УДК 519.876.5.:621.391

Казакова Н.Ф., Щербина Ю.В.

Одесский национальный экономический университет, г. Одесса

ОБОБЩЕНИЕ РЕЗУЛЬТАТОВ ТЕСТИРОВАНИЯ ГЕНЕРАТОРОВ ПСЕВДОСЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Рассмотрены вопросы, связанные с оценкой результатов тестирования выходных псевдослучайных последовательностей, формируемых программными методами для целей моделирования и криптографии.
Ключевые слова: тестирование, генератор псевдослучайных последовательностей, методика.

Постановка проблемы в общем виде и ее связь с важными научными или практическими заданиями. Современный уровень развития вычислительной техники способствует активизации исследовательских работ в таких, казалось бы, не связанных между собой областях, как машинное моделирование и криптография. Общими для них являются необходимость в датчиках псевдослучайных последовательностей (ПСП), распределение вероятностей символов на выходе которых, в максимально возможной степени соответствует равномерному закону. Эта равномерность должна быть такой, чтобы наблюдатель не мог бы предположить значение очередного двоичного символа формируемой последовательности с вероятностью, отличимой от $1/2$, какой бы статистикой о значении предыдущих символов он не располагал. Псевдослучайные последовательности, удовлетворяющие этому требованию, будут неотличимы от истинно «случайных».

Анализ последних исследований и публикаций, в которых положено начало решению проблемы, выделены нерешенные вопросы общей проблемы. Первые результаты в области разработки программных методов формирования случайных чисел были обобщены и сформулированы Дональдом Кнутом в его знаменитой работе «Искусство программирования на ЭВМ» [1]. Наряду с научным обоснованием способов формирования псевдослучайных последовательностей он поднял вопрос об определении их качества. Проблема состояла в том, чтобы найти «шкалу», с помощью которой можно было бы оценивать степень «равномерности» распределения вероятностей чисел, формируемых на выходе генератора ПСП. Однако Дональд Кнут такой шкалы не предложил. Он предложил группу тестов основанных на статистическом критерии c^2. В эту группу входят:

1. *Проверка несцепленных серий.* Последовательность разбивается на m непересекающихся серий и строится распределение χ^2 для частот появления каждой возможной серии.
2. *Проверка интервалов.* Данный тест проверяет равномерность распределения символов в исследуемой последовательности, анализируя длины подпоследовательностей, все элементы которых принадлежат определённому числовому интервалу.
3. *Проверка комбинаций.* Последовательность разбивается на подпоследовательности определённой длины, и исследуются серии, состоящие из различных комбинаций чисел.
4. *Тест собирателя купонов.* Пусть $e_1, e_2, ..., e_n$ – последовательность длины n и размерности m. Исследуется распределение m-разрядных комбинаций в составе подпоследовательностей определённой длины.

5. *Проверка перестановок*. Данный тест проверяет равномерность распределения символов в исследуемой последовательности, анализируя взаимное расположение чисел в подпоследовательностях.

6. *Проверка на монотонность*. Служит для определения равномерности исходя из анализа невозрастающих и неубывающих подпоследовательностей.

7. *Проверка корреляции*. Проверяется взаимонезависимость элементов последовательности.

Перечисленные тесты предполагают определение степени схожести тестируемой последовательности с реальной случайной последовательностью на основании вводимого эталона и критерия. В каждом из них описывается уникальная для данного теста методика определения некоторого количественного показателя P_v, на основании которого делается вывод о том, следует ли считать тестируемую последовательность «случайной».

Формальное определение критерия предполагает задание нулевой гипотезы H_0, в соответствии с которой тестируемая последовательность является случайной. С ней непосредственно связана альтернативная гипотеза H_A, в соответствии с которой эта последовательность не может быть признана случайной. Принимая нулевую гипотезу, экспериментатор с вероятностью a рискует ошибиться (совершить так называемую «ошибку первого рода»). Соответственно, с вероятностью $1-a$ он будет прав. Обычно величину a выбирают в пределах $0.01 < a < 0.001$.

Аналогичные рассуждения могут быть приведены и для ошибки второго рода, когда тестируемая последовательность случайной не является, но на основании значения P_v с вероятностью b делается вывод о том, что она «случайна». В отличие от a, b не имеет фиксированного значения. Если поток данных неслучаен, существует бесконечное число ситуаций, при каждом из которых величина b принимает свое значение. Именно в силу неоднозначности таких ситуаций вычисление ошибки второго рода является сложной задачей и в тестовых алгоритмах не используется.

Достоинством тестового пакета, предложенного Кнутом, является их небольшое количество. Он утверждал, что количество тестов не должно быть строго определено. По его мнению, прохождение генератором каждого следующего теста лишь укрепляет уверенность разработчика в пригодности тестируемого генератора. Отчасти это так.

