

УДК 519.2

**СТАТИСТИЧЕСКАЯ ПОСЛЕДОВАТЕЛЬНАЯ
ПРОВЕРКА ГИПОТЕЗ О ПАРАМЕТРАХ
ВЕРОЯТНОСТНЫХ МОДЕЛЕЙ
ДВОИЧНЫХ СЛУЧАЙНЫХ ДАННЫХ**

А. Ю. Харин

*Белорусский государственный университет, пр. Независимости 4,
220030, г. Минск, Беларусь, KharinAY@bsu.by*

Рассматривается актуальная вероятностно-статистическая задача компьютерного анализа данных – задача последовательной проверки простых гипотез о параметрах распределения вероятностей наблюдаемых бинарных данных. Она решается для двух вероятностных моделей наблюдений: схемы независимых испытаний и однородной цепи Маркова. Получены легко интерпретируемые и удобные для компьютерной реализации явные выражения статистик последовательных тестов (статистических критериев, решающих правил), разработан подход для вычисления их характеристик эффективности – вероятностей ошибочных решений и математических ожиданий случайного числа случайных наблюдений, необходимых для обеспечения требуемой точности. Получены асимптотические разложения для уклонений указанных характеристик эффективности при «засорениях» распределения вероятностей наблюдаемых данных.

Ключевые слова: случайные двоичные данные; статистический последовательный тест; вероятность ошибки; математическое ожидание случайного числа наблюдений; «засорения»; асимптотические разложения.

**STATISTICAL SEQUENTIAL HYPOTHESES TESTING
ON PARAMETERS OF PROBABILITY MODELS
OF RANDOM BINARY DATA**

A. Yu. Kharin

*Belarusian State University, 4 Independence av.,
Minsk 220030, Belarus, KharinAY@bsu.by*

An important probabilistic and statistical problem of computer data analysis – the problem of sequential testing of simple hypotheses on parameters of probability models of observed binary data – is considered in the paper. This problem is being solved for two probability models of observation: for independent observations and for homogeneous Markov chains. Explicit expressions of the sequential tests statistics are derived, valuable for interpretation and convenient for computer realization; an approach is developed to calculate the performance characteristics – error probabilities and mathematical expectations of the random number of random observations required to guarantee the requested accuracy for

decision rules. Asymptotic expansions for the deviations of the mentioned performance characteristics are constructed under “contamination” of the probability models of observed data.

Keywords: random binary data; statistical sequential test; error probability; mathematical expectation of the random number of observations; “contamination”; asymptotic expansions.

Введение

В современном мире хранимых данных становится всё больше, а роль их неуклонно растёт во многих процессах. Анализ данных становится весьма значительной компонентой современной жизни. Его эффективность во многих областях определяет успешность принятия решений во многих областях. Двоичные стохастические данные образуют важный класс по крайней мере по следующим причинам: 1) они наиболее естественны в настоящее время для компьютерной обработки; 2) двоичные данные описывают множество различных ситуаций в градациях наличия/отсутствия, положительного/отрицательного и тому подобных; 3) двоичные данные могут использоваться для описания достаточно богатых семейств распределений вероятностей при групповом рассмотрении. Классические методы статистического анализа часто неприменимы для анализа таких данных, поскольку не выполнены стандартные предположения, или их применение не демонстрирует приемлемую эффективность.

Поскольку детерминированный подход имеет весьма ограниченный потенциал описания происходящих процессов, реальные данные целесообразно полагать случайными, и для их анализа применять вероятностные модели. В рамках таких моделей часто встречается потребность различения двух типовых ситуаций относительно вероятностей распределений двоичных данных. Эти две типовые ситуации могут быть описаны в контексте статистической проверки двух простых гипотез [2].

Во многих ситуациях, в особенности в статистическом контроле качества, в системах автоматического оповещения, в персонализированной медицине, в принятии финансовых решений оказывается чрезвычайно важным использование минимального числа необходимых случайных наблюдений, гарантирующих требуемую точность [3]. Последовательные статистические тесты [1] следуют этому принципу в предположении, что число необходимых наблюдений определяется в самом процессе наблюдения, зависит от самих поступающих случайных наблюдений, и поэтому само является случайной величиной. Поскольку структура последовательного статистического решающего правила является достаточно сложной, обычно теоретическое исследование фактических значе-

ний его характеристик эффективности – вероятностей ошибочных решений и математических ожиданий случайного числа необходимых наблюдений – представляет собой достаточно сложную задачу, которую предпочитают не решать на практике [11].

