УДК 519.234:519.236

А.Л. КОСТЕВИЧ, А.В. ШИЛКИН, И.С. НИКИТИНА

О ПОДХОДЕ К ВЫЯВЛЕНИЮ МНОГОКРАТНОЙ РАЗЛАДКИ В ИСПЫТАНИЯХ БЕРНУЛЛИ С ИСПОЛЬЗОВАНИЕМ CTW-ПРЕДИКТОРА

An approach to statistical test construction for randomness testing of binary sequences on the base of universal predictors is considered. We derive the power of statistical tests constructed on the base of the maximum-likelihood predictors for the models of Bernoulli asymmetric trials and Markov chain. The results are extended to the Context Tree Weighting predictor, which is universal for these models. We show that usage of universal predictor in sliding window allows to detect effectively multiple model changes. Comparison of the proposed approach and traditional methods of multiple changes detection is conducted.

Рассмотрим задачу проверки гипотезы о том, что бинарная последовательность описывается моделью независимых симметричных испытаний Бернулли – гипотезы H_0 о чистой случайности – против альтернативной гипотезы о наличии многократной разладки: в некоторых фрагментах имеется отклонение от 0,5 вероятности успеха в испытаниях Бернулли (гипотеза $H_{1,0}$) или возникла марковская зависимость (гипотеза $H_{1,1}$). Задача проверки гипотезы о чистой случайности возникает в различных областях [1]: криптографии, имитационном моделировании и др. Альтернативная гипотеза о наличии многократной разладки используется для описания моделей наблюдения со «сбоями»: фрагменты регистрируемой последовательности могут описываться либо гипотезой H_0 о чистой случайности, либо одной из альтернативных гипотез $H_{1,i}$ о наличии i-го свойства.

Задаче выявления однократного изменения свойств наблюдаемой последовательности посвящено достаточно большое число публикаций, однако проблема выявления многократной разладки с набором альтернатив $\{H_{1,i}\}$ является более трудной [2]. Для выявления многократной разладки, как правило, рекомендуется разбить наблюдаемую последовательность на пересекающиеся фрагменты-«окна», в каждом из которых рекомендуется применить набор статистических критериев, предназна-ченных для выявления индивидуальной альтернативы $H_{1,i}$, и вынести итоговое решение с помощью процедуры множественной проверки гипотез.

Рассмотрим случай, когда все альтернативные гипотезы $\{H_{1,i}\}$ принадлежат одному классу, для которого существует универсальный предиктор, и предлагается подход к выявлению многократной разладки с использованием этого предиктора. Данный подход рассмотрим на примере универсального предиктора Context Tree Weighting [3] (в дальнейшем для краткости – CTW-предиктор) и названных альтернативных гипотез $H_{1,0}$ и $H_{1,1}$.

Опишем сначала способ применения универсального предиктора для построения статистического критерия, а затем предложим порядок использования CTW-предиктора для выявления многократной разладки.

1. Критерий проверки гипотез на базе универсальных предикторов. Приведем сначала предложенную в [4] схему построения статистического критерия проверки гипотезы H_0 против альтернативы H_1 общего вида (H_1 детализирована ниже) на базе любого универсального предиктора. Пусть наблюдается временной ряд $X_1^t = X_1, X_2, \ldots, X_t$, $X_i \in A = \{0,1\}$, который описывается набором условных вероятностей $\{P\{X_t \mid X_1^{t-1}\}\}$ из некоторого класса моделей M. Универсальный предиктор указывает способ построения статистических оценок неизвестных вероятностей $\{\hat{P}\{X_t \mid X_1^{t-1}\}\}$ и строит

прогноз значения следующего наблюдения согласно максимуму оценки условной вероятности. При этом *универсальность* [5] означает, что для любых условных вероятностей $\{P\{X_t \mid X_1^{t-1}\}\}$ из класса M

$$\min_{a\in A} P\{a \mid x_1^t\} - P\{\arg\min_{a\in A} \hat{P}\{a \mid x_1^t\} \mid x_1^t\} \xrightarrow{P} 0 \text{ при } t\to\infty.$$

