УДК 621.391:004.73

ОЦЕНКА СТАТИСТИЧЕСКИХ СВОЙСТВ ПОСЛЕДОВАТЕЛЬНОСТЕЙ НА ВЫХОДЕ КОМБИНАЦИОННОГО ГЕНЕРАТОРА С ПОМОЩЬЮ ГРАФИЧЕСКИХ ТЕСТОВ

А.А. ЛАВДАНСКИЙ, Э.В. ФАУРЕ

Рассмотрен класс комбинационных генераторов, в котором в качестве комбинирующей функции используется операция суммирования в некотором конечном поле. Исследованы статистические свойства последовательности чисел на выходе комбинационного генератора, где в качестве исходных первичных генераторов использованы таблицы перестановок с взаимно простыми периодами повторения. Рассмотрены графические методы определения статистических свойств последовательностей чисел. Произведен анализ полученных с помощью графических тестов статистических характеристик последовательности на выходе комбинационного генератора с различным заполнением исходных таблиц перестановок (линейный конгруэнтный метод, квантовый генератор случайных чисел), выполнено их сравнение с характеристиками последовательностей на выходе существующих генераторов случайных (оцифрованные радиошумы) и псевдослучайных («Вихрь Мерсенна») чисел. Полученные результаты свидетельствуют об идентичности полученных с помощью графических методов оценки статистических свойств всех исследуемых последовательностей

ВВЕДЕНИЕ

Решение многих практических задач невозможно без использования генераторов случайных или псевдослучайных чисел (ПСЧ). Последовательности и случайных, и ПСЧ широко используются в таких важных задачах как имитационное моделирование, криптография, передача данных и др.

Отметим, что под случайной последовательностью чисел будем понимать последовательность, порожденную процессом, исход которого непредсказуем и не может быть повторно воспроизведен. Под псевдослучайной последовательностью чисел будем понимать последовательность, сформированную с помощью детерминированного алгоритма и обладающую статистическими свойствами, близкими к статистическим свойствам случайных последовательностей.

ПОСТАНОВКА ПРОБЛЕМЫ

Качественная последовательность ПСЧ должна соответствовать следующим требованиям: непредсказуемость, воспроизводимость, равномерное распределение слов последовательности, бесконечный период повторения, отсутствие корреляции между словами (символами) последовательности. Заметим, что достичь бесконечного периода повторения невозможно с помощью

детерминированных алгоритмов, поэтому под бесконечным периодом будем понимать настолько большой период, полное воспроизведение которого будет значительно превышать вычислительные возможности аппаратного обеспечения на текущем этапе развития вычислительной техники.

Всем перечисленным выше требованиям (кроме требования по воспроизводимости последовательности) отвечают последовательности, полученные квантованием естественных случайных физических процессов (например, «белого шума»). По существу, «белый шум» является идеальным источником случайных последовательностей чисел. Но использование его в качестве источника случайных чисел для ряда практических задач крайне затруднено или просто невозможно. Так, для решения криптографических задач требуется обеспечить воспроизводимость случайной последовательности в силу разнесения во времени (или в пространстве) процессов шифрования и дешифрования. Отсутствие же методов формирования псевдослучайных последовательностей чисел, отвечающих всем перечисленным требованиям, стимулирует непрерывный процесс улучшения существующих и поиска новых принципов и подходов в данной области исследований.

Таким образом, разработка и анализ новых методов и устройств формирования псевдослучайных последовательностей чисел, обладающих статистическими свойствами, близкими к статистическим свойствам случайных последовательностей чисел, а также обладающих свойством воспроизводимости, является актуальной проблемой.

ПОСТАНОВКА ЗАДАЧИ

Отдельного внимания среди устройств формирования случайных последовательностей чисел заслуживает класс комбинационных генераторов [1], реализующих комбинацию нескольких исходных генераторов с помощью некоторой комбинирующей функции. Комбинационные генераторы позволяют избавиться от таких слабых сторон генераторов, как малый период повторения, корреляция символов последовательности, неудовлетворительные статистические свойства.

Общим недостатком известных комбинационных генераторов является недостаточная криптографическая стойкость, связанная с наличием корреляции между выходной последовательностью и последовательностями на выходе исходных генераторов.

Комбинирование нескольких случайных процессов является актуальным направлением научных и прикладных исследований, находящих отражение, например, в [2, 3], однако множество вопросов все еще остаются неизученными.

Цель работы:

- анализ эффективности использования операции суммирования в некотором конечном поле в качестве комбинирующей функции комбинационного генератора;
- оценка с помощью графических тестов статистических свойств последовательности на выходе комбинационного генератора и их сравнение со статистическими свойствами случайной последовательностью чисел.

