ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение

высшего профессионального образования

Национальный исследовательский университет

«Высшая школа экономики»

Московский институт электроники и математики

Национального исследовательского университета

«Высшая школа экономики»

Департамент Прикладной Математики

Кафедра «Компьютерная безопасность»

КУРСОВАЯ РАБОТА

по дисциплине «Математическая статистика»

Вариант № 3

Выполнила:

студентка группы СКБ-161

Алемаскина А.С.

Проверил: доцент Иванов А.В.

МОСКВА – 2018

### ЗАДАНИЕ № 2

## Условие:

Рассмотреть 10 файлов формата .jpg. Для каждого файла построить эмпирическое распределение 8-ми, 16-ти и 24-х битовых векторов, вычислить выборочное среднее и выборочную дисперсию соответственно 8-ми, 16-ти и 24-х битовых векторов. Построить аналогичные характеристики для заархивированных файлов и для файлов, зашифрованных алгоритмом блочного шифрования в режиме гаммирования с обратной связью по шифртексту (CFB). Для каждого из файлов (вместе с заархивированными и зашифрованными) проверить гипотезу о равновероятности распределения векторов соответствующей длины. Оценить возможность определить по имеющемуся распределению двоичных векторов в файле, является ли этот файл открытым или зашифрованным.

## Теоретическая часть

**Выборочное среднее** выборки обозначается как и определяется формулой

**Выборочная дисперсия** выборки обозначается как и определяется формулой

**Критерий согласия** **χ2 (Хи-квадрат)**

Критерий для проверки гипотезы о принадлежности выборки некоторому теоретическому закону распределения

Критерий может использоваться при проверке простых гипотез вида

**Проверка**

Область определения случайной величины разбивают на k непересекающихся интервалов граничными точками и подсчитывают число значений из выборки попавших в i-й интервал и вероятности попадания в интервал

соответствующие теоретическому закону с функцией распределения

При проверке простой гипотезы известны как вид закона так и все его параметры (известен скалярный или векторный параметр (наш случай)

В основе статистик, используемых при проверке критерия согласия хи-квадрат лежит измерение отклонений от

Статистика критерия согласия хи-квадрат определяется соотношением

При проверке простой гипотезы при большом n эта статистика подчиняется распределению хи-квадрат с r = k-1 степенями свободы.

При проверке простой гипотезы задается вероятность ошибки первого рода , считается(берется из таблицы) квантиль теоретического распределения хи-квадрат для заданного

Если , то гипотеза принимается.

## Ход эксперимента

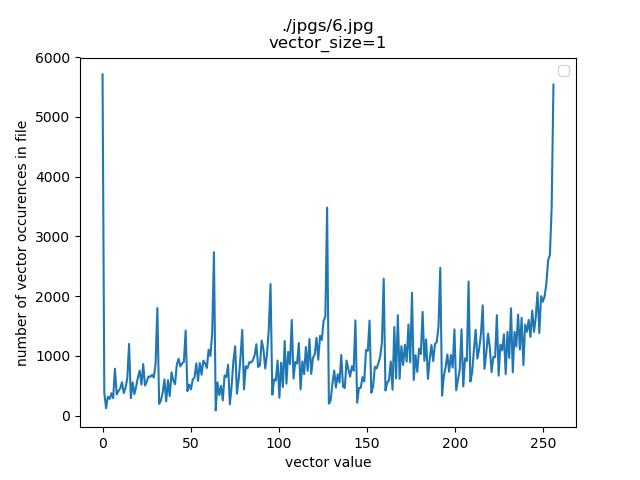
1. Были взяты 10 произвольных .jpg изображений, каждое из которых было зашифрована алгоритмом AES в режиме CFB, а также заархивировано. Далее для каждого из получившихся файлов были выполнены следующие операции:
2. Каждое изображение разбивается на наборы из 1,2 и 3-х байтовых векторов.
3. Для каждой выборки вычисляются выборочное среднее и выборочная дисперсия.
4. Вычисляется экспериментальное значение , для берутся теоретические значения квантилей, проверяется гипотеза о равновероятности распределения векторов соответствующей длины (если экспериментальное значение меньше теоретического, то гипотеза о равномерном распределении векторов принимается, если больше, то отклоняется).

