ИНФОРМАТИКА

2004 **июль-сентябрь** **№3**

УДК 681.3.06.004.239.056(075.8)

Ю.С. Харин, А.И. Петлицкий

СТАТИСТИЧЕСКОЕ ТЕСТИРОВАНИЕ ДВОИЧНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ НА ОСНОВЕ СРАВНЕНИЯ ФРАГМЕНТОВ

Построены алгоритмы тестирования случайных и псевдослучайных последовательностей, основанные на сравнении их фрагментов с помощью статистик скалярного произведения. Получены оценки мощности и быстродействия алгоритмов.

Введение

Генерирование случайных последовательностей с заданным вероятностным законом и проверка их адекватности – одни из важнейших проблем криптологии [1 – 3]. Генераторы случайных последовательностей и программы тестирования используются в системах информационной безопасности для генерации ключевой информации и задания ряда параметров этих систем. Генерация случайной последовательности с произвольным законом распределения сводится с помощью известных методов обратной функции, исключения и композиции [1, 4] к генерации так называемой базовой случайной последовательности – равномерно распределенной двоичной случайной последовательности (РРДСП). РРДСП – это последовательность случайных битов $x_1, x_2, ..., x_t, x_{t+1}, ... \in V = \{0,1\}$, обладающая двумя свойствами (гипотеза H_0): C_1) для любого числа $n \in \mathrm{N}$ и произвольных индексов $1 \le t_1 < ... < t_n$ случайные биты $x_{t_1}, ..., x_{t_n}$ независимы в совокупности; C_2) для любого $t \in \mathrm{N}$ случайный бит имеет равновероятные значения:

$$\mathrm{P}\{\mathrm{x}_t = 0\} = \mathrm{P}\{\mathrm{x}_t = 1\} = 0{,}5 .$$

Статистическое тестирование заключается в проверке этих двух требований, т. е. гипотез H_0 и $H_1 = \overline{H_0}$ на основе наблюдаемой реализации $x_1, ..., x_n \in V$ конечной длительности *n*.

Обзор существующих алгоритмов тестирования представлен в [3]. Большинство существующих тестов основано на применении методов математической статистики для проверки выполнимости специальных свойств двоичной последовательности, вытекающих из C_1, C_2. В [3] предложен подход к тестированию, основанный на разбиении последовательности $\{x_t\}$ на фрагменты по *N* символов и сравнении свойств этих фрагментов. В данной статье на основе этого подхода построена серия тестов и проведен их сравнительный анализ по точности и быстродействию.

Пусть *L*, *N*, *m* – заданные натуральные числа. Осуществим разбиение подлежащей тестированию двоичной последовательности $x_1, ..., x_n \in V$ длины $n = LNm$ на Lm последовательных непересекающихся фрагментов длины *N*:

$$X = \{x_1, x_2, ..., x_{L \cdot m \cdot N}\} = \{X_1^{(1)}, X_2^{(1)}, ..., X_m^{(1)}, X_1^{(2)}, ..., X_m^{(L)}\}; \quad (1)$$

$$X_i^{(l)} = \{x_{(l-1)N \cdot m + (i-1)N + 1}, x_{(l-1)N \cdot m + (i-1)N + 2}, ..., x_{(l-1)N \cdot m + i \cdot N}\},\ i = 1, ..., m,\ l = 1, ..., L .$$

На фрагментах $\{X^{(l)} : l = 1, ..., L\}$ определим статистики скалярных произведений:

$$Y_{ik}^{(l)} = X_i^{(l)T} X_k^{(l)} = \sum_{j=1}^{N} x_{(l-1)N \cdot m + (i-1)N + j} x_{(l-1)N \cdot m + (k-1)N + j},\ i = 1, ..., m-1,\ k = i+1, ..., m . \quad (2)$$

Статистика $Y_{ik}^{(l)}$ характеризует «степень похожести» *i*-го и *k*-го фрагментов.