РЕШЕНИЕ ЗАДАЧИ

Комбинационный генератор псевдослучайных чисел реализует операцию комбинации исходных генераторов. В качестве исходных генераторов в работе рассматриваются генераторы подстановок или предварительно сформированные таблицы подстановок. В этом случае в процессе работы комбинационного генератора исходные значения, подлежащие комбинации, формируются циклически генераторами подстановок или считываются также циклически из таблиц подстановок. Заметим, что закон распределения чисел внутри подстановки является равномерным для всех значений из области определения случайной величины и обладает нулевой ошибкой воспроизведения закона распределения. Ошибка воспроизведения определяется в соответствии с методикой, изложенной в [4].

Таким образом, в состав рассматриваемого комбинационного генератора входят n таблиц подстановок $n \geq 2$, а также блок реализации комбинации. В данной работе в качестве комбинирующей функции используется операция суммирования по некоторому модулю M. Структурная схема генератора представлена на рис. 1.

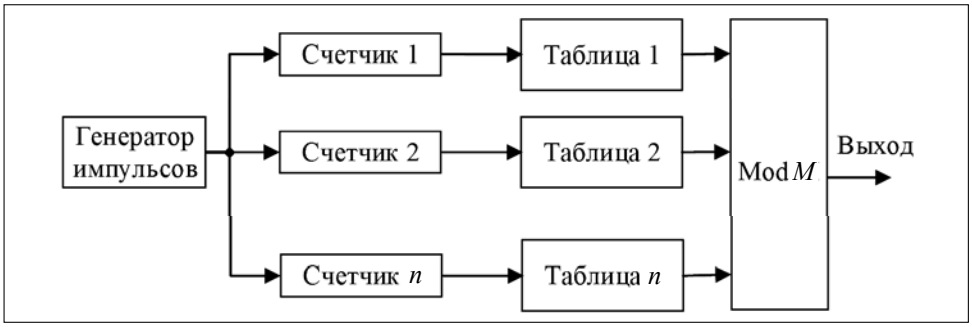

Рис. 1. Структурная схема комбинационного генератора

В *i*-й исходной таблице в произвольном порядке без повторов и пропусков записаны числа от 0 до M_i $(i \in [1, n])$. Заполнение таблиц может производиться как случайными данными (например, с использованием источника случайных чисел), так и с помощью детерминированного алгоритма (линейного конгруэнтного генератора, генератора *M*-последовательности и т.д.). Кроме того, представляется возможным использование различных алгоритмов заполнения для каждой из исходных таблиц.

Период последовательности T на выходе комбинационного генератора определяется мощностями алфавитов исходных генераторов (размерами таблиц) $M_1, M_2, \ldots, M_n$ и равен их наименьшему общему кратному (НОК):

$$T = HOK(M_1, M_2, \ldots, M_n).$$

Для получения максимального периода повторения последовательности на выходе комбинационного генератора мощности алфавитов исходных генераторов должны быть взаимно простыми. Тогда период

$$T = \prod_{i=1}^{n} M_i.$$

Рассмотрим принцип функционирования описанного комбинационного генератора, схема которого изображена на рис. 1.

В *i*-ю таблицу подстановок загружается случайная последовательность чисел от 0 до M_i ($i \in [1, n]$) без повторов и пропусков.

Для каждой таблицы существует счетчик, указывающий на текущее значение, считываемое из таблицы. Начальное значение счетчиков равно нулю. Счетчики всех таблиц работают синхронно. При появлении импульса на выходе генератора импульсов каждый из счетчиков увеличивает свое значение на единицу.

Данные с выходов каждого из счетчиков являются адресными для считывания значений из исходных таблиц подстановок. При этом на выходе каждой из таблиц появляется значение, соответствующее адресу, на который указывает счетчик таблицы. Сформированные *n* слов одновременно поступают на сумматор по модулю *M*. Выход сумматора является выходом устройства формирования псевдослучайной последовательности чисел. Описанная процедура продолжается циклически до выключения устройства.

Каждый из счетчиков при достижении значения, равного размеру соответствующей ему таблицы, обнуляется.

Выполним анализ статистических свойств описанного комбинационного генератора. Для этого определим методики оценки ПСЧ.

Исследование свойств выходной последовательности генератора ПСЧ использует две группы тестов: графические и статистические.

Графические тесты отображают статистические свойства последовательности в графическом виде (графики, гистограммы и т.д.), по виду которых можно судить о случайности последовательности. Важно отметить, что результат прохождения такого теста является субъективным, поскольку графический тест не дает численной оценки. Графические тесты являются первым этапом исследования последовательности псевдослучайных чисел, который позволяет сделать вывод, стоит ли работать с алгоритмом генерации, сформировавшим эту последовательность, в дальнейшем, либо же он требует доработки.

