СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ ПОЛНОСВЯЗНЫХ МНОГОМЕРНЫХ ДВОИЧНЫХ ВРЕМЕННЫХ РЯДОВ

С.А. Шибалко

shibalko2003@bk.ru;
Научный руководитель– Ю.С. Харин, доктор физико-математических наук, профессор

В статье предложена малопараметрическая модель на основе базисных функций многомерного двоичного временного ряда. Для параметров предложенной модели строится состоятельная асимптотически нормальная статистическая оценка. Для построенной оценки найдены асимптотические смещение и вариация. Полученные результаты могут быть использованы для статистического анализа двоичных временных рядов.

Ключевые слова: дискретный временной ряд; малопараметрическая модель; цепь Маркова высокого порядка; асимптотические свойства оценки; статистическое оценивание параметров.

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.

Примеры прикладных задач статистического анализа дискретных временных рядов [1]:

- Генетика (анализ генетических последовательностей);
- Экономика финансы (анализ динамики «рост-падение» биржевых курсов);
- Защита информации(потоки данных в компьютерных системах защиты информации).

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Пусть X_t – это N-мерная однородная двоичная цепь Маркова (N-ДЦМ) порядка $s \geq 1$, определенная на вероятностном пространстве $(\Omega, \mathcal{F}, \mathrm{P})$:

$$X_t = \begin{pmatrix} x_{t1} \\ \vdots \\ x_{tN} \end{pmatrix} \in V^N, t \in Z, \quad (1)$$

где $x_{ti} \in V = \{0,1\}$ – двоичная случайная величина, задающая i-ю компоненту в момент времени t.

Рассмотрим ситуацию, когда при фиксированной s-предыстории:

$$C_t = \{X_{t-1} = J_{t-1}, \dots, X_{t-s} = J_{t-s}\}$$

случайные величины $x_{t1}, \dots, x_{tN}$ зависимы так, что:

$$\mathrm{P}\{X_t = J_t | C_t\} = \mathrm{P}\{x_{t1} = j_{t1} | C_t\} \cdot \mathrm{P}\{x_{t2} = j_{t2} | x_{t1} = j_{t1}, C_t\} \cdot \ldots \cdot \mathrm{P}\{x_{tN} = j_{tN} | x_{t1} = j_{t1}, \dots, x_{t,N-1} = j_{t,N-1}, C_t\}, \qquad (2)$$

где условное распределение i-й компоненты $x_{ti} \in V$ при условии, что фиксирована s-предыстория, представимо в виде:

для первой компоненты:

$$\mathrm{P}\{x_{t1} = j_{t1} | X_{t-1} = J_{t-1}, \ . \ . \ . \ , X_{t-s} = J_{t-s}\} = \begin{cases} p_1(J_{t-s}, \dots, J_{t-1}), & j_{t,i} = 1, \\ 1 - p_1(J_{t-s}, \dots, J_{t-1}), & j_{t,i} = 0. \end{cases} \qquad (3)$$

для остальных компонент ($i = 2, \dots, N$):

$$\mathrm{P}\{x_{ti} = j_{ti} | x_{t1} = j_{t1}, \dots, x_{ti-1} = j_{ti-1}, C_t\} = \begin{cases} p_i(x_{t1}, \dots, x_{ti-1}, J_{t-s}, \dots, J_{t-1}), & j_{ti} = 1, \\ 1 - p_i(x_{t1}, \dots, x_{ti-1}, J_{t-s}, \dots, J_{t-1}), & j_{ti} = 0. \end{cases} \qquad (4)$$

МАЛОПАРАМЕТРИЧЕСКАЯ МОДЕЛЬ

Малопараметрическая модель на основе базисных функций (для упрощения записи индекс $i \in \{1, \dots, N\}$ опущен) для (4) имеет вид:

$$p = p(J_{1:s}) = F(\textstyle\sum_{k=1}^{m} b_k \psi_k(J_{1:s})), J_{1:s} \in V^{Ns}, \qquad (5)$$

где $F(\cdot)$ – некоторая заданная абсолютно непрерывная функция распределения, $B = (b_k) \in R^m$ – вектор-столбец $m \leq 2^{Ns}$ неизвестных коэффициентов N-ДЦМ, $\{\psi_k(J_{1:s})\}$ – базисные функции, $J_{1:s} = (J_1', \dots, J_s')' \in V^{Ns}$ – составной вектор – столбец, задающий s – предысторию.

Пусть наблюдается реализация длины T для N-ДЦМ (1)-(4):

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

Используя FBE (Frequencies-Based Estimation) – метод, предложенный в [2], оценим вектор параметров B модели (5) по наблюдаемой реализации $X_{1:T}$.

