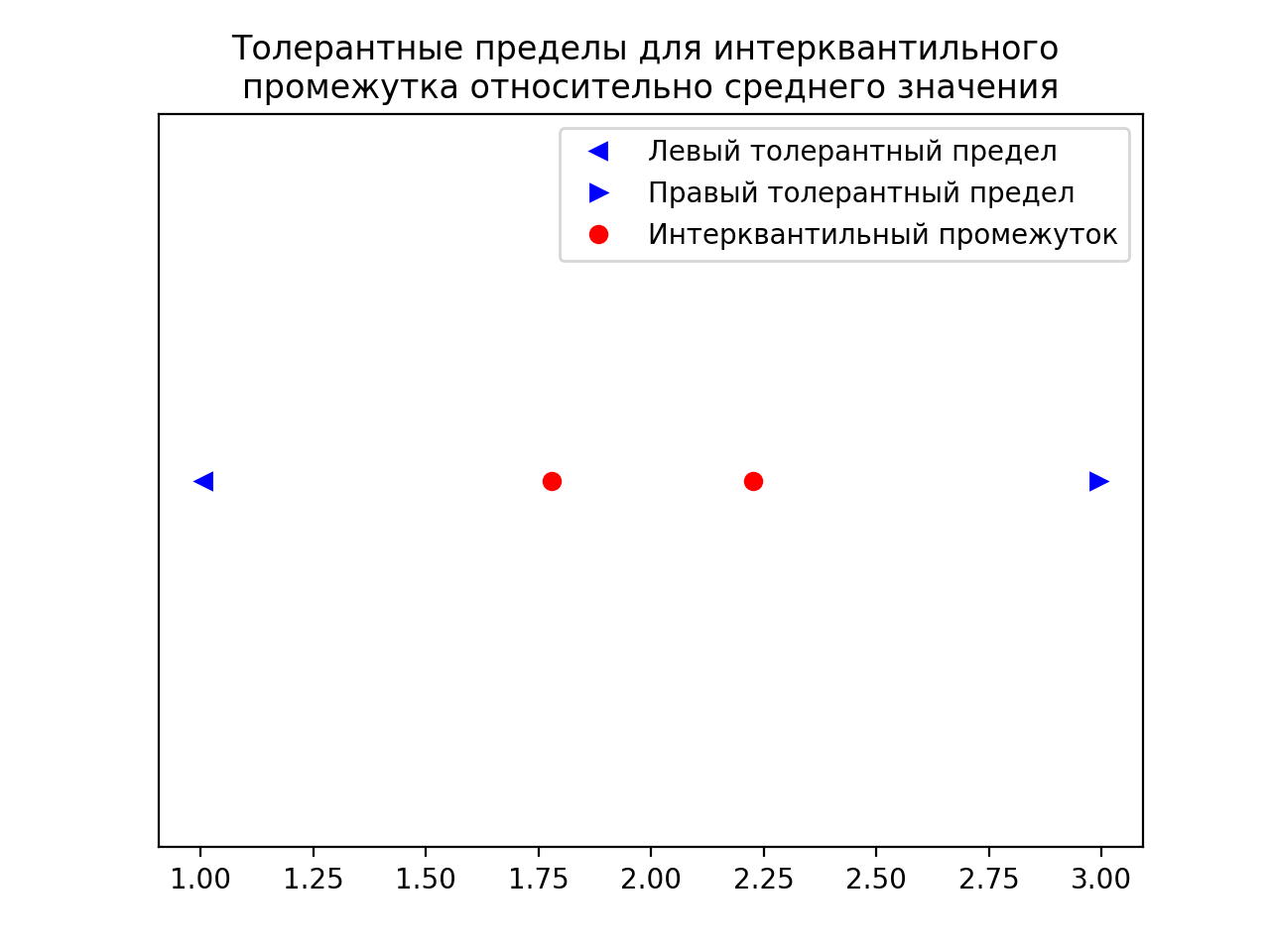
Графики интервальных оценок мат. ожидания и дисперсии  


