ISSN 0002–354X (print)
УДК 519.2

Поступило в редакцию 06.10.2016
Received 06.10.2016

Член-корреспондент Ю. С. Харин, М. В. Мальцев, Н. С. Сологуб

НИИ прикладных проблем математики и информатики Белорусского государственного университета, Минск, Республика Беларусь

ВЕКТОРНАЯ ЦЕПЬ МАРКОВА С ЧАСТИЧНЫМИ СВЯЗЯМИ И СТАТИСТИЧЕСКИЕ ВЫВОДЫ О ЕЕ ПАРАМЕТРАХ

Предложена новая малопараметрическая модель дискретных временных рядов – однородная векторная цепь Маркова s-го порядка с частичными связями, для которой условное распределение вероятностей определяется лишь некоторыми компонентами предыдущих векторов-состояний. Установлены вероятностные свойства модели: критерий эргодичности, условия, при которых стационарное распределение вероятностей является равномерным. Построены состоятельные статистические оценки параметров модели.

Ключевые слова: векторная цепь Маркова с частичными связями, критерий эргодичности, статистическое оценивание параметров.

Corresponding Member Yu. S. Kharin, M. V. Maltsew, N. S. Sologub

Research Institute for Applied Problems of Mathematics and Informatics of the Belarusian State University, Minsk, Republic of Belarus

VECTOR MARKOV CHAIN WITH PARTIAL CONNECTIONS AND STATISTICAL INFERENCES ON ITS PARAMETERS

A new mathematical model of discrete time series is proposed. It is called homogenous vector Markov chain of the order s with partial connections. The conditional probability distribution for this model is determined only by a few components of previous vector states. Probabilistic properties of the model are given: ergodicity conditions and conditions under which the stationary probability distribution is uniform. Consistent statistical estimators for model parameters are constructed.

Keywords: vector Markov chain with partial connections, ergodicity conditions, statistical estimation of parameters.

Введение. Цепь Маркова – широко используемая на практике математическая модель дискретных временных рядов, которая применяется в экономике [1], генетике [2], социологии [3]. Еще одной важной областью приложения этой модели является защита информации [4]. Неотъемлемыми элементами систем защиты информации являются криптографические генераторы – программные, аппаратные или программно-аппаратные устройства, предназначенные для получения случайных или псевдослучайных последовательностей [5]. Для псевдослучайной последовательности каждый ее элемент является функцией предыдущих элементов, поэтому в выходных последовательностях криптографических генераторов присутствуют зависимости, как правило, большой глубины. Для описания таких зависимостей адекватной моделью является цепь Маркова порядка $s \gg 1$ [6]. К сожалению, непосредственно использовать цепь Маркова порядка s зачастую затруднительно, поскольку число параметров этой модели увеличивается экспоненциально с ростом s. В связи с этим в практических приложениях необходимы так называемые малопараметрические (parsimonious) марковские модели, число параметров которых зависит от s полиномиально [7]. К таким моделям относится разработанная в Белорусском государственном университете цепь Маркова порядка s с r частичными связями [8], для которой условное распределение вероятностей зависит не от всех s предыдущих состояний, а только от $r < s$ избранных. Целью данного сообщения является обобщение этой модели для векторной цепи Маркова с $m \geq 2$ компонентами и ее вероятностно-статистический анализ. Такая ситуация часто возникает в приложениях, связанных с анализом динамики многомерных стохастических данных.