К графическим тестам можно отнести следующие: гистограмма распределения элементов последовательности; распределение на плоскости; проверка серий (распределение *k*-грамм); автокорреляционная функция; профиль линейной сложности; графический спектральный тест.

Статистические свойства исследуемой последовательности могут быть описаны вероятностью встречи определенного шаблона бит (байт) в этой последовательности. Для пояснения возьмем, например, распределение бит в последовательности. Зная длину последовательности в битах, а также посчитав количество бит, равных «1» и «0», возможно вычислить вероятность появления в потоке единицы либо нуля. Для случайной последовательности эти данные известны априорно (очевидно, что для квантованного «белого шума» вероятность появления единицы (как и нуля) стремится к 0,5). Предположение о случайности тестируемой последовательности строится на основании сравнения результатов тестирования с соответствующими результатами случайной последовательности чисел. Это сравнение и положено в основу статистических тестов. Результатом работы статистического теста

есть число (обычно доверительная вероятность подтверждения гипотезы о случайности последовательности), по которому можно однозначно судить о прохождении теста.

Рассмотрим графические тесты подробнее.

Гистограмма распределения элементов последовательности. Данный тест позволяет оценить соответствие распределения слов заданному закону распределения. Тест также полезен для выявления часто (редко) встречающихся, либо отсутствующих слов в последовательности. Построение гистограммы выполняется следующим образом. Производится подсчет символов, встречающихся в выборке размером V слов. Посчитанное количество отображается на графике в точке, соответствующей определенному символу. Для случайной последовательности (равномерное распределение) все возможные слова должны быть отображены на графике с примерно одинаковым значением, стремящимся к V/M, где M — мощность алфавита слов (количество различных слов в последовательности). Также данный тест можно интерпретировать как статистический, оценив распределение слов с помощью критерия хи-квадрат, что, впрочем, и производится в пакетах статистического тестирования.

Распределение на плоскости. Тест позволяет выявить зависимость между элементами последовательности. Последовательность слов группируется парами, которые рассматриваются как координаты на двумерном графике, т.е. если представить последовательность слов как $\{a_0, a_1, a_2, \ldots, a_V\}$, где V — размер выборки, то координаты графика можно представить как $x = a_k$, $y = a_{k+1}$, $0 \le k \le V-1$. Отображение этих точек на плоскости размером $2^R - 1 \times 2^R - 1$ точек (где R — разрядность чисел последовательности), является результатом теста. Для случайной последовательности расположение точек на плоскости будет хаотичным, а при росте выборки плоскость полностью будет заполнена точками. Признаком неслучайной последовательности является наличие на полученном изображении «узоров» (явно выраженных вертикальных либо горизонтальных линий, периодических рисунков и т.д.).

Проверка серий (распределение *k*-грамм). Последовательность из k слов (бит), при $k > 1$, называется k-граммой. Количество возможных k-грамм вычисляется как M^k, где M — мощность алфавита. Для примера, при $M = 2$ количество битовых биграмм равно $2^2 = 4$. Это биграммы $00, 01, 10, 11$. Для случайной последовательности количество k-грамм для одного k должно быть примерно одинаковым и стремиться к значению $\frac{V}{k \times M^k}$, где V — размер выборки. Тест позволяет оценить равномерность распределения символов в последовательности на основе k-грамм. Для этого производится подсчет количества различных k-грамм в последовательности с последующим отображением результата на гистограмме. Успешным прохождением теста является равная высота столбцов на гистограмме. Данный тест (как и гистограмму распределения элементов последовательности) возможно оценить с помощью критерия χ-квадрат, что позволит однозначно

судить о прохождении теста. Согласно [1], *b*-ичная последовательность длины N случайна, если она *k*-распределена для всех положительных целых чисел $k \le \log_2 N$.

Автокорреляционная функция. Тест предназначен для определения зависимостей внутри последовательности. Для этого определяется степень корреляции между сдвинутыми копиями последовательности. Обозначим бит последовательности как a_k, где k — номер бита в последовательности, $0 \le k < V$, V — размер выборки. Произведем нормирование битового представления исследуемой последовательности следующим образом: $1 \to 1,\ 0 \to -1$. Значение автокорреляционной функции для сдвига *i* обозначим как ρ_i. Тогда значение корреляции для *i*-того сдвига равно:

$$\rho_i = \frac{\sum_{k=0}^{V-1}\left(a_k \times a_{|k+i|_V}\right)}{\sum_{k=0}^{V-1} a_k^2}.$$

Для построения графика автокорреляционной функции рассчитываются значения ρ_i для $0 \le i < V$. Результатом выполнения теста будет отображение точек ρ_i на графике. При этом значение ρ_0 всегда должно быть равно единице. Чем ближе к нулю значения ρ_i в других точках, тем ближе исследуемая последовательность к случайной. Наличие пиков на графике (исключая пик в нулевой точке) свидетельствует про внутреннюю корреляцию бит последовательности.