Легко установить, что если верна гипотеза H_0, то распределение статистики $Y_{ik}^{(l)}$ имеет вид

$$\mathrm{P}_{H_0}\left\{Y_{ik}^{(l)} = y\right\} = 2^{-2N} 3^{N-y} \binom{N}{y},\ y \in \{0,1,...,N\},\ 1 \le i < k \le m,$$

а математическое ожидание и дисперсия принимают значения

$$\mathrm{E}_{H_0}\left\{Y_{ik}^{(l)}\right\} = \frac{N}{4},\ D_{H_0}\left\{Y_{ik}^{(l)}\right\} = \frac{3N}{16},\ 1 \le i < k \le m.$$

Далее верхний индекс *l* в тех случаях, когда он теоретически не существенен, будет полагаться равным 1 и опускаться, так как $Y_{ik}^{(1)}, Y_{ik}^{(2)}, ..., Y_{ik}^{(L)}$ одинаково распределены.

1. Тест на основе среднего арифметического статистик скалярного произведения X_1 и X_k

Данный тест основан на среднем арифметическом скалярных произведений фрагментов:

$$\overline{Y} = (m-1)^{-1} \sum_{k=2}^{m} Y_{1k}\ . \tag{3}$$

Теорема 1. *Если верна гипотеза H_0, то распределение статистики $\overline{Y}$ определяется следующим образом:*

$$p_0 = \mathrm{P}_{H_0}\left\{\overline{Y} = 0\right\} = 2^{-N}\left(1 + 2^{1-m}\right)^N;$$

$$p_y = \mathrm{P}_{H_0}\left\{\overline{Y} = \frac{y}{m-1}\right\} = 2^{-N} \sum_{j=0}^{N} 2^{(1-m)j} \binom{N}{j} \left(\sum_{j_1+...+j_{m-1} \le y} \binom{j}{j_1} \cdot ... \cdot \binom{j}{j_{m-1}} - \sum_{j_1+...+j_{m-1} \le y-1} \binom{j}{j_1} \cdot ... \cdot \binom{j}{j_{m-1}} \right),$$

где $y \in \{1,2,...,(m-1)N\}$,

или по эквивалентным формулам:

$$p_y = \mathrm{P}_{H_0}\left\{\overline{Y} = \frac{y}{m-1}\right\} = \sum_{a_1+...+a_N = y} q_{a_1} \cdot ... \cdot q_{a_N}\ ,$$

где $q_{a_j} = 2^{-m} \binom{m-1}{a_j} + \frac{1}{2}\delta_{0a_j}$, $a_j \in \{0,1,...,m-1\}$, $j = 1,...,N$, $y \in \{0,1,...,(m-1)N\}$.

Доказательство. Первое утверждение теоремы следует из того, что

$$p_0 = \mathrm{P}_{H_0}\left\{\sum_{k=2}^{m}\sum_{r=1}^{N} x_r x_{(k-1)N+r} = 0\right\} = \prod_{r=1}^{N} \mathrm{P}_{H_0}\left\{\sum_{k=2}^{m} x_r x_{(k-1)N+r} = 0\right\};$$

$$p_y = \mathrm{P}_{H_0}\left\{\overline{Y} = \frac{y}{m-1}\right\} = \mathrm{P}_{H_0}\left\{\overline{Y} \le \frac{y}{m-1}\right\} - \mathrm{P}_{H_0}\left\{\overline{Y} \le \frac{y-1}{m-1}\right\}.$$

При доказательстве второго утверждения используется тот факт, что

$$p_y = \mathrm{P}_{H_0}\left\{\sum_{r=1}^{N}\eta_r = y\right\} = \sum_{a_1+\ldots+a_N=y}\mathrm{P}_{H_0}\{\eta_1 = a_1\}\cdot\ldots\cdot\mathrm{P}_{H_0}\{\eta_N = a_N\},$$

где $\eta_r = \sum_{k=2}^{m} x_r x_{(k-1)N+r}$, $\mathrm{P}_{H_0}\{\eta_r = 0\} = \mathrm{P}_{H_0}\{x_r = 0\} + \mathrm{P}_{H_0}\{x_r = 1\}\mathrm{P}_{H_0}\left\{\sum_{k=2}^{m} x_{(k-1)N+r} = 0\right\}$;

$\mathrm{P}_{H_0}\{\eta_r = b\} = \mathrm{P}_{H_0}\{x_r = 1\}\mathrm{P}_{H_0}\left\{\sum_{k=2}^{m} x_{(k-1)N+r} = b\right\}$, $b = 1,\ldots,m-1$.