## Результаты

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Наименование файла | Размер вектора, байт | Выборочное среднее | Выборочная дисперсия | Хи2 полученный | Хи2 теоретический | Проверка гипотезы |
| img/1.jpg | 1 | 125.8391429 | 5550.01909255 | 592.685377999 | 292.17485804729586 | N0 |
| cfb/1.jpg.enc | 1 | 127.617323072 | 5407.4524719 | 195.912403803 | 292.17485804729586 | Yes |
| zipped/1.zip | 1 | 130.160755276 | 5383.21667777 | 3236.97267075 | 292.17485804729586 | N0 |
| img/1.jpg | 2 | 32231.5719028 | 364570106.683 | 78290.6165686 | 66130.62639516569 | N0 |
| cfb/1.jpg.enc | 2 | 32597.3306172 | 352895606.787 | 64881.0104119 | 66130.62639516569 | Yes |
| zipped/1.zip | 2 | 33431.7418183 | 352709032.894 | 122294.557282 | 66130.62639516569 | N0 |
| img/1.jpg | 3 | 8278567.88185 | 2.40196404856e+13 | 17798113.4438 | 16786743.146288473 | N0 |
| cfb/1.jpg.enc | 3 | 8368875.07696 | 2.33443863506e+13 | 16780391.9364 | 16786743.146288473 | Yes |
| zipped/1.zip | 3 | 8564257.79643 | 2.30910759506e+13 | 17044614.2112 | 16786743.146288473 | N0 |
| img/2.jpg | 1 | 124.703854335 | 5325.00303203 | 2560.52514252 | 292.17485804729586 | N0 |
| cfb/2.jpg.enc | 1 | 126.860767464 | 5478.50027737 | 263.236790607 | 292.17485804729586 | Yes |
| zipped/2.zip | 1 | 125.983091149 | 5472.30573409 | 1115.34718815 | 292.17485804729586 | N0 |
| img/2.jpg | 2 | 32104.7206548 | 348259364.786 | 91844.3714392 | 66130.62639516569 | N0 |
| cfb/2.jpg.enc | 2 | 32548.1224858 | 362279779.51 | 65410.0107545 | 66130.62639516569 | Yes |
| zipped/2.zip | 2 | 32392.5819589 | 358294825.767 | 79556.6708813 | 66130.62639516569 | N0 |
| img/2.jpg | 3 | 8195196.28909 | 2.2706369329e+13 | 17084794.1101 | 16786743.146288473 | N0 |
| cfb/2.jpg.enc | 3 | 8339244.63005 | 2.33700844482e+13 | 16783319.8296 | 16786743.146288473 | Yes |
| zipped/2.zip | 3 | 8293360.72361 | 2.33750534838e+13 | 16829271.7999 | 16786743.146288473 | N0 |
| img/3.jpg | 1 | 124.071179636 | 5353.61818046 | 1670.67948253 | 292.17485804729586 | N0 |
| cfb/3.jpg.enc | 1 | 127.675523462 | 5444.62435504 | 243.970334848 | 292.17485804729586 | Yes |
| zipped/3.zip | 1 | 126.710760456 | 5425.77960192 | 688.83400789 | 292.17485804729586 | N0 |
| img/3.jpg | 2 | 31829.1190476 | 350289232.923 | 97671.2893944 | 66130.62639516569 | N0 |
| cfb/3.jpg.enc | 2 | 32752.6724657 | 356446621.856 | 65548.934259 | 66130.62639516569 | Yes |
| zipped/3.zip | 2 | 32580.9913247 | 356321230.674 | 72538.7408167 | 66130.62639516569 | N0 |
| img/3.jpg | 3 | 8218352.28429 | 2.294470342e+13 | 18763481.997 | 16786743.146288473 | N0 |
| cfb/3.jpg.enc | 3 | 8390049.58916 | 2.33480612839e+13 | 16784904.4501 | 16786743.146288473 | Yes |
| zipped/3.zip | 3 | 8343533.265 | 2.33341602254e+13 | 16856792.484 | 16786743.146288473 | N0 |
| img/4.jpg | 1 | 129.471003222 | 5228.74465969 | 1715.31443728 | 292.17485804729586 | N0 |
| cfb/4.jpg.enc | 1 | 126.555799356 | 5478.1373986 | 269.287134763 | 292.17485804729586 | Yes |