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

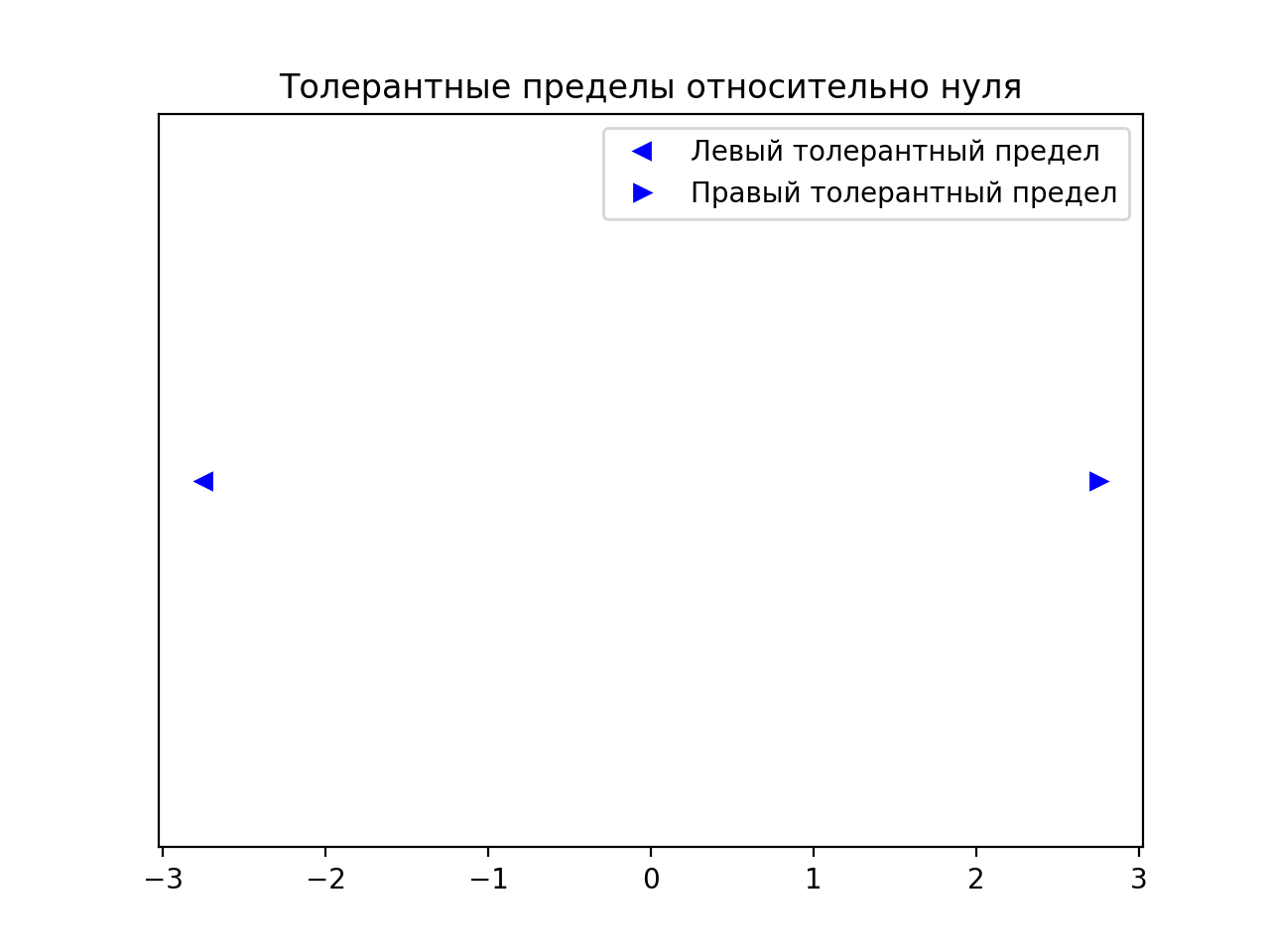
Непараметрические толеранные пределы для всей выборки симметричные относительно среднего значения.

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

Толерантные пределы всей выборки симметричные относительно среднего значения  
[1.00645, 2.9926]

График  
****

Толерантные пределы всей выборки симметричные относительно нуля  
[-2.75199, 2.75199]