Пусть регистрируется обучающая выборка $Z=(z_1,\ldots,z_m)$ объема m, по которой будет происходить построение оценок $\{\hat{P}\{X_t \mid X_1^{t-1}\}\}$ универсальным предиктором. Пусть независимо от Z наблюдается выборка $X=(x_1,\ldots,x_n)$ объема n, по которой для каждого $t=1,\ldots,n-1$ по первым t наблюдениям будем строить прогноз для x_{t+1} с использованием предиктора, затем вычислять индикатор успеха прогноза

$$Y_t = I\{\hat{X}_t = x_t\}, \quad t=1,\ldots,n.$$

Если для тестируемой выборки верна гипотеза H_0 о случайности, то последовательность индикаторов $\{Y_t\}$ также будет описываться гипотезой H_0 о чистой случайности. Если же для тестируемой выборки верна гипотеза H_1 (т. е. существуют $i_1,\ldots,i_{t-1},i_t \in A$, что $P\{i_t \mid i_1^{t-1}\} = 1/2 + \varepsilon_{i_1,\ldots,i_{t-1}}$ и $0 < |\varepsilon_{i_1,\ldots,i_{t-1}}| < 1/2$) и предиктор является универсальным, то последовательность индикаторов успеха прогноза будет иметь некоторое (зависящее от альтернативы и предиктора) совместное распределение, но со следующими маргинальными вероятностями:

$$H_1^{(Y)}:\ P\{Y_t = 1\} = \frac{1}{2} + \varepsilon, \text{ причем } \varepsilon = \varepsilon(m) > 0 \text{ при } m\to\infty. \tag{1}$$

В качестве статистического критерия для проверки H_0 против уже альтернативы $H_1^{(Y)}$ (и исходной альтернативы H_1 соответственно) в [4] предложен следующий статистический критерий:

$$\text{принять } \begin{cases} H_0, \text{ если } 2\sqrt{n}\,(S-\frac{1}{2}) < \Phi^{-1}(1-\alpha), \\ H_1, \text{ иначе,} \end{cases} \quad S = \frac{1}{n}\sum_{t=1}^{n} Y_t\ , \tag{2}$$

где $\Phi(\cdot)$ – функция распределения стандартного нормального закона, α – уровень значимости.

В [4] показано, что критерий (2) имеет заданный уровень значимости и в случае использования универсального для H_1 предиктора является состоятельным.

2. Предиктор CTW. Предиктор на основе контекстных деревьев CTW [3] является универсальным для класса C_D – множества источников ограниченной памяти древовидной структуры [3]. Далее рассмотрим случай $D = 1$. В классе C_1 содержится модель независимых испытаний Бернулли (альтернатива $H_{1,0}$) и модель цепи Маркова 1-го порядка (альтернатива $H_{1,1}$).

Построение прогноза с использованием CTW-предиктора в каждый момент времени $t=1,2,\ldots,n$ эквивалентно вычислению разности

$$\Lambda(X_1^t) = \hat{P}_{\mathrm{CTW}}\{1 \mid X_1^t\} - \hat{P}_{\mathrm{CTW}}\{0 \mid X_1^t\}, \tag{3}$$

где $\hat{P}_{\mathrm{CTW}}\{a \mid X_1^t\}$ – оценка условной вероятности, вычисленная предиктором CTW [3] (для удобства изложения предполагается, что имеется «предыстория» X_{1-D}^0). Если разность (3) положительна, то предиктор CTW делает прогноз «1», иначе – прогноз «0».

Рассмотрим случай $X_t=1$, тогда разность (3) имеет вид

$$\Lambda(X_1^t) = \frac{1}{2}P_e(n_0,n_1)\frac{n_1-n_0}{n_1+n_0+1} + \frac{1}{2}P_e(n_{00},n_{01})P_e(n_{10},n_{11})\frac{n_{11}-n_{10}}{n_{11}+n_{10}+1}, \tag{4}$$

где $n_i = \sum_{k=1}^{t} I\{X_k = i\}$, $n_{ij} = \sum_{k=1}^{t-1} I\{X_k = i, X_{k+1} = j\}$ – число символов i и биграмм (i, j) в последовательности X_1^t соответственно; $P_e(a,b) = \prod_{i=1}^{a}(i-0,5)/i \ \prod_{j=1}^{b}(j-0,5)/(a+j), \quad a,b>0$. Случай $X_t=0$ рассматривается аналогично.