Профиль линейной сложности. Линейной сложностью конечной двоичной последовательности называется число, равное длине самого короткого регистра сдвига с линейной обратной связью, который генерирует последовательность, имеющую в качестве первых n членов значения этой двоичной последовательности. Последовательность слов рассматривается как битовая последовательность. Для построения графика линейной сложности берутся первые n бит последовательности, $n > 1$. Для этих n бит производится расчет значения линейной сложности. Определение значения линейной сложности производится с помощью алгоритма Берлекэмпа-Мэсси [5–6]. Линейная сложность для n бит откладывается на графике. Успешным прохождением теста является малое отклонение полученного графика от графика $f(n) = n/2$ [7].

Графический спектральный тест. Тест предназначен для определения равномерности распределения нулей и единиц на основе анализа высоты выбросов преобразования Фурье. Для этого битовая последовательность нормируется $(1 \to 1, 0 \to -1)$, после чего к ней применяется дискретное преобразование Фурье. Первая половина полученной последовательности гармоник отображается на гистограмме. Для случайной последовательности чисел число гармоник, значительно превышающих среднюю высоту гармоник на гистограмме, должно стремиться к нулю.

Проведем сравнительное тестирование комбинационного генератора с помощью описанных выше графических тестов.

Для тестирования будем использовать два варианта заполнения исходных таблиц — с помощью линейного конгруэнтного метода и с помощью квантового генератора случайных чисел [8]. Для удобства тестирования будем использовать мощность алфавита $M = 256$, что позволяет исследовать как последовательность слов, так и последовательность бит в выборке (путем последовательной конкатенации битового представления слов в единый поток бит). Количество исходных таблиц в комбинационном генераторе установим равным 4, определив следующие размеры этих таблиц: 251, 241, 239, 229 слов без повторов и пропусков в каждой таблице соответственно. Такая комбинация исходных таблиц даст максимальный период (поскольку размеры таблиц взаимно простые), равный $T = 251 \times 241 \times 239 \times 229 = 3310732921$ слов. Параметры использованных конгруэнтных генераторов следующие:

$$K_1 = 34, C_1 = 67, M_1 = 251, S_{01} = 0;$$

$$K_2 = 158, C_2 = 5, M_2 = 241, S_{02} = 0;$$

$$K_3 = 24, C_3 = 222, M_3 = 239, S_{03} = 0;$$

$$K_4 = 143, C_4 = 47, M_4 = 229, S_{04} = 0.$$

В случае отсутствия цикла максимальной длины производится формирование сверхцикла путем конкатенации циклов генератора до достижения максимального периода, равного M_i слов [9].

Для сравнения с другими методами формирования последовательностей псевдослучайных чисел в тестирование также включен генератор «Вихрь Мерсенна» [10].

С целью сравнения статистических параметров последовательности на выходе комбинационного генератора с параметрами случайной последовательности чисел анализ также будет произведен для последовательности, сформированной с помощью оцифрованных радиошумов [11]. Будем называть эту последовательность случайной последовательностью слов.

Рассмотрим графики огибающей для гистограммы распределения элементов последовательности (рис. 2). Для тестирования взята выборка размером в 2^{25} слов.

Как следует из рис. 2, различные варианты заполнения исходных таблиц дают одинаковый результат — обеспечивают равномерное распределение слов в исследуемой выборке. Распределение слов сравнимо с таковым для генератора «Вихрь Мерсенна» и случайной последовательности слов.

Проведем оценку полученных результатов с помощью критерия χ^2 [12] с доверительной вероятностью 0,05. Для 255 степеней свободы критическое значение χ^2 равно $\chi^2_{0,05;255} = 293{,}2478$. Результаты оценки гистограммы распределения следующие: заполнение исходных таблиц с помощью линейного конгруэнтного метода ($\chi^2 = 229{,}3256$); заполнение исходных таблиц с помощью квантового генератора случайных чисел

($\chi^2 = 235{,}7351$); генератор «Вихрь Мерсенна» ($\chi^2 = 281{,}6868$); случайная последовательность слов ($\chi^2 = 252{,}6039$).