| zipped/4.zip | 1 | 128.076265932 | 5572.43476222 | 357.100585601 | 292.17485804729586 | N0 |
| img/4.jpg | 2 | 33269.7416398 | 343635741.582 | 78957.1817576 | 66130.62639516569 | N0 |
| cfb/4.jpg.enc | 2 | 32579.7939118 | 354855378.702 | 65662.3356294 | 66130.62639516569 | Yes |
| zipped/4.zip | 2 | 33103.5990631 | 361570646.223 | 70430.5009646 | 66130.62639516569 | N0 |
| img/4.jpg | 3 | 8475654.31211 | 2.24658199303e+13 | 17209735.1098 | 16786743.146288473 | N0 |
| cfb/4.jpg.enc | 3 | 8280328.01741 | 2.35019560666e+13 | 16773600.2089 | 16786743.146288473 | Yes |
| zipped/4.zip | 3 | 8355407.42757 | 2.43762815722e+13 | 17049742.6457 | 16786743.146288473 | N0 |
| img/5.jpg | 1 | 126.95995419 | 5220.47638832 | 1699.01450659 | 292.17485804729586 | N0 |
| cfb/5.jpg.enc | 1 | 127.50975377 | 5477.61436373 | 217.497423172 | 292.17485804729586 | Yes |
| zipped/5.zip | 1 | 128.459121808 | 5439.64204567 | 355.242329598 | 292.17485804729586 | N0 |
| img/5.jpg | 2 | 32700.2753092 | 344966732.011 | 77003.8689876 | 66130.62639516569 | N0 |
| cfb/5.jpg.enc | 2 | 32797.3162315 | 359522069.875 | 65746.1387998 | 66130.62639516569 | Yes |
| zipped/5.zip | 2 | 33316.3330067 | 356825350.19 | 69746.6037967 | 66130.62639516569 | N0 |
| img/5.jpg | 3 | 8401676.01706 | 2.22815252002e+13 | 17356415.6369 | 16786743.146288473 | N0 |
| cfb/5.jpg.enc | 3 | 8397252.10811 | 2.35199881015e+13 | 16768482.9995 | 16786743.146288473 | Yes |
| zipped/5.zip | 3 | 8419955.41962 | 2.32502242616e+13 | 17129324.2576 | 16786743.146288473 | N0 |
| img/6.jpg | 1 | 146.920587378 | 5786.2395693 | 98158.5821517 | 292.17485804729586 | N0 |
| cfb/6.jpg.enc | 1 | 127.308834309 | 5464.35095184 | 224.855398088 | 292.17485804729586 | Yes |
| zipped/6.zip | 1 | 130.201231479 | 5415.05933036 | 1939.32342575 | 292.17485804729586 | N0 |
| img/6.jpg | 2 | 37677.821225 | 379451352.76 | 4070691.9414 | 66130.62639516569 | N0 |
| cfb/6.jpg.enc | 2 | 32742.2556259 | 357922804.454 | 64811.468704 | 66130.62639516569 | Yes |
| zipped/6.zip | 2 | 33466.1907618 | 355483456.332 | 95784.3322368 | 66130.62639516569 | N0 |
| img/6.jpg | 3 | 9680578.04991 | 2.47757579636e+13 | 27571442.9826 | 16786743.146288473 | N0 |
| cfb/6.jpg.enc | 3 | 8387197.37768 | 2.35384812107e+13 | 16763250.8459 | 16786743.146288473 | Yes |
| zipped/6.zip | 3 | 8539766.51662 | 2.3380742589e+13 | 16885500.0919 | 16786743.146288473 | N0 |
| img/7.jpg | 1 | 129.471870479 | 5507.01878479 | 12552.1365309 | 292.17485804729586 | N0 |
| cfb/7.jpg.enc | 1 | 127.554976097 | 5460.81134214 | 215.426970406 | 292.17485804729586 | Yes |
| zipped/7.zip | 1 | 127.156691168 | 5393.37009261 | 339.038589429 | 292.17485804729586 | N0 |
| img/7.jpg | 2 | 33281.5462128 | 360970128.933 | 381390.087157 | 66130.62639516569 | N0 |
| cfb/7.jpg.enc | 2 | 32818.6523072 | 356605692.104 | 65312.7263311 | 66130.62639516569 | Yes |