Математическая модель. Примем обозначения: $\mathbb{N}$ – множество натуральных чисел; $A = \{0, 1, ..., N - 1\}$ – множество мощности $|A| = N \geq 2$; $m \in \mathbb{N}$ – число, которое будем называть размерностью цепи Маркова, $J_i = (j_{i1},..., j_{im}) \in A^m$, $i = 1, 2, ...,$ – m-мерный целочисленный вектор; $J_a^b = (J_a, ..., J_b)$, $a, b \in \mathbb{N},\ a \le b,$ – упорядоченный набор из $b-a+1$ m-мерных векторов; $\{x_t = (x_{t1},\ldots,x_{tm}) \in A^m : t \in \mathbb{N}\}$ – заданная на вероятностном пространстве (Ω, F, P) однородная векторная цепь Маркова порядка s с пространством состояний A^m, начальным распределением вероятностей

$$\pi^{(0)}_{J_1^s} = P\{x_1 = J_1,..., x_s = J_s\},\ J_1,\ldots,J_s \in A^m, \tag{1}$$

и матрицей вероятностей одношаговых переходов $P = (p_{J_1^s, J_{s+1}})$,

$$p_{J_1^s, J_{s+1}} = P\{x_t = J_{s+1} \mid x_{t-1} = J_s, ..., x_{t-s} = J_1\},\ J_1,\ldots,J_{s+1} \in A^m,\ t = s+1,\ s+2, ... \tag{2}$$

Такую цепь Маркова будем обозначать ВЦМ(s) – векторная цепь Маркова порядка s.

Число независимых параметров матрицы P с учетом условия нормировки равно

$$D_s = N^{ms}(N^m - 1).$$

В табл. 1 для различных порядков s при $m = 8$ указано число параметров двоичной ВЦМ(s).

Т а б л и ц а 1. **Число параметров двоичной ВЦМ(*s*)**
T a b l e 1. **The number of parameters for the binary VMC(*s*)**

s	1	2	4	8	16
D_s	65280	16711680	$\approx 1{,}095 \cdot 10^{12}$	$\approx 4{,}704 \cdot 10^{21}$	$\approx 8{,}677 \cdot 10^{40}$

Число параметров ВЦМ(s) возрастает экспоненциально, и непосредственное применение этой модели на практике возможно лишь при небольших значениях мощности N, размерности m и порядка цепи Маркова s. В связи с этим, развивая [8], построим модификацию этой модели, для которой условное распределение вероятностей определяется лишь некоторыми «значимыми» компонентами предыдущих векторов-состояний. Обозначим:

$$M_r = \{(k_1, l_1), (k_2, l_2), \ldots, (k_r, l_r)\} \subseteq M_* = \{(k, l) : 1 \le k \le s, 1 \le l \le m\} \ –$$

шаблон-множество, представляющее собой упорядоченный в лексикографическом порядке набор $1 \le r \le sm$ различных значений пар индексов, причем $k_1 = 1$; $\mathbf{M}_r$ – множество всевозможных таких шаблонов; $S_{M_r}(J_t, ..., J_{t+s-1}) = (j_{t+k_1-1, l_1}, ..., j_{t+k_r-1, l_r})$, $t = 1, 2,...,$ – функция-селектор, которая в соответствии с шаблон-множеством M_r «вырезает» r компонент из множества ms компонент $\{j_{u,l} : t \le u \le t+s-1,\ 1 \le l \le m\} : S_{M_r} : A^{ms} \to A^r$; $Q = (q_{(i_1,...,i_r), I_{r+1}})$ – некоторая стохастическая $N^r \times N^m$-матрица, $i_1,\ldots,i_r \in A,\ I_{r+1} \in A^m$.

О п р е д е л е н и е. *Если вероятности одношаговых переходов (2) допускают следующее малопараметрическое представление:*

$$p_{J_1^s, J_{s+1}} = q_{S_{M_r}(J_1,...,J_s), J_{s+1}} = q_{(j_{k_1,l_1},...,j_{k_r,l_r}), J_{s+1}},\ J_1,\ldots,J_{s+1} \in A^m, \tag{3}$$

то ВЦМ(s) назовем векторной цепью Маркова с r частичными связями и шаблоном связей M_r; обозначать такую модель будем ВЦМ(s, r).

Из приведенного определения следует, что условное распределение вероятностей состояния x_t временного ряда в момент времени t зависит не от всех ms компонент s прошлых состояний, а только от r избранных компонент, которые определяются шаблон-множеством M_r. Если $r = sm$, то $M_r = M_*$, и в этом случае приходим к полносвязной векторной цепи Маркова s-го порядка: ВЦМ(s, sm) ≡ ВЦМ(s). Если $m = 1$, то ВЦМ(s, r) превращается в ранее разработанную модель ЦМ(s, r) [8].