По мере развития технологий, в составе которых используются псевдослучайные последовательности, особенно, например, в криптографии, требования к равномерности распределения двоичных символов в их составе растут. Растет и число новых тестов, позволяющих производить дополнительную оценку их качества. Так, например, в Internet можно найти набор статистических тестов под названием Diehard [2], разработанный Джорджем Марсальей (George Marsaglia). Он включает 12 тестов и доступен по адресу http://stat.fsu.edu/pub/diehard/. По адресу http://www.isi.qut.edu.au/resources/cryptx/ можно связаться с разработчиками пакета тестов Crypt-X [3] и получить программное обеспечение и руководство по их применению. Создатели этих пакетов, к сожалению, также как и Д. Кнут, не предложили способа обобщения результатов, полученных при использовании всего пакета предлагаемых тестов. По их мнению, следует ограничиться простейшей качественной шкалой с двумя значениями: последовательность «случайна» – «не случайна». При этом случайной она признается, если все входящие в пакет тесты признают ее случайной.

Впервые методика обобщающей количественной оценки результатов тестирования была предложена в пакете, рекомендованном NIST «A Statistical Test Suite for the Validation of Random Number Generators and Pseudo Random Number Generators for Cryptographic Applications» [4,5]. Этот пакет и руководство к нему имеется в свободном

доступе на сайте института (http://csrc.nist.gov/publications/PubsSPs.html). Вопрос заключается в том, ***является ли данная методика пригодной только для тестов, входящих в пакет, предложенный NIST, или же она носит универсальный характер***.

Целью работы является анализ методики, изложенной в руководстве к пакету тестов NIST, распространение ее применимости на иные произвольные наборы тестов. Ее официальное описание приводится в документе [4]. Оценка результатов тестирования, выполняемая на ее основе, предполагает выполнение следующих этапов.

1. Формирование тестируемым генератором m последовательностей, длина каждой из которых составляет n двоичных символов.

$$\begin{aligned} e^{(1)} &= e_1^{(1)}, e_2^{(1)}, e_3^{(1)}, ..., e_n^{(1)}, \\ e^{(2)} &= e_1^{(2)}, e_2^{(2)}, e_3^{(2)}, ..., e_n^{(2)}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ e^{(m)} &= e_1^{(m)}, e_2^{(m)}, e_3^{(m)}, ..., e_n^{(m)}. \end{aligned} \tag{1}$$

Эти последовательности должны быть сформированы тестируемым генератором с разными ключами, чтобы их можно было считать «независимыми», то есть они не должны быть «нарезаны» из одной псевдослучайной последовательности. Количество последовательностей m определяется величиной выбранного значения ошибки первого рода a. Например, если $a = 0,01$, значит, количество последовательностей лучше выбирать равным 100 или 1000 с тем, чтобы удобно было подсчитывать долю последовательностей, не прошедших тест. Что касается длины последовательности n, то она определяется методикой проведения конкретного теста.