Здесь представлен подход для вычисления и анализа характеристик эффективности последовательных тестов для двоичных данных. На практике фактические распределения вероятностей случайных данных часто отклоняются от гипотетических предположений о них – гипотетическое распределение вероятностей получается “искажённым” [4], [5], поэтому такую ситуацию мы также рассматриваем в работе, и исследуем отклонения характеристик эффективности при появлении таких искажений.

1. Модель независимых двоичных случайных наблюдений

Обозначим через $\mathbf{N}, \mathbf{Z}, \mathbf{R}$ соответственно множества натуральных, целых и действительных чисел; $\mathbf{Z}_+ = \mathbf{N} \cup \{0\}$. Пусть на вероятностном пространстве $(\Omega, \mathcal{F}, P)$ наблюдаются независимые одинаково распределённые K -мерные двоичные случайные векторы

$$x_t = (x_{ti}) \in U = \{u^1, \dots, u^{2^K}\} = \left\{ \begin{pmatrix} u_1 \\ \vdots \\ u_K \end{pmatrix}, u_i \in B = \{0, 1\}, i = 1, \dots, K \right\}, t \in \mathbf{N} \quad (1)$$

с независимыми компонентами. Распределение вероятностей случайных векторов (1) зависит от неизвестного истинного значения вектора параметров

$$\theta = \begin{pmatrix} p_1 \\ \vdots \\ p_K \end{pmatrix} \in \Theta = \{\theta_0, \theta_1\}, \text{ где } \theta_0 = \begin{pmatrix} p_1^0 \\ \vdots \\ p_K^0 \end{pmatrix}, \theta_1 = \begin{pmatrix} p_1^1 \\ \vdots \\ p_K^1 \end{pmatrix}, \theta_0 \neq \theta_1. \text{ Такая}$$

модель на практике часто применяется для принятия решений в пользу одной из двух типичных альтернативных ситуаций. Компоненты вектора параметров $p_i \in \{p_i^0, p_i^1\}$, $p_i^0, p_i^1 \in (0, 1)$, представляют собой вероятности случайных событий $\{x_{ti} = 1\}$, $t \in \mathbf{N}$, $i = 1, \dots, K$.

Пусть выполнены следующие предположения для распределения вероятностей случайных векторов (1):

$$P(u; \theta) = P_0\{x_t = u\} = a^{-J(u; \theta)}, t \in \mathbf{N}, u \in U, \quad (2)$$

где $a \in \mathbf{R}$, $a > 1$; функция $J(u; \theta): U \times \Theta \rightarrow \mathbf{Z}_+$ удовлетворяет следующему условию:

$$\sum_{u \in U} a^{-J(u; \theta)} = 1. \quad (3)$$

Имеются две простые гипотезы относительно значения вектора параметров θ распределения вероятностей (2):

$$H_0: \theta = \theta_0, \quad H_1: \theta = \theta_1. \quad (4)$$

Обозначим накопленную логарифмическую статистику отношения правдоподобия:

$$\Lambda_n = \Lambda_n(x_1, \dots, x_n) = \sum_{t=1}^n \lambda_t, \quad (5)$$

где

$$\lambda_t = \lambda(x_t) = \log_a (P(x_t; \theta_1) / P(x_t; \theta_0)) - \quad (6)$$

логарифм статистики отношения правдоподобия для вектора двоичных наблюдений x_t .

Теорема 1. Для модели (1) – (3) случайная последовательность значений статистики (5) в задаче проверки гипотез (4) образует однородную цепь Маркова с дискретным временем, и представима в виде:

$$\Lambda_n = \Lambda_n(x_1, \dots, x_n) = \sum_{k=1}^{2^K} n_k \cdot (J(u^k; \theta_0) - J(u^k; \theta_1)) \in \mathbf{Z}, \quad n \in \mathbf{N}, \quad (7)$$

где n_k означает число векторов, равных u^k , среди наблюдаемых n двоичных случайных векторов $\{x_1, \dots, x_n\}$, $\sum_{k=1}^{2^K} n_k = n$.