Таким образом, ВЦМ(s, r) размерности m определяется следующими параметрами: $s \ge 1$ – порядок цепи Маркова; $r \in \{1, ..., sm\}$ – число связей; $M_r \in \mathbf{M}_r$ – шаблон связей; Q – стохастическая $N^r \times N^m$-матрица. Число независимых параметров для ВЦМ(s, r) равно

$$d = N^r(N^m - 1) + 2r - 1.$$

В табл. 2 для различных значений порядка s и числа связей r при $m = 8$ указано число параметров двоичной ВЦМ(s, r).

Т а б л и ц а 2. **Число параметров двоичной ВЦМ(*s*, *r*)**
T a b l e 2. **The numbers of parameters for the binary VMC(*s*, *r*)**

(s, r)	(1, 2)	(2, 4)	(4, 6)	(8, 8)	(16, 10)	(32, 16)
d	1023	4087	16331	65295	261139	16711711

Из сравнения табл. 1 и 2 виден существенный выигрыш в числе параметров модели ВЦМ(s, r) по сравнению с полносвязной моделью ВЦМ(s).

Т е о р е м а 1. *Однородная векторная цепь Маркова с частичными связями ВЦМ*(s, r) *является эргодической тогда и только тогда, когда найдется такое* $c \in \mathbb{N}$, *что выполняется неравенство*

$$\min_{J_1^s,\, J_{c+1}^{c+s} \in A^{ms}} \sum_{J_{s+1}^c \in A^{m(c-s)}} \prod_{t=1}^{c} q_{S_{M_r}(J_t,\ldots,J_{t+s-1}),J_{t+s}} > 0. \tag{4}$$

Д о к а з а т е л ь с т в о. Для доказательства рассмотрим эквивалентную x_t цепь Маркова первого порядка с расширенным пространством состояний:

$$\overline{x}_t = (x_{t,1},\ldots,x_{t,m}, x_{t+1,1},\ldots,x_{t+1,m},\ldots,x_{t+s-1,1},\ldots,x_{t+s-1,m}) \in A^{ms},$$

матрица вероятностей одношаговых переходов которой имеет вид

$$\overline{P} = \left(\overline{p}_{I_1^s,J_1^s}\right),\ \overline{p}_{I_1^s,J_1^s} = \mathrm{I}\{I_2^s = J_1^{s-1}\} p_{I_1^s,J_s}, \tag{5}$$

где $\mathrm{I}\{C\}$ – индикаторная функция события C, а $p_{I_1^s,J_s}$ определяется (3). Тогда эргодичность x_t эквивалентна эргодичности $\overline{x}_t$. Согласно критерию эргодичности, для цепи Маркова первого порядка [9] $\overline{x}_t$ является эргодической тогда и только тогда, когда найдется такое $c \in \mathbb{N}$, что выполняется неравенство:

$$\min_{J_1^s,\, J_{c+1}^{c+s} \in A^{ms}} \overline{p}^{(c)}_{J_1^s,J_{c+1}^{c+s}} > 0,$$

где $\overline{p}^{(c)}_{J_1^s,J_{c+1}^{c+s}}$ – вероятность перехода из J_1^s в J_{c+1}^{c+s} для цепи Маркова $\overline{x}_t$ за c шагов. В работе [10] получено следующее представление для этой вероятности:

$$\overline{p}^{(c)}_{J_1^s,J_{c+1}^{c+s}} = \sum_{J_{s+1}^c \in A^{m(c-s)}} \prod_{t=1}^{c} p_{J_t^{t+s-1},J_{t+s}},$$

откуда, используя равенство (3), получаем требуемый в теореме результат (4). □

С л е д с т в и е 1. *Если все элементы матрицы* Q *положительны, то ВЦМ*(s, r) *эргодична.*