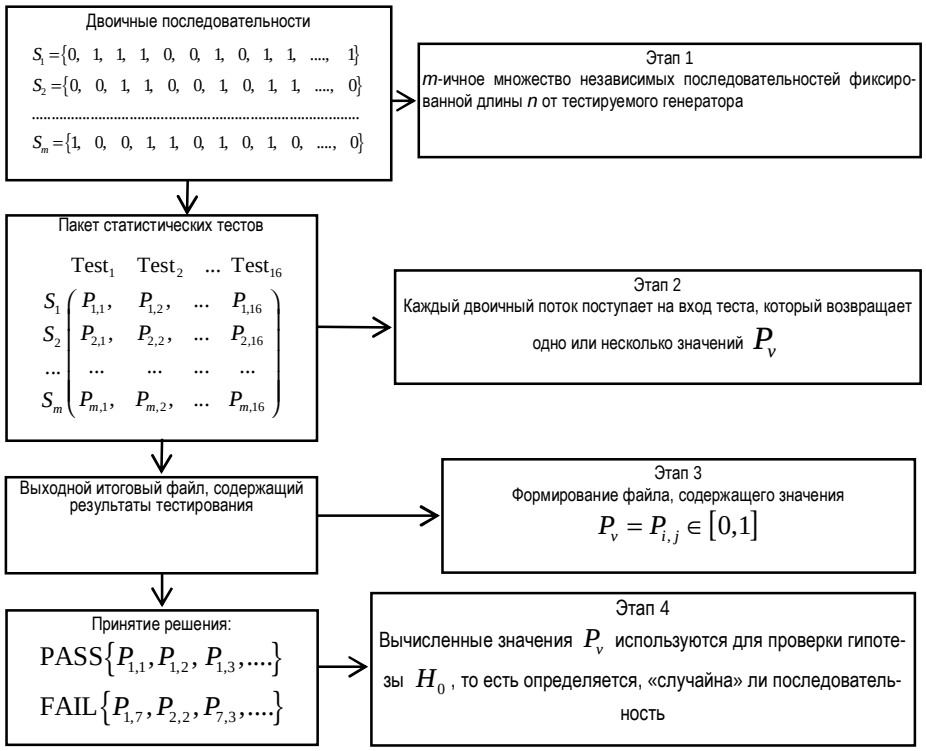

Рис. 1. Архитектура статистического тестирования NIST

2. Тестирование набором статистических тестов. Каждая из последовательностей набора (1) подвергается испытанию тестами, входящими в набор, и результаты тестирования $s(obs)$ запоминаются.

Для каждой из m последовательностей вычисляется значение параметра P_v, под которым понимается вероятность того, что совершенный генератор случайных чисел произвел бы последовательность менее случайную, чем исследуемая последовательность. При этом для каждого теста выбирается уровень значимости a. Если значение P_v больше либо равно a, то последовательность считается случайной. Методика определения величины P_v в каждом тесте своя. Полученные результаты могут быть занесены в таблицу. Если число тестов равно t, то таблица будет выглядеть так:

Таблица 1

Результаты вычисления значений P_v для тестовых статистик $s(obs)$

Последовательность	Тест 1	Тест 2	…	Тест *t*
$e^{(1)}$	$P_v{}^{(1)}{}_1$	$P_v{}^{(1)}{}_2$	…	$P_v{}^{(1)}{}_m$
$e^{(2)}$	$P_v{}^{(2)}{}_1$	$P_v{}^{(2)}{}_2$	…	$P_v{}^{(2)}{}_m$
…	…	…	…	…
$e^{(m)}$	$P_v{}^{(m)}{}_1$	$P_v{}^{(m)}{}_2$	…	$P_v{}^{(m)}{}_m$

При этом общее число тестовых статистик будет равно mt.

3. Формирование файла, содержащего результаты вычислений из табл. 1, в котором величины $P_v{}^{(j)}{}_i$ представлены как вероятностные характеристики. Этот файл предназначен для автоматизации процесса оценки полученных результатов испытаний.

4. Оценка результатов каждого теста. Ожидается, что при фиксированном уровне определенный процент последовательностей будет оценен как «неслучайные». Например, если уровень значимости выбирается 0,01 (т.е. $a = 0,01$), то, как ожидается, около 1% последовательностей будут признаны «неслучайными». Всякий раз когда значения $P_v \geq a$, последовательность признается «случайной».

Перечисленные этапы тестирования иллюстрирует рис. 1.

Далее, основываясь на результатах работ, выполненных в соответствии с условиями этапа 4, определяют долю последовательностей, прошедших испытание в каждом тесте g_i, $i = 1, 2, ..., t$, как отношение числа последовательностей, признанных «случайными», к числу последовательностей, признанных неслучайными. Учитывая эмпирические результаты для конкретных статистических испытаний, вычислить долю последовательностей, которые проходят испытания. Например, если в некотором *i*-том тесте были протестированы 1000 двоичных последовательностей (т. е. $m = 1000$, уровень значимости $a = 0.01$), и для 996 из них $P_v \geq 0.01$, то этот показатель составляет $g_i = 996/1000 = 0.9960$.

В идеальном случае будет забраковано ровно a от общего числа тестируемых последовательностей, а в реальном случае величины g_i будут рассеяны около этого значения. При этом границы доверительного интервала предлагается ограничить тремя среднеквадратическими отклонениями, считая, что распределение результатов испытаний подчиняется нормальному закону:

$$g_{\min(\max)} \pm 3\sqrt{\frac{p(1-p)}{m}}, \qquad (2)$$

где $p = 1 - a$, m – размер выборки.

Выход величины за пределы этого интервала говорит о том, что испытуемая последовательность признается данным тестом «неслучайной». Пусть, например, используется тысяча последовательностей от испытуемого генератора ($m = 1000$) при $a = 0,01$. Тогда пределы доверительного интервала будут равны

$$g_{\min(\max)} = 0.99 \pm \sqrt[3]{\frac{0.99(0.01)}{1000}} = 0.99 \pm 0.0094392.$$

Приведенные рассуждения иллюстрирует рис. 2.

