Исследование встроенного в графическую среду разработки LabView генератора случайных чисел

О. А. Микус[1], Е. С. Сулоева[2]
Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)
[1]sadalphamiks@gmail.com, [2]suloewa@list.ru

***Аннотация.* Исследуется встроенный в графическую среду разработки LabView генератор случайных чисел на соответствие заявленным в документации характеристикам. Определяется необходимое количество повторов для серии экспериментов, а также проводится машинный эксперимент для различных объемов выборок. При различных сочетаниях параметров строятся гистограммы распределения случайных величин с целью проверки соответствия их заданному закону распределения. Произведено исследование генератора с установлением различной разрядности выходных значений .**

Ключевые слова: генератор случайных чисел; LabView; случайная величина; плотность распределения вероятности; метрологический анализ; гистограмма

I. Введение

Воспроизведение входных величин является начальным этапом имитационного моделирования при метрологическом синтезе измерительного канала. Однако этот этап является самым важным, поскольку от достоверности воспроизведённых величин будут зависеть не только результаты экспериментов, но и правильность установления точностных характеристик результатов [1]. Рост требований к увеличению достоверности результатов измерений и переход к цифровым моделям в метрологической практике ставит перед собой задачу обеспечения необходимого уровня точности воспроизведения значений величин при имитационном моделировании.

II. Машинный эксперимент

Объектом данного исследования является встроенный в графическую среду разработки LabView генератор квазислучайных случайных чисел Random number 0-1, поскольку для формирования аддитивных помех входных воздействий будет использоваться именно он.

В общем виде входное воздействие имеет вид:

$$\gamma_j(t) = \gamma'_j(t) + n_j(t),$$

где $n_j(t)$-помеха, носящая аддитивных характер и имеющая соответствующий закон распределения вероятности в качестве априори установленных данных

В технической документации заявлено, что Random number генерирует случайную величину с плавающей точкой формата double (т. е. двойной точности) в диапазоне [0;1], число больше или равно нулю, но строго меньше единицы, а также что распределение величины равномерное.

Для проверки данных параметров был разработан виртуальный инструмент в среде разработки LabView, который представлял собой генератор случайных чисел с возможностью выбора диапазона измерений, чтобы расширить функционал и проверить работу встроенного генератора на различных диапазонах, объемах выборки, а также серии экспериментов с заданным числом выборок. Также в разработанной подпрограмме есть возможность регуляции разрядности выходной величины, максимум которой требовалось определить экспериментально, поскольку этого не было указано в документации.

Помимо вышесказанного разработанный виртуальный инструмент проверял выборки на повторяемость, чтобы определить цикличность псевдослучайных величин, выдаваемых выстроенным генератором.

III. Результаты экспериментов

Эмпирическим путём в рамках машинного эксперимента было доказано, что разрядность встроенного генератора случайных чисел не превышает 18-ти разрядов. Однако, периодичность случайных величин выявлена не была, однако, т.к. даже при уменьшенной в 2 раза разрядности относительно максимально возможной и при выборке n=10000 на для 100 экспериментов не было выявлено повторяющихся результатов. Повторы начинаются с случае, когда разрядность принудительно уменьшается до 5 – 6 разрядов. Среднее количество повторов на эксперимент R рассчитывается по формуле:

$$\boldsymbol{R} = \frac{\sum \boldsymbol{r}}{\boldsymbol{m}},$$

где r – количество повторов в выборке, m– количество экспериментов. При разрядности 6 имеем $\boldsymbol{R} \approx \mathbf{0.1}$, при разрядности 5 $\boldsymbol{R} \approx \mathbf{3}$ при выборке в 10000 элементов.

Далее эксперименты проводились для наиболее часто используемых комбинаций диапазонов входных значений с одинаковым размером выборки *n*=10000 и разрядностью равной 9 разрядов. Результаты для интервала существования аддитивной помехи $n_j(t)$ [0;1] представлены на рис. 1, для увеличенного 10 кратно диапазона от 0 до 10 на рис. 2. Данные, полученные в ходе машинного эксперимента, говорят о том, что плотность распределения является равномерной и совпадает с заявленной при объеме выборки, равной n=10000.