При помощи теоремы 1 на основе χ^2-статистики строится тест T_1 среднего арифметического скалярного произведения X_1 и X_k.

1. Последовательность разбивается на Lm непересекающихся фрагментов согласно формуле (1).

2. По этим фрагментам $\left\{X^{(l)} : l = 1,\ldots,L\right\}$ согласно выражениям (2), (3) вычисляются статистики $\left\{Y_{1k}^{(l)} : k = 2,\ldots,m\right\}$ и $\overline{Y}^{(l)}$.

3. По совокупности статистик $\left\{\overline{Y}^{(l)} : l = 1,\ldots,L\right\}$ вычисляется χ^2-статистика:

$$\chi^2 = \sum_{y=0}^{(m-1)N}\frac{\left(v_y - Lp_y\right)^2}{Lp_y},\ v_y = \sum_{l=1}^{L}\mathrm{I}\left\{\overline{Y}^{(l)} = \frac{y}{m-1}\right\}.$$

4. Выносится решение с помощью статистического правила: если $P > \varepsilon$, то принимается гипотеза H_0, иначе принимается H_1, где $P = 1 - G_K(\chi^2)$ – так называемое P-значение, $G_K(\cdot)$ – стандартная функция χ^2-распределения с $K = (m-1)N$ степенями свободы [5], а $\varepsilon \in (0,1)$ – задаваемый уровень значимости теста.

Если некоторые из элементов вероятностей $\{p_y\}$ достаточно малы (на практике если $Lp_y < 5$), то целесообразно проводить группировку соответствующих значений [3].

2. Тест на основе среднего арифметического статистик скалярного произведения X_i и X_k

Определим статистику

$$\overline{Y} = \frac{2}{m(m-1)}\sum_{1\le i<k\le m} Y_{ik}, \qquad (4)$$

в этом случае тест является обобщением теста T_1.

Теорема 2. Если верна гипотеза H_0, то распределение статистики $\overline{Y}$ определяется формулами*

$$p_y = \mathrm{P}_{H_0}\left\{\overline{Y} = \frac{2y}{m(m-1)}\right\} = \sum_{a_1+\ldots+a_N=y} q_{a_1}\cdot\ldots\cdot q_{a_N},$$

где $a_j \in \left\{0, \binom{2}{2}, \binom{3}{2}, \ldots, \binom{m}{2}\right\}$, $q_{a_j} = 2^{-m}\left(\binom{m}{d_j} + \delta_{0a_j}\right)$, $d_j = 0.5 + \sqrt{0.25 + 2a_j}$, $j = 1,\ldots,N$,

$y \in U = \{u_0, u_1, \ldots, u_K\} \subset R$ *– конечное множество K+1 упорядоченных (в лексикографическом порядке) всевозможных комбинаций элементов* $a_1, a_2, \ldots, a_N$.

* Теоремы 2, 3 доказываются с использованием методов комбинаторики аналогично теореме 1.

При помощи теоремы 2 на основе χ^2-статистики строится тест T_2 среднего арифметического скалярного произведения X_i и X_k.

1. Последовательность разбивается на Lm непересекающихся фрагментов согласно формуле (1).

2. По фрагментам $\left\{X^{(l)} : l = 1,..., L\right\}$ согласно выражениям (2), (4) вычисляются статистики $\left\{Y_{ik}^{(l)} : i = 1,..., m-1,\ \ k = i+1,..., m\right\}$ и $\overline{Y}^{(l)}$.

3. Вычисляется χ^2-статистика: $\chi^2 = \sum_{r=0}^{K} \frac{\left(v_r - Lp_{u_r}\right)^2}{Lp_{u_r}}$, $v_r = \sum_{l=1}^{L} \mathrm{I}\left\{\overline{Y}^{(l)} = \frac{2u_r}{m(m-1)}\right\}$.

4. Выносится решение с помощью статистического правила: если $P > \varepsilon$, то принимается гипотеза H_0, иначе принимается H_1, где $P = 1 - G_K(\chi^2)$.