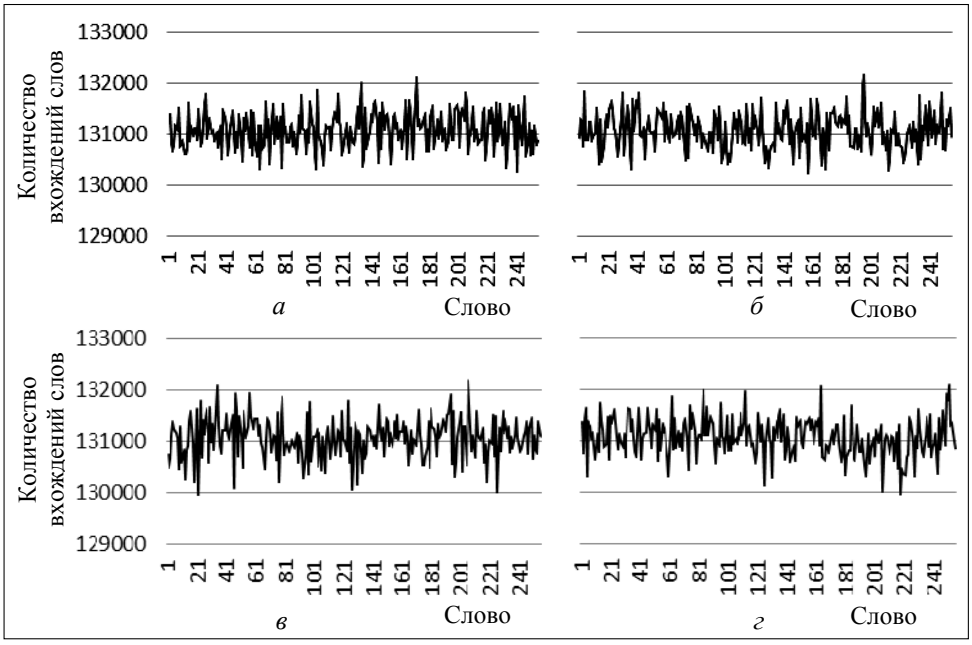

Рис. 2. График огибающей для гистограммы распределения слов последовательности: *а* — комбинационный генератор, заполнение исходных таблиц с помощью линейного конгруэнтного метода; *б* — комбинационный генератор, заполнение исходных таблиц с помощью квантового генератора случайных чисел; *в* — генератор «Вихрь Мерсенна»; *г* — случайная последовательность слов

Все полученные значения χ^2 меньше критического значения $\chi^2_{0{,}05;255}$, что подтверждает равномерность распределения слов в исследуемых последовательностях псевдослучайных чисел.

Результаты теста распределения на плоскости изображены на рис. 3. Выборка для каждой из исследуемых последовательностей постепенно увеличивается и составляет 2^{14}, 2^{15}, 2^{16}, 2^{18} слов, соответственно.

На полученных изображениях не наблюдается различных «узоров», а с ростом выборки плоскость полностью покрываются точками. Результаты тестирования свидетельствуют о присутствии всех возможных биграмм и отсутствии закономерности в их распределении для всех исследуемых последовательностей чисел.

Результаты определения профиля линейной сложности приведены на рис. 4. Для тестирования взяты первые 2^{11} бит каждой из исследуемых последовательностей.

Результаты анализа свидетельствуют о равномерном увеличении линейной сложности по мере увеличения размера выборки для всех исследуемых в работе последовательностей чисел. Полученные графики имеют малое отклонение от линии $f(n) = n/2$, что является свидетельством успешного прохождения теста.