3. Свойства предиктора CTW для гипотез $H_{1,0}$ и $H_{1,1}$. В случае альтернативы $H_{1,0}$ – последовательности независимых испытаний Бернулли с вероятностью успеха $p>0,5$ – знак разности (4) определяется только значением p и достаточно рассмотреть частоты n_0 и n_1. Тогда предиктор CTW асимптотически эквивалентен предиктору максимального правдоподобия для испытаний Бернулли:

$$\hat{X}_{t+1} = \begin{cases} 1, \text{ если } \hat{p} \ge 0,5, \\ 0, \text{ иначе,} \end{cases} \quad \hat{p} = \frac{1}{m}\sum_{t=1}^{m} I\{z_t = 1\}, \quad S_M = \sum_{t=1}^{n} I\{x_t = 1\}, \tag{5}$$

где S_M – достаточная статистика для альтернативы $H_{1,0}$. Отметим, что статистика S критерия (2) является функцией от статистики S_M.

Теорема 1. *При верной гипотезе $H_{1,0}$ в асимптотике m, n→∞ критерий* (2) *на базе предиктора* (5) *является состоятельным и его мощность имеет вид*

$$W_{m,n}^{(1,0)} \approx \left(1-\Phi\left(\frac{\sqrt{m}(0,5-p)}{\sqrt{p(1-p)}}\right)\right)\left(1-\Phi\left(\frac{\Phi^{-1}(1-\alpha)}{2\sqrt{p(1-p)}}-\frac{\sqrt{n}}{\sqrt{p(1-p)}}(p-0,5)\right)\right)+$$

$$+\Phi\left(\frac{\sqrt{m}(0,5-p)}{\sqrt{p(1-p)}}\right)\left(1-\Phi\left(\frac{\Phi^{-1}(1-\alpha)}{2\sqrt{p(1-p)}}+\frac{\sqrt{n}}{\sqrt{p(1-p)}}(p-0,5)\right)\right).$$

Доказательство следует из независимости выборок X и Z, а также асимптотической гауссовости оценки максимального правдоподобия вероятности успеха испытаний Бернулли по выборке объема t: $\sqrt{t}(\hat{p}-p)\sim N(0,p(1-p))$. ■

Рассмотрим альтернативу $H_{1,1}$ – последовательность описывается моделью цепи Маркова первого порядка с дважды стохастической матрицей вероятностей одношаговых переходов:

$$P=(p_{ij})=\begin{pmatrix}1-p & p\\ p & 1-p\end{pmatrix},\quad 0<p<1. \tag{6}$$

Так как стационарное распределение цепи Маркова (6) является равномерным, то знак разности (4) определяется только вторым слагаемым и использование предиктора CTW с параметром D=1 в случае $H_{1,1}$ эквивалентно использованию предиктора максимального правдоподобия для цепи Маркова:

$$\hat{X}_{t+1}=\hat{f}(x_t)=\arg\max_{j=0,1}\hat{p}_{x_t,j},\quad \hat{p}_{ij}=\frac{1}{m-1}\sum_{t=2}^{m}I\{z_{t-1}=i,z_t=j\},\quad S_Z=\sum_{t=2}^{n}I\{x_{t-1}\neq x_t\}, \tag{7}$$

где S_Z – достаточная статистика для альтернативы $H_{1,1}$ (6).

Из (7) можно видеть, что прогноз марковского предиктора определяется позицией максимума в строке матрицы $\hat{P}$. Следовательно, всего существует 4 варианта расположения максимумов и соответственно 4 класса предикторов:

$$M_1:\hat{p}_{00}\geq 0,5,\ \hat{p}_{10}<0,5;\quad M_2:\hat{p}_{00}<0,5,\ \hat{p}_{10}<0,5,$$

$$M_3:\hat{p}_{00}\geq 0,5,\ \hat{p}_{10}\geq 0,5;\quad M_4:\hat{p}_{00}<0,5,\ \hat{p}_{10}\geq 0,5.$$

Отметим [6], что в случае односторонних альтернатив p>0,5 (p<0,5) статистика S критерия (2) совпадет со статистикой числа знакоперемен S_Z (7).

