ISSN 1561-8323 (Print)
ISSN 2524-2431 (Online)

## *МАТЕМАТИКА*

### *MATHEMATICS*

УДК 519.2
https://doi.org/10.29235/1561-8323-2024-68-4-271-281

Поступило в редакцию 23.04.2024
Received 23.04.2024

**Академик Ю. С. Харин[1], С. А. Шибалко[2]**

*[1]Научно-исследовательский институт прикладных проблем математики и информатики Белорусского государственного университета, Минск, Республика Беларусь*
*[2]Белорусский государственный университет, Минск, Республика Беларусь*

# СТАТИСТИЧЕСКИЙ АНАЛИЗ МНОГОМЕРНЫХ ДВОИЧНЫХ ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ НЕЙРОСЕТЕВОЙ МОДЕЛИ

**Аннотация.** Рассматривается задача статистического анализа $N$-мерных двоичных временных рядов. Предлагается подход к решению этой задачи на основе малопараметрической нейросетевой модели эргодической цепи Маркова порядка $s$. Построены состоятельные статистические оценки параметров модели и алгоритмы компьютерного анализа данных с использованием нейросетевой модели: алгоритм оценивания параметров и алгоритм прогнозирования. Приведены результаты компьютерных экспериментов на модельных и реальных данных.

**Ключевые слова:** многомерный двоичный временной ряд, цепь Маркова порядка $s$, нейросетевая модель, статистическое оценивание параметров, статистическое прогнозирование

**Для цитирования.** Харин, Ю. С. Статистический анализ многомерных двоичных временных рядов на основе нейросетевой модели / Ю. С. Харин, С. А. Шибалко // Докл. Нац. акад. наук Беларуси. – 2024. – Т. 68, № 4. – С. 271–281. https://doi.org/10.29235/1561-8323-2024-68-4-271-281

**Academician Yuriy S. Kharin[1], Sierhei A. Shibalko[2]**

*[1]Research Institute for Applied Problems of Mathematics and Informatics of the Belarusian State University, Minsk, Republic of Belarus*
*[2]Belarusian State University, Minsk, Republic of Belarus*

# STATISTICAL ANALYSIS OF MULTIVARIATE BINARY TIME SERIES BASED ON A NEURAL NETWORK MODEL

**Abstract.** This article is devoted to the statistical analysis of multivariate binary time series. For solving this problem a parsimonious neural network model of Markov's ergodic chain of order $s$ was determined. Consistent statistical estimators for model parameters and estimation algorithms of parameters and forecasting algorithms of future states of time series were developed. The results of computer experiments on simulated and real data are presented.

**Keywords:** multivariate binary time series, Markov chain of order $s$, neural network-based model, statistical estimation, statistical forecasting

**For citation.** Kharin Yu. S., Shibalko S. A. Statistical analysis of multivariate binary time series based on a neural network model. *Doklady Natsional'noi akademii nauk Belarusi = Doklady of the National Academy of Sciences of Belarus*, 2024, vol. 68, no. 4, pp. 271–281 (in Russian). https://doi.org/10.29235/1561-8323-2024-68-4-271-281

**Введение.** При исследовании и прогнозировании динамики сложных процессов и явлений часто возникает проблема статистического анализа временных рядов, т. е. упорядоченных по времени последовательностей наблюдений $\{x_t \in X\colon t = 1, 2, \dots\}$ за некоторым случайным процессом. Теория статистического анализа временных рядов [1] глубоко развита для «непрерывных» моделей, когда множество наблюдаемых значений $X = \mathbb{R}$ – множество действительных чисел. Цифровизация общества делает необходимым рассмотрение более адекватных «дискретных» моделей, когда $X$ – дискретное множество.

Один из наиболее развитых подходов к вероятностному моделированию дискретных временных рядов основан на обобщенной линейной модели регрессии GLM (Generalized Linear Model) [2], являющейся некоторым развитием классической линейной регрессии. В [3] приведена модель авторегрессии – скользящего среднего ARMA для дискретных временных рядов. Также для анализа дискретных временных рядов применяются скрытые цепи Маркова HMM (Hidden Markov Model) [4], а также их разновидности MSDR (Markov Switching Dynamic Regression) [5] и MSAR (Markov Switching AutoRegression) [6].

Обзор современного состояния в области статистического анализа многомерных дискретных временных рядов представлен в [7]. В настоящем сообщении подход, предложенный в [8], развивается для многомерных двоичных рядов, которые являются адекватной вероятностной моделью данных в генетике, защите информации, медицине, экономике и других приложениях.

**Вероятностная модель многомерных двоичных временных рядов.** Примем обозначения: $\mathbb{Z}$ – множество целых чисел, $\mathbb{R}^k$ – k-мерное евклидово пространство, $V = \{0, 1\}$ – двоичный алфавит, $\mathbb{N}$ – множество натуральных чисел, штрих у матрицы – символ транспонирования.

Определим на вероятностном пространстве $(\Omega, \mathcal{F}, \mathrm{P})$ $N$-мерный ($N \in \mathbb{N}$) двоичный временной ряд $X_t = (x_{t1}, ..., x_{tN})' \in V^N$, порожденный семейством условных распределений вероятностей

$$\mathrm{P}\{X_t = J_t \mid \mathcal{F}_{t-1}\} = \mathrm{P}\{X_t = J_t \mid X_{t-1} = J_{t-1}, \ldots, X_{t-s} = J_{t-s}\}, t \in \mathbb{Z}, \tag{1}$$

где $x_{tl} \in V$ – двоичная (бинарная) случайная величина, задающая компоненту номер $l$ временного ряда в момент времени $t \in \mathbb{Z}$, $\mathcal{F}_{t-1} = \sigma\{X_\tau : \tau \le t-1\}$ – σ-алгебра случайных событий, порожденных указанными в скобках случайными векторами, $J_t = (j_{tl}) \in V^N$ – значение двоичного случайного вектора $X_t$ в момент времени $t \in \mathbb{Z}$, $s$ – глубина предыстории (памяти процесса), $s \in \mathbb{N}$.