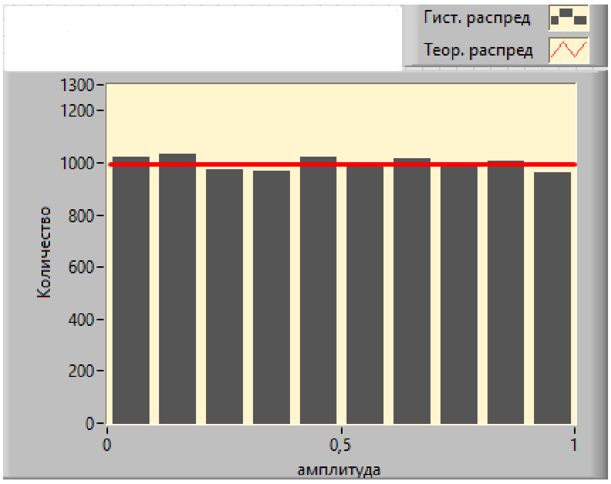

Рис. 1. Гистограмма распределения для диапазона 0-1

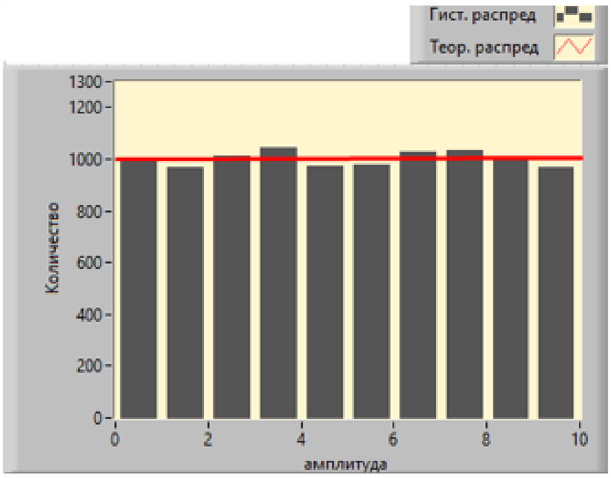

Рис. 2. Гистограмма распределения для диапазона 0-10

На основе рассматриваемого генератора формируется результат распределения аддитивной помехи $n_j(t)$ для увеличенного объема выборки. На рис. 3 представлена гистограмма распределения, которая указывает, что чем больше выборка, например, n=100000, (т. е. в 10 раз больше рассмотренных ранее примеров), тем ближе экспериментально полученный результат к теоретическому равномерному закону распределения.

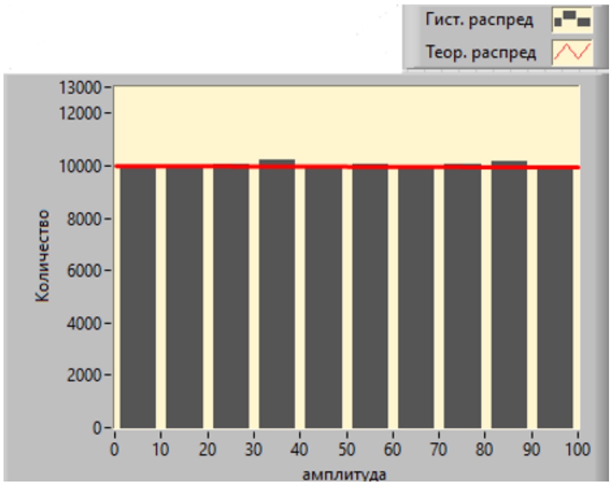

Рис. 3. Размер выборки 100000

Далее были проведены эксперименты с различным количеством интервалов гистограмм: k=31; k=46; k=100; k=20, рассчитанным по формулам, приведённым ниже.