| zipped/7.zip | 2 | 32569.7990316 | 351964125.448 | 67470.1001049 | 66130.62639516569 | N0 |
| img/7.jpg | 3 | 8510541.27098 | 2.3627793664e+13 | 18213982.9591 | 16786743.146288473 | N0 |
| cfb/7.jpg.enc | 3 | 8384383.61004 | 2.34515745354e+13 | 16767710.3135 | 16786743.146288473 | Yes |
| zipped/7.zip | 3 | 8341587.6877 | 2.29417397437e+13 | 16947828.8842 | 16786743.146288473 | N0 |
| img/8.jpg | 1 | 128.353800285 | 5388.72665961 | 2499.44657132 | 292.17485804729586 | N0 |
| cfb/8.jpg.enc | 1 | 127.724106641 | 5459.73011656 | 235.415503608 | 292.17485804729586 | Yes |
| zipped/8.zip | 1 | 129.557621801 | 5429.6742691 | 474.544736657 | 292.17485804729586 | N0 |
| img/8.jpg | 2 | 32927.9958519 | 355055391.269 | 110647.538816 | 66130.62639516569 | N0 |
| cfb/8.jpg.enc | 2 | 32908.9529598 | 357782459.287 | 65657.4734265 | 66130.62639516569 | Yes |
| zipped/8.zip | 2 | 33446.2334485 | 354568648.647 | 67633.0946909 | 66130.62639516569 | N0 |
| img/8.jpg | 3 | 8454814.52042 | 2.33297864495e+13 | 17450011.4361 | 16786743.146288473 | N0 |
| cfb/8.jpg.enc | 3 | 8402375.02104 | 2.33550516967e+13 | 16772920.3585 | 16786743.146288473 | Yes |
| zipped/8.zip | 3 | 8500567.03971 | 2.32067470842e+13 | 16829277.3584 | 16786743.146288473 | N0 |
| img/9.jpg | 1 | 129.796816804 | 5700.40555505 | 5285.065288 | 292.17485804729586 | N0 |
| cfb/9.jpg.enc | 1 | 127.613555652 | 5460.31111984 | 243.656886098 | 292.17485804729586 | Yes |
| zipped/9.zip | 1 | 128.307541926 | 5445.19208782 | 1598.86684236 | 292.17485804729586 | N0 |
| img/9.jpg | 2 | 33359.4440017 | 375324366.054 | 433866.907751 | 66130.62639516569 | N0 |
| cfb/9.jpg.enc | 2 | 32803.3652663 | 359292313.076 | 65871.8466869 | 66130.62639516569 | Yes |
| zipped/9.zip | 2 | 33089.6379875 | 357122313.696 | 89158.2771442 | 66130.62639516569 | N0 |
| img/9.jpg | 3 | 8544968.94712 | 2.46821318261e+13 | 62507359.6771 | 16786743.146288473 | N0 |
| cfb/9.jpg.enc | 3 | 8367172.23652 | 2.33006537657e+13 | 16779258.6657 | 16786743.146288473 | Yes |
| zipped/9.zip | 3 | 8456281.66922 | 2.3522468455e+13 | 16851436.5925 | 16786743.146288473 | N0 |
| img/10.jpg | 1 | 129.429539115 | 5499.72767572 | 2026.85974233 | 292.17485804729586 | N0 |
| cfb/10.jpg.enc | 1 | 127.181830256 | 5467.9277028 | 250.815513348 | 292.17485804729586 | Yes |
| zipped/10.zip | 1 | 127.508054942 | 5479.52998281 | 4174.02051613 | 292.17485804729586 | N0 |
| img/10.jpg | 2 | 33143.6782128 | 360986671.376 | 126594.858205 | 66130.62639516569 | N0 |
| cfb/10.jpg.enc | 2 | 32781.603092 | 357162888.58 | 65450.2122934 | 66130.62639516569 | Yes |
| zipped/10.zip | 2 | 32791.0037853 | 359710917.224 | 136624.810348 | 66130.62639516569 | N0 |
| img/10.jpg | 3 | 8496178.58355 | 2.33941479291e+13 | 17574849.4633 | 16786743.146288473 | N0 |
| cfb/10.jpg.enc | 3 | 8454532.42768 | 2.35865863495e+13 | 16772687.6177 | 16786743.146288473 | Yes |
| zipped/10.zip | 3 | 8379501.13564 | 2.35447228769e+13 | 16961138.6578 | 16786743.146288473 | N0 |