3. Тест на основе экстремальной статистики скалярного произведения X_{2i-1} и X_{2i}

На статистиках $\{Y_{2i-1\ 2i}\}$ строим статистику максимального скалярного произведения:

$$Y_{\max} = \max\left\{Y_{12}, Y_{34},..., Y_{2[m/2]-1\ 2[m/2]}\right\}. \tag{5}$$

Теорема 3. *Если верна гипотеза H_0, то распределение статистики Y_{max} имеет вид*

$$p_0 = \mathrm{P}_{H_0}\left\{Y_{\max} = 0\right\} = \left(\frac{3}{4}\right)^{N\cdot[m/2]};$$

$$p_y = \mathrm{P}_{H_0}\left\{Y_{\max} = y\right\} = \left(2^{-N}\sum_{j=0}^{N} 2^{-j}\binom{N}{j}\sum_{s=0}^{y}\binom{j}{s}\right)^{[m/2]} - \left(2^{-N}\sum_{j=0}^{N} 2^{-j}\binom{N}{j}\sum_{s=0}^{y-1}\binom{j}{s}\right)^{[m/2]},\ \ y \in \{1,2,..., N\}.$$

При помощи теоремы 3 на основе χ^2-статистики строится тест T_3 максимального скалярного произведения X_{2i-1} и X_{2i}.

1. Последовательность разбивается на Lm непересекающихся фрагментов согласно формуле (1).

2. По этим фрагментам $\left\{X^{(l)} : l = 1,..., L\right\}$ согласно выражениям (2), (5) вычисляются статистики $\left\{Y_{2i-1\ 2i}^{(l)} : i = 1,..., [m/2]\right\}$ и $Y_{\max}^{(l)}$.

3. По статистикам $\left\{Y_{\max}^{(l)} : l = 1,..., L\right\}$ вычисляется χ^2-статистика:

$$\chi^2 = \sum_{y=0}^{N} \frac{\left(v_y - Lp_y\right)^2}{Lp_y},\ v_y = \sum_{l=1}^{L} \mathrm{I}\left\{Y_{\max}^{(l)} = y\right\}.$$

4. Выносится решение с помощью статистического правила: если $P > \varepsilon$, то принимается гипотеза H_0, иначе принимается H_1, где $P = 1 - G_N(\chi^2)$.

4. Тест, основанный на экстремальной статистике скалярного произведения X_i и X_k

Данный тест является обобщением теста T_3 и основывается на статистике

$$Y_{\max} = \max_{1 \le i < k \le m} \left\{Y_{ik}\right\}. \tag{6}$$

В данном случае распределение статистики $Y_{\max}$ при верной гипотезе H_0 вычисляется приближенно с помощью метода Монте-Карло [4]. Алгоритм вычисления имеет следующий вид:

1. Генерируется случайная последовательность $\{z_t\}$ длины SmN, где S – достаточно большое натуральное число (определяющее точность метода Монте-Карло).

2. Данная последовательность разбивается на Sm последовательных непересекающихся фрагментов: $Z = \{z_1, z_2, ..., z_{S\cdot m\cdot N}\} = \{Z_1^{(1)}, Z_2^{(1)}, ..., Z_m^{(1)}, Z_1^{(2)}, ..., Z_m^{(S)}\}$.

3. По фрагментам $\{Z^{(s)} : s = 1, ..., S\}$ вычисляются статистики

$$T_{ik}^{(s)} = Z_i^{(s)T} Z_k^{(s)} = \sum_{j=1}^{N} z_{(s-1)N\cdot m+(i-1)N+j} z_{(s-1)N\cdot m+(k-1)N+j}\ ,\ i = 1, ..., m-1\ ,\ k = i+1, ..., m\ ;$$

$$T_{\max}^{(s)} = \max_{1\le i<k\le m} \{T_{ik}^{(s)}\}.$$

4. Вычисляется $p_y = \frac{1}{S}\sum_{s=1}^{S} \mathrm{I}\{T_{\max}^{(s)} = y\},\ y = 1, ..., N$.

Замечание 1. *Теоретическое значение получено только для* $p_0 = \left(2^{-m}(1+m)\right)^N$.

Таким образом, тест T_4 максимального скалярного произведения X_i и X_k имеет следующий вид.

1. Последовательность разбивается на Lm непересекающихся фрагментов согласно формуле (1).

2. По этим фрагментам $\{X^{(l)} : l = 1, ..., L\}$ согласно выражениям (2), (6) вычисляются статистики $\{Y_{ik}^{(l)} : i = 1, ..., m-1,\ k = i+1, ..., m\}$ и $Y_{\max}^{(l)}$.