Рассмотрим ситуацию, когда при фиксированной $s$-предыстории $X_{t-1} = J_{t-1}, \ldots, X_{t-s} = J_{t-s}$ случайные величины $x_{t1}, \ldots, x_{tN}$ условно независимы:

$$\begin{gathered}\mathrm{P}\{X_t = J_t \mid X_{t-1} = J_{t-1}, \ldots, X_{t-s} = J_{t-s}\} = \\ = \prod_{l=1}^{N} \mathrm{P}\{x_{tl} = j_{tl} \mid X_{t-1} = J_{t-1}, \ldots, X_{t-s} = J_{t-s}\}, J_t = (j_{tl}) \in V^N,\end{gathered} \tag{2}$$

где условное распределение $l$-го бита $x_{tl}$ при условии, что фиксирована $s$-предыстория, представимо в виде

$$\mathrm{P}\{x_{tl} = j_{tl} \mid X_{t-1} = J_{t-1}, \ldots, X_{t-s} = J_{t-s}\} = \begin{cases} p_l(J_{t-1:t-s}), & j_{tl} = 1, \\ 1 - p_l(J_{t-1:t-s}), & j_{tl} = 0; \end{cases} \tag{3}$$

где $J_{t-1:t-s} = (J'_{t-1}, J'_{t-2}, ..., J'_{t-s})' \in V^{Ns}$ – составной двоичный вектор-столбец $s$-предыстории. Заметим, что условная независимость (2) в общем случае не эквивалентна безусловной независимости, так что компоненты $x_{t1}, \ldots, x_{tN}$ зависимы между собой через общую для них предысторию.

Для $l$-й ($l = 1, \ldots, N$) компоненты аналогично [7] определим нейросетевую модель (индекс компоненты $l$ опущен для упрощения обозначений в силу (2), (3)):

$$p = p(J_{s:1}) = F\left(\sum_{k=1}^{m} b_k F_k \left(\sum_{l=1}^{Ns} a_{kl} j_l\right)\right), \tag{4}$$

где $F(\cdot)$ – некоторая дважды непрерывно дифференцируемая функция распределения $0 < F(x) < 1$; $F_1, \ldots, F_m$ – некоторые заданные абсолютно непрерывные функции распределения; $B = (b_k) \in \mathbb{R}^m$ и $A_k = (a_{k,1}, \ldots, a_{k,Ns})' \in \mathbb{R}^{Ns}$, $k = 1, \ldots m$, – неизвестные вектор-столбцы коэффициентов (параметров) модели. Таким образом, модель для $l$-й компоненты описывается двухслойной искусственной нейронной сетью с $Ns$ входами, одним выходом, $m$ нейронами на первом слое и одним нейроном на втором слое; при этом функции $F_1, \ldots, F_m$ называют функциями активации.

Общее число параметров модели (1)–(4) равно $m(Ns + 1)$. Так как любую функцию $Ns$ двоичных переменных можно задать $2^{Ns}$ значениями, то величина $m$ удовлетворяет неравенству

$$1 \le m \le m^+ = \frac{2^{Ns}}{Ns+1}.$$

Представим модель (4) в матричном виде:

$$p = p(J_{s:1}) = F(B'\boldsymbol{F}(A'J_{s:1})), \, J_{s:1} \in V^{Ns}, \tag{5}$$

где $\boldsymbol{F} = (F_1, \ldots, F_m)' : \mathbb{R}^m \to \mathbb{R}^m$ – векторная функция, осуществляющая функциональное преобразование покомпонентно, элементами которой являются функции $\{F_k\}$, $A = (A_1, \ldots, A_m) \in \mathbb{R}^{Ns \times m}$ – составная матрица коэффициентов (параметров) модели.

Т е о р е м а 1. *Векторный двоичный временной ряд $x_t$, описываемый моделью (1)–(4), является N-мерной однородной эргодической цепью Маркова порядка s.*

Д о к а з а т е л ь с т в о. По построению модели (1)–(4) переходное распределение вероятностей зависит лишь от предыстории глубины $s$, что равносильно обобщенному марковскому свойству $s$-го порядка. Поскольку $0 < F(x) < 1$, $x \in \mathbb{R}$, то все элементы матрицы вероятностей одношаговых переходов (3) строго положительны. Это обеспечивает выполнение достаточных условий эргодичности [9].

Два набора параметров $\theta^{(1)} = (B^{(1)}, A_1^{(1)}, ..., A_m^{(1)}) \in \mathbb{R}^{m(Ns+1)}$ и $\theta^{(2)} = (B^{(2)}, A_1^{(2)}, ..., A_m^{(2)}) \in \mathbb{R}^{m(Ns+1)}$ являются эквивалентными для модели (1)–(4), если $F(B^{(1)'}\boldsymbol{F}(A^{(1)'}J_{s:1})) \equiv F(B^{(2)'}\boldsymbol{F}(A^{(2)'}J_{s:1}))$, $J_{s:1} \in V^{Ns}$.

Л е м м а 1. *Для любого набора параметров $\theta^{(1)} = (B, A_1, ..., A_m) \in \mathbb{R}^{m(Ns+1)}$ модели (1)–(4) существует m! эквивалентных ему наборов параметров, отличающихся перестановкой пар $(b_1, A_1), \ldots, (b_m, A_m)$.*

Д о к а з а т е л ь с т в о. При перестановке слагаемых $b_k F_k \left( \sum_{l=1}^{Ns} a_{kl} j_l \right)$ в (4) сумма под знаком $F(\cdot)$ не изменяется, а количество возможных перестановок равно $m!$.

С л е д с т в и е. *Набор параметров модели $(b_1, A_1), \ldots, (b_m, A_m)$ идентифицируем с точностью до множества эквивалентных параметров $(b_{\pi_1}, A_{\pi_1}), ..., (b_{\pi_m}, A_{\pi_m})$, где $(\pi_1, \ldots, \pi_m) \in \Pi_m$ – произвольная подстановка на множестве* $(1, \ldots, m)$.

**Статистическое оценивание параметров модели.** Для оценивания вектор-столбцов параметров модели (1)–(4) применим FBE-метод [10], основанный на многомерных частотах.

