УДК 681.3: 004.021: 519.683.8

ГЕНЕРАЦИЯ ПСЕВДОСЛУЧАЙНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ С ЗАДАННЫМ ЗАКОНОМ РАСПРЕДЕЛЕНИЯ НА ОСНОВЕ ГЕНЕРАТОРА БЕЛОГО ШУМА

Бабицкий К.П.

Белорусский государственный университет информатики и радиоэлектроники, г. Минск, Республика Беларусь

Научный руководитель: Ролич О.Ч. – канд. техн. наук, доцент, доцент кафедры ПИКС

Аннотация. Экспериментально изучен алгоритм Неймана для генерации псевдослучайных величин с заданным законом распределения. Исследован его принцип работы и характер генерируемых чисел: корреляция с законом распределения, быстродействие, количество генерируемых чисел при конечном количестве входных данных.

Ключевые слова: белый шум, закон распределения, псевдослучайное число, генерация псевдослучайных чисел.

Введение. Важным вопросом науки и техники является генерация случайных чисел. Сфера применения случайных величин велика: моделирование результатов экспериментов, исследование устойчивости сигналов и алгоритмов анализа данных к шумам, применение в криптографии. Разные задачи могут предъявлять разные требования к случайно сгенерированным числам: непредсказуемость или возможность «угадывать» числа в последовательности псевдослучайных величин, определённый закон распределения псевдослучайных значений. Алгоритмы генерации псевдослучайных чисел с заданным законом распределения будут рассмотрены и сравнены в статье [1].

Основная часть. Для дальнейших рассуждений обозначим, что:

– для разработки программ будем использовать язык программирования C/C++ в *IDE Microsoft Visual studio Community* 2022;

– в качестве генератора белого шума будем представлять сигналы АЦП, подаваемые на контакты микроконтроллера (процессора);

– алгоритмы программ будут разработаны таким образом, чтобы быть универсальными для всех распределений, и «подгоняться» для конкретных законов распределения не будут. При этом в качестве законов распределения будем принимать такие, которые задают вероятности генерации непрерывно распределённых величин;

– в качестве законов распределения для оценки алгоритмов будем использовать нормальное, экспоненциальное распределения и распределение Максвелла;

– сравнивать алгоритмы будем по затраченному времени на чтение входных данных и генерацию чисел, корреляции с заданным распределением, отношения удачных генераций случайных величин (далее – СВ) к количеству входных значений при равномерном периодическом, равномерном непериодическом и неравномерном непериодическом распределении входных значений.

Универсальных методов для генерирования СВ не много, основными являются методы Неймана и обратной функции, они позволяют генерировать СВ с непрерывным распределением по заданным плотностям распределения и функциями вероятности соответственно. Метод взятия обратной функции работает с функцией распределения вероятностей и связан с аналитическим вычислением обратной функции, что является очень сложной задачей, а для некоторых интегралов невозможной, поэтому будем оценивать возможности алгоритма Неймана.

Для оценки попробуем сгенерировать дискретные значения по закону для непрерывного распределения. Подобный эксперимент, не смотря на кажущуюся противоречивость, имеет место быть: непрерывность величин в программировании является относительной. Если записывать число на бумаге, то количество знаков после запятой можно считать бесконечным: при необходимости можно дописать ещё несколько знаков. В программировании же

любое число (переменная) строго ограничена количеством памяти, выделенной для него и просто «дописать» несколько знаков после запятой, если вся память уже занята, не представляется возможным.

Алгоритм основан на том, что часть попадающих на вход чисел отбрасывается, а часть становится результатом генерации с заданным распределением. Алгоритм Неймана для генерации СВ работает с плотностью вероятности: на вход подаётся пара равномерно распределённых чисел: одно на отрезке генерируемых чисел, второе – значений плотности вероятности, тем самым являясь координатами на графике плотности вероятности. Если точка с такими координатами находится под графиком плотности, то генерация СВ прошла успешно, иначе эта пара чисел отбрасывается.

Для генерации дискретны СВ в качестве входных данных примем: 131072 пар чисел, где первое (координата) число, равномерно распределённое дискретное, кратное 1, значение от 0 до 127 с периодом 128, второе число (шанс) – равномерно распределённое число на [0;100) с точностью до сотых и периодом 10000.