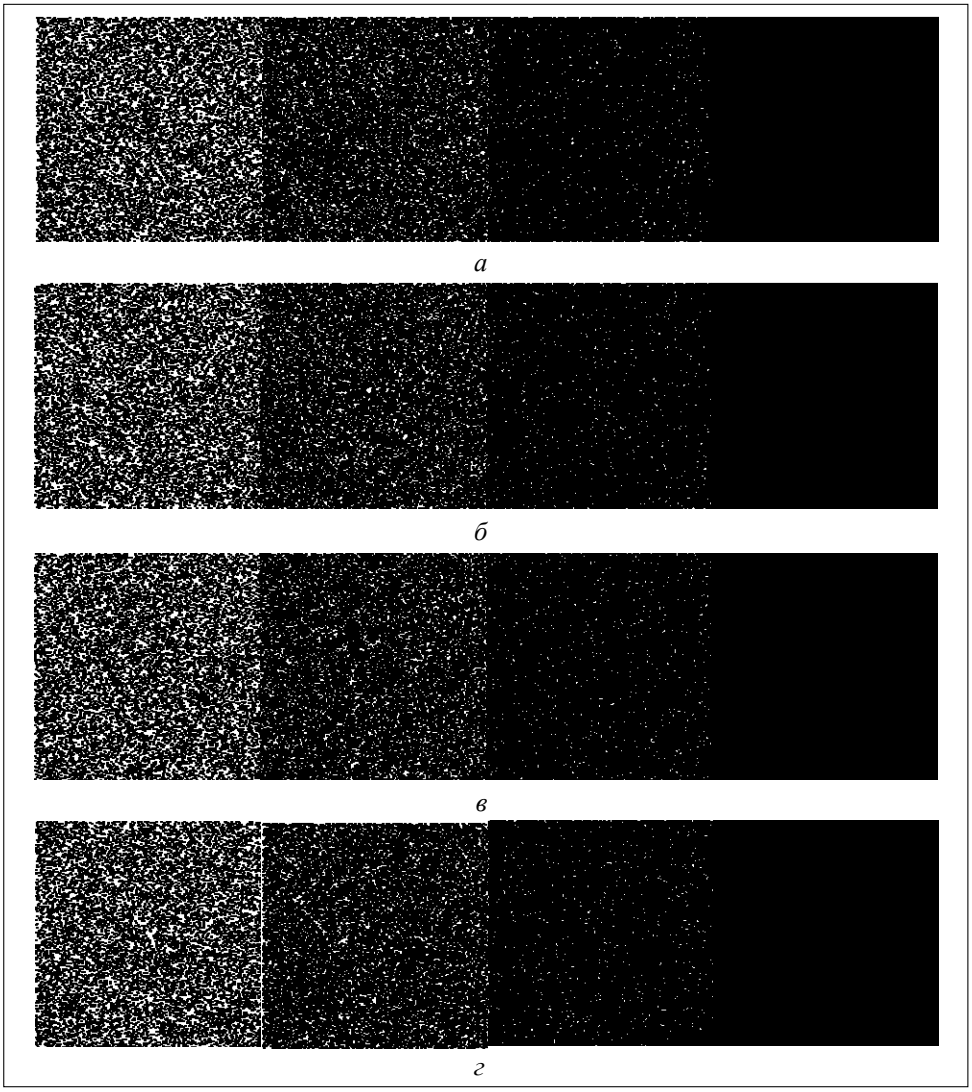

Рис. 3. Распределение слов на плоскости: *а* — комбинационный генератор, заполнение исходных таблиц с помощью линейного конгруэнтного метода; *б* — комбинационный генератор, заполнение исходных таблиц с помощью квантового генератора случайных чисел; *в* — генератор «Вихрь Мерсенна»; *г* — случайная последовательность слов

Результаты прохождения графического спектрального теста представлены на рис. 5. Для проведения теста взяты первые 2^9 бит каждой из исследуемых последовательностей. Средняя длина гармоник для каждой из последовательностей соответствует горизонтальной линии.

На гистограммах, приведенных на рис. 5, *б, в, г*, не наблюдается гармоник, значительно превышающих среднюю длину. Для варианта заполнения таблиц комбинационного генератора с помощью линейного конгруэнтного метода (рис. $5,a$), на гистограмме присутствуют пики гармоник, превышающие среднюю длину более чем в три раза, чего не наблюдается для остальных исследуемых выборок. Такой результат тестирования показывает незначительное отличие последовательности, сформированной с помощью

заполнения исходных таблиц результатом работы линейного конгруэнтного метода, от случайной последовательности и требует дальнейшего исследования.

