УДК 519.2

МАЛОПАРАМЕТРИЧЕСКАЯ МОДЕЛЬ НА ОСНОВЕ БАЗИСНЫХ ФУНКЦИЙ ДЛЯ СТАТИСТИЧЕСКОГО АНАЛИЗА МНОГОМЕРНЫХ ДВОИЧНЫХ ВРЕМЕННЫХ РЯДОВ

С. А. ШИБАЛКО
(Белорусский государственный университет, г. Минск)

Аннотация. *В докладе рассматривается задача статистического анализа многомерных двоичных временных рядов. Для решения этой задачи предлагается малопараметрическая модель эргодической цепи Маркова порядка s на основе базисных функций. Построены состоятельные статистические оценки параметров модели и подстановочный алгоритм статистического прогнозирования. Приведены результаты компьютерных экспериментов на модельных и реальных данных.*

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

Введение. Цифровизация экономики и всего окружающего мира ведет к увеличению данных в дискретном пространстве состояний с дискретным временем. Для математического описания таких данных используются дискретные, в том числе двоичные, временные ряды. Обзор современного состояния в области статистического анализа многомерных дискретных временных рядов представлен в [1].

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 (Ω, $\mathcal{F}$, P) N-мерный ($N \in \mathbb{N}$) двоичный временной ряд $X_t = (x_{t1}, \ldots, x_{tN})' \in V^N$, порожденный семейством условных распределений вероятностей:

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

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), $s \in \mathbb{N}$.

В докладе рассматриваются два случая.

1) Случай условно независимых компонент

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:

$$P\{X_t = J_t \mid X_{t-1} = J_{t-1}, \ldots, X_{t-s} = J_{t-s}\} =$$

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

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

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

здесь $J_{t-1:t-s} = (J'_{t-1}, J'_{t-2}, \ldots, J'_{t-s})' \in V^{Ns}$ – 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.

2) Случай полносвязных компонент

При фиксированной s-предыстории $H = \{X_{t-1} = J_{t-1}, \ldots, X_{t-s} = J_{t-s}\}$ случайные величины $x_{t1}, \ldots, x_{tN}$ зависимы так, как приведено на формуле:

$$P\{X_t = J_t \mid H\} = P\{x_{t1} = j_{t1} \mid H\} \cdot P\{x_{t2} = j_{t2} \mid x_{t1} = j_{t1}, H\} \times \ldots$$

$$\times P\{x_{tN} = j_{tN} \mid x_{t1} = j_{t1}, \ldots, x_{tN-1} = j_{tN-1} H\}, \quad J_t = (j_{tl}) \in V^N, \tag{4}$$

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

– для первой компоненты

$$P\{x_{t1} = j_{t1} \mid H\} == \begin{cases} p_1(J_{t-1:t-s}), & j_{t1} = 1, \\ 1 - p_1(J_{t-1:t-s}), & j_{t1} = 0; \end{cases} \tag{5}$$

– для остальных компонент

$$P\{x_{tl}=j_{tl}\mid x_{t1}=j_{t1},\dots,x_{tl-1}=j_{tl-1,H}\}=\begin{cases}p_l(x_{t1},\dots,x_{tl-1},J_{t-1:t-s}), & j_{tl}=1,\\ 1-p_l(x_{t1},\dots,x_{tl-1},J_{t-1:t-s}), & j_{tl}=0;\end{cases} \quad (6)$$

такую модель будем называть полносвязной [2].

Модели (2), (3) и (4)–(6) требуют большое количество параметров для поточечного задания. Из-за «проклятия размерности» для уменьшения потребления памяти компьютера и времени работы алгоритмов оценивания параметров модели необходимо использовать малопараметрические модели [3].

Малопараметрическая модель на основе базисных функций. Для l-й $(l=1,\dots,N)$ компоненты в докладе определена следующая малопараметрическая модель на основе базисных функций (индекс компоненты l опущен для упрощения обозначений).

$$p=p(J_{t:t-s+1})=F(\textstyle\sum_{k=1}^{m}b_k\psi_k(J_{t:t-s+1})),\ J_{t:t-s+1}\in V^{Ns}, \quad (7)$$

где F – некоторая дважды непрерывно дифференцируемая функция распределения $0<F(x)<1$;

$B=(b_k)\in\mathbb{R}^m$ – вектор столбец неизвестных параметров модели;

$\psi_k(J_{t:t-s+1})$ – семейство линейно независимых на V^{Ns} базисных функций.

Модель (7) имеет всего $m<m^+=2^{Ns}$ неизвестных параметров.

Теорема 1. *Векторный двоичный временной ряд* X_t, *описываемый моделью* (7) *является N-мерной однородной эргодической цепью Маркова порядка s.*

Доказательство. По построению модели (7) 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 вероятности одношаговых переходов (3) или (5), (6) строго положительны. Это обеспечивает выполнение достаточных условий эргодичности.

Статистическое оценивание параметров модели. Для оценивания вектор-столбца параметров модели (1)–(7) применим FBE-метод [4], основанный на многомерных частотах.