Теорема 2. *При верной гипотезе $H_{1,1}$ в асимптотике m, n→∞ критерий* (2) *на базе предиктора* (7) *является состоятельным, мощность критерия имеет вид*

$$W_{m,n}^{(1,1)}=\sum_{k=1}^{4}P\{\hat{P}\in M_k\}W(\cdot\,|\,\hat{P}\in M_k),$$

$$P\{\hat{P}\in M_1\}\approx\left(1-\Phi(\frac{\sqrt{m}}{\sqrt{2p_{00}(1-p_{00})}})(\frac{1}{2}-p_{00})\right)\left(\Phi(\frac{\sqrt{m}}{\sqrt{2p_{10}(1-p_{10})}})(\frac{1}{2}-p_{10})\right),$$

$$P\{\hat{P}\in M_2\}\approx\left(\Phi(\frac{\sqrt{m}}{\sqrt{2p_{00}(1-p_{00})}})(\frac{1}{2}-p_{00})\right)\left(\Phi(\frac{\sqrt{m}}{\sqrt{2p_{10}(1-p_{10})}})(\frac{1}{2}-p_{10})\right),$$

$$P\{\hat{P}\in M_3\}\approx\left(1-\Phi(\frac{\sqrt{m}}{\sqrt{2p_{00}(1-p_{00})}})(\frac{1}{2}-p_{00})\right)\left(1-\Phi(\frac{\sqrt{m}}{\sqrt{2p_{10}(1-p_{10})}})(\frac{1}{2}-p_{10})\right),$$

$$P\{\hat{P}\in M_4\}\approx\left(\Phi(\frac{\sqrt{m}}{\sqrt{2p_{00}(1-p_{00})}})(\frac{1}{2}-p_{00})\right)\left(1-\Phi(\frac{\sqrt{m}}{\sqrt{2p_{10}(1-p_{10})}})(\frac{1}{2}-p_{10})\right),$$

$$W(\cdot\,|\,\hat{P}\in M_k)\approx 1-\Phi(\Phi^{-1}(1-\alpha)\frac{\sigma_0}{\sigma_1}-\frac{1}{\sigma_1}(\mu_1-\mu_0)),\quad \mu_0=0,5,\sigma_0^2=\frac{1}{4n},$$

$$\mu_1=\sum_{i=0,1}\pi_i p_{i,f^{(k)}(i)},\ \sigma_1^2=\frac{1}{n}(D\{Y_1\}+2\sum_{t=2}^{\infty}\mathrm{Cov}\{Y_1Y_t\})<\infty,$$

$$\mathrm{Cov}\{Y_1 Y_t\} = \sum_{ij\in\{0,1\}^2} \pi_i p_{i,f^{(k)}(i)} (p^{(t-1)}_{f^{(k)}(i),j} - \pi_j) p_{j,f^{(k)}(j)},$$

$$D\{Y_1\} = \frac{1}{n}\sum_{i=0,1} \pi_i p_{i,f^{(k)}(i)} (1 - \sum_{i=0,1} \pi_i p_{i,f^{(k)}(i)}),\ \ f^{(k)}(i) = \arg\max_{j=0,1} \hat{p}_{ij} \text{ при } \hat{P} = (\hat{p}_{ij}) \in M_k,$$

где через $p^{(t-1)}_{i,j}$ *обозначена вероятность перехода из состояния i в состояние j за t–1 шаг.*

Доказательство. Так как выборки X и Z предполагаются независимыми, предиктор (7) принадлежит к одному из классов M_k, $k = 1...4$, поэтому для вычисления мощности $W^{(1,1)}_{m,n}$ может быть применена теорема умножения. Условная мощность $W(\cdot \mid \hat{P} \in M_k)$ и вероятностные характеристики статистики S были найдены в [6]. Для нахождения условных вероятностей $P\{\hat{P} \in M_k\}$ используется свойство асимптотической гауссовости статистик $\hat{p}_{i0} : \sqrt{m}(\hat{p}_{i0} - p_{i0}) \sim N(0, 2p_{i0}(1-p_{i0}))$, $i = 0,1$, и их независимость $P(\hat{p}_{00} \le x, \hat{p}_{10} \le y) = P(\hat{p}_{00} \le x) P(\hat{p}_{10} \le y)$. ■

Замечание. Если верны гипотезы $H_{1,i}$, то в модели (1) для последовательности Y_t отклонение вероятности успеха прогноза $\varepsilon = \varepsilon^{(1,i)}(m) \xrightarrow{P} |p - 0,5|$ при $m \to \infty$.