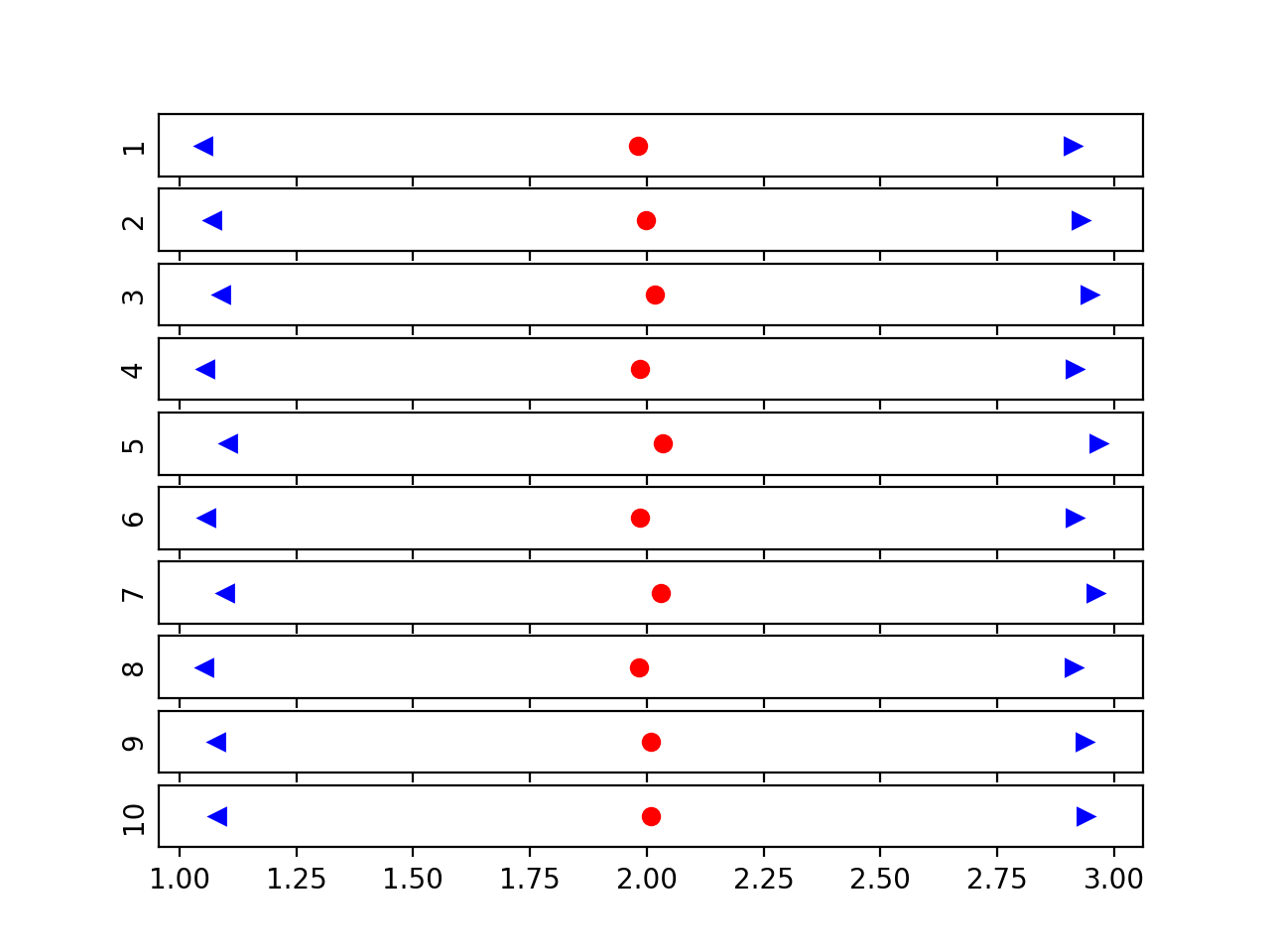
График  


Параметрические толерантные пределы для подбыборок

Параметр

Таблица результатов

|  |  |  |
| --- | --- | --- |
| Номер подвыборки | Левый | Правый |
| 1 | 1.0513200532060945 | 2.911085595008191 |
| 2 | 1.0696148384162243 | 2.927435847298061 |
| 3 | 1.088768497263326 | 2.9473748813081024 |
| 4 | 1.055748803279926 | 2.9148320681486455 |
| 5 | 1.1043413873421906 | 2.9662561198006663 |
| 6 | 1.0570377234836585 | 2.9159466311591986 |
| 7 | 1.0988646877860468 | 2.9593040014996674 |
| 8 | 1.053502522443965 | 2.9129159418417494 |
| 9 | 1.0791558232711833 | 2.9369651899431024 |
| 10 | 1.0802414240167324 | 2.9380982795546964 |

График  
****

Красные точки на графиках – это мат. ожидания подвыборок.

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

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

**2 Идентификация закона**

**2.1 Начальный выбор распределения**

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

**2.2 Определение параметров теоретических распределений**

**Метод моментов**

*Для Нормального распределения:*

*Для распределения Лапласа:*

*Для Гамма-распределения:*

**Метод максимального правдоподобия**

Пусть дана некоторая функция. Возьмем от нее логарифм.