Пусть выполнен критерий эргодичности, сформулированный в теореме 1. Тогда, согласно [6; 9], существует стационарное распределение вероятностей, которое будем обозначать $(\pi_{J_1^s})$, $J_1^s \in A^{ms}$; его маргинальные распределения обозначим

$$\pi_s^{M_r}(i_1,\ldots,i_r) = \mathrm{P}\{S_{M_r}(x_t,\ldots,x_{t+s-1}) = (i_1,\ldots,i_r)\},$$

$$\pi_{s+1}^{M_r}(i_1,\ldots,i_r,I_{r+1}) = \mathrm{P}\{S_{M_r}(x_t,\ldots,x_{t+s-1}) = (i_1,\ldots,i_r), x_{t+s} = I_{r+1}\} = \pi_s^{M_r}(i_1,\ldots,i_r) q_{(i_1,\ldots,i_r),I_{r+1}}.$$

Стационарное распределение вероятностей является решением системы линейных алгебраических уравнений

$$\begin{cases} \sum\limits_{J_1^s \in A^{ms}} \overline{p}_{J_1^s,I_1^s} \pi_{J_1^s} = \pi_{I_1^s},\ I_1^s \in A^{ms}, \\ \sum\limits_{J_1^s \in A^{ms}} \pi_{J_1^s} = 1. \end{cases}$$

Установим условия, при которых стационарное распределение вероятностей ВЦМ(s, r) является равномерным:

$$\pi_{J_1^s} = N^{-ms},\ J_1^s \in A^{ms}.$$

Т е о р е м а 2. *Если для матрицы Q эргодической ВЦМ(s, r) выполняется условие*

$$\sum_{I_1 \in A^m} q_{S_{M_r}(I_1,\ldots,I_s),\, I_{s+1}} = 1,\ I_2,\ldots,I_{s+1} \in A^m, \tag{6}$$

то стационарное распределение вероятностей этой модели является равномерным.

Д о к а з а т е л ь с т в о. Как и в теореме 1, для доказательства будем использовать эквивалентную x_t цепь Маркова первого порядка $\bar{x}_t$ с матрицей вероятностей одношаговых переходов (5). Известно [11], что стационарное распределение вероятностей цепи Маркова первого порядка является равномерным тогда и только тогда, когда ее матрица вероятностей одношаговых переходов бистохастическая:

$$\sum_{I_1^s \in A^{ms}} \bar{p}_{I_1^s,\, J_1^s} = 1,\ J_1^s \in A^{ms}. \tag{7}$$

Преобразуем (7) с учетом (3) и (5):

$$\sum_{I_1^s \in A^{ms}} \bar{p}_{I_1^s,\, J_1^s} = \sum_{I_1^s \in A^{ms}} \mathrm{I}\{I_2^s = J_1^{s-1}\} p_{I_1^s,\, J_s} = \sum_{I_1 \in A^m} q_{S_{M_r}(I_1, J_1,\ldots, J_{s-1}),\, J_s},$$

что соответствует (6). □

Статистическое оценивание параметров. Для статистического оценивания параметров ВЦМ(s, r) построим вначале функцию правдоподобия. Введем обозначения:

$X^{(n)} = (x_1,\ldots, x_n) \in A^{mn}$ – наблюдаемая реализация длины n векторной цепи Маркова ВЦМ(s, r);

$$\nu_{s+1}^{M_r}(i_1,\ldots, i_r, I_{r+1}) = \sum_{t=1}^{n-s} \mathrm{I}\{S_{M_r}(x_t,\ldots,x_{t+s-1}) = (i_1,\ldots, i_r),\, x_{t+s} = I_{r+1}\},$$

$$\nu_s^{M_r}(i_1,\ldots, i_r) = \sum_{I_{r+1} \in A^m} \nu_{s+1}^{M_r}(i_1,\ldots,i_r, I_{r+1}),\ (i_1,\ldots,i_r) \in A^r,\, I_{r+1} \in A^m\ –$$

частотные статистики ВЦМ(s, r), вычисленные по реализации $X^{(n)}$.