4. Выявление многократной разладки. Предположим, что наблюдаемая выборка X состоит из неизвестного числа N фрагментов неизвестной длины, каждый из которых описывается одной из гипотез H_0, $H_{1,0}$ или $H_{1,1}$:

$$X = X_1^{m_1}, X_{m_1+1}^{m_2}, \ldots, X_{m_{N-1}+1}^{m_N},\ m_1 + \sum_{j=2}^{N} (m_j - m_{j-1}) = n.$$

4.1. Построение процедуры множественной проверки гипотез. Рассмотрим сначала традиционный подход к выявлению многократной разладки. Равномерно наиболее мощными критериями выявления гипотез $H_{1,0}$ и $H_{1,1}$ являются критерии знаков и знакоперемен соответственно. Критерий знаков основан на статистике S_M (5), критерий знакоперемен – на статистике S_Z (7).

Легко видеть, что если в выборке имеются симметричные отклонения от 0,5 вероятностей успеха и вероятностей одношаговых переходов, то применение критериев знаков и знакоперемен ко всей выборке не приведет к выявлению разладки. Поэтому будем применять каждый из критериев по принципу «скользящего окна»: по пересекающимся фрагментам длины w со сдвигом s.

Таким образом, процедура множественной проверки гипотез для обнаружения разладки состоит из $2K$ критериев: K критериев знаков и K критериев знакоперемен, где $K=[(n–w)/s]$ – число фрагментов, к которым применяется каждый из критериев. Гипотеза H_0 принимается процедурой только в том случае, если каждый критерий принял гипотезу H_0.

В теоремах 3 и 4 устанавливается совместное распределение статистик критериев процедуры при истинной нулевой гипотезе. Определим статистики критериев знаков и знакоперемен по k-му фрагменту длины w следующим образом:

$$S_{M,k} = \sum_{i=(k-1)s+1}^{(k-1)s+w} \mathbf{I}\{x_i = 1\},\ \ S_{Z,k} = \sum_{i=(k-1)s+2}^{(k-1)s+w} \mathbf{I}\{x_{i-1} \ne x_i\},\ \ k = \overline{1,K}.$$

Теорема 3. *При истинной гипотезе H_0 верны следующие утверждения:*

1) коэффициент корреляции между статистиками критериев знаков и знакоперемен для всех k и l равен 0: $\mathrm{Corr}\{S_{M,k}, S_{Z,l}\} = 0,\ \forall k,l \in \{1,...,K\}$;

2) коэффициент корреляции между статистиками критериев знаков для различных k и l имеет вид: $\rho_{M,kl} = \mathrm{Corr}\{S_{M,k}, S_{M,l}\} = d_{kl}/w$;

3) коэффициент корреляции между статистиками критериев знакоперемен для различных k и l равен

$$\rho_{Z,kl} = \mathrm{Corr}\{S_{Z,k}, S_{Z,l}\} = \begin{cases} \dfrac{d_{kl} - 1}{w - 1}, & \text{если } s\,|k-l| < w, \\ 0, & \text{иначе}, \end{cases}$$

где d_{kl} – длина пересечения k-го и l-го окна, $k,l \in \{1,...,K\}$.

Доказательство каждого утверждения теоремы проводится по определению коэффициента корреляции случайных величин X, Y с использованием представления статистик критерия знаков и знакоперемен как суммы индикаторных функций (5), (7) и свойства индикаторной функции от события A: $E\{I\{A\}\} = P\{A\}$. ■

Теорема 4. *При истинной гипотезе H_0 вектор нормированных статистик*

$$\left(\frac{2S_{M,1}-w}{\sqrt{w}},\ldots,\frac{2S_{M,m}-w}{\sqrt{w}},\frac{2S_{Z,1}-(w-1)}{\sqrt{w-1}},\ldots,\frac{2S_{Z,m}-(w-1)}{\sqrt{w-1}}\right)$$

имеет предельное в асимптотике $w\to\infty$ ($n=cw$) $2K$-мерное нормальное распределение с нулевым математическим ожиданием и ковариационной матрицей

$$\Sigma=\begin{vmatrix}\Sigma_M & 0_{K\times K}\\ 0_{K\times K} & \Sigma_Z\end{vmatrix},\ \Sigma_M=(\rho_{M,kl})_{k,l=1}^{K},\ \Sigma_Z=(\rho_{Z,kl})_{k,l=1}^{K}.$$