*Для нормального распределения:*

Профиццеренцируем по c и найдем его значение

Теперь 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

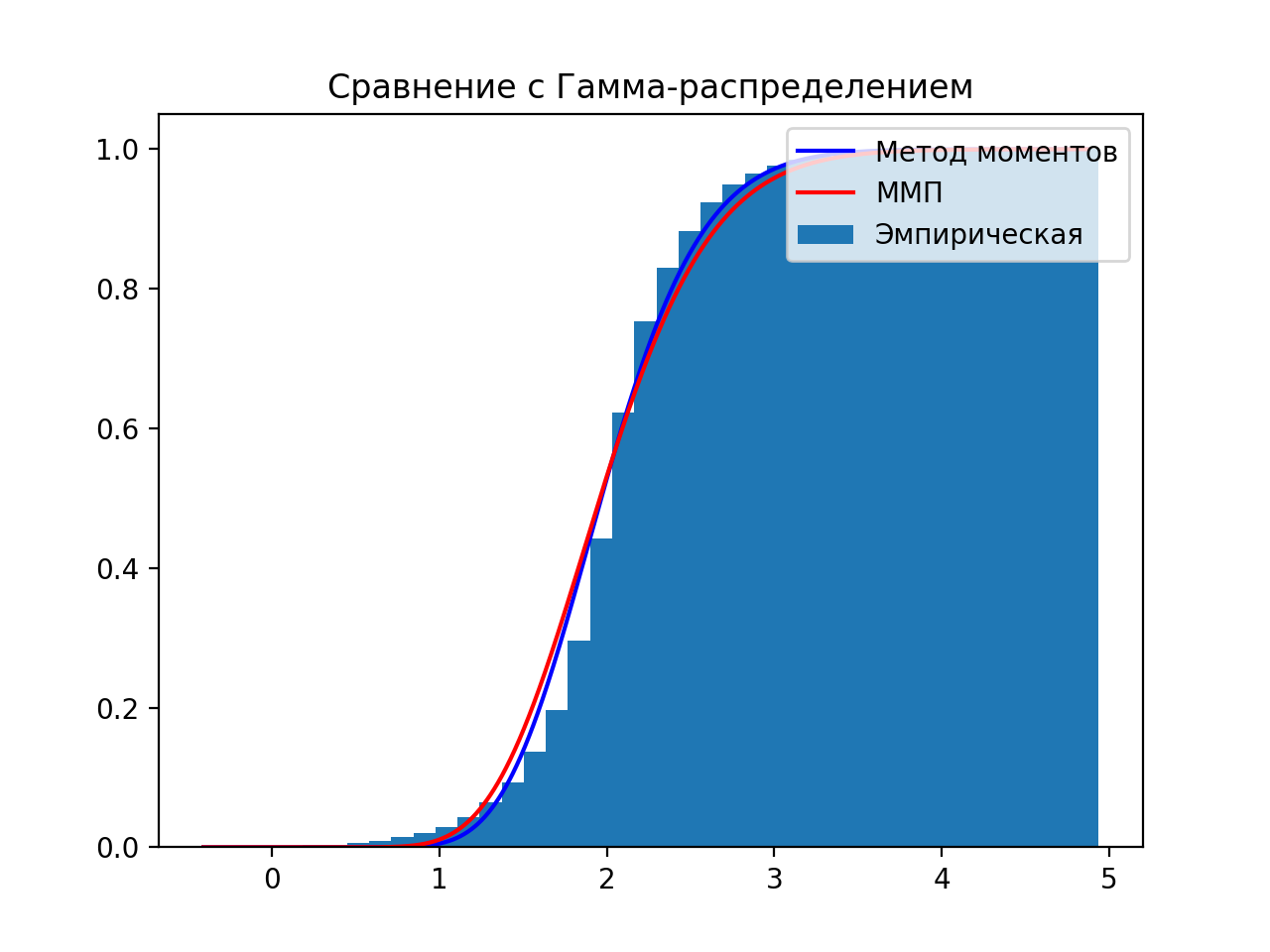
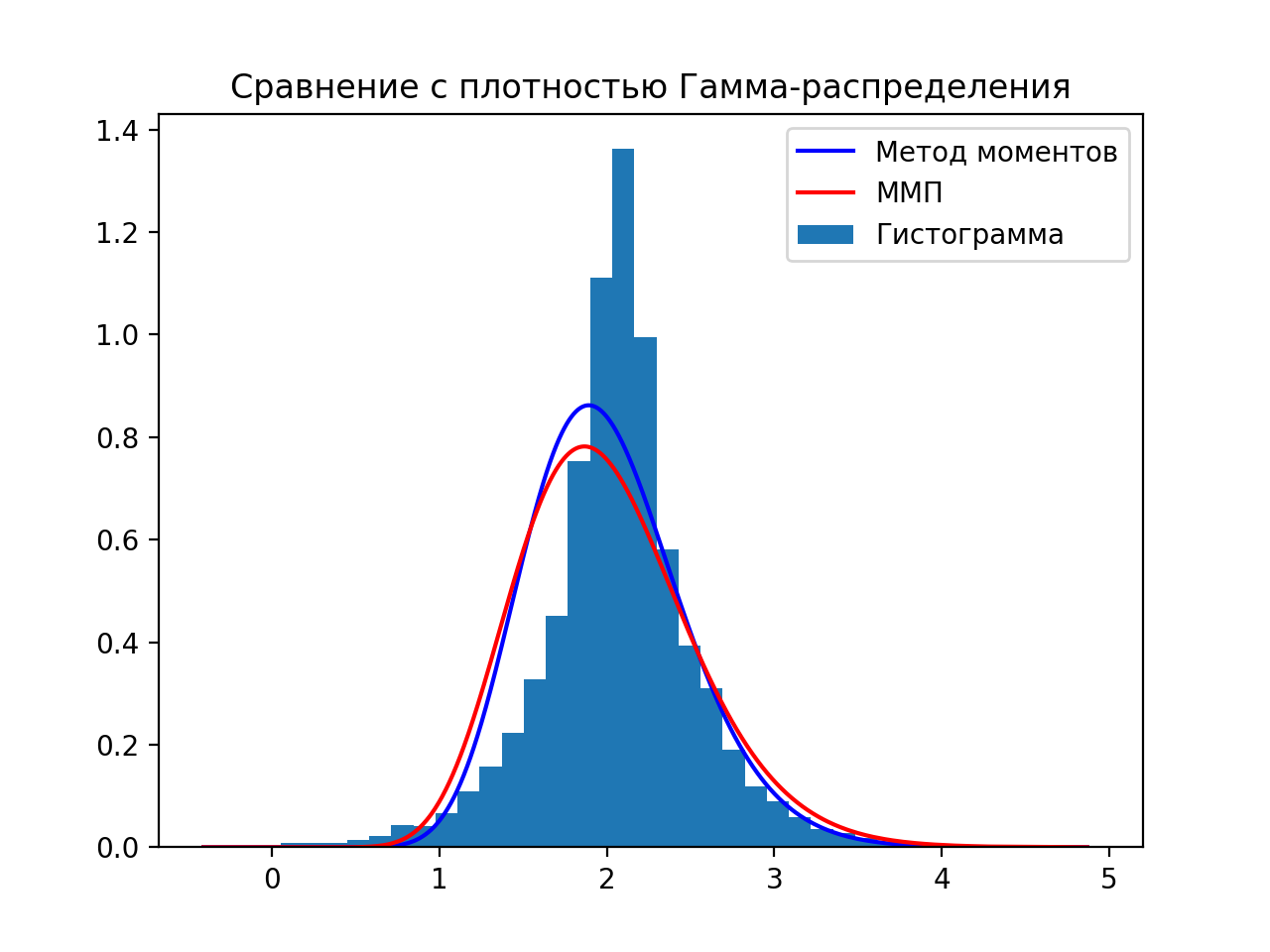