Пусть наблюдается реализация длины $T$ двоичного $N$-мерного временного ряда:

$$X_{1:T} = (X_1, ..., X_T) \in V^{TN}.$$

Примем следующие обозначения: $\mathbf{1}\{C\}$ = {1, если $C$ истинно, 0 в противном случае} – индикаторная функция события $C$, $F^{-1}(\cdot) : (0, 1) \to \mathbb{R}^1$ – квантильная функция, $(1; J_{s:1}) = (1, J_s' J_{s-1}' \ldots J_1')' \in V^{Ns+1}$ – «расширенный» двоичный вектор-столбец,

$$\nu_s^T(J_{s:1}) = \sum_{t=s}^{T} 1\{X_t = J_s, ..., X_{t-s+1} = J_1\},$$

$$\nu_{s+1}^T(1; J_{s:1}) = \sum_{t=s}^{T-1} 1\{x_{t+1,l} = 1, X_t = J_s, ..., X_{t-s+1} = J_1\},$$

где $J^{(s)} = \{J_{s:1} \in V^{Ns}, t = s, s+1, ..., T : \nu_s^T(J_{s:1}) > 0\} \subseteq V^{Ns}$ – подмножество $s$-предысторий, имеющих ненулевые частоты в $X_{1:T}$.

Построим частотную оценку условной вероятности

$$\hat{p}(J_{s:1}) = \begin{cases} \dfrac{T-s}{T-s+1} \cdot \dfrac{\nu_{s+1}^T(1; J_{s:1})}{\nu_s^T(J_{s:1})}, & J_{s:1} \in J^{(s)}, \\[2ex] \dfrac{1}{2}, & J_{s:1} \notin J^{(s)}; \end{cases} \tag{6}$$

и статистики

$$\hat{u}(J_{s:1}) = F^{-1}(\hat{p}(J_{s:1})) \in \mathbb{R}, \tag{7}$$

$$E = (e_k) = \sum_{J_{s:1} \in J^{(s)}} \hat{u}(J_{s:1}) \boldsymbol{F}'(A' J_{s:1}) \in \mathbb{R}^{m \times 1},$$
$$D = (d_{ij}) = \sum_{J_{s:1} \in J^{(s)}} \boldsymbol{F}(A' J_{s:1}) \boldsymbol{F}'(A' J_{s:1}) \in \mathbb{R}^{m \times m}, D^{-1} = (\bar{d}_{ij}) \in \mathbb{R}^{m \times m}. \tag{8}$$

Т е о р е м а 2. *Если в (8) определитель $|D| \neq 0$ и известны параметры $A_i = (a_{i,1}, \ldots, a_{i,ns})' \in R^{Ns}$, $i = 1, \ldots, m$, то FBE-оценка для B имеет вид*

$$\hat{B} = (b_k) = D^{-1} E; \tag{9}$$

*при $T \to \infty$ она состоятельна, т. е. сходится по вероятности к истинному значению $B^0$:*

$$\hat{B} \xrightarrow{\mathrm{P}} B^0.$$

Д о к а з а т е л ь с т в о. Во-первых, построим FBE-оценку для *B*. Определим квадратичную функцию потерь

$$W(B, A) = \sum_{J_{s:1} \in J^{(s)}} (\hat{u}(J_{s:1}) - B' \boldsymbol{F}(A' J_{s:1}))^2 \geq 0. \tag{10}$$

Чтобы построить статистические оценки $\hat{A}, \hat{B}$ параметров *A*, *B* модели (1)–(4), необходимо решить следующую экстремальную задачу:

$$W(B, A) \to \min_{B, A}. \tag{11}$$

Найдем минимум функции (10) при фиксированной матрице *A*:

$$W(B, A) \to \min_{B}.$$

Используя градиент

$$\nabla_B W(B, A) = \sum_{J_{s:1} \in J^{(s)}} (-2\hat{u}(J_{s:1}) \boldsymbol{F}(A' J_{s:1}) + 2 \boldsymbol{F}(A' J_{s:1}) \boldsymbol{F}'(A' J_{s:1}) \hat{B}),$$

построим FBE-оценку параметра *B* при произвольном фиксированном *A*:

$$\hat{B} = \left( \sum_{J_{s:1} \in J^{(s)}} \boldsymbol{F}(A' J_{s:1}) \boldsymbol{F}'(A' J_{s:1}) \right)^{-1} \left( \sum_{J_{s:1} \in J^{(s)}} \hat{u}(J_{s:1}) \boldsymbol{F}(A' J_{s:1}) \right) = D^{-1} E \in \mathbb{R}^{m \times 1}, \tag{12}$$

что совпадает с (9).

Во-вторых, докажем состоятельность построенной оценки $\hat{B}$ (12). Так как оценка $\hat{p}(J_{s:1})$ (6) является состоятельной [9], т. е. сходится по вероятности к истинному значению $p^0 = p^0(J_{s:1}) = F(B^{0\prime} \boldsymbol{F}(A^{0\prime} J_{s:1})), J_{s:1} \in V^{Ns}$, где $B^0$, $A^0$ – истинные значения параметров модели, и функция *F* является дважды непрерывно дифференцируемой функцией распределения, то в силу теоремы о функциональном преобразовании сходящихся случайных последовательностей [11] оценка $\hat{u}(J_{s:1})$ (7) также является состоятельной: $\hat{u}(J_{s:1}) \xrightarrow{\mathrm{P}} u^0 = B^{0\prime} \boldsymbol{F}(A^{0\prime} J_{s:1}) = \boldsymbol{F}'(A^{0\prime} J_{s:1}) B^0$. Следовательно, с учетом (8):

$$E = \sum_{J_{s:1} \in J^{(s)}} \boldsymbol{F}(A^{0\prime} J_{s:1}) \hat{u}(J_{s:1}) \xrightarrow{\mathrm{P}} \left( \sum_{J_{s:1} \in J^{(s)}} \boldsymbol{F}(A^{0\prime} J_{s:1}) \boldsymbol{F}'(A^{0\prime} J_{s:1}) \right) B = DB \Rightarrow B^0 = D^{-1} E,$$

отсюда получим: $\hat{B} \xrightarrow{\mathrm{P}} B^0$.

На втором этапе решения задачи (11) для построения статистической оценки $\hat{A}$ подставим оценку $\hat{B}$ в функцию потерь (10) и получим экстремальную задачу

$$w(A) \to \min_A,$$
$$w(A) ::= W(\hat{B}, A) = \sum_{J_{s:1} \in J^{(s)}} \hat{u}^2(J_{s:1}) - E'D^{-1}E. \qquad (13)$$

Нам понадобятся вспомогательные леммы для решения задачи (13).