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

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

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 $(1; J'_{t:t-s+1})' = (1, J'_{t-1} J'_{t-2} \dots J'_{t-s})' \in V^{Ns+1}$ – «расширенный» вектор-столбец, $\Psi(J_{t:t-s+1}) = (\psi_k(J_{t:t-s+1})) \in \mathbb{R}^{m\times 1}$ – вектор-столбец, состоящий из базисных функций $\psi_k(J_{t:t-s+1})$.

1) Случай условно независимых компонент

$$v_s^T(J_{t:t-s+1}) = \sum_{t=s}^{T} 1\{X_t = J_t, \dots, X_{t-s+1} = J_{t-s+1}\},$$

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

2) Случай полносвязных компонент

$$v_s^T(J_{t:t-s+1}) = \sum_{t=s}^{T} 1\{x_{t+1,l-1} = j_{t+1,l-1}, \dots, x_{t+1,1} = j_{t+1,1}, X_t = J_t, \dots, X_{t-s+1} = J_{t-s+1}\},$$

$$v_{s+1}{}^T(1; J_{t:t-s+1}) = \sum_{t=s}^{T} 1\{x_{t+1,l} = 1, x_{t+1,l-1} = j_{t+1,l-1}, \dots, x_{t+1,1} =$$

$$= j_{t+1,1}, X_t = J_t, \dots, X_{t-s+1} = J_{t-s+1}\},$$

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

Построим частотную оценку условной вероятности:

$$\hat{p}(J_{t:t-s+1}) = \begin{cases} \dfrac{T-s}{T-s+1} \cdot \dfrac{v_{s+1}^T(1; J_{t:t-s+1})}{v_s^T(J_{t:t-s+1})}, & J_{t:t-s+1} \in J^{(s)}, \\ \dfrac{1}{2}, & J_{t:t-s+1} \notin J^{(s)}; \end{cases} \tag{8}$$

и статистики:

$$\hat{u}(J_{t:t-s+1}) = F^{-1}(\hat{p}(J_{t:t-s+1})) \in \mathbb{R}, \tag{9}$$

$$E = (e^k) = \sum\nolimits_{J_{t:t-s+1} \in J^{(s)}} \hat{u}(J_{t:t-s+1}) \Psi'(J_{t:t-s+1}) \in \mathbb{R}^{m\times 1}, \tag{10}$$

$$D = (d_{ij}) = \sum\nolimits_{J_{t:t-s+1} \in J^{(s)}} \Psi(J_{t:t-s+1}) \Psi'(J_{t:t-s+1}) \in \mathbb{R}^{m\times m}, D^{-1} = (\bar{d}_{ij}) \in \mathbb{R}^{m\times m}. \tag{11}$$

Теорема 2. *Если определитель* $|D| \neq 0$ *(11), то FBE-оценка для B имеет вид*

$$\hat{B} = \left(b_k\right) = D^{-1}E,$$

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

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

Подстановочный алгоритм статистического прогнозирования. Используя статистическую оценку $\hat{B}$ параметра модели (7), подстановочный алгоритм оптимального прогнозирования на один шаг определяется явным соотношением:

$$\hat{x}_{tl} = \mathbf{1}\left\{\hat{p}\left(X_{t-1:t-s}\right) - \frac{1}{2} > 0\right\}, \tag{12}$$

где:

$$\hat{p}\left(X_{t-1:t-s}\right) = F\left(\sum_{k=1}^{m} \hat{b}_k \psi_k\left(X_{t-1:t-s}\right)\right). \tag{13}$$

Прогнозирование $\hat{x}_{t+1,l}$ на следующий шаг осуществляется аналогично (12) , только фрагмент $X_{t-1:t-s} = \left(X_{t-1}, \ldots, X_{t-s}\right)$ в (13) заменяется на $\left(\hat{X}_t, \ldots, X_{t-s+1}\right)$, и т.д.

ЛИТЕРАТУРА

1. Fokianos K., Fried R., Kharin Yu., Voloshko V. Statistical analysis of multivariate discrete-valued time series / K. Fokianos, R. Fried, Yu. Kharin, V. Voloshko // J. Multivar, Anal, – 2022. - Vol. 188. – https://doi.org/10.1016 /j.jmva.2021.104805.
2. Шибалко, С.А. Статистическое оценивание параметров полносвязных многомерных двоичных временных рядов / С.А. Шибалко // 80-я научная конференция студентов и аспирантов Белорусского государственного университета [Электронный ресурс] : материалы конф., Минск, 10–20 марта 2023 г.: в 3 ч., ч. 1 / Белорус. гос. ун-т ; редкол.: А. В. Блохин (гл. ред.) [и др.]. – Минск : БГУ, 2023. – 1 электрон. опт. диск (CD-ROM). – С. 419-423.
3. Харин, Ю.С. Оптимальность и робастность в статистическом прогнозировании / Ю. С. Харин – Минск: БГУ, 2008. – 263 с.
4. Kharin, Yu. Robust estimation for binomial conditionally nonlinear autoregressive time series based on multivariate conditional frequencies / Yu. Kharin, V. Voloshko // J. Multivar. Anal. – 2021. – Vol. 185. – https://doi.org/10.1016/j.jmva.2021.104777.