3. По совокупности статистик $\{Y_{\max}^{(l)} : l = 1, ..., L\}$ вычисляется χ^2-статистика:

$$\chi^2 = \sum_{y=0}^{N} \frac{(v_y - Lp_y)^2}{Lp_y},\ v_y = \sum_{l=1}^{L} \mathrm{I}\{Y_{\max}^{(l)} = y\}.$$

4. Выносится решение с помощью статистического правила: если $P > \varepsilon$, то принимается гипотеза H_0, иначе принимается H_1, где $P = 1 - G_N(\chi^2)$.

При реализации тестов для вычисления P-значения функции χ^2-распределения с K степенями свободы применялась следующая формула, полученная с использованием [6]:

$$P_K(t) = \begin{cases} 2\left(1 - \Phi\left(\sqrt{t}\right) + \frac{1}{\sqrt{2\pi}} e^{-\frac{t}{2}} \sum_{i=1}^{(K-1)/2} \frac{t^{i-\frac{1}{2}}}{(2i-1)!!}\right), & \text{если } K - \text{нечетное}, \\ \left(1 + \sum_{i=1}^{(K-2)/2} \frac{t^i}{(2i)!!}\right) e^{-\frac{t}{2}}, & \text{если } K - \text{четное}, \end{cases}$$

где $\Phi(\cdot)$ – функция стандартного нормального распределения, $t > 0$.

Замечание 2. *Для вычисления* $\Phi(z)$ *можно воспользоваться формулой из* [6]*: если* $z > 0$, *то* $\Phi(z) = 1 - \left(2\pi e^{z^2}\right)^{-0.5}\left(b_1 r + b_2 r^2 + b_3 r^3 + b_4 r^4 + b_5 r^5\right) + \varepsilon(z)$, *где* $r = (1 + 0.2316419z)^{-1}$, $b_1 = 0.31938153$, $b_2 = -0.356563782$, $b_3 = 1.781477937$, $b_4 = -1.821255978$, $b_5 = 1.330274429$, $|\varepsilon(z)| < 7.5\cdot 10^{-8}$*; если* $z \le 0$, *то используется свойство* $\Phi(z) + \Phi(-z) = 1$.

5. Мощность статистических тестов

Была исследована мощность разработанных тестов и теста Т на основе экстремальной статистики скалярного произведения, предложенного в работе [3]. Мощность теста w есть ве-

роятность принятия альтернативы H_1 при условии, что она верна и характеризует точность различения альтернативы H_1 от гипотезы H_0. Для исследуемых тестов мощность определяется асимптотически при $n \to \infty$ следующим образом [5, 7]:

$$w = \mathrm{P}_{H_1}\left\{1 - G_K(\chi^2) \le \varepsilon\right\} = \mathrm{P}_{H_1}\left\{G_K(\chi^2) \ge 1 - \varepsilon\right\} = \mathrm{P}_{H_1}\left\{\chi^2 \ge G_K^{-1}(1-\varepsilon)\right\} \approx 1 - F\left(G_K^{-1}(1-\varepsilon)\right), \qquad (7)$$

где $F(\cdot)$ – функция нецентрального распределения χ^2 с K степенями свободы и параметром нецентральности $a = \sum_{y=0}^{K} \frac{L}{p_y}\left(p_{1y} - p_y\right)^2$; $\{p_{1y}\}$ – распределение статистики при справедливости альтернативы H_1.

Исследованы два случая альтернатив H_1, в первом из которых нарушается свойство C_2, а во втором – свойство C_1.