Программа будет работать по следующему алгоритму:

1. Считывание пар значений;
2. Скалирование значений второго числа в паре в соответствии с максимальным значением плотности вероятности;
3. Генерация св при помощи сравнения второго числа со значением плотности вероятности в точке;
4. Запись результатов в файл.

Распределения будут задаваться формулами 1, 2 и 3:

$$p_M(x) = 0.2 * (x - 10)^2 * \exp\left(-(x - 10)^2 * 0.2\right) \quad (1)$$

$$p_N(x) = \frac{1}{2.506 * 8} * \exp\left(\frac{-(x - 64)^2}{2 * 64}\right) \quad (2)$$

$$p_{\exp}(x) = 0.3 * \exp\left(-0.3 * x\right) \quad (3)$$

Формула (1) задаёт плотность вероятности для распределения Максвелла, (2) – нормального распределения, (3) – экспоненциального [2].

Степень корреляции сгенерированных СВ и плотности распределения для дискретных равномерно распределённых входных значений составила более 0,99 для всех видов распределения при генерировании методом Неймана. При этом иногда алгоритм «пропускал» значения, в соответствующих координатах которым плотность вероятности составляла ~0,00. Это связано с алгоритмической особенностью программного исполнения метода Неймана: точки, у которых шанс равен плотности вероятности в точке соответствующей координаты, так же пропускались. Подобной проблемы можно избежать, если алгоритм будет «пропускать» строго те точки, которые лежат под графиком плотности вероятности, а точки, которые лежат выше или на графике, – пропускать, но тогда снизится количество генерируемых значений. Исправлять эту алгоритмическую особенность в дальнейших рассуждениях не будем. Результаты генерации представлены в таблице 1.

Таблица 1 – Оценка метода Неймана для дискретных равномерно распределённых периодических входных величин.

Распределение	Максвелла	Экспоненциальное	Нормальное
Время генерации, мс	474	469	480
Кол-во генераций	5683	4027	20488
Успешных генераций, %	4,34	3,07	15,63
Корреляция	0,99956006	0,999159596	0,9994313

Далее проверим метод Неймана для непериодических равномерно распределённых входных величин, соответствующим образом изменив входные координату и шанс.

Здесь же проверим для непериодических неравномерно распределённых величин. Это необходимо сделать в соответствии с ранее принятым условием: в качестве генератора белого шума выступает АЦП. В таком случае предугадать, какие значения будут выступать в качестве входных не представляется возможным, при этом иметь случайное распределение. При этом у каждого значения координаты будет хотя бы 10 соответствующих ей шансов. Результаты этих экспериментов приведены в таблицах 2 и 3.

Таблица 2 – Оценка метода Неймана для дискретных равномерно распределённых непериодических входных величин.

Распределение	Максвелла	Экспоненциальное	Нормальное
Время генерации, мс	472	473	471
Кол-во генераций	5685	3955	20575
Успешных генераций, %	4,34	3,02	15,70
Корреляция	0,9998742	0,999762424	0,99967688

Таблица 3 – Оценка метода Неймана для дискретных неравномерно распределённых входных величин.

Распределение	Максвелла	Экспоненциальное	Нормальное
Время генерации, мс	472	473	471
Кол-во генераций	5685	3955	20575
Успешных генераций, %	4,34	3,02	15,70
Корреляция	0,9998742	0,999762424	0,99967688

Как видно из таблиц, разница между периодическими и непериодическим равномерно распределёнными входными данными практически отсутствует и находится в пределах некоторое погрешности. Для неравномерно распределённых величин, в свою очередь, можно отметить рост времени генерации. При этом количество сгенерированных по экспоненциальному закону величин заметно увеличилось: на ~0,2%, распределение Максвелла и нормальное в этом плане почти никак не изменились. Важно отметить значительное падение корреляции, что означает бóльшую отклонённость от заданного распределения, чем у равномерно распределённых величин [3].

Помимо этого, метод Неймана, как и метод обратной функции, можно использовать для генерирования СВ с заданными функциями вероятности. Для этого программно можно применять следующий алгоритм:

1. Аппроксимация функции вероятности методом трапеций с заданной точностью;
2. Вычисление производной функции вероятности. В соответствии с её статистическим смыслом получим значения функции плотности вероятности;
3. Считывание пар входных чисел;
4. Скалирование значений второго числа в паре в соответствии с максимальным значением плотности вероятности;
5. Генерация св при помощи сравнения второго числа со значением плотности вероятности в точке;
6. Запись результатов в файл.

