**Получение выборки**

**Задача**

1. Исследовать методы воспроизведения числовых последовательностей с известными точечными и интервальными характеристиками
2. Сформировать алгоритмическое обеспечение и разработать программную систему воспроизведения числовых последовательностей с характеристиками

ВВЕДЕНИЕ

На данный момент имеется большое количество литературы, посвященной генерированию случайных величин и моделированию последовательности случайных процессов [1-5]. Особенно часто моделирование случайных последовательностей используется в радиотехнике [3].

Способы получения случайных чисел (генераторы) разделяются на аппаратурные и программные, позволяющие получать случайные и псевдослучайные числа соответственно. Аппаратурные способы основаны на использовании различных естественных источников первичных сигналов, по своей природе являющихся случайными, например, радиоактивный распад, шумы электронных и полупроводниковых приборов и т.д. Программные способы предполагают использование компьютеров для формирования (генерирования) так называемых псевдослучайных чисел по детерминированным алгоритмам. Последовательности чисел называют псевдослучайными, так как они генерируются по определенному алгоритму и в процессе формирования повторяется определенный цикл. Программные способы в наибольшей степени удовлетворяют требованиям, предъявляемым к качеству моделируемой последовательности случайных чисел. Важнейшими требованиями являются следующие: высокая скорость получения псевдослучайных чисел; для заданного алгоритма проверенное качество получаемых псевдослучайных чисел (независимость, закон распределения, числовые характеристики, интервал периодичности) остается неизменным, что позволяет применять этот алгоритм для решения различных задач. Большинство программных генераторов работают на основе линейно-конгруэнтного метода.

Линейные конгруэнтные генераторы созданы Лемером в 1949 г. В них последовательность целых чисел  определяется по рекурсивной формуле:



где – модуль,  *–* множитель,  – приращение,  – начальное значение. Параметры , , ,  – неотрицательные целые числа. По указанной выше формуле для получения  нужно разделить на , т.е.  будет остатком этого деления. Тогда за искомое случайное число в интервале принимается.

В зависимости от условия,  или различают смешанные и мультипликативные генераторы. Полное описание методов генерирования можно найти в работе [1].

Во многих публикациях приводятся параметры данных, для которых проводили расчёты. Проверка проведённых расчётов зачастую требует восстановления исходных данных, так каких получение оказывается затруднительным или вовсе невозможным.

Цель данной работы: восстановить случайную величину (СВ) с заданными параметрами. Так как в большинстве работ указаны наборы параметров, позволяющие получить математическое ожидание и среднеквадратическое отклонение , что указывает на нормальное распределение значений СВ, а также указан диапазон значений случайной величины , требуется алгоритм генерации СВ, распределённой по нормальному закону с известными параметрами.

Существуют различные способы получения квазислучайной величины с заданным законом распределения.

Метод обратного преобразования позволяет получить искомую СВ путём модификации равномерно распределённой СВ. Закон преобразования представляет собой обратную функцию распределения вероятностей. В данном случае этот метод неудобен, так как требует вычисления обратной функции нормального распределения



Метод полярных координат не требует сложных численных расчётов. Данный метод позволяет получить искомую СВ после нескольких этапов преобразования координат некоторой случайной точки, находящейся в пределах единичной окружности.

На практике также применяется алгоритм «Зиккурат». Для получения СВ фигуру, ограниченную графиком  разделяют на сегменты равной площади. Для каждого сегменту рассчитывается частота попадания отсчётов искомой СВ на интервал, соответствующий его основанию.

Метод полярных координат наиболее прост для реализации. Однако алгоритм «Зиккурат» гораздо быстрее с вычислительной точки зрения, он часто применяется в случаях, где требуется большое количество случайных чисел. Его использование позволит выводить выборки разной длины с достаточно высокой скоростью, причём полученный алгоритм будет готов к работе с разными законами распределения, что и требуется для данного исследования.