Стоит отметить, что практически все рекомендации по выбору числа интервалов исходят из того, чтобы при данном объеме выборки n как можно лучше приблизить плотность распределения ее непараметрической оценкой (гистограммой). В данной работе выбор числа интервалов k рассматривается с позиций построения наиболее мощного критерия согласия при близких конкурирующих гипотезах. Естественно, что определение количества интервалов связывается с объемом выборки. В [2] на основании различных источников приводится целый ряд рекомендаций по выбору числа интервалов k. Так, при выборе интервалов равной длины определяющим является требование, чтобы количество наблюдений, попавших в интервалы, было не слишком малым и сравнимым. При этом наиболее часто рекомендуется, чтобы количество наблюдений, попавших в интервал, было не менее 10.

Во многих источниках можно найти упоминание эвристической формулы Старджесса для определения «оптимального» числа интервалов [3].

$$k = \log_2 n + 1 = 3.3 \lg n + 1$$

Для выборки с n=10000 результаты представлены на рис. 4.

В [4] для определения «оптимального» числа интервалов рекомендуется формула Брукса и Каррузера

$$k = 5 \lg n.$$

Для выборки с n=10000 результаты представлены на рис. 5.

В [5] рекомендуют соотношение.

$$k = \sqrt{n}$$

Для выборки с n=10000 результаты представлены на рис. 6.

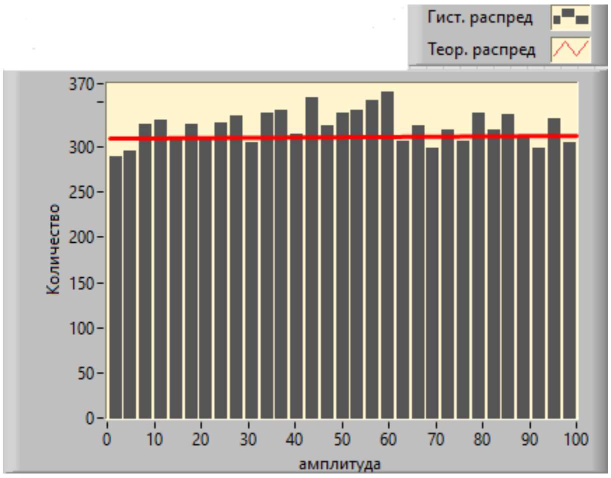

Рис. 4. Гистограмма с количеством интервалов по формуле Старджесса

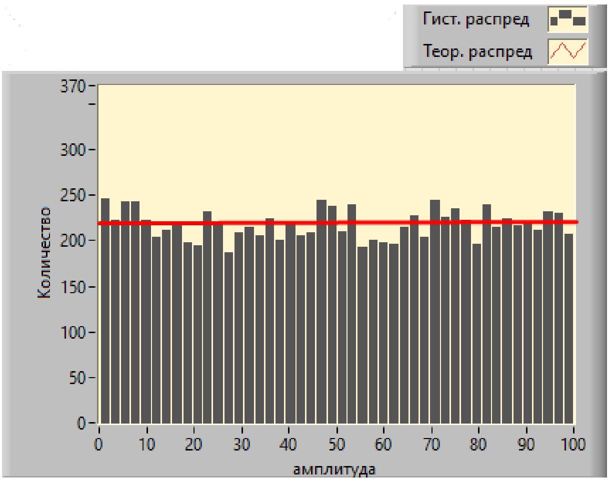

Рис. 5. Гистограмма с количеством интервалов по формуле Брукса и Каррузера

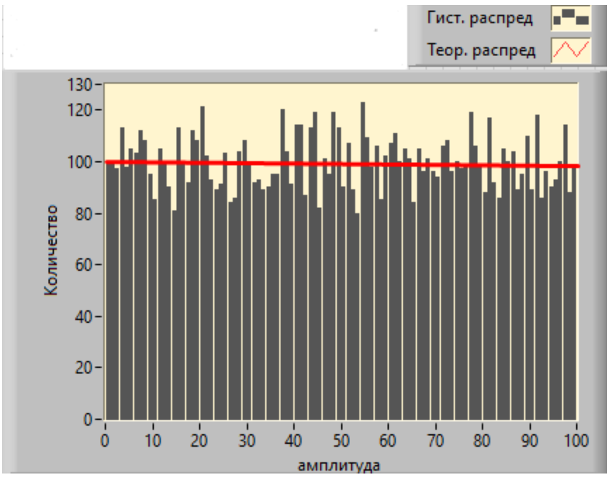

Рис. 6. Гистограмма с количеством интервалов по формуле Хейнхолда и Гайеде

В [6] для равновероятных интервалов их количество устанавливается порядка

$$k \approx 4\sqrt[5]{2}\left(\frac{n}{t}\right)^{0.4},$$

где t – квантиль стандартного нормального распределения для заданного уровня значимости. В ряде работ приводятся модификации данной формулы.