Важным отличием такого алгоритма является невозможность изначально задать плотность вероятности аналитически. Это означает, что для записи плотности вероятности в интересующих нас точках придётся использовать массив значений. А если точек будет очень много то, возможно, будет необходимо найти оригинал решётчатого изображения, что так же является непростой задачей. Ниже приведены таблицы 4, 5 и 6 результата такой генерации для функций вероятности законов распределения, заданных плотностями вероятности ранее (формулы (1), (2) и (3)), корреляция вычислялась для плотностей вероятности [4].

Таблица 4 – Оценка метода Неймана для дискретных равномерно распределённых периодических входных величин в случае заданной функции вероятности.

Распределение	Максвелла	Экспоненциальное	Нормальное
Время генерации, мс	498	499	496
Кол-во генераций	5523	3876	19998
Успешных генераций, %	4,21	2,96	15,26
Корреляция	0,910245644	0,9245678910	0,95567954

Таблица 5 – Оценка метода Неймана для дискретных равномерно распределённых непериодических входных величин в случае заданной функции вероятности.

Распределение	Максвелла	Экспоненциальное	Нормальное
Время генерации, мс	495	493	496
Кол-во генераций	5489	3906	20154
Успешных генераций, %	4,19	2,98	15,38
Корреляция	0,910065438	0,9111524004	0,94461638

Таблица 6 – Оценка метода Неймана для дискретных неравномерно распределённых входных величин в случае заданной функции вероятности.

Распределение	Максвелла	Экспоненциальное	Нормальное
Время генерации, мс	512	504	514
Кол-во генераций	5479	4035	19879
Успешных генераций, %	4,18	3,08	15,17
Корреляция	0,893275981	0,880038942	0,90603758

Из таблиц видно, что количество и качество генераций таким методом упало, так же понизилось быстродействие. Но такой метод, тем не менее, позволяет генерировать СВ по ещё одному варианту закона распределения. При этом точность вычисления можно повысить, изменяя параметры аппроксимации и взятия производной.

Вывод. Был изучен алгоритм генерации СВ методом Неймана. На практике показана его реальная способность генерации СВ с довольно высоким значением корреляции. Однако результативновность (процент успешных генераций) довольно мала, хоть и зависит от плотностей распределения. Так же имеет возможность программной реализации разными способами, что является положительным атрибутом, но в неопытных руках может привести к алгоритмическим ошибкам и, как итог, проблемами в качестве генерации. Скорость генерации чисел находятся на удовлетворительном уровне: до полсекунды на 4 тысячи генераций. Ещё одной положительной чертой можно назвать простую программную реализацию и возможность использовать разные законы распределения.

Список литературы

1. Моделирование случайной величины с заданным законом распределения. [Электронный ресурс] – Режим доступа: http://stratum.ac.ru/education/textbooks/modelir/lection24.html – Дата доступа 24.12.2022.

2. Экспоненциальное распределение и его свойства. [Электронный ресурс] – Режим доступа: http://statistica.ru/theory/eksponentsialnoe-raspredelenie/ – Дата доступа 24.12.2022.

3. Распределение Максвелла. [Электронный ресурс] – Режим доступа: https://ru.wikipedia.org/wiki/Распределение_Максвелла – Дата доступа 24.12.2022.

4. Нормальное распределение. [Электронный ресурс] – Режим доступа: http://statistica.ru/theory/normalnoe-raspredelenie/ – Дата доступа 24.12.2022.

UDC 681.3: 004.021: 519.683.8

PSEUDORANDOM GENERATION OF SEQUENCES WITH A GIVEN LAW OF DISTRIBUTIONS BASED ON A WHITE NOISE GENERATOR

Babitsky K.P.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Rolich O.Ch. – PhD, associate professor, associate professor of the Department of ICSD

Annotation. The Neumann algorithm for generating pseudo-random variables with a given distribution law has been experimentally studied. Its principle of operation and the nature of the generated numbers are investigated: correlation with the distribution law, speed, the number of generated numbers with a finite amount of input data.

Keywords: white noise, distribution law, pseudorandom number, pseudorandom number generation.