Случай 1. Альтернатива H_1: $\{x_t\}$ удовлетворяют свойству C_1, а свойство C_2 нарушено ($\mathrm{P}\{x_t = 1\} = b \ne 0{,}5$). Для тестов T_1, T_2 и T_3 определены $\{p_{1y}\}$:

– для теста T_1

$$p_{1y} = \mathrm{P}_{H_1}\left\{\overline{Y} = \frac{y}{m-1}\right\} = \sum_{a_1+\ldots+a_N=y} q_{a_1} \cdot \ldots \cdot q_{a_N},$$

$$q_{a_j} = b^{a_j+1}(1-b)^{m-(a_j+1)}\binom{m-1}{a_j} + (1-b)\delta_{0a_j},\ a_j \in \{0,1,\ldots,m-1\},\ j = 1,\ldots,N,\ y \in \{0,1,\ldots,(m-1)N\};$$

– для теста T_2

$$p_{1y} = \mathrm{P}_{H_1}\left\{\overline{Y} = \frac{2y}{m(m-1)}\right\} = \sum_{a_1+\ldots+a_N=y} q_{a_1} \cdot \ldots \cdot q_{a_N},$$

где $a_j \in \left\{0, \binom{2}{2}, \binom{3}{2}, \ldots, \binom{m}{2}\right\}$, $q_{a_j} = \binom{m}{d_j} b^{d_j}(1-b)^{d_j} + (1-b)^m \delta_{0a_j}$, $d_j = 0.5 + \sqrt{0.25 + 2a_j}$,
$j = 1,\ldots,N$, $y \in \{u_0, u_1, \ldots, u_K\}$;

– для теста T_3

$$p_{10} = \left(1-b^2\right)^{N\cdot[m/2]};$$

$$p_{1y} = \left(\sum_{j=0}^{N}\sum_{s=0}^{y}\binom{N}{j}\binom{j}{s} b^{j+s}(1-b)^{2N-(j+s)}\right)^{[m/2]} - \left(\sum_{j=0}^{N}\sum_{s=0}^{y-1}\binom{N}{j}\binom{j}{s} b^{j+s}(1-b)^{2N-(j+s)}\right)^{[m/2]},$$

$$y \in \{1,2,\ldots,N\}.$$

Таким образом, согласно формуле (7) для тестов T, T_1, T_2 и T_3 могут быть теоретически вычислены асимптотические значения мощности w; по методу Монте-Карло для всех исследуемых тестов определены экспериментальные значения мощности.

Полученные результаты исследований показали, что мощность увеличивается при увеличении длины последовательности n и при увеличении отклонения параметра b от 0,5, а уменьшается при увеличении значений параметров m, N. Зависимость экспериментальных значений мощности w от n при b =0,51, $m = 8$, $N = 4$, $\varepsilon = 0{,}1$ показана на рис. 1, а на рис. 2 представлены эти зависимости для T и T_3 вместе с теоретическими кривыми мощности. Влияние параметра b

на экспериментальные значения мощности в случае n=100 000, $m = 8$, $N = 4$, $\varepsilon = 0{,}1$ изображено на рис. 3.

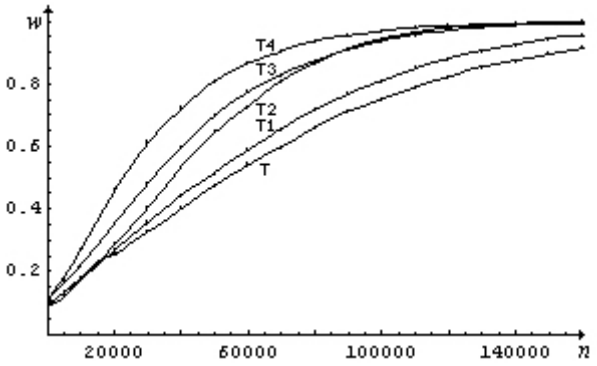

Рис. 1. График зависимости экспериментальных значений мощности w от n

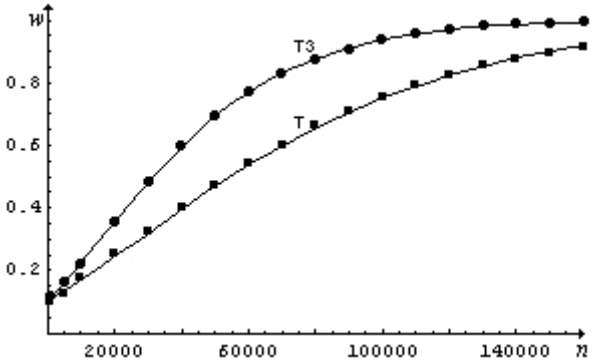

Рис. 2. График зависимости w от n (кривые – теоретическая мощность, кружочки – экспериментальные значения мощности для T_3, квадраты – экспериментальные значения w для T)