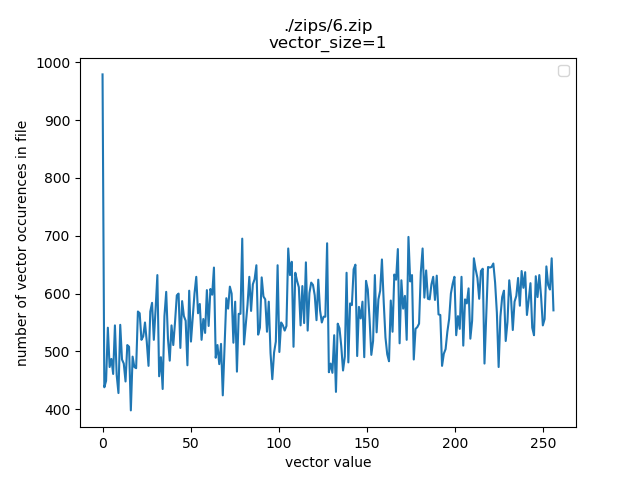
Частоты встречаемости векторов:

Для обычного изображения



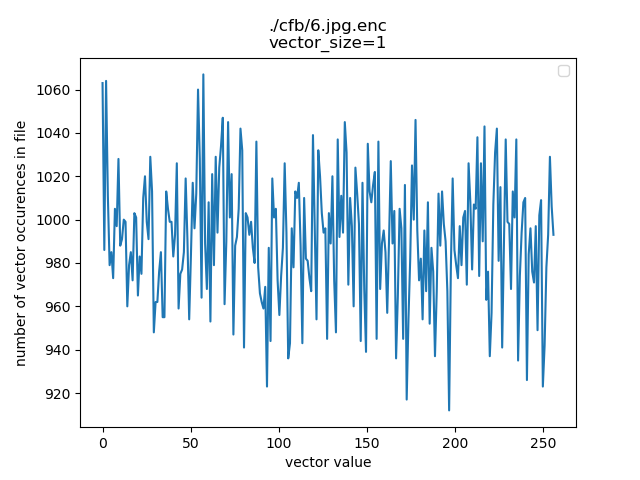
Разброс встречаемости между векторами большой (около 5 тысяч )