Примем дополнительные обозначения: $\delta_{ij}$ – символ Кронекера, $\otimes$ – произведение Кронекера, $I_m \in \mathbb{R}^{m \times m}$ – единичная матрица, $G_f$ – матрица Якоби векторной функции $\boldsymbol{F}$, $f_i = \dfrac{dF_i(z)}{dz}$, составной вектор-столбец $s$-предыстории $J_{s:1}$ для краткости обозначим $J$, $K_{n,m}$ – так называемая коммутационная матрица [12]. Блочная матрица $K_{n,m} \in \mathbb{R}^{nm \times nm}$, состоящая из $n \times m$ блоков, называется коммутационной, если в $(i, j)$ блоке элемент $(j, i)$ равен 1, а все остальные элементы в этом блоке равны 0 ($i \in \{1, \ldots, n\}, j \in \{1, \ldots, m\}$), например, при $n = 2$, $m = 3$:

$$K_{2,3} = \begin{pmatrix} 1 & 0 & \vdots & 0 & 0 & \vdots & 0 & 0 \\ 0 & 0 & \vdots & 1 & 0 & \vdots & 0 & 0 \\ 0 & 0 & \vdots & 0 & 0 & \vdots & 1 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 1 & \vdots & 0 & 0 & \vdots & 0 & 0 \\ 0 & 0 & \vdots & 0 & 1 & \vdots & 0 & 0 \\ 0 & 0 & \vdots & 0 & 0 & \vdots & 0 & 1 \end{pmatrix}.$$

Л е м м а 2. *В принятых обозначениях справедливы следующие соотношения для элементов матриц E и D, определенных в (8)* ($i, k, l, q, r \in \{1, \ldots, m\}$, $v \in \{1, \ldots, Ns\}$):

$$\frac{\partial e_k}{\partial a_{iv}} = \delta_{ik} \sum_{J \in J^{(s)}} (\hat{u}(J) f_i(A'J)) j_v, \qquad (14)$$

$$\frac{\partial d_{qr}}{\partial a_{iv}} = \sum_{J \in J^{(s)}} f_i(A'J)(\delta_{iq} F_r(A'J) + \delta_{ir} F_q(A'J)) j_v, \qquad (15)$$

$$\frac{\partial \bar{d}_{kl}}{\partial d_{qr}} = -\bar{d}_{kq} \bar{d}_{rl}. \qquad (16)$$

Д о к а з а т е л ь с т в о. В силу (8) получаем

$$e_k = \sum_{J \in J^{(s)}} \hat{u}(J) F_k(J), \; d_{qr} = \sum_{J \in J^{(s)}} F_q(A'_q J) F_r(A'_r J),$$

откуда следуют (14), (15).

Для доказательства соотношения (16) воспользуемся формулой из [12]:

$$\left( \frac{dD^{-1}}{dD} \right)_{(k,l),(q,r)} = \frac{d(D^{-1})_{kl}}{d(D)_{qr}} = -(D^{-1})_{kq}((D^{-1})')_{lr} = -\bar{d}_{kq} \bar{d}_{lr}.$$

Л е м м а 3. *В принятых обозначениях частные производные первого порядка целевой функции (13) вычисляются по формулам* ($i \in \{1, \ldots, m\}$, $v \in \{1, \ldots, s\}$):

$$\frac{\partial w(A)}{\partial a_{iv}} = -2\hat{b}_i \sum_{J \in J^{(s)}} j_v f_i(A'J)(\hat{u}(J) - \hat{B}'\boldsymbol{F}(A'J)).$$

Д о к а з а т е л ь с т в о. Из (13) следует:

$$\frac{\partial w(A)}{\partial a_{iv}} = -\left(\sum_{k,l=1}^{m} e_k e_l \frac{\partial \bar{d}_{kl}}{\partial a_{iv}} + \bar{d}_{kl}\left(\frac{\partial e_k}{\partial a_{iv}} e_l + \frac{\partial e_l}{\partial a_{iv}} e_k\right)\right). \tag{17}$$

Рассмотрим второе слагаемое использовав соотношения (14), (15) и учтем свойство симметричности матрицы $D$:

$$\begin{aligned}\sum_{k,l=1}^{m} \bar{d}_{kl}\left(\frac{\partial e_k}{\partial a_{iv}} e_l + \frac{\partial e_l}{\partial a_{iv}} e_k\right) &= \sum_{J\in J^{(s)}} \hat{u}(J) f_i(A'J) j_v \sum_{k,l=1}^{m} \bar{d}_{kl}(\delta_{ik} e_l + \delta_{il} e_k) = \\ &= \sum_{J\in J^{(s)}} \hat{u}(J) f_i(A'J) j_v \left(\sum_{l=1}^{m} e_l \bar{d}_{il} + \sum_{l=1}^{m} e_k \bar{d}_{ki}\right) = 2\hat{b}_i \sum_{J\in J^{(s)}} j_v \hat{u}(J) f_i(A'J).\end{aligned} \tag{18}$$

Теперь рассмотрим первое слагаемое. Согласно соотношению (16) из леммы 2 получаем

$$\begin{aligned}\frac{\partial \bar{d}_{kl}}{\partial a_{iv}} &= -\sum_{q,r=1}^{m} \bar{d}_{kq} \bar{d}_{rl} \sum_{J\in J^{(s)}} j_v (\delta_{iq} F_r(A'_r J) + \delta_{ir} F_q(A'_q J)) f_i(A'_i J) = \\ &= -\sum_{J\in J^{(s)}} j_v f_i(A_i{}' J)(\bar{d}_{ik}(D^{-1}G)_k + \bar{d}_{il}(D^{-1}G)_l).\end{aligned} \tag{19}$$

Подставим (18) и (19) в (17) и преобразуем

$$\begin{aligned}\frac{\partial w}{\partial a_{iv}} &= -2\hat{b}_i \sum_{J\in J^{(s)}} j_v f_i(A'_i J)\hat{u}(J) + \sum_{J\in J^{(s)}} j_v f_i(A_i{}' J) \sum_{k.l=1}^{m} e_k e_l (\bar{d}_{ik}(D^{-1}G)_k + \bar{d}_{il}(D^{-1}G)_l) = \\ &= -2\hat{b}_i \sum_{J\in J^{(s)}} j_v f_i(A'J)(\hat{u}(J) - \hat{B}'\boldsymbol{F}(A'J)).\end{aligned}$$