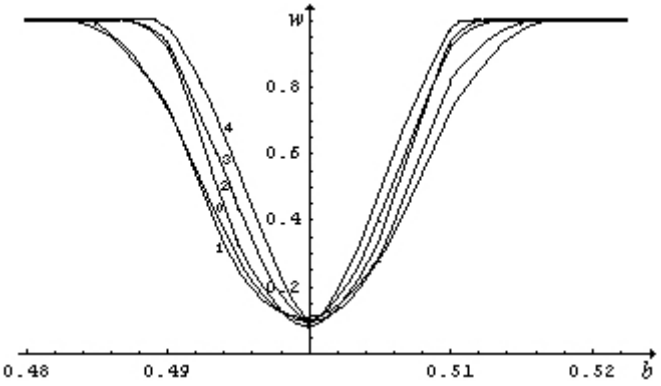

Рис. 3. График зависимости экспериментальных значений мощности w от параметра b (0 – T, 1 – T_1, 2 – T_2, 3 – T_3, 4 – T_4)

Случай 2. Альтернатива H_1: $\{x_t\}$ удовлетворяет свойству C_2, а свойство C_1 нарушено ($\{x_t\}$ – двоичная цепь Маркова с начальным распределением $\pi = (0.5 \;\; 0.5)^{\mathrm{T}}$ и матрицей вероятностей одношаговых переходов $\mathrm{P} = \begin{pmatrix} 0.5 + a & 0.5 - a \\ 0.5 - a & 0.5 + a \end{pmatrix}$, $a \neq 0$). Из-за существенной нелинейности статистик теоретические оценки мощности w в этом случае получить не удалось, но были получены экспериментальные значения мощности. Так же, как и в предыдущем случае, увеличение параметра n и отклонения параметра a от 0 увеличивают мощность (рис. 4). При увели-

чении параметров *m* и *N* мощность *w* уменьшается для всех тестов, за исключением T_3 и T_4, для которых происходит колебание значений *w* (рис. 5).

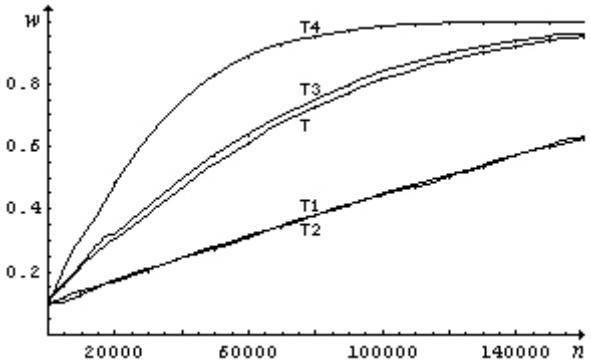

Рис. 4. График зависимости экспериментальных значений мощности *w* от *n* при *a*=0,02, *m*=8, *N*=4, $\varepsilon = 0{,}1$

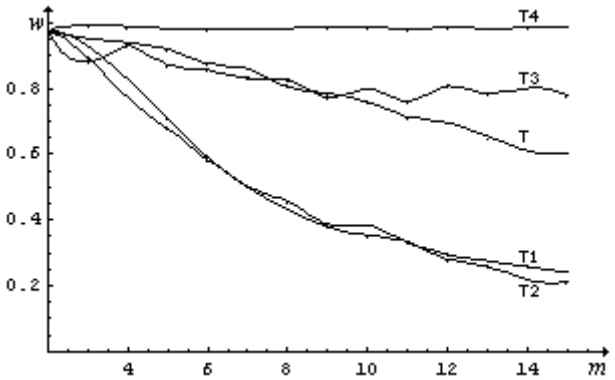

Рис. 5. График зависимости экспериментальных значений *w* от параметра *m* при *n*=100 000, *a*=0,02, *N*=4, $\varepsilon = 0{,}1$

Таким образом, наибольшую мощность для обоих семейств альтернатив среди исследуемых тестов имеет тест T_4. Тест T_3 близок по мощности к T_4; тесты T_1, T_2 предпочтительнее использовать для обнаружения нарушения свойства C_2.