Л е м м а. *Если известны порядок цепи Маркова s, число связей r и шаблон-множество M_r, то функция правдоподобия для ВЦМ(s, r), построенная по реализации $X^{(n)}$, имеет вид*

$$L_n(X^{(n)}, Q) = \pi_{x_1,\ldots,x_s}^{(0)} \prod_{t=s}^{n-1} q_{S(x_{t-s+1},\ldots,x_t),\, x_{t+1}}. \tag{8}$$

Д о к а з а т е л ь с т в о. Соотношение (8) следует из представления для n-мерного распределения вероятностей, которое получаем, используя обобщенную формулу умножения вероятностей, марковское свойство и определение ВЦМ(s, r):

$$\mathrm{P}\{X_1^n = J_1^n\} = \mathrm{P}\{X_1^s = J_1^s\} \prod_{t=s}^{n-1} \mathrm{P}\{x_{t+1} = J_{t+1} | X_1^t = J_1^t\} = \pi_{J_1,\ldots,J_s}^{(0)} \prod_{t=s}^{n-1} q_{S_{M_r}(J_{t-s+1},\ldots,J_t),\, J_{t+1}}.\ \square$$

С л е д с т в и е 2. *Логарифмическая функция правдоподобия для ВЦМ(s, r) имеет вид*

$$l_n(X^{(n)}, Q) = \ln \pi_{J_1,\ldots,J_s}^{(0)} + \sum_{\substack{i_1,\ldots,i_r \in A,\\ I_{r+1} \in A^m}} \nu_{s+1}^{M_r}(i_1,\ldots,i_r, I_{r+1}) \ln q_{(i_1,\ldots,i_r),\, I_{r+1}}. \tag{9}$$

Построим теперь оценку $\hat{Q} = (\hat{q}_{(i_1,\ldots,i_r),\, I_{r+1}})$ максимального правдоподобия (ОМП) матрицы Q.

Т е о р е м а 3. *Если известны порядок цепи Маркова s, число связей r и шаблон-множество M_r, то ОМП вероятностей одношаговых переходов $q_{(i_1,\ldots,i_r),\, I_{r+1}}$ имеют вид*

$$\hat{q}_{(i_1,\ldots,i_r),\,I_{r+1}}=\begin{cases}\dfrac{\nu_{s+1}^{M_r}(i_1,\ldots,i_r,I_{r+1})}{\nu_s^{M_r}(i_1,\ldots,i_r)}, \text{ если } \nu_s^{M_r}(i_1,\ldots,i_r)>0,\\ 1/N^m, \text{ если } \nu_s^{M_r}(i_1,\ldots,i_r)=0.\end{cases} \tag{10}$$

Д о к а з а т е л ь с т в о. Для нахождения ОМП требуется решить задачу на условный экстремум:

$$\begin{cases} l_n(X^{(n)},Q)\to \max\limits_{Q},\\ \sum\limits_{I_{r+1}\in A^m} q_{(i_1,\ldots,i_r),I_{r+1}}=1,\ i_1,\ldots,i_r\in A.\end{cases}$$

Эта задача распадается на N^r подзадач отыскания условного максимума для каждого набора $(i_1, ..., i_r)$:

$$\begin{cases}\sum\limits_{I_{r+1}\in A^m}\nu_{s+1}^{M_r}(i_1,\ldots,i_r,\,I_{r+1})\ln q_{(i_1,\ldots,i_r),I_{r+1}}\to \max\limits_{q_{(i_1,\ldots,i_r),I_{r+1}}},\\ \sum\limits_{I_{r+1}\in A^m} q_{(i_1,\ldots,i_r),I_{r+1}}=1,\ \ i_1,\ldots,i_r\in A.\end{cases} \tag{11}$$

Решив полученные задачи (11) методом множителей Лагранжа, приходим к оценкам (10). □