Доказательство. Марковское свойство [9] случайной последовательности (5) вытекает из независимости её приращений (6) в силу независимости случайных наблюдений (1). Выражение (7) получается эквивалентными преобразованиями (5), (6) при выполнении предположений (2), (3). ■

Следствие 1. Если в условиях теоремы 1 наблюдаются одномерные величины ($K=1$), то критериальная статистика (5) вычисляется следующим образом:

$$\Lambda_n = \Lambda_n(x_1, \dots, x_n) = n_0(J(0; p^0) - J(0; p^1)) + (n - n_0)(J(1; p^0) - J(1; p^1)), \quad n \in \mathbf{N},$$

где n_0 означает число наблюдений, равных 0.

С использованием статистики (7) последовательный критерий отношения вероятностей (ПКОВ, SPRT) [1] для гипотез (4) строится сле-

дующим образом: решение, принимаемое после n полученных наблюдений ($n = 1, 2, \dots$), равно

$$d = d(n) = \mathbf{1}_{[C_+, +\infty)}(\Lambda_n) + 2 \cdot \mathbf{1}_{(C_-, C_+)}(\Lambda_n), \quad (8)$$

где $\mathbf{1}_D(\cdot)$ означает индикаторную функцию множества D . Решения $d = 0$ и $d = 1$ означают остановку процесса наблюдения и принятие гипотезы H_0 или H_1 соответственно после наблюдения n двоичных случайных векторов; $d = 2$ означает, что требуется наблюдать $(n+1)$ -й вектор; $C_-, C_+ \in \mathbf{Z}$, $C_- < C_+$ являются параметрами решающего правила (8) и называются его порогами; на практике их принято вычислять в соответствии с [1]:

$$C_- = \lceil \log_a(\beta_0 / (1 - \alpha_0)) \rceil, \quad C_+ = \lceil \log_a((1 - \beta_0) / \alpha_0) \rceil, \quad (9)$$

где α_0, β_0 – допустимые значения вероятностей ошибок I рода (принять гипотезу H_1 , когда на самом деле верна H_0) и II рода (верна гипотеза H_1 , а принята H_0); $\lceil \cdot \rceil$ означает взятие целой части аргумента по математическим правилам. При использовании порогов (9) фактические значения вероятностей ошибок I и II рода могут существенно отличаться от значений α_0, β_0 , и здесь является актуальной задача вычисления фактических значений характеристик эффективности (включая вероятности ошибочных решений) для последовательных тестов.

Примем обозначения: $\delta_{i,j}$ – дельта-символ Кронекера; $\mathbf{I}_k$ – единичная матрица размерности k ; $\mathbf{0}_{m \times n}$ – матрица размера $(m \times n)$, все элементы которой равны 0; $\mathbf{1}(\cdot)$ – функция Хэвисайда (единичного скачка); $\mathbf{1}_k$ – k -вектор (столбец), все компоненты которого равны 1. Через $t^{(k)}$ обозначим математическое ожидание случайного числа наблюдений (случайного объёма выборки), когда справедлива гипотеза H_k , $k \in \{0, 1\}$, через α, β – фактические значения вероятностей ошибок I и II рода для критерия (8);

$N = C_+ - C_-$; пусть $P^{(k)} = \begin{pmatrix} \mathbf{I}_2 & \vdots & \mathbf{0}_{2 \times N} \\ \text{---} & \vdots & \text{---} \\ R^{(k)} & \vdots & Q^{(k)} \end{pmatrix}$ означает матрицу размерности $(N+2) \times (N+2)$, блоки $R^{(k)}, Q^{(k)}$ которой определяются поэлементно:

$$p_{ij}^{(k)} = \begin{cases} \sum_{u \in U} \delta_{J(u; \theta_0) - J(u; \theta_1), j-i} P(u; \theta_k), & i, j \in (C_-, C_+) \\ \sum_{u \in U} \mathbf{1}(C_- - i + J(u; \theta_1) - J(u; \theta_0)) P(u; \theta_k), & i \in (C_-, C_+), j = C_-, \\ \sum_{u \in U} \mathbf{1}(J(u; \theta_0) - J(u; \theta_1) + i - C_+) P(u; \theta_k), & i \in (C_-, C_+), j = C_+. \end{cases}$$