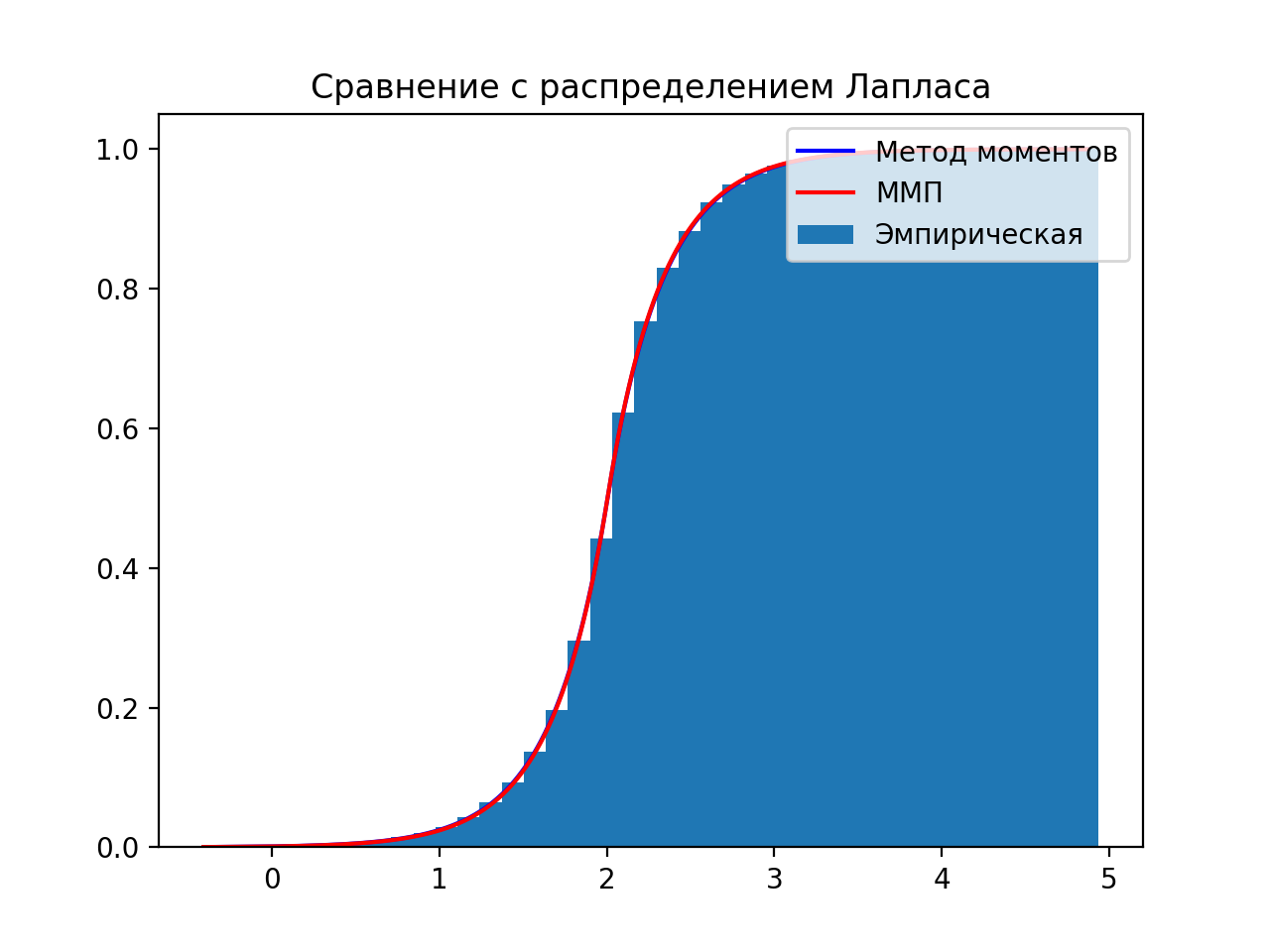
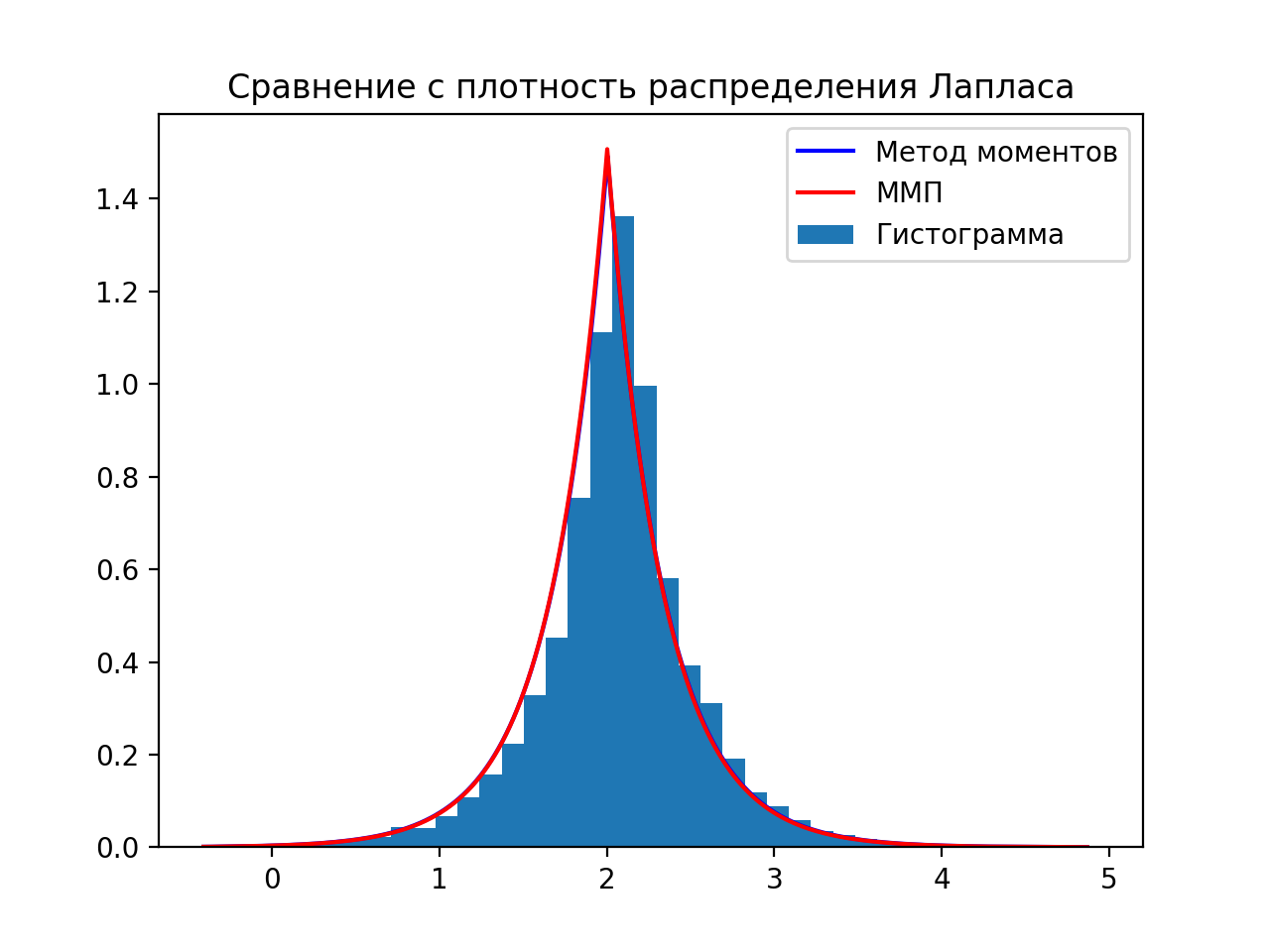