Л е м м а 4. *В принятых обозначениях справедливы следующие матричные соотношения:*

$$\frac{d\boldsymbol{F}(A'J)}{dA} = (I_m \otimes J)K_{1,m}G_F, \tag{20}$$

$$\frac{d\boldsymbol{F}'(A'J)}{dA} = \frac{d\boldsymbol{F}(A'J)}{dA} K_{1,m}, \tag{21}$$

$$\frac{dE}{dA} = \sum_{J\in J^{(s)}} \hat{u}(J) \frac{d\boldsymbol{F}(A'J)}{dA}, \tag{22}$$

$$\frac{dD}{dA} = \sum_{J\in J^{(s)}} \frac{d\boldsymbol{F}(A'J)}{dA}(\boldsymbol{F}'(A'J)\otimes I_m) + \frac{d\boldsymbol{F}'(A'J)}{dA}(I_m \otimes \boldsymbol{F}'(A'J)), \tag{23}$$

$$\frac{dD^{-1}}{dA} = -\frac{dD}{dA} D^{-1} \otimes D^{-1}. \tag{24}$$

Д о к а з а т е л ь с т в о. Для доказательства соотношений (20)–(24) воспользуемся следующими формулами матричного дифференцирования из [12; 13]:

$$\begin{aligned}&\frac{dA'X}{dX} = I_m \otimes A, \frac{dZ(Y(X))}{dX} = \frac{dY}{dX}\frac{dZ}{dY}, \frac{dX'}{dX} = K_{n,m}, \\ &\frac{d(Y(X)Z(X))}{dX} = \frac{dY}{dX}(Z\otimes I_n) + \frac{dZ}{dX}(I_m \otimes Y'), \frac{dX^{-1}}{dX} = -X^{-1}\otimes (X')^{-1}.\end{aligned} \tag{25}$$

Для доказательства (20), (21) воспользуемся формулами (25):

$$\frac{d\boldsymbol{F}(A'J)}{dA} = \frac{dJ'A}{dA}\frac{d(A'J)'}{dA'J}\frac{d\boldsymbol{F}(A'J)}{dA'J} = (I_m \otimes J)K_{1,m}G_F,$$

$$\frac{d\boldsymbol{F}'(A'J)}{dA}=\frac{d\boldsymbol{F}(A'J)}{dA}\frac{d\boldsymbol{F}'(A'J)}{d\boldsymbol{F}(A'J)}=\frac{d\boldsymbol{F}(A'J)}{dA}K_{1,m}.$$

Соотношение (22) следует из (8):

$$\frac{dE}{dA}=\frac{d\sum\limits_{J\in J^{(s)}}\hat{u}(J)\boldsymbol{F}(A'J)}{dA}=\sum_{J\in J^{(s)}}\hat{u}(J)\frac{d\boldsymbol{F}(A'J)}{dA}.$$

Для доказательства (23), (24) применим формулы (25):

$$\frac{dD}{dA}=\sum_{J\in J^{(s)}}\frac{d\boldsymbol{F}(A'J)\boldsymbol{F}'(A'J)}{dA}=\sum_{J\in J^{(s)}}\frac{d\boldsymbol{F}(A'J)}{dA}(\boldsymbol{F}'(A'J)\otimes I_m)+\frac{d\boldsymbol{F}'(A'J)}{dA}(I_m\otimes\boldsymbol{F}'(A'J)),$$

$$\frac{dD^{-1}}{dA}=\frac{dD}{dA}\frac{dD^{-1}}{dD}=-\frac{dD}{dA}D^{-1}\otimes D^{-1}.$$

Л е м м а 5. *В принятых обозначениях частные производные первого порядка целевой функции (13) вычисляются по формулам (в матричном виде)*

$$\frac{dw}{dA}=-\left(\frac{dE'}{dA}+\frac{dE}{dA}\right)\hat{B}-\frac{dD^{-1}}{dA}(E\otimes I_m)E. \tag{26}$$

Д о к а з а т е л ь с т в о. Соотношение (26) доказывается с помощью формул (25)

$$\frac{dw}{dA}=-\frac{dE'D^{-1}E}{dA}=-\frac{dE'}{dA}(D^{-1}E\otimes I_1)-\frac{dD^{-1}E}{dA}(I_1\otimes E)=-\frac{dE'}{dA}\hat{B}+$$

$$-\left(\frac{dD^{-1}}{dA}(E\otimes I_m)+\frac{dE}{dA}(I_1\otimes D^{-1})\right)E=-\left(\frac{dE'}{dA}+\frac{dE}{dA}\right)\hat{B}-\frac{dD^{-1}}{dA}(E\otimes I_m)E.$$

Используя леммы 1–5, построим итерационный алгоритм градиентного спуска для статистического оценивания параметров модели (1)–(4):

1) инициализируем начальное значение матричного параметра $A$: $\hat{A}^{(0)}$;

2) вычисляем оценку $\hat{B}$ согласно (9): $\hat{B}^{(\tau-1)}=D^{-1}(\hat{A}^{(\tau-1)})E(\hat{A}^{(\tau-1)}),\ \tau=1,2,\ldots,$ где $D^{-1}(\cdot)$, $E(\cdot)$ вычисляются по формуле (8);

3) обновляем значение оценки $\hat{A}$ используя (17), (26):

$$\hat{A}^{(\tau)}=\hat{A}^{(\tau-1)}-\alpha\frac{dw(\hat{B}^{(\tau-1)},\hat{A}^{(\tau-1)})}{dA}, \tag{27}$$

где $\hat{B}^{(\tau)}$ – оценка параметра $B$ на $\tau$-й итерации; $\hat{A}^{(\tau)}$ – оценка параметра $A$ на $\tau$-й итерации; $\alpha$ – задаваемая величина шага итерации.

Для нахождения глобального минимума в (13) необходимо осуществить итерационный процесс (27) для $L\in\mathbb{N}$ различных вариантов задания начального значения $\hat{A}^{(0)}$.