Доказательство. Статистики $\{S_{M,k}\}$, $(\{S_{Z,l}\})$ вычислены по различным фрагментам выборки и не являются линейно зависимыми, матрицы Σ_M и Σ_Z являются теплицевыми симметричными. Утверждение теоремы следует из многомерного случая центральной предельной теоремы и найденных в теореме 3 параметров совместного распределения статистик критериев.■

Процедурой множественной проверки гипотез также определяются уровни значимости критериев α_l таким образом, чтобы вероятность ошибки первого рода всей процедуры не превосходила заданного уровня α. Уровни значимости могут быть вычислены с использованием неравенства Бонферрони $\alpha_l=\alpha/(2K)$ [7].

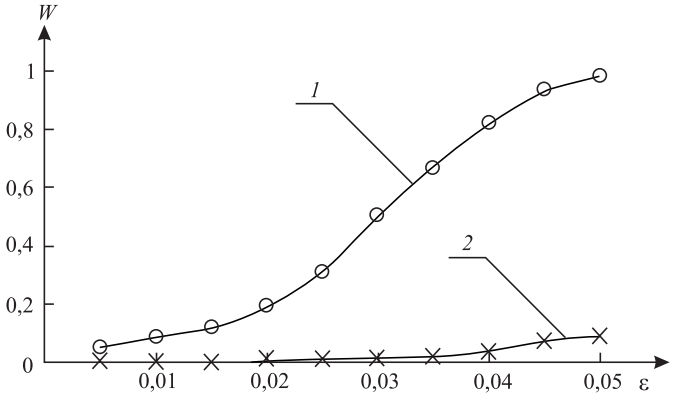
Для более точного нахождения α_l можно использовать неравенство Бонферрони второго порядка, учитывая параметры совместного распределения статистик критериев. В [7] предлагается процедура нахождения уточненного уровня значимости критериев α_l из уравнения $\alpha=G(\alpha_l)$, где $G(\cdot)$ – уточненная оценка сверху вероятности ошибки первого рода процедуры, зависящая от попарных корреляций, найденных в теоремах 3, 4. При анализе решения уравнения $\alpha=G(\alpha_l)$ для различных значений параметров w и s было установлено, что значительное увеличение мощности выявления разладки при использовании уточненного уровня значимости критериев достигается, когда длина «окна» w сравнима с длиной всей выборки, например $w=n/2$. Однако при таком большом значении w критерии не будут обнаруживать разладки малой длины. При меньших значениях обычно используемых на практике длин окна w применение уточненной процедуры Бонферрони не дает значимого выигрыша в мощности по сравнению с классической процедурой Бонферрони с $\alpha_l=\alpha/(2K)$.

Из построения процедуры множественной проверки гипотез можно видеть, что вероятность выявления разладки определяется мощностью критериев знаков и знакоперемен (в зависимости от типа произошедшей разладки), которые применяются по w наблюдениям с уровнем значимости $\alpha_l=\alpha/(2K)$. Таким образом, традиционная процедура множественной проверки гипотез обладает двумя недостатками: решение в пользу гипотез $H_{1,i}$ принимается по наблюдениям, число которых меньше объема выборки, в критериях используется уровень значимости α_l, близкий к 0.

4.2. Использование предиктора CTW. Как и в случае традиционного подхода, разобьем выборку X на пересекающиеся фрагменты-«окна» длины w со сдвигом s=1. Каждый такой фрагмент будем рассматривать в качестве обучающей выборки Z. С помощью предиктора будем прогнозировать следующий за «окном» символ, а затем сдвигать окно. К построенной последовательности индикаторов успехов прогнозов будем применять подход, описанный в п. 1–3. Вероятность успеха прогнозов определяется вероятностными характеристиками фрагмента выборки X, попавшей в «окно». Если в «окне» находится фрагмент длины w, который описывается гипотезой H_0, то согласно модели (1) отклонение вероятности успеха прогноза от 0,5 равно 0 ($\varepsilon=\varepsilon(w)=0$), если для фрагмента верна гипотеза $H_{1,i}$, то отклонение вероятности успеха прогноза от 0,5 больше 0 $\left(\varepsilon=\varepsilon^{(1,i)}(w)>0\right)$. Таким образом, если длина «окна» $w<\min\{m_1,m_2-m_1,\ldots,m_N-m_{N-1}\}$, то статистика S критерия (2) обладает математическим ожиданием $E\{S\}>0{,}5$ и конечной дисперсией, критерий (2) будет состоятельным и позволит установить факт многократной разладки.