Обозначим ещё:

$$\pi^{(k)} = (\pi_i^{(k)}), \pi_i^{(k)} = \sum_{u \in U} \delta_{J(u; \theta_0) - J(u; \theta_1), i} P(u; \theta_k), i \in \{C_- + 1, \dots, C_+ - 1\};$$

$$\pi_{C_+}^{(k)} = \sum_{i \geq C_+} \sum_{u \in U} \delta_{J(u; \theta_0) - J(u; \theta_1), i} P(u; \theta_k), \pi_{C_-}^{(k)} = \sum_{i \leq C_-} \sum_{u \in U} \delta_{J(u; \theta_0) - J(u; \theta_1), i} P(u; \theta_k);$$

$S^{(k)} = I_N - Q^{(k)}, B^{(k)} = (S^{(k)})^{-1} R^{(k)}$; пусть $W_{(i)}$ означает i -й столбец матрицы W .

Теорема 2. Если в модели (1) – (6) $|S^{(k)}| \neq 0, k \in \{0, 1\}$, то характеристики эффективности последовательного решающего правила (8) вычисляются в явном виде:

$$t^{(k)} = (\pi^{(k)})' (S^{(k)})^{-1} \mathbf{1}_N + 1, \alpha = (\pi^{(0)})' B_{(2)}^{(0)} + \pi_{C_+}^{(0)}, \beta = (\pi^{(1)})' B_{(1)}^{(1)} + \pi_{C_-}^{(1)}.$$

Доказательство основано на теории конечных однородных цепей Маркова с дискретным временем. Случайная последовательность

$$\zeta_n = C_- \cdot \mathbf{1}_{(-\infty, C_-]}(\Lambda_n) + C_+ \cdot \mathbf{1}_{[C_+, +\infty)}(\Lambda_n) + \Lambda_n \cdot \mathbf{1}_{(C_-, C_+)}(\Lambda_n)$$

является однородной цепью Маркова с $N+2$ состояниями, при этом состояния C_- , C_+ являются поглощающими. ■

Ситуация, когда распределение вероятностей случайных наблюдений “засорено”, рассмотрено в следующем разделе для более общей модели, в которой наблюдаемые двоичные случайные данные образуют цепь Маркова.

2. Модель многомерных случайных наблюдений, образующих однородную цепь Маркова

Пусть наблюдаются случайные векторы $x_1, x_2, \dots$ с двоичными компонентами, эти векторы принимают значения на множестве

$$U = \{u^1, \dots, u^{2^K}\} = \left\{ \begin{pmatrix} u_1 \\ \vdots \\ u_K \end{pmatrix}, u_i \in B = \{0, 1\}, i = 1, \dots, K \right\}. \text{ Для упрощения обозначений введём в рассмотрение множество } V = \{0, 1, \dots, M-1\}, M = 2^K, \text{ биективно соответствующее множеству } U. \text{ Для наблюдаемой цепи Маркова обозначим вектор вероятностей начальных состояний } \pi = (\pi_i), i \in V, \text{ и матрицу вероятностей одношаговых переходов } P = (p_{ij}), i, j \in V:$$

ний введём в рассмотрение множество $V = \{0, 1, \dots, M-1\}, M = 2^K$, биективно соответствующее множеству U . Для наблюдаемой цепи Маркова обозначим вектор вероятностей начальных состояний $\pi = (\pi_i), i \in V$, и матрицу вероятностей одношаговых переходов $P = (p_{ij}), i, j \in V$:

$$P\{x_1 = i\} = \pi_i, \quad P\{x_n = j \mid x_{n-1} = i\} = p_{ij}, \quad i, j \in V.$$

Как и для модели независимых случайных векторов с двоичными компонентами, будем рассматривать две простые альтернативные гипотезы относительно значений параметров цепи Маркова:

$$H_0: \pi = \pi^{(0)}, P = P^{(0)}; \quad H_1: \pi = \pi^{(1)}, P = P^{(1)}, \quad (10)$$

где $\pi^{(0)} = (\pi_i^{(0)})$, $\pi^{(1)} = (\pi_i^{(1)})$ – два соответствующих значения вектора вероятностей начальных состояний, $P^{(0)} = (p_{ij}^{(0)}) \neq P^{(1)} = (p_{ij}^{(1)})$ – заданные матрицы вероятностей одношаговых переходов при справедливости соответствующих гипотез.