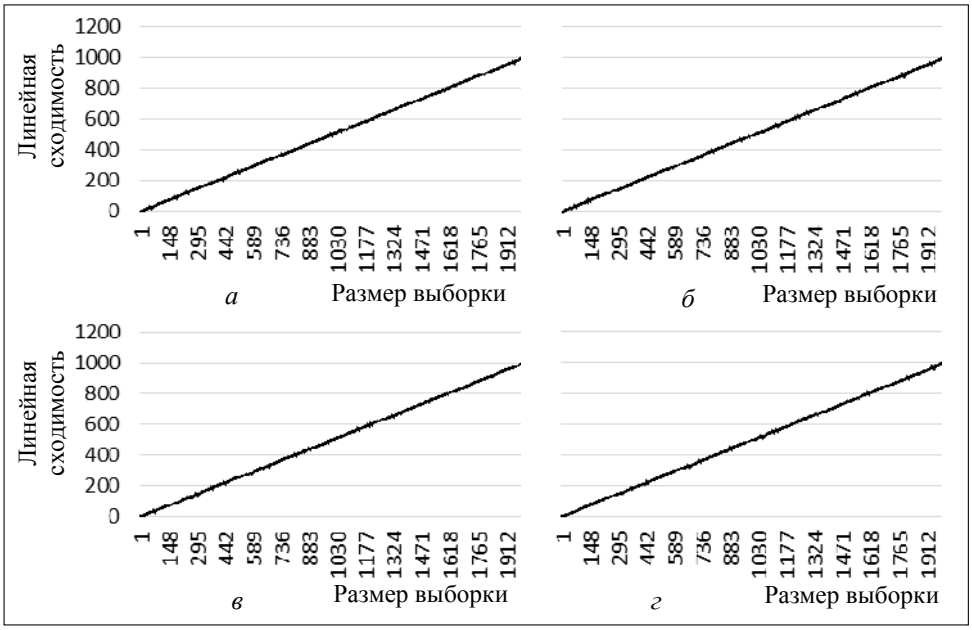

Рис. 4. Профиль линейной сложности: *а* — комбинационный генератор, заполнение исходных таблиц с помощью линейного конгруэнтного метода; *б* — комбинационный генератор, заполнение исходных таблиц с помощью квантового генератора случайных чисел; *в* — генератор «Вихрь Мерсенна»; *г* — случайная последовательность слов

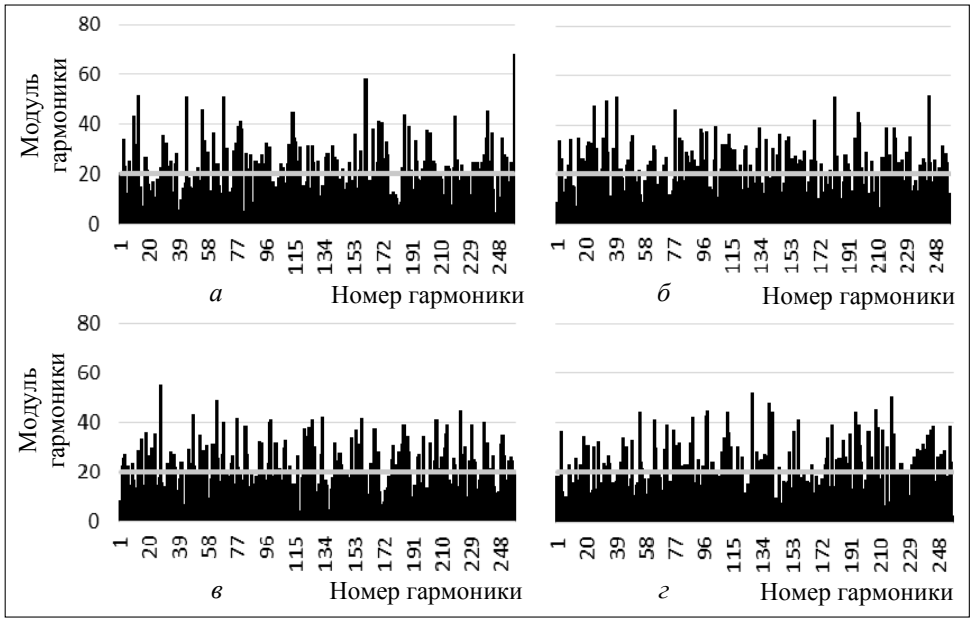

Рис. 5. Графический спектральный тест: *а* — комбинационный генератор, заполнение исходных таблиц с помощью линейного конгруэнтного метода; *б* — комбинационный генератор, заполнение исходных таблиц с помощью квантового генератора случайных чисел; *в* — генератор «Вихрь Мерсенна»; *г* — случайная последовательность слов

График автокорреляционной функции представлен на рис. 6. Для построения графика используется битовое представление исследуемых последовательностей. Для тестирования взяты первые 2^{15} бит каждой из исследуемых последовательностей.