Т е о р е м а 4. *Если ВЦМ(s, r) является стационарной, то при $n\to\infty$ оценки (10) являются состоятельными в смысле сходимости по вероятности*:

$$\hat{q}_{(i_1,\ldots,i_r),\,I_{r+1}}\xrightarrow{\mathrm{P}} q_{(i_1,\ldots,i_r),\,I_{r+1}},\ i_1,\ldots,i_r\in A,\, I_{r+1}\in A^m.$$

Д о к а з а т е л ь с т в о. Для стационарной цепи Маркова порядка s при $n\to\infty$ справедлива сходимость нормированных частот состояний к стационарному распределению [10]:

$$\hat{\pi}_{J_1^{s+1}}=\frac{1}{n-s}\sum_{t=1}^{n-s}\mathrm{I}\{x_t=J_1,\ldots,x_{t+s}=J_{s+1}\}\xrightarrow{\mathrm{P}}\pi_{J_1^{s+1}}=\pi_{J_1^s}p_{J_1^s,J_{s+1}}.$$

Поскольку частоты $\nu_{s+1}^{M_r}(i_1,\ldots,i_r,I_{r+1})$, $\nu_s^{M_r}(i_1,\ldots,i_r)$, входящие в ОМП (10), – суммы частот $(s + 1)$-грамм, то также справедлива аналогичная сходимость при $n\to\infty$:

$$\hat{\pi}_s^{M_r}(i_1,\ldots,i_r)=\frac{1}{n-s}\nu_s^{M_r}(i_1,\ldots,i_r)\xrightarrow{\mathrm{P}}\pi_s^{M_r}(i_1,\ldots,i_r),$$

$$\hat{\pi}_{s+1}^{M_r}(i_1,\ldots,i_r,\,I_{r+1})=\frac{1}{n-s}\nu_{s+1}^{M_r}(i_1,\ldots,i_r,\,I_{r+1})\xrightarrow{\mathrm{P}}\pi_{s+1}^{M_r}(i_1,\ldots,i_r,\,I_{r+1})=\pi_s^{M_r}(i_1,\ldots,i_r)q_{(i_1,\ldots,i_r),\,I_{r+1}},$$

откуда с учетом теоремы о функциональном преобразовании сходящихся по вероятности случайных последовательностей [13] получаем требуемый в теореме результат. □

Строить статистическую оценку шаблон-множества, как и матрицы Q, при известных значениях s, r будем с помощью метода максимального правдоподобия. Обозначим: $\mathrm{M}_r^+(M_{r-1})$ – множество шаблонов, которые получаются за счет расширения шаблона M_{r-1} на одну компоненту, которая выбирается из множества $M_* \setminus M_{r-1}$, $r = 2, 3,...$;

$$\widehat{H}(M_r)=-\sum_{\substack{i_1,\ldots,i_r\in A,\\ I_{r+1}\in A^m}}\nu_{s+1}^{M_r}(i_1,\ldots,i_r,I_{r+1})\ln\frac{\nu_{s+1}^{M_r}(i_1,\ldots,i_r,I_{r+1})}{\nu_s^{M_r}(i_1,\ldots,i_r)}\ -$$

построенная по подстановочному принципу оценка условной энтропии Шеннона для будущего вектора x_{t+1} при условии состояния шаблона $i_1,...,i_r$.

Т е о р е м а 5. *Если известны порядок цепи Маркова s и число связей r, то ОМП шаблон-множества M_r имеет вид*

$$\widehat{M}_r=\arg\min_{M_r\in\mathbf{M}_r}\widehat{H}(M_r).$$

Д о к а з а т е л ь с т в о. Требуемое утверждение следует из представления логарифмической функции правдоподобия (9). □