Пусть $(B^0, A^0)$ – истинные значения параметров, $(B_\pi^0, A_\pi^0)$ – перестановочный набор параметров: $B_\pi^0=(b_{\pi_1}^0,\ldots,b_{\pi_m}^0)$, $A_\pi^0=(A_{\pi_1}^0,\ldots,A_{\pi_m}^0)$, где $\pi=(\pi_1,\ldots,\pi_m)\in\Pi_m$ – произвольная подстановка из множества $m!$ всевозможных подстановок на $\{1,\ldots,m\}$.

Т е о р е м а 3. *Если имеет место модель (1)–(4) с истинными значениями параметров $B^0$, $A^0$, $|D|\neq 0$ и $(\hat{B},\hat{A})$ – статистические оценки параметров $(B^0, A^0)$, построенные с помощью алгоритма (27), то при $L\to+\infty$ и $T\to+\infty$ имеет место сходимость*

$$(\hat{B},\hat{A})\xrightarrow{\text{P}}(B_\pi^0, A_\pi^0)$$

*для некоторой подстановки $\pi\in\Pi_m$.*

Д о к а з а т е л ь с т в о. Как уже отмечалось в ходе доказательства теоремы 2, $\hat{p}(J_{s:1}) \xrightarrow{\mathrm{P}} p^0(J_{s:1})$, $J_{s:1} \in V^{Ns}$, поэтому с учетом (5), (7) и теоремы о функциональном преобразовании сходящихся случайных последовательностей [1] при истинных значениях параметров $B^0$, $A^0$ и $T \to +\infty$:

$$\hat{u}(J_{s:1}) \xrightarrow{\mathrm{P}} F^{-1}(F(B^{0\prime}\boldsymbol{F}(A^{0\prime}J_{s:1}))) = B^{0\prime}\boldsymbol{F}(A^{0\prime}J_{s:1}).$$

При этом согласно (10) $W(B^0, A^0) \xrightarrow{\mathrm{P}} 0.$ Аналогично, в силу леммы 1 для любой подстановки $\pi \in \Pi_m$

$$W(B_\pi^0, A_\pi^0) \xrightarrow{\mathrm{P}} 0. \tag{28}$$

При $L \to +\infty$ начальная точка $(B^0, A^0)$ алгоритма (27) попадает в окрестность одной из $m!$ точек $(B_\pi^0, A_\pi^0)$, поэтому при увеличении числа итераций $\tau \to +\infty$ с учетом (28) получаем $W(B^{(\tau)}, A^{(\tau)}) \xrightarrow{\mathrm{P}} 0.$ Отсюда следует, что $(\hat{B}, \hat{A}) \xrightarrow{\mathrm{P}} (B_\pi^0, A_\pi^0)$ для некоторой подстановки $\pi \in \Pi_m$, т. е. статистическая оценка $(\hat{B}, \hat{A})$ состоятельна с точностью до множества перестановок, указанного в лемме 1.

**Подстановочный алгоритм прогнозирования.** Используя статистические оценки $\hat{A}, \hat{B}$ параметров модели (1)–(4), подстановочный алгоритм оптимального прогнозирования на один шаг определяется явным соотношением [14]:

$$\hat{x}_{tl} = \mathbf{1}\left\{\hat{p}_l(X_{t-1:t-s}) - \frac{1}{2} > 0\right\}, l = 1, \ldots, N, \tag{29}$$

где

$$\hat{p}_l(X_{t-1:t-s}) = F(\hat{B}'_{(l)}\boldsymbol{F}(\hat{A}'_{(l)}X_{t-1:t-s})), \tag{30}$$

где $(\hat{B}_{(l)}, \hat{A}_{(l)})$ – оценки параметров, полученные по алгоритму (27) для $l$-го бита в (3), (4).

Прогнозирование $\hat{x}_{t+1,l}$ на следующий шаг осуществляется аналогично (29), только фрагмент $X_{t-1:t-s} = (X_{t-1}, \ldots, X_{t-s})$ в (30) заменяется на $X_{t:t-s+1} = (\hat{X}_t, \ldots, X_{t-s+1})$ и т. д.

**Эксперименты на модельных данных.** Использовались следующие значения параметров модели: $N = 2$, $(m, s) \in \{(2, 4), (3, 4), (5, 5), (10, 5)\}$, число реализаций смоделированного временного ряда $M = 10$, длина временного ряда $T = 1000$, величина шага $\alpha = 0{,}1$. Оценивалась зависимость среднеквадратичной ошибки оценивания вероятностной функции $p^0(J_{s:1})$ от числа итераций $I$, числа функций $m$ и глубины предыстории $s$:

$$\hat{\Delta}^2_{p^0} = \frac{1}{M}\sum_{v=1}^{M}\left(\frac{1}{N}\sum_{J_{s:1}\in J^{(s)}}\left\|p_v^0(J_{s:1}) - F(\hat{B}'_v\boldsymbol{F}(\hat{A}'_vJ_{s:1}))\right\|^2\right),$$

где $\hat{B}_v, \hat{A}_v$ – статистические оценки, полученные по алгоритму (27) для реализации номер $v$.

Графики зависимости $\hat{\Delta}^2_{p^0}$ от числа итераций (27), представленные на рис. 1, иллюстрируют состоятельность построенных оценок $(\hat{B}, \hat{A})$; «всплески» на графиках объясняются выбором значения шага $\alpha$.

**Эксперименты на реальных данных.** Для экспериментов использовались данные $x_t = (x_{ti})$ двух курсов валют ($N = 2$): российского рубля ($i = 1$) и доллара США ($i = 2$) по отношению к белорусскому рублю с 01.01.2020 по 01.12.2022 ($T = 1064$). Исследовались приращения курсов валют: если курс валюты увеличился в сравнении со «вчерашним», то значение $x_{ti} = 1$, если курс упал или остался прежним, то значение $x_{ti} = 0$ ($i = 1, 2$). Оценивалась зависимость нормированной функции потерь: $\tilde{W}(B, A) = \dfrac{W(B, A)}{|J^{(s)}|}$, где $|J^{(s)}|$ – мощность множества $J^{(s)}$, от числа итераций $I$. Использовались следующие значения параметров: $m = 10$, $s = 5$, $\alpha = 0{,}1$. На рис. 2 приведены графики зависимости $\tilde{W}(B, A)$ от числа итераций $I$, иллюстрирующие процесс статистического оценивания параметров модели.