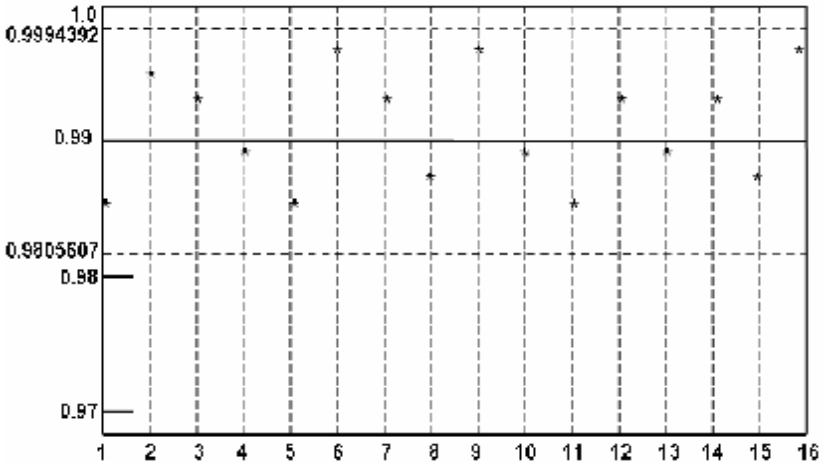

Рис. 2. Распределение g_i

Другим способом определения качества испытуемой последовательности является проверка равномерности распределения величин P_v на интервале $[0,1]$. Эта проверка может быть осуществлена с применение критерия соответствия c^2. Для результатов, полученных с использованием каждого теста, строится гистограмма. Интервал $[0,1]$ разбивается на 10 равных участков и определяется число попаданий величины в каждый из этих участков v_i., а затем вычисляется показатель

$$c_j^2 = \sum_{k=1}^{10} \frac{(v_i - m/10)^2}{m/10}.$$

Далее вычисляется показатель

$$P_{v_T} = igamc\left(\frac{9}{2}, \frac{\chi}{2}\right).$$

Если в данном тесте выполняется условие $P_{v_T} \geq 0{,}0001$, то тестируемая последовательность признается данным тестом «случайной».

Наконец, если все последовательности вида (1) во всех тестах показали частоты γ_i, попадающие в доверительный интервал (2), а величины P_{v_T} в каждом тесте не превышают значения 0,0001, испытуемый генератор считают пригодным для прикладных задач.

Выводы. В заключение следует отметить, что данная методика успешно применялась авторами при обобщении результатов, полученных с помощью универсального теста Маурера, входящего в пакет NIST. Как известно, этот тест, фактически, представляет собой семейство из 10 отдельных тестов, каждый из которых анализирует псевдослучайные последовательности на предмет интервалов между группами символов с длинами от 6 до 16 бит. В принципе эта методика может быть успешно применена к любой другой группе тестов. Ее достоинство в том, что она дает количественный показатель качества испытуемого генератора, взамен не особенно конкретного правила, предложенного Д. Кнутом.

Литература

1. Кнут Д. Искусство программирования на ЭВМ. – Т.2. – М.: Мир, 1977. – 727 с.
2. The Marsaglia Random Number CDROM including the Diehard Battery of Tests of Randomness// http://www.stat.fsu.edu/pub/diehard/
3. Statistical test suite Crypt-X //http://www.isi.qut.edu.au/resources/cryptx/
4. A Statistical Test Suite for the Validation of Random Number Generators and Pseudo Random Number Generators for Cryptographic Applications. NIST Special Publication 800-22. May 15, 2001.
5. Иванов М.А., Чугунков И.В. Теория, применение и оценка качества генераторов псевдослучайных последовательностей. – М.: КУДИЦ-ОБРАЗ, 2003. – 240 с.

Надійшла до редколегії 20.04.2012

Рецензент: д.т.н., пр*оф*. І.П. Панфілов, Одеська національна академія зв'язку ім.О.С.Попова

Казакова Н.Ф., Щербина Ю.В.

УЗАГАЛЬНЕННЯ РЕЗУЛЬТАТІВ ТЕСТУВАННЯ ГЕНЕРАТОРА ПСЕВДОВИПАДКОВИХ ПОСЛІДОВНОСТЕЙ

Розглянуто питання, пов'язані з оцінкою результатів тестування вихідних псевдовипадкових послідовностей, що формуються програмними методами з метою моделювання та криптографії.

Ключові слова: тестування, генератор псевдовипадкових послідовностей, методика

Kazakova N.F., Shcherbina Yu.V.

GENERALIZATION OF TESTING RESULTS GENERATORS OF PSEUDOCASUAL SEQUENCES

The problems associated with assessing the results of testing the output pseudo-random sequences generated programmatically for the purposes of modeling, and cryptography were considered.

Key words: testing, generator of pseudocasual sequences, method