Примем следующие обозначения: $\mathbf{1}\{C\}$ –идикаторная функция события C;

для первой компоненты:

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

$$v_{s+1}^T(J_{1:s}; 1) \sum_{t=s}^{T} \mathbf{1}\left\{x_{t+1,1} = 1, X_t = J_s, \dots, X_{t-s+1} = J_1\right\}$$

для остальных компонент ($i = 2, \dots, N$):

$$v_s^T(J_{1:s}) = \sum_{t=s}^{T-1} \mathbf{1}\begin{Bmatrix} x_{t+1,i-1} = j_{t+1,i-1}, \dots, x_{t+1,1} = j_{t+1,1}, \\ X_t = J_s, \dots, X_{t-s+1} = J_1 \end{Bmatrix},$$

$$v_{s+1}^T(J_{1:s}; 1) \sum_{t=s}^{T} \mathbf{1}\begin{Bmatrix} x_{t+1,i} = 1, x_{t+1,i-1} = j_{t+1,i-1}, \dots, x_{t+1,1} = j_{t+1,1}, \\ X_t = J_s, \dots, X_{t-s+1} = J_1 \end{Bmatrix} -$$

частоты s-грамм $J_{1:s}$ и $(J_{1:s}; 1)$;

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

Где

$$\boldsymbol{J}^{(s)} = \{J_{1:s} \in V^{Ns} : \ v_s^T(J_{1:s}) > 0\} \subseteq V^{Ns} -$$

подмножество s-грамм, имеющих ненулевые частоты в $X_{1:T}$;

$$\hat{u}(J_{1:s}) = F^{-1}\big(\hat{p}(J_{1:s})\big) \in R^1;$$
$$D = \sum_{J_{1:s} \in J^{(s)}} \Psi(J_{1:s})\Psi^T(J_{1:s}) \ \in \ R^{m\times m}, \tag{6}$$

где $F^{-1}(\cdot)$ −квантильная функция,

$$\Psi(J_{1:s}) = \{\psi_k(J_{1:s})\} \in \ R^{m\times 1},$$

$$E = \sum_{J_{1:s} \in J^{(s)}} \hat{u}(J_{1:s})\Psi(J_{1:s}) \ \in \ R^{m\times 1}.$$

АСИМПТОТИЧЕСКИЕ СВОЙСТВА ОЦЕНКИ $\widehat{B}$

Теорема 1. Если N-ДЦМ эргодична и определитель матрицы D определяемой (6), $|D| \neq 0$, то FBE-оценка имеет вид:

$$\hat{B} = \big(\widehat{b_k}\big) = D^{-1}E$$

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

$$\hat{B} \xrightarrow{P} B^0,$$

является асимптотически несмещенной:

$$E\{\hat{B} - B^0\} = \ E\{\hat{B}\} - B^0 \underset{T\to\infty}{\longrightarrow} 0,$$

а ее асимптотическая ковариационная матрица в случае $i-$ой компоненты ($i = 1, \dots, N$):

$$E\left\{T\left(\hat{B}-B^0\right)\left(\hat{B}-B^0\right)'\right\}\underset{T\to\infty}{\longrightarrow} D^{-1}\Psi\bar{F}\left(\sum\nolimits_p\right)^{-1}\bar{F}'\Psi'\left(D^{-1}\right)',$$

Где

$$\sum\nolimits_p = diag\left(p_i(1-p_i)\right) \in R^{2^{s+(i-1)}\times 2^{s+(i-1)}},$$

$$\bar{F} = \left(F^{-1'}(p_i)\right) \in R^{2^{s+(i-1)}}.$$

Теорема 2. В условиях Теоремы 1 FBE-оценка $\hat{B}$ имеет асимптотически нормальное распределение:

$$\sqrt{T}(\hat{B}-B)\underset{T\to\infty}{\longrightarrow}\mathcal{N}_m(0,\Sigma),$$

$$\Sigma = D^{-1}\,\Psi\bar{F}\sum\nolimits_p\bar{F}'\Psi'\left(D^{-1}\right)'.$$

Заключение

В работе получены следующие результаты:

1. для многомерных двоичных временных рядов предложена малопараметрическая модель на основе базисных функций;
2. построена состоятельная асимптотически нормальная статистическая оценка параметров предложенной модели;

для построенной оценки найдены асимптотические смещение и вариация.

Полученные результаты могут быть использованы для решения, указанных во введении, прикладных задач статистического анализа дискретных временных рядов.

Библиографические ссылки

1. *Харин, Ю.С.* Оптимальность и робастность в статистическом прогнозировании / Ю. С. Харин - Минск: БГУ, 2008. 263 с
2. *Yu. S. Kharin, V. A. Voloshko, E. A. Medved*, Statistical Estimation of Parameters for Binary Conditionally Nonlinear Autoregressive Time Series / Mathematical Methods of Statistics. -2018. Vol. 26. No 2. P 103-118