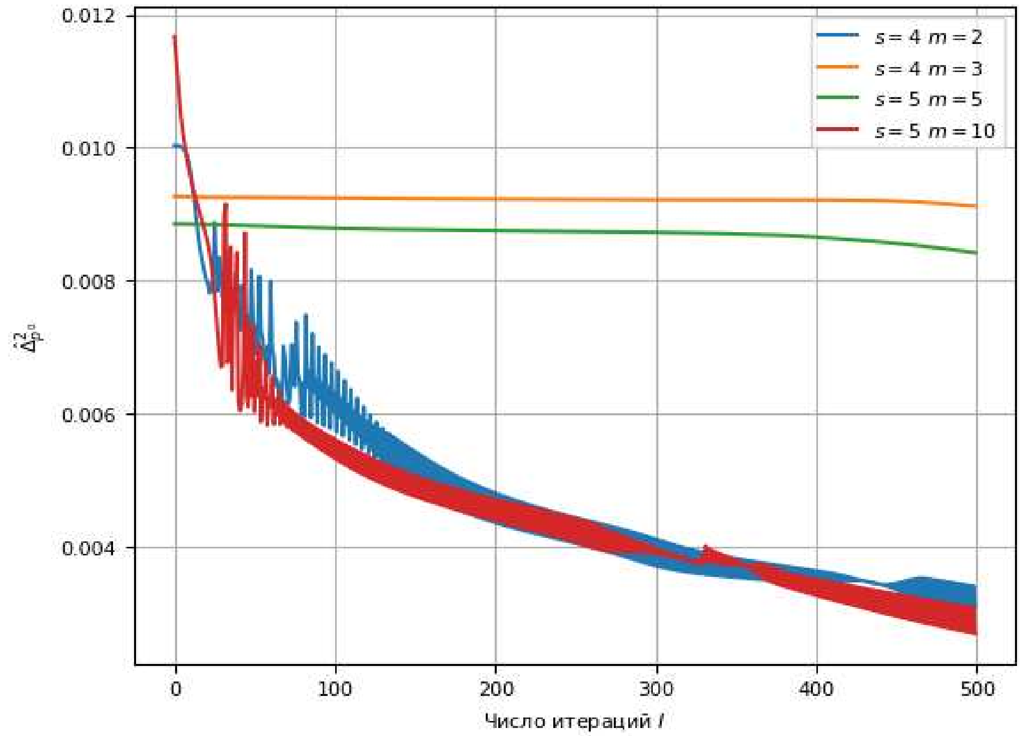


Рис. 1. Зависимость среднеквадратичной ошибки оценивания вероятности $p^0$ от числа итераций $I$, числа функций $m$ и глубины предыстории $s$

Fig. 1. Dependence of the mean square error of probability $p^0$ estimation on the number of iterations $I$, the number of functions $m$ and the depth of the history $s$

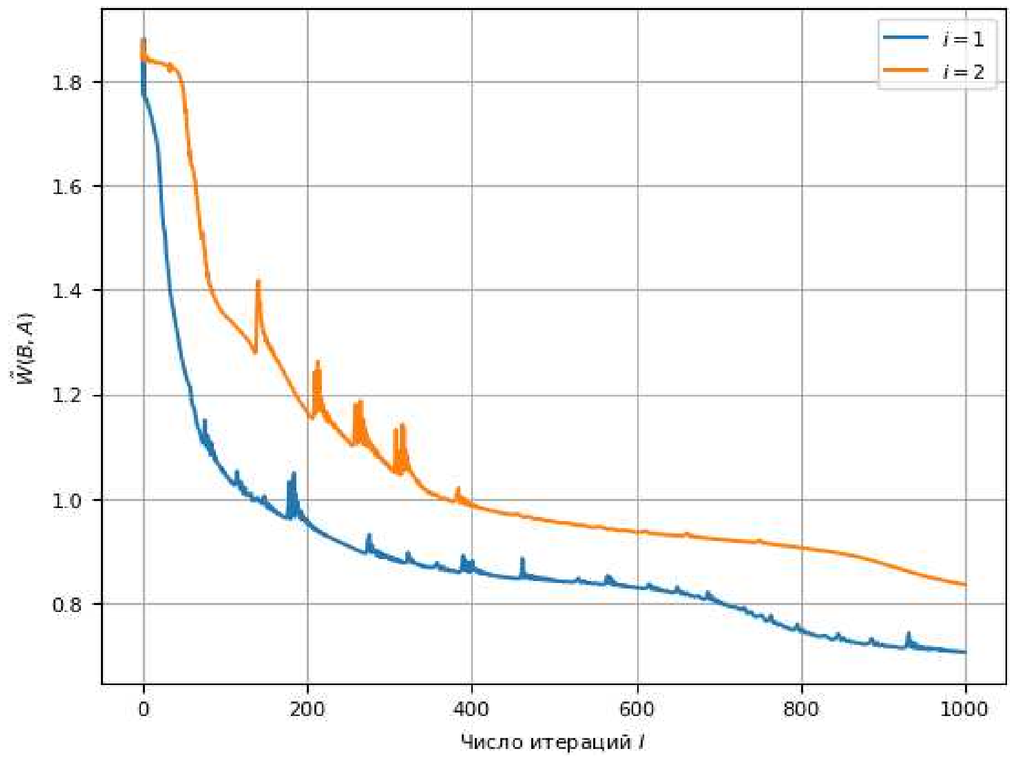


Рис. 2. Зависимость нормированной функции потерь $\tilde{W}(B, A)$ от числа итераций $I$

Fig. 2. Dependence of the normed loss function $\tilde{W}(B, A)$ on the number of iterations $I$

**Заключение.** В работе получены следующие основные результаты: для многомерных двоичных временных рядов предложена малопараметрическая нейросетевая модель; построена состоятельная статистическая оценка параметров нейросетевой модели; разработан подстановочный алгоритм прогнозирования многомерного двоичного временного ряда; проведены компьютерные эксперименты на модельных и реальных данных, иллюстрирующие применимость полученных результатов при решении прикладных задач.