Построение оценки $\widehat{M}_r$ путем полного перебора всех возможных значений шаблон-множества требует $O(nmN^{r+m}(sm)^{r-1})$ операций, и использование полного перебора поэтому возможно только при малом числе связей r. Для уменьшения вычислительной сложности предлагается алгоритм пошагового увеличения шаблона, который заключается в последовательном построении оценок:

$$\widehat{M}_{r_-}, \widehat{M}_{r_-+1}, \ldots, \widehat{M}_r,$$

где $1 \le r_- \le r$. Начальный шаблон $\widehat{M}_{r_-}$ строится с помощью алгоритма полного перебора; начальное значение числа связей r_- определяется исходя из имеющихся вычислительных ресурсов. Пошаговое наращивание шаблона для $r' = r_- + 1, \ldots, r$, осуществляется по правилу

$$\widehat{M}_{r'} = \arg \min_{M_{r'} \in \mathrm{M}^{+}_{r'}(M_{r'-1})} \widehat{H}(M_{r'}).$$

Вычислительная сложность этого алгоритма имеет меньший порядок $O(nmN^{r+1}(sm)^{r_- -1})$.

Использование метода максимального правдоподобия для оценивания порядка цепи Маркова s и числа связей r приводит к проблеме, известной как over-fitting – чрезмерной «подгонке» модели под наблюдаемые данные [14]. Поэтому для построения оценок s и r предлагается использовать байесовский информационный критерий BIC [15], позволяющий преодолевать указанный недостаток. Оценки $(\hat{s}, \hat{r})$ определяются при решении задачи на минимум:

$$(\hat{s}, \hat{r}) = \arg \min_{2 \le s' \le s_+,\ 1 \le r' \le r_+} BIC(s', r'),$$

$$BIC(s', r') = -l_n(X^{(n)}, Q) + 2d \ln(n - s') =$$

$$-\sum_{\substack{i_1, \ldots, i_{r'} \in A, \\ I_{r'+1} \in A^m}} \nu_{s'+1}^{M_{r'}}(i_1, \ldots, i_{r'}, I_{r'+1}) \ln \frac{\nu_{s'+1}^{M_{r'}}(i_1, \ldots, i_{r'}, I_{r'+1})}{\nu_{s'}^{M_{r'}}(i_1, \ldots, i_{r'})} + (N^{r'}(N^m - 1) + 2r' - 1) \ln(n - s'),$$

где $s_+ \ge 1$, $1 \le r_+ \le ms_+$ – максимально допустимые значения параметров s и r соответственно, задаваемые априорно.

Численные результаты. Проиллюстрируем свойства построенных статистических оценок с помощью компьютерного моделирования. В экспериментах использовалась двоичная векторная цепь Маркова с частичными связями с параметрами: $m = 4$, $s = 4$, $r = 6$, $M_6 = \{(1, 1), (2, 2), (2, 4), (3, 1), (3, 2), (4, 3)\}$, элементы матрицы Q размерности $2^6 \times 2^4$ генерировались как случайные величины с равномерным распределением вероятностей на отрезке [0, 1] (с учетом условия нормировки). Моделировались независимые реализации ВЦМ(s, r), состоящие из n двоичных m-векторов, $n \in \{10^5, 2 \cdot 10^5, \ldots, 100 \cdot 10^5\}$. По (10) строились статистические оценки вероятностей одношаговых переходов, затем вычислялась среднеквадратическая ошибка оценивания всех 1024 элементов матрицы Q:

$$\Delta_n = \sum_{i_1, \ldots, i_6 \in \{0,1\}} \sum_{I \in \{0,1\}^4} \left(q_{(i_1, \ldots, i_6), I} - \hat{q}_{(i_1, \ldots, i_6), I}\right)^2.$$

График зависимости Δ_n от длины наблюдаемой последовательности n представлен на рисунке, иллюстрирующем состоятельность статистических оценок (10).