В отличие от универсальных предикторов Лемпеля – Зива [8] и Sampled Pattern Matching предиктора [9] предиктор CTW может быть эффективно реализован в рамках описанного подхода. В частности, построение прогноза, обновление частот и пересчет (4) при D=1 в момент t выполняется за 35 операций типа сложение/вычитание, 14 операций логарифмирования, 4 операции умножения и 4 – возведения в степень. При этом критерий на базе предиктора CTW при D=1 в случае любой из односторонних альтернатив $H_{1,0}$ с $p=0{,}5+\varepsilon$ (будем обозначать $H_{1,0+}$), $H_{1,0}$ с $p=0{,}5-\varepsilon$ (или $H_{1,0-}$), $H_{1,1}$ с $p=0{,}5+\varepsilon$ (или $H_{1,1+}$), $H_{1,1}$ с $p=0{,}5-\varepsilon$ (или $H_{1,1-}$), где $\varepsilon>0$, будет основан на тех же статистиках, что и критерии знаков и знакоперемен, которые являются равномерно наиболее мощными для выявления гипотез $H_{1,0}$ и $H_{1,1}$ соответственно.

5. Результаты вычислительного эксперимента. Рассмотрим следующую модель многократной разладки: регистрируется выборка длиной 20 000 бит, выборка представляет собой последовательность N=8 фрагментов длиной 2500 бит, данная последовательность описывается гипотезами H_0, $H_{1,0+}$, $H_{1,1+}$, H_0, H_0, $H_{1,0-}$, $H_{1,1-}$, H_0 соответственно. На рисунке приведена оценка мощностей критерия (2) с использованием предиктора CTW в окне длины 200 и процедуры множественной проверки гипотез на базе критериев знаков и знакоперемен, построенная методом Монте-Карло по 1000 экспериментов для ε=0,005,…,0,05. Из рисунка видно, что использование критерия (2) на базе предиктора CTW в окне позволяет эффективно выявлять многократную разладку, в то время как традиционный подход является неэффективным.



Оценка мощности: *1* – предиктор CTW,
2 – критерий знаков и знакоперемен

1. NIST Special Publication 800-22: A statistical test suite for random and pseudorandom number generators for cryptographic applications. 2001.

2. Basseville M., Nikiforov I.V. Detection of Abrupt Changes: Theory and Application. Englewood Cliffs, 1993.

3. Willems F.M.J., Shtarkov Y.M., Tjalkens Tj.J. // IEEE Trans. on Inform. Theory. 1995. P. 653.

4. Kostevich A.L., Shilkin A.V. // Proceedings of the Eighth International Conference «Computer Data Analysis and Modeling: Complex Stochastic Data and Systems». Minsk, 2007. Vol. 1. P. 256.

5. Suzuki J. // Systems and Computers. 2003. Vol. 34 (6). P. 1.

6. Костевич А.Л., Шилкин А.В. // Проблемы теоретической и прикладной математики: Тр. 39-й Всерос. молодежной конф. Екатеринбург, 2008. С. 364.

7. Костевич А.Л., Милованова И.С. // Обозрение прикладной и промышленной математики. 2004. Т. 11. Вып. 2. С. 242.

8. Feder M., Merhav N., Gutman M. // IEEE Trans. on Inform. Theory. 1992. Vol. 38 (4). P. 1258.

9. Jacquet P., Szpankowski W., Apostol I. // IEEE Trans. on Inform. Theory. 2002. Vol. 48. P. 1462.

Поступила в редакцию 15.01.10.

Андрей Леонидович Костевич – кандидат физико-математических наук, ведущий научный сотрудник НИЛ математических методов защиты информации НИИППМИ.

Антон Владимирович Шилкин – младший научный сотрудник НИЛ математических методов защиты информации НИИППМИ.

Ирина Сергеевна Никитина – аспирант кафедры математического моделирования и анализа. Научный руководитель – А.Л. Костевич.