**Генератор случайных чисел**

Пусть некоторая случайная последовательность принимает значения в указанном диапазоне , она имеет нормальное распределение с заданными параметрами  длина искомой выборки .

Плотность распределения СВ задаётся формулой



Отсчёты  в диапазоне  содержатся в векторе cut.

После получения предполагаемого закона распределения рассчитываются границы интервалов и частоты. Фигура, ограниченная графиком  и предельными значениями СВ делится на сегменты равной площади. Количество сегментов можно менять, в данном алгоритме их вдвое больше, чем столбцов в гистограмме, построенной для выборки длины , т.е. . Границы интервалов выводятся во втором столбце матрицы getI. Частоты freqs определяются как длины средних линий прямоугольных трапеций, длины оснований которых содержатся в первом столбце той же матрицы. Затем их значения переводятся в доли от , что позволяет определить количество значений искомой СВ на каждом интервале для любых значений . Таким образом, получаем правило формирования искомой СВ.

После расчёта частот для каждого интервала генератор формирует выборку, преобразуя равномерно распределённую СВ, в соответствии с полученным правилом по следующему алгоритму:

* Генерация случайной точки, лежащей на интервале ;
* Если точка лежит в пределах *i*-го сегмента, проверяется количество ранее сгенерированных точек, лежащих на интервале .
* Случайная точка записывается в массив, если количество c1*i* точек на интервале  не превышает количество значений , которое может содержаться на данном интервале freqs *i* .

**Проверка**

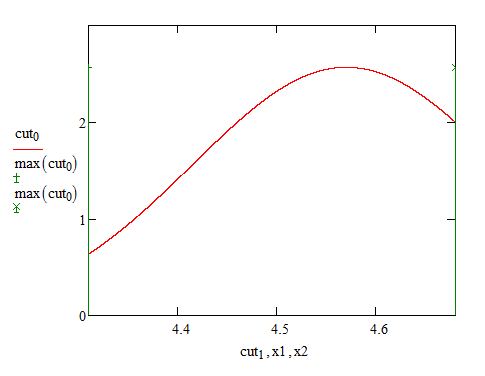
**Отладка**

Алгоритм отлажен на значениях параметров реальных выборок, при этом использованы СВ, заданные диапазоном , математическим ожиданием и коэффициентом вариации (), диапазоном  математическим ожиданием и его доверительным интервалом ().

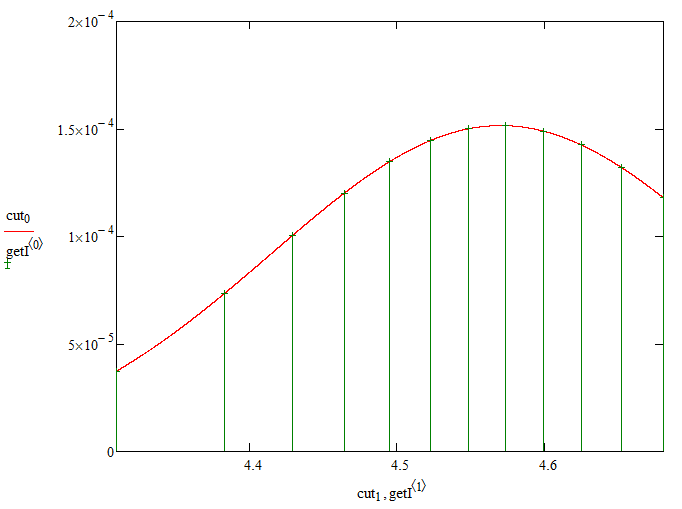
Для набора параметров



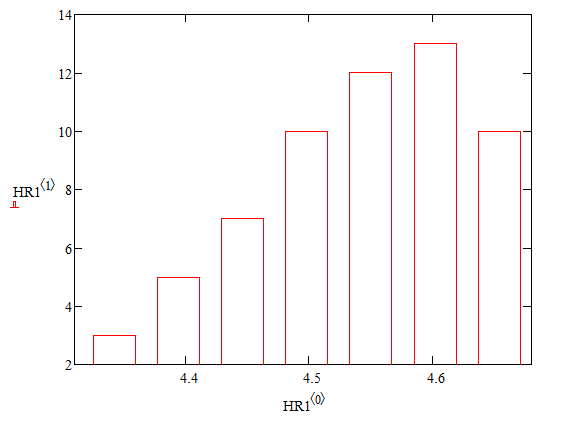
имеем плотность распределения cut:



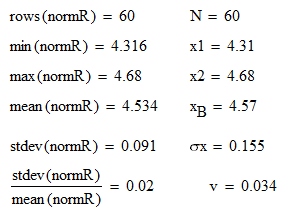
После разделения на сегменты получаем матрицу getI:



Матрица getI позволяет рассчитать частоты freqs и сформировать искомую выборку. Распределение полученной СВ представлено на графике:



Полученная случайная величина проверяется на соответствие заданным значениям параметров:



Чтобы проверить качество работы алгоритма были получены несколько выборок для одинаковых наборов параметров. Каждый набор полученных выборок проверен на однородность.

По результатам проведённых проверок можно сделать вывод о качестве полученного генератора. На выходе алгоритма получаем различные выборки, которые могут быть получены из одной генеральной совокупности с вероятностью 99,9%. Таким образом данный генератор может быть использован для дальнейших исследований.

**Работа с реальными данными**

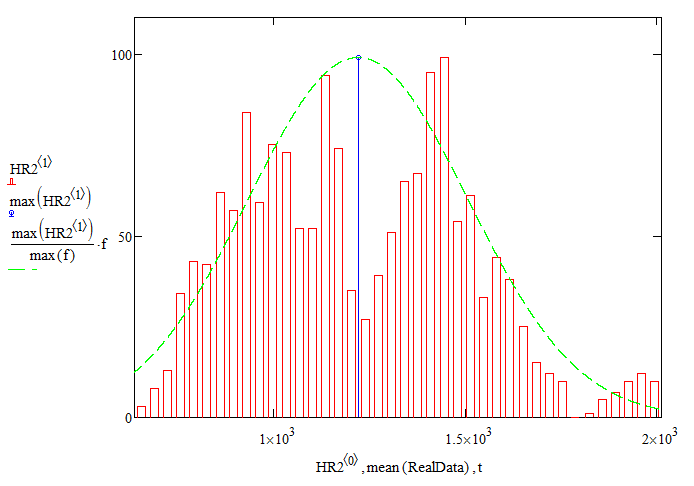
Сгенерированы СВ, распределённые по нормальному закону, параметры которого рассчитаны по реальным данным. Для расчётов использованы значения биржевых индексов RTS и KOSPI, ширина третьего тергита медоносных пчёл.

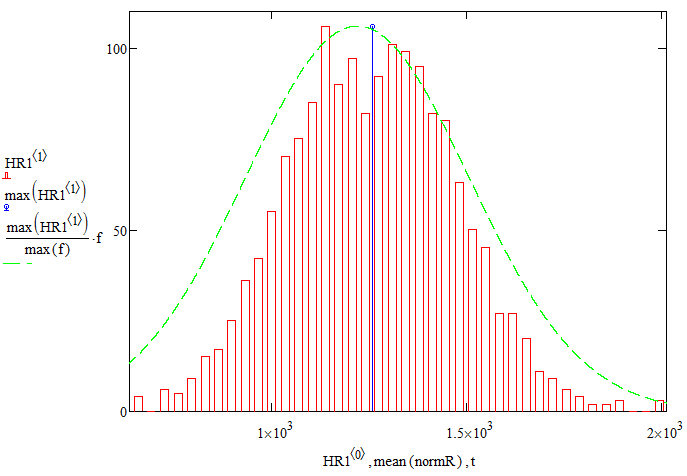
*Индекс RTS*

В алгоритм введены данные о значении индекса на момент закрытия бирж. По имеющимся данным получены следующие параметры для генератора:



На графике показаны распределение исходных данных и вид плотности нормального распределения с полученными ранее параметрами, синим цветом выделено выборочное математическое ожидание.