Для построения последовательного статистического теста проверки гипотез (10) в рассматриваемой модели наблюдений обозначим:

$$\lambda_1 = \log \frac{\pi_{x_1}^{(1)}}{\pi_{x_1}^{(0)}}, \quad \lambda_k = \log \frac{p_{x_{k-1}, x_k}^{(1)}}{p_{x_{k-1}, x_k}^{(0)}}, \quad k > 1; \quad \Lambda_n = \sum_{k=1}^n \lambda_k, \quad n \in \mathbf{N}. \quad (11)$$

Последовательный статистический тест для рассмотренной модели и гипотез (10) строится в соответствии с решающим правилом (8) при замене (5) – (7) на (11). В соответствии с таким статистическим последовательным тестом, для заданных значений порогов (например, согласно (9)) $C_-, C_+ \in \mathbf{R}$, $C_- < 0$, $C_+ > 0$, принимается гипотеза H_0 после n наблюдений, если $\Lambda_n \leq C_-$, гипотеза H_1 принимается, когда $\Lambda_n \geq C_+$, в обоих этих случаях процесс наблюдения останавливается и тест завершается, иначе тест продолжается и требуется наблюдать $(n+1)$ -й случайный вектор с двоичными компонентами. Случайная последовательность $(K+1)$ -мерных случайных векторов $(\Lambda_n, x_n)'$, $n \in \mathbf{N}$, является цепью Маркова по определению:

$$P\{\Lambda_n, x_n \mid \Lambda_{n-1}, \Lambda_{n-2}, \dots, \Lambda_1, x_{n-1}, x_{n-2}, \dots, x_1\} = P\{\Lambda_n, x_n \mid \Lambda_{n-1}, x_{n-1}\}.$$

Пусть дополнительно $\pi^{(0)}, P^{(0)}$, $\pi^{(1)}, P^{(1)}$ удовлетворяют следующим предположениям:

$$\exists a \in \mathbf{R}, \quad m_i, m_{ij} \in \mathbf{Z}, \quad i, j \in V: \quad \log \frac{\pi_i^{(1)}}{\pi_i^{(0)}} = m_i a, \quad \log \frac{p_{ij}^{(1)}}{p_{ij}^{(0)}} = m_{ij} a. \quad (12)$$

Без потери общности в такой ситуации предположим, что для статистического теста (5), (8), (11) пороги $C_-, C_+ \in \mathbf{Z}$, и обозначим: $t^{(k)}$ – математическое ожидание случайного числа наблюдений до принятия одной из гипотез, при условии справедливости H_k , $k \in \{0,1\}$;

$$W^{(k)} = \left(w_{ij}^{(k)} \right) = \begin{pmatrix} \mathbf{I}_2 & \vdots & \mathbf{0}_{2 \times MN} \\ \text{---} & \vdots & \text{---} \\ R^{(k)} & \vdots & Q^{(k)} \end{pmatrix} -$$

матрица размерности $(MN+2)(MN+2)$ с блоками $R^{(k)}, Q^{(k)}$, заданными поэлементно ($s, t \in V$) следующим образом:

$$\begin{aligned} w_{Mi+s, Mj+t}^{(k)} &= \delta_{m_{st}, j-i} p_{st}^{(k)}, i, j \in (C_-, C_+), \\ w_{Mi+s, Mj+t}^{(k)} &= \begin{cases} \sum_{t \in V} \mathbf{1}(i + m_{st} - C_+), j = C_+, i \in (C_-, C_+), \\ \sum_{t \in V} \mathbf{1}(C_- - i - m_{st}), j = C_-, i \in (C_-, C_+); \end{cases} \end{aligned} \quad (13)$$

как и в случае независимых наблюдаемых векторов с двоичными компонентами, обозначим матрицы

$$S^{(k)} = \mathbf{I}_{MN} - Q^{(k)}, B^{(k)} = \left(S^{(k)} \right)^{-1} R^{(k)}, k \in \{0,1\};$$

векторы

$$\omega^{(k)} = \left(\omega_i^{(k)} \right), i \in \{MC_- + 1, \dots, MC_+ - 1\}: \omega_{Mi+s}^{(k)} = \delta_{m_s, i} \pi_s^{(k)}, i \in (C_-, C_+), \quad (14)$$

и две априорные вероятности поглощения:

$$\omega_{MC_-}^{(k)} = \sum_{s \in V} \mathbf{1}(C_- - m_s) \pi_s^{(k)}, \omega_{MC_+}^{(k)} = \sum_{s \in V} \mathbf{1}(m_s - C_+) \pi_s^{(k)}. \quad (15)$$