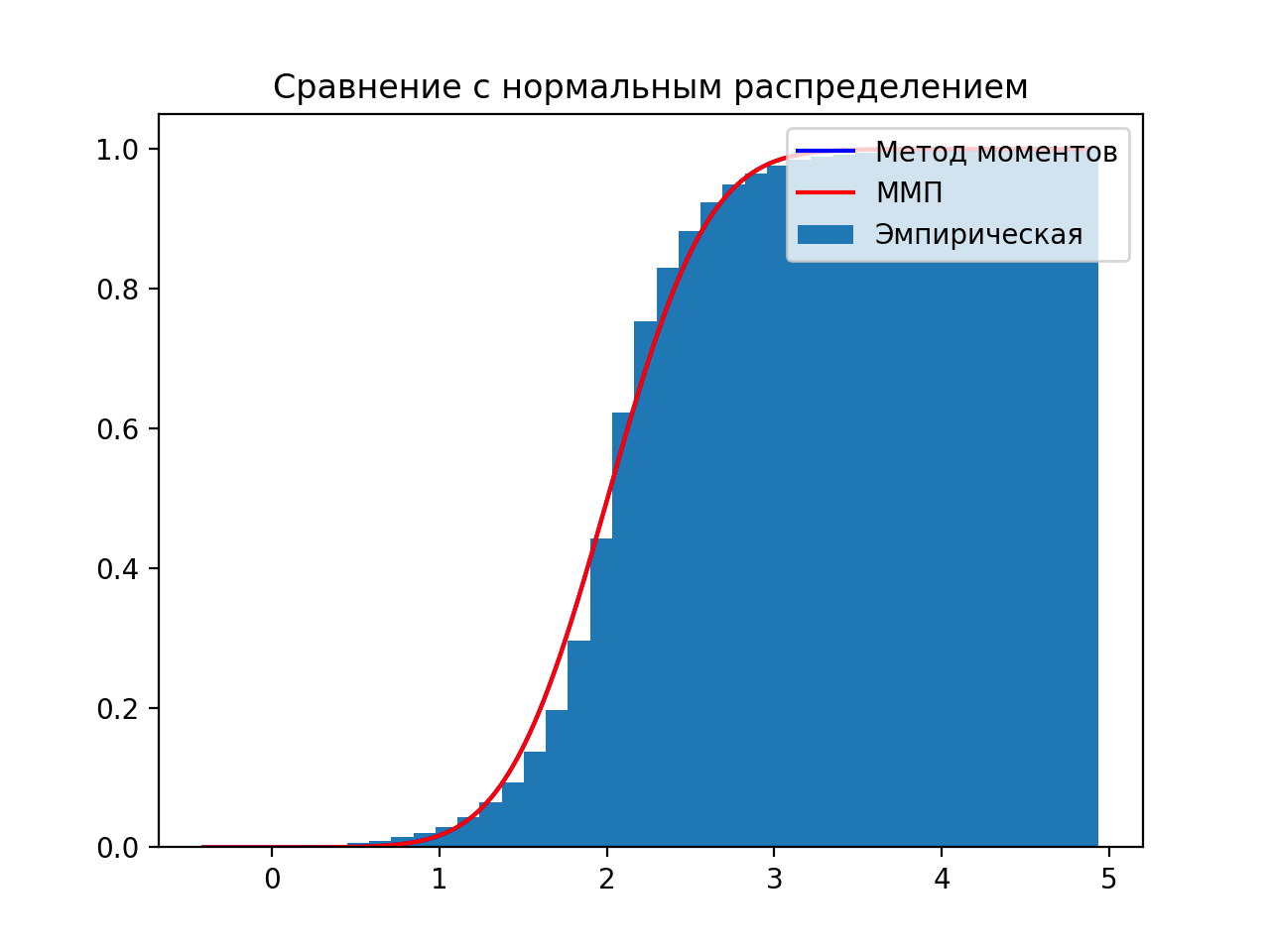
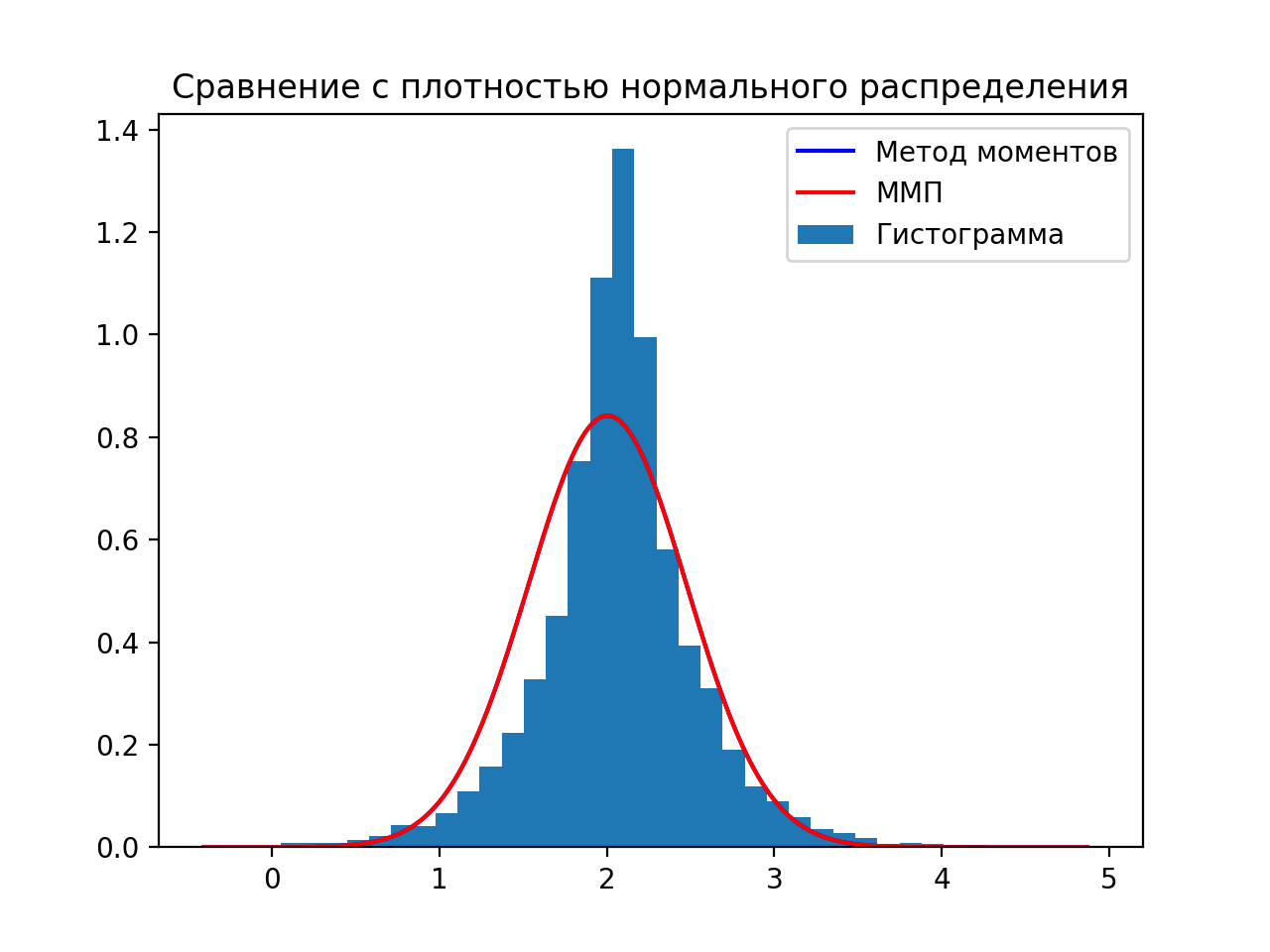
Продифференцируем по