Такой же график построен для сгенерированной выборки: 

Проверка полученных выборок на однородность показала, что они не могут быть взяты из одной генеральной совокупности с вероятностью 99,9%.

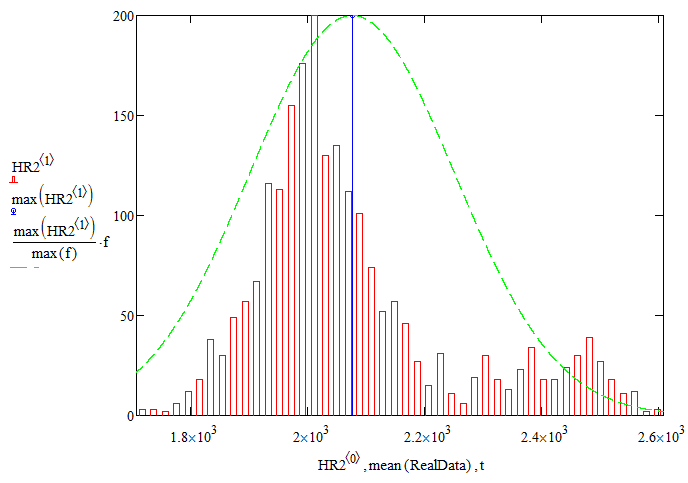
*Индекс KOSPI*

Аналогичные расчёты проведены для индекса KOSPI.

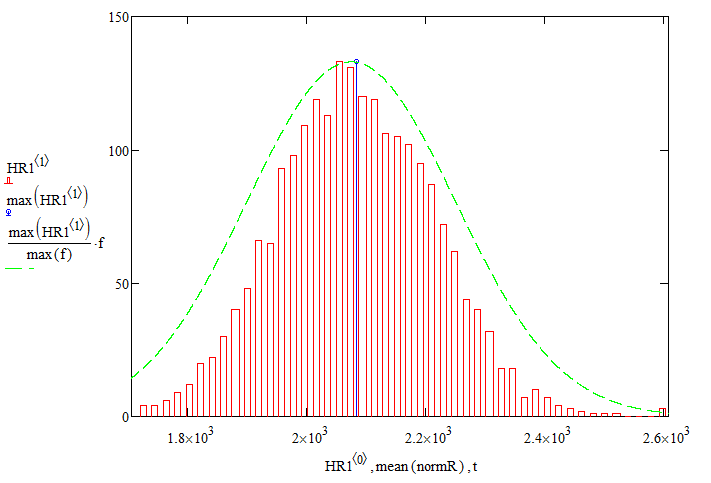
Параметры распределения:



Распределение исходных данных



Распределение СВ на выходе генератора



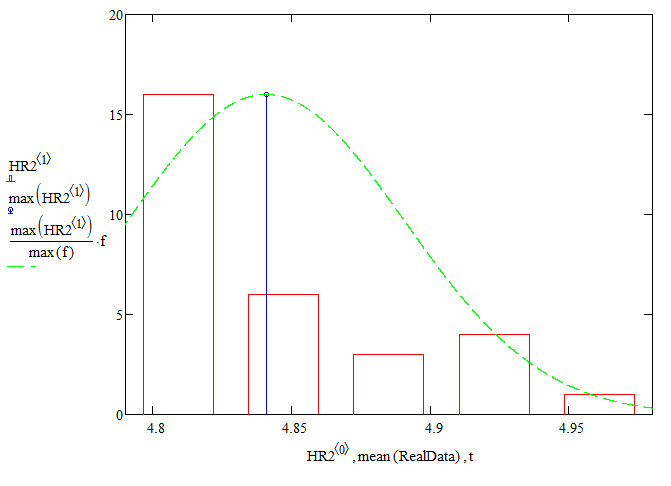
Проверка полученных выборок на однородность показала, что они не могут быть взяты из одной генеральной совокупности с вероятностью 99,9%.