При больших объемах выборок разброс значений k, задаваемых различными формулами, достаточно велик. Поэтому на практике совместно с рекомендациями ВНИИ Метрологии [7] было приятно решение при выборе числа интервалов больше руководствоваться разумными соображениями, выбирая число интервалов так, чтобы в интервалы попадало число наблюдений не менее 5–10. Так, в зависимости от n предлагаются следующие величины k.

ТАБЛИЦА I. РЕКОМЕНДУЕМОЕ КОЛИЧЕСТВО ИНТЕРВАЛОВ В ЗАВИСИМОСТИ ОТ ОБЪЁМА ВЫБОРКИ

Объем выборки n	**Количество интервалов гистограммы *k***
40-100	7-9
100-500	8-12
500-1000	10-16
1000-10000	12-22

Следуя указаниям из табл. 1, для объема выборки n=10000 была построена гистограмма с 20-ю интервалами, представленная на рис. 7.

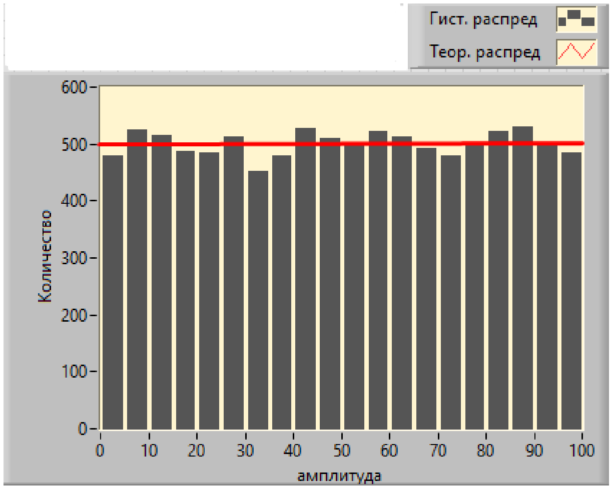

Рис. 7. Гистограмма с k=20 интервалов

Из результатов исследования различных подходов к выбору количества интервалов гистограммы можно сделать вывод, что наиболее точный подход обеспечивает гистограмму, имеющую наименьшую дисперсию относительно теоретических значений по сравнению с рассмотренными выше методами. Это обусловлено особенностями зависимости дисперсии от размера выборки и количества интервалов гистограммы, выявленной в [8]: в рамках нескольких выборок дисперсия при увеличении количества интервалов увеличивается; при увеличении количества значений в рамках одной выборки – наоборот уменьшается.

Далее аналогичный машинный эксперимент был проделан для распределения Симпсона. Если исходить из теории, то получается, что, если ξ_1 и ξ_2 – это случайные, равномерно распределённые величины, то случайная величина $\xi = \xi_1 + \xi_2$ должна иметь треугольное распределение (распределение Симпсона). Если в случае со сложением результатов генерируемых значений стандартным генератором это подтвердиться, то это будет косвенно подтверждать тот факт, что случайные величины были равномерно распределены. Результаты данного эксперимента представлены на рис. 8 и 9 с k равным 10 и 20 интервалов соответственно.