Для заархивированного



Разность около 300

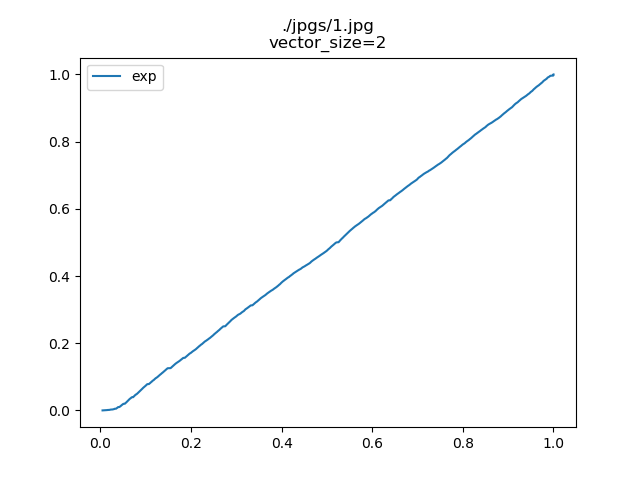
Для зашифрованного AES



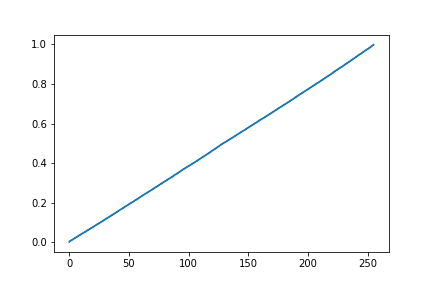
Разность наименьшая- около 150

Графики эмпирической функции распределения для 8-битных векторов:

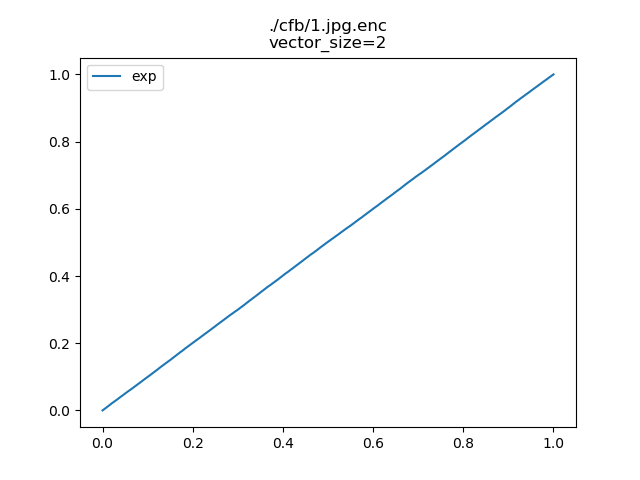
Для обычного изображения



Для заархивированного



Для зашифрованного AES



Графики эмпирических функций распределения для обычного и заархивированного изображений имеют некоторые шероховатости, а график зашифрованного изображения гладкий (что соответствует равномерному распределению).

## Вывод

1. Гипотеза о равномерном распределении принимается только для зашифрованных AES файлов и отвергается для остальных.
2. Для зашифрованных с помощью алгоритма AES изображений функция распределения наиболее близка к функции равномерного распределения.
3. С помощью критерия **χ2** можно определить, является файл открытым или зашифрованным. Для такой проверки необходимо: разбить файл на векторы равной длины, проверить гипотезу о том, что эти векторы распределены равномерно. Если гипотеза принимается, то файл зашифрован, если отвергается, то файл открытый.

### ЗАДАНИЕ № 4

## Условие:

Имеется файл, являющийся фрагментом базы данных. Известно, что распределение двоичных векторов (4-х байтовых для абсолютно непрерывных распределений) имеет экспоненциальное (показательное) распределение. Плотность вероятности .