*Ширина третьего тергита, пасека 1*

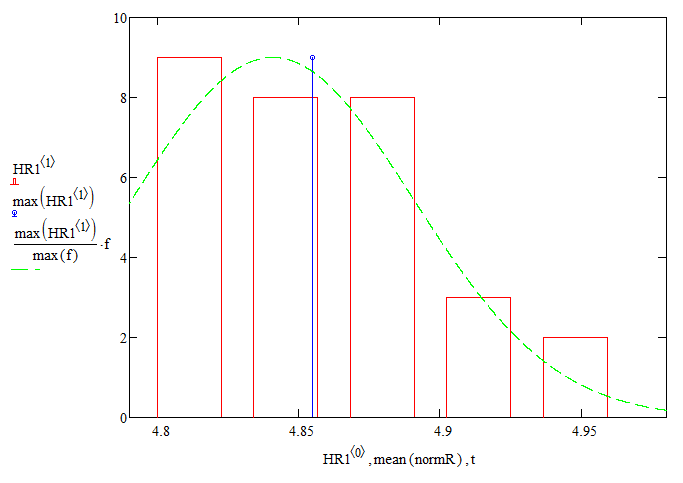
Параметры распределения:



Распределение исходных данных



Распределение СВ на выходе генератора



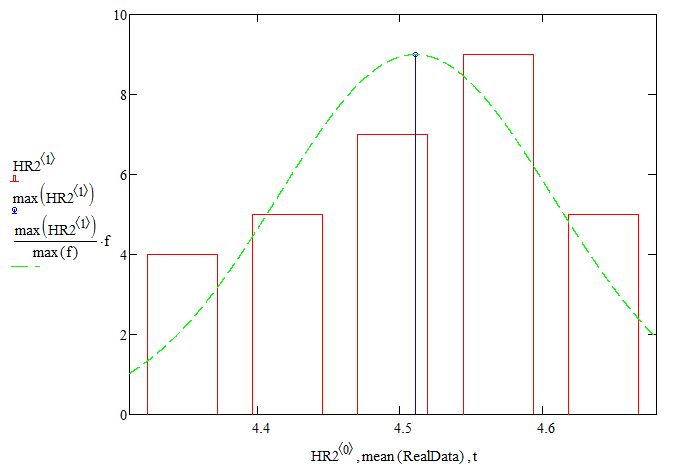
Проверка полученных выборок на однородность показала, что они не могут быть взяты из одной генеральной совокупности с вероятностью 99,9%.

*Ширина третьего тергита, пасека 3*

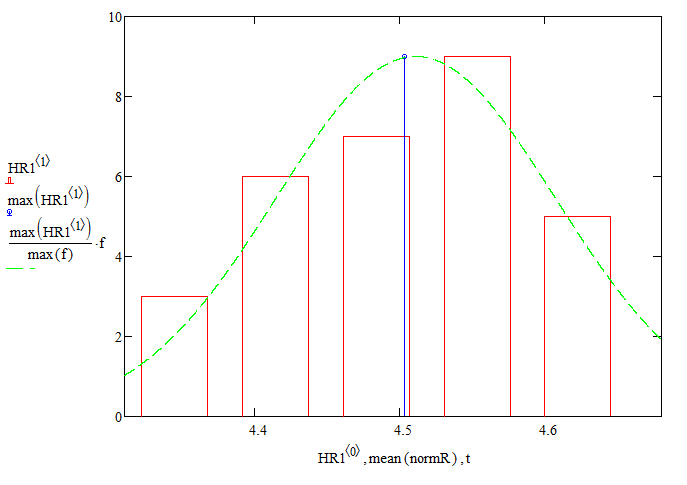
Параметры распределения:



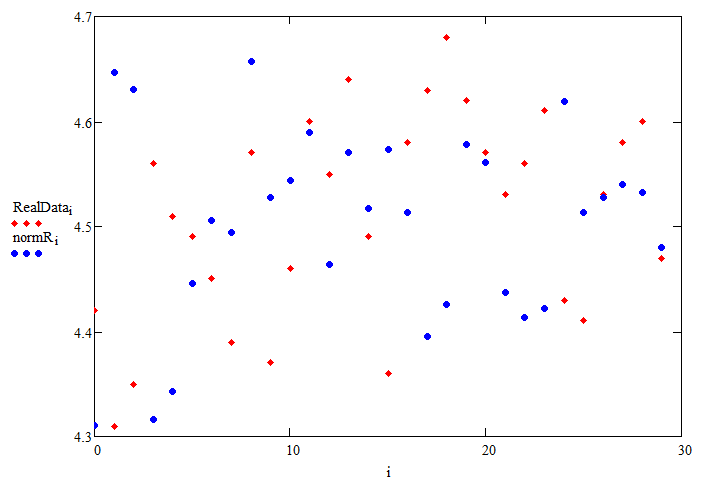
Распределение исходных данных



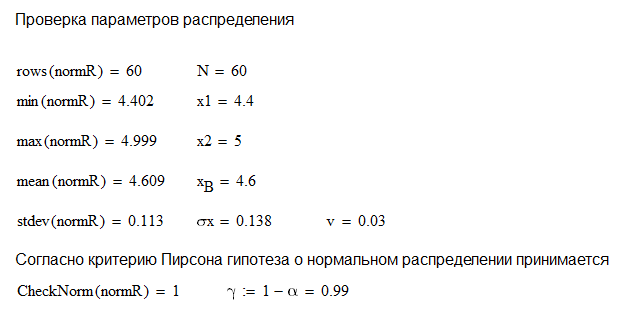
Распределение СВ на выходе генератора

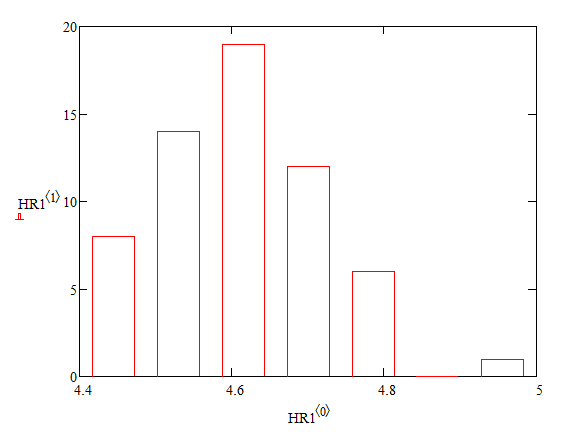


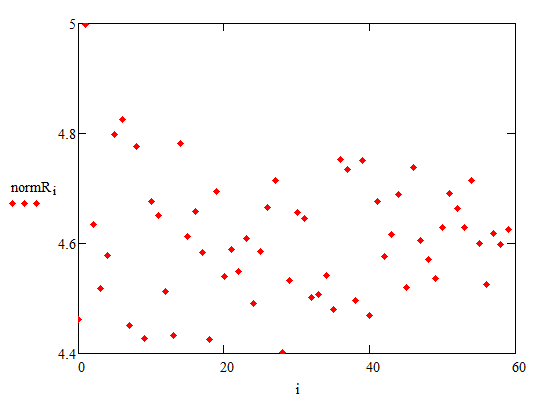
Исходные данные и СВ на выходе генератора



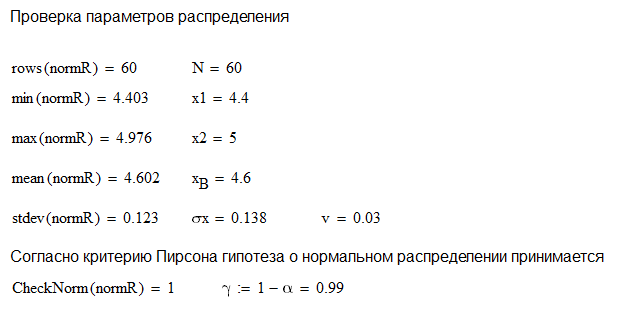
Проверка полученных выборок на однородность показала, что они взяты из одной генеральной совокупности с вероятностью 99,9%.