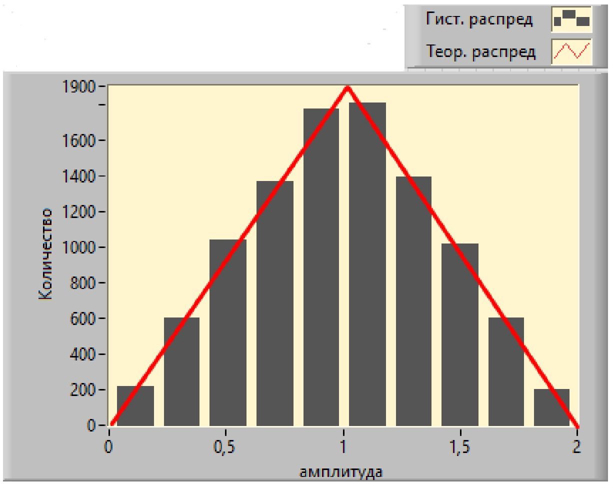

Рис. 8. Распределение Симпсона 10 интервалов

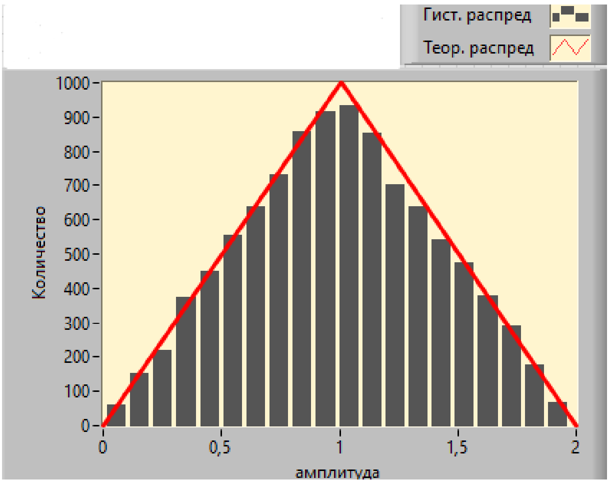

Рис. 9. Распределение Симпсона 20 интервалов

Можно видеть, что при сравнени результатов машинного эксперимета и теоретических с увеличением количества интервалов в случае распределения Симпсона, эффект описанный в [8] не наблюдался.

IV. Заключение

В рамках данной работы были представлены результаты исследования стандартного генератора случайных чисел, встроенного в графическую среду разработки LabView. После проведения машинного эксперимента можно сказать, что максимальная разрядность составляет 18 знаков; при выборке 100000 не выявлено повторов за 100 экспериментов, повторы на размерах выборки до 100000 наблюдаются только при искусственном сокращении разрядности встроенного генератора случайных чисел до 5 значащих цифр. При построении гистограмм результатов машинного эксперимента был обозначен подход к выбору количества интервалов гистограммы для конкретного объема выборки, имеющий наименьшую дисперсию относительно теоретических значений. Дополнительно была исследована операция сложения двух сгенерированных равномерно распределенных величин для получения треугольного закона распределения, результаты моделирования совпадают с законом распределения Симпсона, что косвенно подтверждает соответствие заявленным характеристикам по плотности распределения рассматриваемого генератора и подтверждает возможность его использования для решения метрологических задач.

Список литературы

[1] Tsvetkov E.I., Suloeva E.S. Analysis of the parameters that determine the reliability of the results of a verification of measuring instruments // Measurement Techniques. 2018. Т. 61. № 9. С. 872-877.

[2] Новицкий П.В., Зограф И.А. Оценка погрешностей результатов измерений. Л.: Энергоатомиздат, 1991. 303 с.

[3] Sturgess H.A. The choice of classic intervals // J. Am. Statist. Assoc. march 1926. 47 p.

[4] Шторм Р. Теория вероятностей. Математическая статистика. Статистический контроль качества. М.: Мир, 1970. 368 с.

[5] Heinhold I., Gaede K.W. Ingeniur statistic. München; Wien, Springler Verlag, 1964. 352 p.

[6] Mann H.B., Wald A. On the choice of the number of class intervals in the application of the chi square test // Ann. Math. Stat., 1942. V. 13. P. 306-317.

[7] Бурдун Г.Д., Марков Б.Н. Основы метрологии. М.: Изд-во стандартов, 1985. 120 с.

[8] Mikus O.A., Tsvetkov E.I. Identification of the Probability Density Distribution of a Random Value // Proceedings of 2021 24th International Conference on Soft Computing and Measurements, SCM 2021, 2021, pp. 12–14, 9507140