6. Быстродействие алгоритмов тестирования

Вычислительный процесс каждого из исследуемых тестов можно представить в виде двух этапов: вспомогательного (вычисление значений $\{p_y\}$) и основного (вычисление $\{v_y\}$, χ^2-статистики и принятие решения).

На Athlon XP 1600+ был проведен анализ быстродействия алгоритмов [8] для каждого этапа. По результатам анализа получена следующая оценка затрат машинного времени на первом этапе:

для $T - t^{(1)}_{маш} \approx 1.45\cdot 10^{-8}\left(10N^3 + 58N^2 + 51N + 9\right)$ с;

$T_1 - t^{(1)}_{маш} \approx 1.14\cdot 10^{-7}\left((14m^2 - 13m)(N-1) + 62m + 26N - 62\right)$ с;

$T_2 - t^{(1)}_{маш} \approx 1.19\cdot 10^{-6}\left((23m^2 + 12m + 3.5N + 52)(N-1) + 72m - 30\right)$ с;

$T_3 - t^{(1)}_{маш} \approx 1.44\cdot 10^{-8}\left(10N^3 + 52N^2 + 63N + 13\right)$ с;

$T_4 - t^{(1)}_{маш} \approx 5.70\cdot 10^{-9}\left(2m^2N + 4mN + m^2 - m + 2\right)S$ с.

На первом этапе, который может быть выполнен до поступления самой последовательности $\{x_t\}$, тест T_4 проигрывает остальным тестам по времени выполнения из-за приближенного вычисления $\{p_y\}$ методом Монте-Карло.

На втором этапе оценка машинного времени следующая:

для $T - t_{маш}^{(2)} \approx 1.13 \cdot 10^{-9}\left(22N(m-1)+3m\right)L$ с;

$T_1 - t_{маш}^{(2)} \approx 1.13 \cdot 10^{-9}\left(22N(m-1)+3\right)L$ с;

$T_2 - t_{маш}^{(2)} \approx 1.11 \cdot 10^{-9}\left(14N(m^2-m)+3\right)L$ с;

$T_3 - t_{маш}^{(2)} \approx 7.96 \cdot 10^{-10}\left([m/2](35N+3)+3\right)L$ с;

$T_4 - t_{маш}^{(2)} \approx 5.07 \cdot 10^{-10}\left((m^2-m)(28N+3)+6\right)L$ с.

На первом и на втором этапах тесты имеют полиномиальную вычислительную сложность, причем преимущество в быстродействии имеет тест T_3. Так, например, при реализации этого теста для N=64, m=16, L=10 000, $t_{маш}^{(1)} = 0{,}06$ с, $t_{маш}^{(2)} = 0{,}15$ с.

Заключение

В статье построена серия тестов $T_1 - T_4$ для случайных и псевдослучайных последовательностей, основанных на сравнении фрагментов этих последовательностей. Проведен сравнительный анализ точности и быстродействия алгоритмов, реализующих данные тесты. Установлены условия предпочтительного применения разработанных тестов.

Список литературы

1. Кнут Д. Искусство программирования: В 3 т. – М.: Мир, 1992.
2. Luby M. Pseudorandomness and Cryptographic applications. – Princeton: Princeton University Press, 1996. – 234 p.
3. Математические и компьютерные основы криптологии / Ю.С. Харин, В.И. Берник, Г.В. Матвеев, С.В. Агиевич – Мн.: Новое знание, 2003. – 382 с.
4. Харин Ю.С. и др. Имитационное и статистическое моделирование. – Мн.: Изд-во БГУ, 1992.
5. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1983. – 416 с.
6. Вадзинский Р.Н. Справочник по вероятностным распределениям. – СПб.: Наука, 2001. – 294 с.
7. Ивченко Г.И., Медведев Ю.И. Математическая статистика. – М.: Высш. шк., 1984. – 248 с.
8. Котов В.М. Теория алгоритмов. – Мн.: Изд-во БГУ, 2001.

Поступила 14.09.04

Белорусский государственный университет,
Минск, пр. Ф. Скорины, 4
e-mail: kharin@bsu.by

Yu.S. Kharin, A.I. Piatlitski

THE STATISTICAL TESTING OF BINARY SEQUENCES BASED ON COMPARISON OF FRAGMENTS

Algorithms for testing random and pseudorandom sequences are proposed. Power and computer complexity of each algorithm are evaluated.