– 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 отрицительное значения подбирались вручную.

Графики сравнения результатов

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

Для нормального распределения и распределения Лапласа оценки параметров, полученные методом моментов и методом максимального правдоподобия очень близки. Для Гамма-распределения оценки параметров отличаются. Судя по графику оценки, полученные методом моментов, оказались ближе к гистограмме, чем оценки, полученные методом максимального правдоподобия. Связано это скорее всего с тем, что данные имеют другое распределение.

Таблица оценок

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Параметры | Нормальное распределение | Распределение Лапласа | Гамма-распределение |
| Метод моментов |  | 2.00344 | 2.00168 | 17.87162 |
|  | 0.47390 | 2.98415 | 0.11210 |
| ММП |  | 2.00344 | 2.00169 | 14.53531 |
|  | 0.47387 | 3.01423 | 0.13784 |

**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. Во всех выбранных распределениях 2 параметра, поэтому везде будет равно 2;

– вероятность попадания в k-ый интервал.

Для всех распределений . Следовательно, и будет одинаковое. Найдем его с помощью функции chi2inv(0.8, 38) в *MATLAB*.

Таблица результатов

|  |  |  |  |
| --- | --- | --- | --- |
|  | Нормальное распределение | Распределение Лапласа | Гамма-распределение |
| Метод моментов | 227085.72656973847 | 43.397882626863236 | 149632624.78285778 |
| ММП | 227760.48432647527 | 44.766153597385596 | 2403799.0540702366 |

Гипотеза о том, что плостноть распределения есть плотность распределения Лапласа с параметрами, полученными методом моментов и ММП подтверждается, а гипотезы о нормальном и Гамма-распределении откланяются.

**Критерий Колмогорова-Смирнова**

Критическое значение при больших вычисляется по формуле

При получаем

Статистика Комогорова-Смирнова вычисляется по формуле

Таблица результатов

|  |  |  |  |
| --- | --- | --- | --- |
|  | Нормальное распределение | Распределение Лапласа | Гамма-распределение |
| Метод моментов | 0.07246623559624266 | 0.01432423011238787 | 0.09271927291110924 |
| ММП | 0.07244981386557925 | 0.011294222268783627 | 0.11473619952682487 |

По данному критерию подтверждается только гипотеза о распределении Лапласа с параметрами, полученными ММП. Однако гипотеза о распределении Лапласа с параметрами, полученными методом моментов, была очень близка к принятию. Остальные гипотезы были отвергнуты.

**Критерий Мизеса**

Критическое значение при равно

Статистика критерия вычисляется по следующей формуле

Таблица результатов

|  |  |  |  |
| --- | --- | --- | --- |
|  | Нормальное распределение | Распределение Лапласа | Гамма-распределение |
| Метод моментов | 12.144325088880466 | 0.17536394606218816 | 15.053071907372733 |
| ММП | 12.138332744857296 | 0.09766194785542075 | 23.87612720795618 |

По критерию Мизеса прошла проверку только гипотезы о том, что распределение является распределением Лапласа с параметрами, полученными методом моментом и ММП.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Название | Нормальное распределение | | | Распределение Лапласа | | | Гамма-распределение | | |
| Фор-ла плотности |  | | |  | | |  | | |
|  | Метод моментов | ММП - аналитич | ММП – числ. | Метод моментов | ММП - аналитич | ММП – числ. | Метод моментов | ММП - аналитич | ММП – числ. |
|  | 2.00344 | 2.00344 | - | 2.00168 | 2.00169 | - | 17.87162 | - | 14.53531 |
|  | 0.47390 | 0.47387 | - | 2.98415 | 3.01423 | - | 0.11210 | - | 0.13784 |
| Хи-квадрат статистика | 227086 | 227760 | - | 43.39788 | 44.76615 | - | 149632625 | - | 2403799 |
| Хи-квадрат критич. знач. |  | | | | | | | | |
| Хи-квадрат вывод | no | no | - | yes | yes | - | no | - | no |
| Колм. - Смирн. статистика | 0.07247 | 0.07245 | - | 0.01432 | 0.01129 | - | 0.09272 | - | 0.11474 |
| Колм. - Смирн. критич. знач. |  | | | | | | | | |
| Колм. - Смирн. вывод | no | no | - | no | yes | - | no | - | no |
| Мизеса - статистика | 12.14433 | 12.13833 | - | 0.17536 | 0.09766 | - | 15.05307 | - | 23.87613 |
| Мизеса критич. знач. |  | | | | | | | | |
| Мизеса вывод | no | no | - | yes | yes | - | no | - | no |

**Вывод**

В данной лабораторныой работе была проанализирована некоторая выборка случайных величин. Были получены точечные и интервальные оценки. С помощью точеченых оценок и гистограммы были сделаны гипотезы и различных распределениях. Далее гипотезы были проверены с помощью различных критериев. В итоге оказалось, что изучаемый набор случайных величин имеет распределение Лапласа с параметрами

.