Параметрическая функция

Необходимо построить и обосновать оценку для указанной параметрической функции.

## Теоретическая часть

Экспоненциальное (показательное) распределение.

Плотность вероятности .

Параметрическая функция

Функция распределения:

Математическое ожидание:

Дисперсия:

Оценка максимальным правдоподобием

Функция правдоподобия:

Логарифмическая функции правдоподобия:

=n\*ln

ln (x,) = = 0, тогда - ММП - оценка.

Проверим несмещенность оценки: E - значит оценка является несмещённой.

Проверим эффективность оценки:

необходимо и достаточно, чтобы

, и

ln (x,) == -n+

Следовательно, , и

Неравенство Рао-Крамера-Фреше:

где ) - информация Фишера, содержащаяся в одном наблюдении.

)=

=

Неравенство верно, следовательно, оценка является несмещённой и эффективной для параметрическая функции .

Проверим состоятельность оценки:

, следовательно, оценка является состоятельной.

## Вычисления

Вычислим значение оценки параметра .

Выдвинем гипотезу о том, что распределение имеет вид

Для проверки гипотезы вычислим число интервалов по формуле Стерджесса:

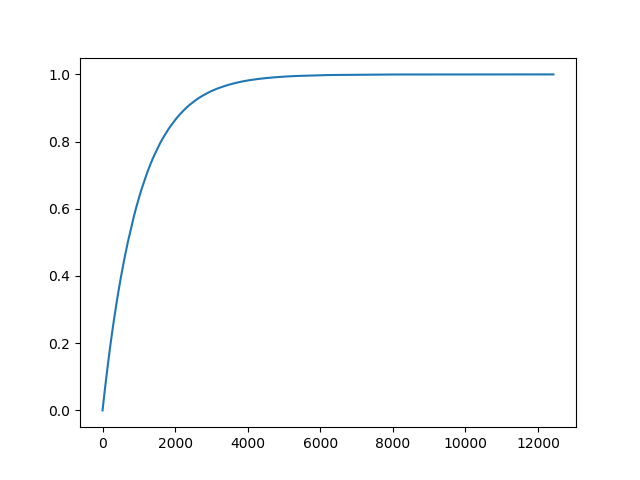
, откуда число степеней свободы n = k-1 = 19.

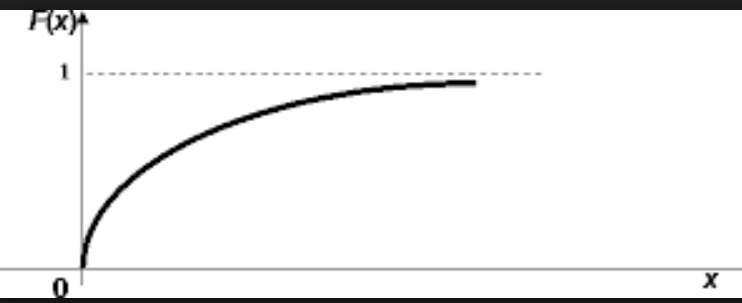
Квантиль распределения хи-квадрат для при 19 степенях свободы равен 30.14352720564616.

Область значений x [0, +). Для разбиения её на интервалы используем следующий алгоритм: отрезок [0, max(xi) + 1] разделим на 19 равных по длине интервалов, а в качестве 20 интервала возьмем [max(xi), +). Используя такое разбиение, вычислим критерий хи-квадрат: = 25.79375870158245.

Вследствие того, что полученное экспериментально значение хи-квадрат: меньше теоретического значения квантиля, гипотезу о распределении для 4-байтовых векторов следует **принять** при .

Рассмотрев график эмпирической функции распределения (Рис1), мы также можем убедиться, что он совпадает с теоретической (Рис2):

 Рис1

Рис 2