## Список используемых источников

1. Anderson, T. W. The Statistical Analysis of Time Series / T. W. Anderson. – New York, 1971. – 704 p. https://doi.org/10.1002/9781118186428

2. Nelder, J. Generalized linear models / J. Nelder, R. Wedderburn // J. R. Stat. Soc. Ser. A. – 1972. – Vol. 135, N 3. – P. 370–384. https://doi.org/10.2307/2344614

3. Biswas, A. Discrete-valued ARMA processes / A. Biswas, P. X.-K. Song // Statist. Probab. Lett. – 2009. – Vol. 79, N 17. – P. 1884–1889. https://doi.org/10.1016/j.spl.2009.05.025

4. Cameron, A. C. Regression Analysis of Count Data / A. C. Cameron, P. K. Trivedi. – Cambridge, 2013. – 567 p. https://doi.org/10.1017/cbo9781139013567

5. Kim, C. Dynamic linear models with Markov-switching / C. Kim // J. Econometrics. – 1994. – Vol. 60, N 1–2. – P. 1–22. https://doi.org/10.1016/0304-4076(94)90036-1

6. Hamilton, J. D. Time Series Analysis / J. D. Hamilton. – Princeton, NJ, 1994. – 799 p. https://doi.org/10.1515/9780691218632

7. Statistical analysis of multivariate discrete-valued time series / K. Fokianos [et al.] // J. Multivariate Anal. – 2022. – Vol. 188. – Art. 104805. https://doi.org/10.1016/j.jmva.2021.104805

8. Харин, Ю. С. Нейросетевые модели биномиальных временных рядов в задачах анализа данных / Ю. С. Харин // Докл. Нац. акад. наук Беларуси. – 2021. – Т. 65, № 6. – С. 654–660. https://doi.org/10.29235/1561-8323-2021-65-6-654-660

9. Billingsley, P. Statistical methods in Markov chains / P. Billingsley // Ann. Math. Stat. – 1961. – Vol. 32, N 1. – P. 12–40. https://doi.org/10.1214/aoms/1177705136

10. Kharin, Yu. Robust estimation for binomial conditionally nonlinear autoregressive time series based on multivariate conditional frequencies / Yu. Kharin, V. Voloshko // J. Multivariate Anal. – 2021. – Vol. 185. – Art. 104777. https://doi.org/10.1016/j.jmva.2021.104777

11. Ширяев, А. Н. Вероятность: в 2 кн. / А. Н. Ширяев. – М., 2004.

12. Kollo, T. Advanced Multivariate Statistics and Matrices / T. Kollo, D. Rosen. – Dordrecht, 2005. – 506 p. https://doi.org/10.1007/1-4020-3419-9

13. Magnus, J. Matrix Differential Calculus with Applications in Statistics and Econometrics / J. Magnus, H. Neudecker. – New York, 2019. – 482 p.

14. Харин, Ю. С. Оптимальность и робастность в статистическом прогнозировании / Ю. С. Харин. – Минск, 2008. – 263 с.

## References

1. Anderson T. W. *The Statistical Analysis of Time Series*. New York, 1971. 704 p. https://doi.org/10.1002/9781118186428

2. Nelder J., Wedderburn R. Generalized linear models. *Journal of the Royal Statistical Society. Series A*, 1972, vol. 135, no. 3, pp. 370–384. https://doi.org/10.2307/2344614

3. Biswas A., Song P. X.-K. Discrete-valued ARMA processes. *Statistics and Probability Letters*, 2009, vol. 79, no. 17, pp. 1884–1889. https://doi.org/10.1016/j.spl.2009.05.025

4. Cameron A. C., Trivedi P. K. Regression Analysis of Count Data. Cambridge, 2013. 567 p. https://doi.org/10.1017/cbo9781139013567

5. Kim C. Dynamic linear models with Markov-switching. *Journal of Econometrics*, 1994, vol. 60, no. 1–2, pp. 1–22. https://doi.org/10.1016/0304-4076(94)90036-1

6. Hamilton J. D. *Time Series Analysis.* Princeton, NJ, 1994. 799 p. https://doi.org/10.1515/9780691218632

7. Fokianos K., Fried R., Kharin Yu., Voloshko V. Statistical analysis of multivariate discrete-valued time series. *Journal of Multivariate Analysis*, 2022, vol. 188, art. 104805. https://doi.org/10.1016/j.jmva.2021.104805

8. Kharin Yu. S. Neural network-based models of binomial time series in data analysis problems. *Doklady Natsional'noi akademii nauk Belarusi = Doklady of the National Academy of Sciences of Belarus*, 2021, vol. 65, no. 6, pp. 654–660 (in Russian). https://doi.org/10.29235/1561-8323-2021-65-6-654-660

9. Billingsley P. Statistical methods in Markov chains. *The Annals of Mathematical Statistics*, 1961, vol. 32, no. 1, pp. 12–40. https://doi.org/10.1214/aoms/1177705136

10. Kharin Yu., Voloshko V. Robust estimation for binomial conditionally nonlinear autoregressive time series based on multivariate conditional frequencies. *Journal of Multivariate Analysis*, 2021, vol. 185, art. 104777. https://doi.org/10.1016/j.jmva.2021.104777

11. Shiryaev A. N. *Probability, in 2 books*. Moscow, 2004 (in Russian).

12. Kollo T., Rosen D. *Advanced Multivariate Statistics and Matrices*. Dordrecht, 2005. 506 p. https://doi.org/10.1007/1-4020-3419-9

13. Magnus J., Neudecker H. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York, 2019. 482 p.

14. Kharin Yu. S. *Optimality and Robustness in Statistical Forecasting*. Minsk, 2008. 263 p. (in Russian).

## Информация об авторах

*Харин Юрий Семенович* – академик, д-р физ.-мат. наук, профессор, директор. НИИ прикладных проблем математики и информатики БГУ (пр. Независимости, 4, 220030, Минск, Республика Беларусь). E-mail: kharin@bsu.by.

*Шибалко Сергей Анатольевич* – студент. Белорусский государственный университет (пр. Независимости, 4, 220030, Минск, Республика Беларусь). E-mail: shibalko2003@bk.ru.

## Information about the authors

*Kharin Yuriy S.* – Academician, D. Sc. (Physics and Mathematics), Professor, Director. Research Institute for Applied Problems of Mathematics and Informatics of the Belarusian State University (4, Nezavisimosti Ave., Minsk, 220030, Republic of Belarus). E-mail: kharin@bsu.by.

*Shibalko Siarhei A.* – Student. Belarusian State University (4, Nezavisimosti Ave., Minsk, 220030, Republic of Belarus). E-mail: shibalko2003@bk.ru.