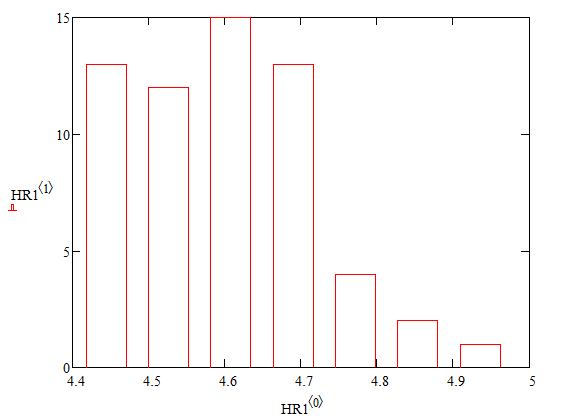


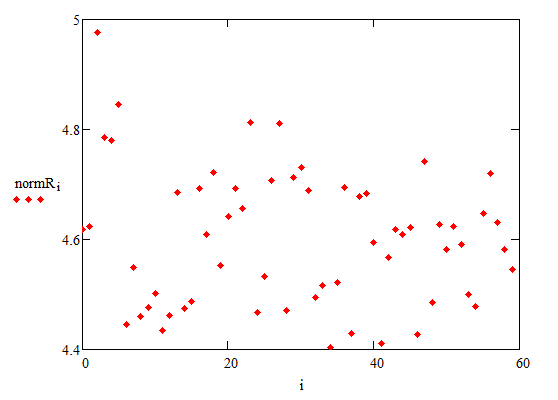




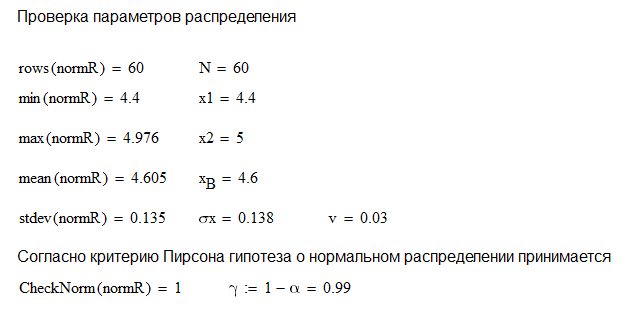
Значения  можно умножить на число, большее количества точек в выборке. Для  имеем:

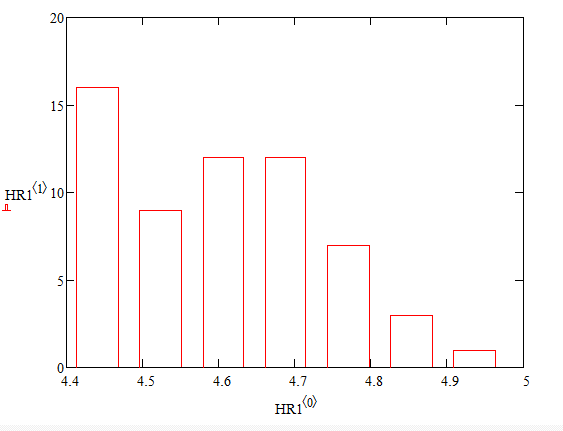


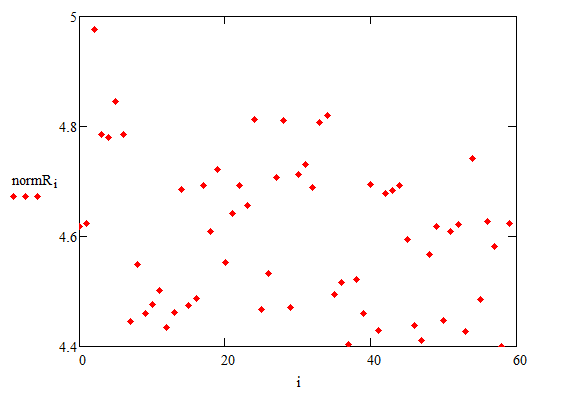




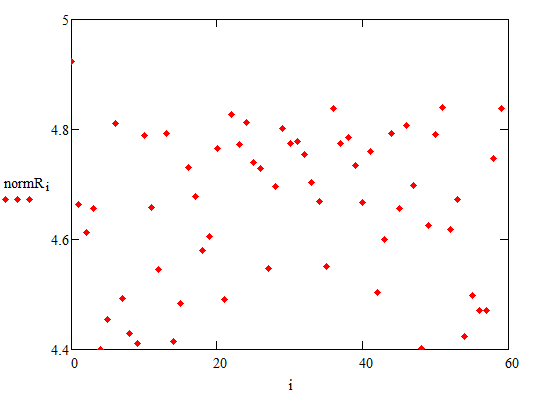
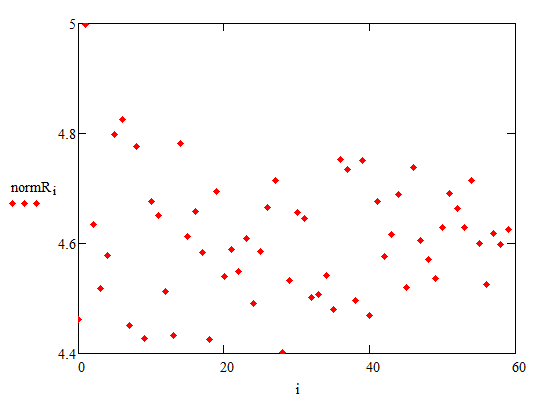
Для :







Увеличение  способствует росту дисперсии генерируемой случайной величины и изменению вида её гистограммы, при этом последующие отсчёты получаемой случайной величины не стремятся к её математическому ожиданию, как это происходит при малых .



Литература.

1. Арзамасцев, А.А. Универсальный генератор случайных чисел для имитационного моделирования / А.А. Арзамасцев, Т.Ю. Китаевская, И.В. Азаров // Вестник Тамбовского гос. ун-та. Сер.: Естесств. и техн. науки. – 2000. – Т.5. – №1. – С. 131-133.
2. Бакалов, В.П. Цифровое моделирование случайных процессов / В.П. Бакалов. – М.: МАИ, 2001. – 84 с.
3. Мокрушин, Л.А. Генерация псевдослучайных числовых последовательностей высокого качества на основе линейного конгруэнтного метода / Л.А. Мокрушин // Известия ЛЭТИ: Сб. научн. трудов: Вопросы проектирования измерительных систем. Ленингр. электротехн. ин-т им. В.И. Ульянова (Ленина). – 1973. – Вып.446 – С. 71-82.
4. Петров, Ю.В. Методы математического моделирования радиотехнических систем / Ю.В. Петров. – СПб.: Балт. гос. техн. ун-т, 2005. – 111 с.
5. Прикладной анализ случайных процессов / С.А. Прохоров, А.В. Графкин, В.В. Графкин и [др.]; Самарский науч. центр Рос. акад. наук. – Самара: СНЦ РАН, 2007. – 582с.: ил.
6. Прохоров, С.А. Моделирование и анализ случайных процессов: лабораторный практикум / С.А. Прохоров. – Самара: Самар. гос. аэрокосм. ун-т, СНЦ РАН, 2001. – 191 с.: ил.