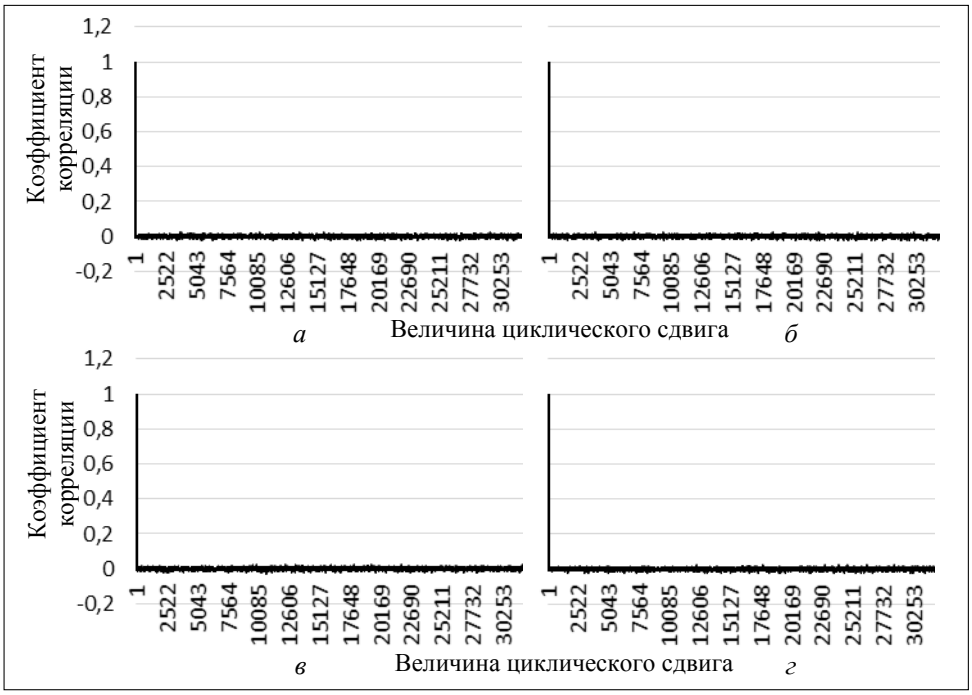

Рис. 6. Автокорреляционная функция: *а* — комбинационный генератор, заполнение исходных таблиц с помощью линейного конгруэнтного метода; *б* — комбинационный генератор, заполнение исходных таблиц с помощью квантового генератора случайных чисел; *в* — генератор «Вихрь Мерсенна»; *г* — случайная последовательность слов

Результаты тестирования показывают отсутствие всплесков корреляции, что является успешным прохождением теста и доказывает отсутствие внутренней корреляции для всех исследуемых последовательностей.

ВЫВОДЫ

В процессе тестирования последовательностей на выходе комбинационного генератора получены следующие результаты: гистограмма распределения слов последовательности подтверждает равномерное распределение слов формируемой генератором последовательности; анализ теста распределения на плоскости не показал каких-либо узоров на полученном изображении; отсутствуют всплески боковых лепестков на автокорреляционной функции, что свидетельствует об отсутствии корреляции между символами последовательности; графический спектральный тест показывает отсутствие значительных всплесков гармоник; профиль линейной сложности показывает линейное увеличение сложности последовательностей по мере увеличения размера выборки.

Проведенное в работе исследование статистических свойств псевдослучайных последовательностей чисел показало, что последовательности, порожденные комбинационным генератором, имеют большой период повторения, успешно проходят графические тесты, являются непредсказуемыми и воспроизводимыми.

ЛИТЕРАТУРА

1. *Кнут Д.Э.* Искусство программирования. В 4-х т. Том 2. Получисленные алгоритмы. — М.: Вильямс, 2007. — 832 с.
2. *Митянкина Т.В.*, *Швыдкий В.В., Фауре Э.В.* Преобразование дискретных случайных процессов комбинационным автоматом // Вісник ЧДТУ. — 2004. — № 3. — С. 67–69.
3. *Фауре Э.В.* Нелинейные преобразования дискретных случайных процессов // Радіоелектронні і комп'ютерні системи. — 2006. — № 6(18). — С. 200–205.
4. *Фауре Э.В., Береза А.С., Ярославская Е.А.* Оценка точности воспроизведения закона распределения дискретной случайной величины при ее преобразовании // Вісник Хмельницького національного університету. — 2012. — № 5. — С. 176–182.
5. *Elwyn R. Berlekamp*. Algebraic Coding Theory. — CA: Aegean Park Press, 1984. — 474 p.
6. *Massey J.L.* Shift-register synthesis and BCH decoding // IEEE Trans. Information Theory. — 1969. — IT-15. — P. 122–127.
7. *Niederreiter H.* Sequences with almost perfect linear complexity profile. In D. Chaum and W.L. Price, editors, Advances in Cryptology – EUROCRYPT '87, volume 304 of Lect. Notes Comput. Sci., pages 37–51, Berlin, 1987. Springer.
8. *QRNG* Service. — http://qrng.physik.hu-berlin.de/.
9. *Береза А.С., Лавданский А.А., Швыдкий В.В., Фауре Э.В.* Генерация конгруэнтных последовательностей чисел с заданными свойствами // Вісник Черкаського державного технологічного університету. — 2012. — № 2. — С. 3–8.
10. *Matsumoto M., Nishimura T.* Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator // ACM Trans. Model. Comput. Simulat. —1998. — **8**. — P. 3–30.
11. *True* Random Number Service.— http:// random.org/.
12. *Иванов М.А., Чугунков И.В.* Теория, применение и оценка качества генераторов псевдослучайных последовательностей. — М.: КУДИЦ-ОБРАЗ, 2003. — 240 с.

Поступила 24.06.2014