Теорема 3. Для рассмотренной выше модели случайных наблюдений, образующих однородную векторную цепь Маркова с двоичными компонентами, если $|S^{(k)}| \neq 0$, $k \in \{0,1\}$, и выполнено (12), то характеристики эффективности последовательного теста (5), (8), (11) вычисляются в явном виде:

$$t^{(k)} = \left(\omega^{(k)} \right)' \left(S^{(k)} \right)^{-1} \mathbf{1}_{MN} + 1, \alpha = \left(\omega^{(0)} \right)' B_{(2)}^{(0)} + \omega_{MC_+}^{(0)}, \beta = \left(\omega^{(1)} \right)' B_{(1)}^{(1)} + \omega_{MC_-}^{(1)}.$$

Доказательство. Введём в рассмотрение случайную последовательность

$$\xi_n = MC_- \cdot \mathbf{1}_{(-\infty, C_-]} \left(\frac{\Lambda_n}{a} \right) + MC_+ \cdot \mathbf{1}_{(C_+, +\infty)} \left(\frac{\Lambda_n}{a} \right) + \left(\frac{\Lambda_n}{a} M + \chi_n \right) \cdot \mathbf{1}_{(C_-, C_+)} \left(\frac{\Lambda_n}{a} \right),$$

$n \in \mathbf{N}$, которая является однородной цепью Маркова с $MN+2$ состояниями; из которых два ($\xi_n = MC_-$ и $\xi_n = MC_+$) являются поглощающими. Матрица вероятностей одношаговых переходов определяется (13), вектор начальных вероятностей состояний, не являющихся поглощающими, задан (14), а начальные вероятности поглощающих состояний вычисляются согласно (15). ■

Теперь рассмотрим часто встречающуюся на практике ситуацию, когда описанная выше гипотетическая модель “засорена” в смысле распределения вероятностей в силу “засорения” вектора вероятностей начальных состояний и матрицы вероятностей одношаговых переходов. Пусть вместо гипотетических значений, фактические (искажённые) вектор вероятностей начальных состояний и матрица вероятностей одношаговых переходов имеют вид:

$$\bar{\pi}^{(k)} = (1 - \varepsilon)\pi^{(k)} + \varepsilon\tilde{\pi}^{(k)}, \quad \bar{P}^{(k)} = (1 - \varepsilon)P^{(k)} + \varepsilon\tilde{P}^{(k)}, \quad k = 0, 1, \quad (16)$$

где $\tilde{\pi}^{(k)}$ и $\tilde{P}^{(k)}$ – вектор вероятностей начальных состояний и матрица вероятностей одношаговых переходов для “засоряющей” цепи Маркова, $P^{(k)} \neq \tilde{P}^{(k)}$, $k = 0, 1$, и $\varepsilon \in [0, \frac{1}{2}]$ – вероятность “засорения” (её также принято называть уровнем искажения).

Для “засорённой” модели (16) обозначим аналогично гипотетическому случаю: $\tilde{W}^{(k)}$, $\tilde{Q}^{(k)}$, $\tilde{R}^{(k)}$, $\tilde{\omega}^{(k)}$, $\tilde{\omega}_{MC\pm}^{(k)}$, $k = 0, 1$, с заменой гипотетического распределения вероятностей на “засоряющее” one. Обозначим также: $\tilde{S}^{(k)} = \mathbf{I}_{MN} - Q^{(k)} - \varepsilon \cdot (\tilde{Q}^{(k)} - Q^{(k)})$.

Теорема 4. Если гипотетическая модель наблюдений, образующих однородную векторную цепь Маркова с двоичными компонентами, искажена в соответствии с (16), а также выполнено условие, аналогичное (12), для искажённой модели, то фактические вероятности ошибок I и II рода $\bar{\alpha}$, $\bar{\beta}$, и условные математические ожидания случайного числа наблюдений $\bar{t}^{(k)}$, $k = 0, 1$, отклоняются от гипотетических значений соответствующих характеристик эффективности на величины порядка $O(\varepsilon)$:

$$\begin{aligned}
\bar{\alpha} - \alpha &= \varepsilon((\omega^{(0)})'((S^{(0)})^{-1}((\tilde{Q}^{(0)} - Q^{(0)})(S^{(0)})^{-1}R^{(0)} + \tilde{R}^{(0)} - R^{(0)}))_{(2)} + \\
&+ (\tilde{\omega}^{(0)} - \omega^{(0)})'B_{(2)}^{(0)} + \tilde{\omega}_{MC+}^{(0)} - \omega_{MC+}^{(0)}) + O(\varepsilon^2), \\
\bar{\beta} - \beta &= \varepsilon((\omega^{(1)})'((S^{(1)})^{-1}((\tilde{Q}^{(1)} - Q^{(1)})(S^{(1)})^{-1}R^{(1)} + \tilde{R}^{(1)} - R^{(1)}))_{(1)} + \\
&+ (\tilde{\omega}^{(1)} - \omega^{(1)})'B_{(1)}^{(1)} + \tilde{\omega}_{MC-}^{(1)} - \omega_{MC-}^{(1)}) + O(\varepsilon^2); \\
\bar{t}^{(k)} - t^{(k)} &= \varepsilon((\tilde{\omega}^{(k)} - \omega^{(k)})' + (\omega^{(k)})'(S^{(k)})^{-1}(\tilde{Q}^{(k)} - Q^{(k)}))(S^{(k)})^{-1}\mathbf{1}_{MN} + O(\varepsilon^2).
\end{aligned}$$

Доказательство. При наличии искажений (16) вектор вероятностей начальных состояний и матрица вероятностей одношаговых переходов случайной последовательности ξ_n имеют форму соответствующих смесей; оставшаяся часть доказательства состоит в выполнении тождественных преобразований. ■

Заключение

Таким образом, в работе рассмотрен подход для вычисления и анализа характеристик эффективности последовательных статистических тестов для двоичных данных в рамках двух моделей наблюдений: для независимых случайных векторов с двоичными компонентами и для однородных векторных цепей Маркова с двоичными компонентами. Отдельно рассмотрена ситуация, когда фактическая модель случайных данных отклоняется от гипотетической модели, и проведён асимптотический анализ разностей фактических и гипотетических значений характеристик эффективности по уровню искажения. Результаты могут применяться для построения робастных (устойчивых к искажениям модели) последовательных статистических тестов для двоичных случайных данных [6], а также для модели случайных последовательностей с трендом [7]. Результаты применимы для ситуации, когда гипотез больше, чем две [10], для проверки сложных гипотез [8] и для анализа усечённых последовательных статистических тестов [12].

Исследования частично поддержаны грантом Белорусского республиканского фонда фундаментальных исследований № Ф23УЗБ-080.

Библиографические ссылки

4. Wald A. Sequential analysis. New York : John Wiley and Sons, 1947.
5. Mukhopadhyay N., B. de Silva Sequential methods and their applications. New York : Marcel Dekker, 2009.
6. Lai T. Sequential analysis: Some classical problems and new challenges. Statistica Sinica, 2001. Vol. 11. P. 303–408.

7. *Huber P, Ronchetti E.* Robust Statistics. New York : Wiley, 2009.
8. *Maevskii V.V., KharinYu.S.* Robust regressive forecasting under functional distortions in a model. Automation and RemoteControl. 2002. 63(11). P. 1803–1820.
9. *Харин А.Ю.* Робастность байесовских и последовательных статистических решающих правил. Минск: БГУ; 2013. 207 с.
10. *Kharin A., Tu T.T.* Performance and robustness analysis of sequential hypotheses testing for time series with trend. Austrian Journal of Statistics. 2017. 46(3-4). P. 23-36.
11. *Kharin A.Y.* An approach to asymptotic robustness analysis of sequential tests for composite parametric hypotheses. Journal of Mathematical Sciences. 2017. 227(2). P. 196–203.
12. *Kemeny J.G., Snell J.L.* Finite Markov Chains. New York : Springer, 1960.
13. *Tu T.T., Kharin A.Y.* Sequential probability ratio test for many simple hypotheses on parameters of time series with trend. Журнал Белорусского государственного университета. Математика. Информатика. 2019(1). P. 35–45.
14. *Айвазян С.А.* Сравнение оптимальных свойств критериев Неймана–Пирсона и Вальда // Теория вероятностей и ее применения. 1959. № 4 (1). С. 86–93.
15. *Kharin A.Y., Tu T.T.* On error probability calculation for the truncated sequential probability ratio test Журнал Белорусского государственного университета. Математика. Информатика. 2018(1). С. 68–76.