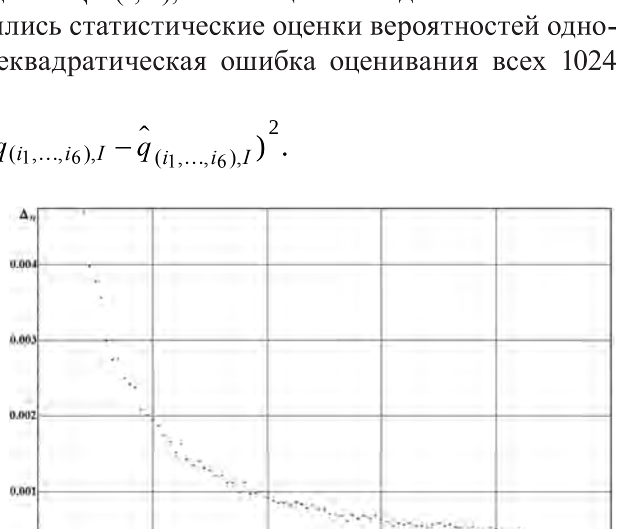

Иллюстрация состоятельности оценок (10)
Illustration of the consistency of estimators (10)

Заключение. В сообщении получены следующие основные результаты. Разработана новая малопараметрическая модель дискретных временных рядов – векторная цепь Маркова порядка s с частичными связями, для которой условное распределение вероятностей определяется лишь некоторыми компонентами предыдущих векторов-состояний. Доказан критерий эргодичности данной модели. Определены условия, при которых стационарное распределение векторной цепи Маркова с частичными связями является равномерным. Построены состоятельные статистические оценки матрицы вероятностей переходов, шаблона связей, порядка цепи Маркова и числа связей.

Список использованных источников

1. Кемени, Дж. Г. Конечные цепи Маркова / Дж. Г. Кемени, Дж. Л. Снелл. – М.: Наука, 1970. – 272 с.
2. Уотермен, М. С. Математические методы для анализа последовательностей ДНК / М. С. Уотермен. – М.: Мир, 1999. – 350 с.
3. Bonacich, P. Asymptotics of a matrix valued Markov chain arising in sociology / P. Bonacich, T. M. Liggett // Stochastic Processes and their Applications. – 2003. – Vol. 104, Issue 1. – P. 155–171. doi:10.1016/S0304-4149(02)00231-4.
4. Криптология / Ю. С. Харин [и др.]. – Минск: БГУ, 2013. – 512 с.
5. Зубков, А. М. Датчики псевдослучайных чисел и их применения / А. М. Зубков // Московский университет и развитие криптографии в России: материалы конф. в МГУ. – М.: МЦНМО, 2003. – С. 200–206.
6. Дуб, Дж. Л. Вероятностные процессы / Дж. Л. Дуб. – М.: Изд-во иностранной литературы, 1956. – 605 с.
7. Kharin, Yu. Parsimonious models for high-order Markov chains and their statistical analysis / Yu. Kharin // VIII World Congress on Probability and Statistics. – Istanbul: Publ. House of Koc. Univ., 2012. – P. 168–169.
8. Харин, Ю. С. Цепи Маркова с *r*-частичными связями и их статистическое оценивание / Ю. С. Харин // Докл. НАН Беларуси. – 2004. – Т. 48, № 1. – С. 40–44.
9. Ширяев, А. Н. Вероятность / А. Н. Ширяев. – Изд. 2-е – М.: Наука, 1989. – 640 с.
10. Харин, Ю. С. Алгоритмы статистического анализа цепей Маркова с условной глубиной памяти / Ю. С. Харин, М. В. Мальцев // Информатика. – 2011. – № 1. – С. 34–43.
11. Боровков, А. А. Теория вероятностей / А. А. Боровков. – М.: Наука, 1986. – 432 с.
12. Basawa, I. V. Statistical inference for stochastic processes / I. V. Basawa, B. L. S. Prakasa Rao. – London: Academic Press, 1980. – 438 p.
13. Крамер, Г. Математические методы статистики / Г. Крамер. – М.: Мир, 1975. – 648 с.
14. Cawley, G. C. On over-fitting in model selection and subsequent selection bias in performance evaluation / G. C. Cawley, N. L. C. Talbot // The Journal of Machine Learning Research. – 2010. – Vol. 11. – P. 2079–2107.
15. Csiszar, I. The consistency of the BIC Markov order estimator / I. Csiszar, P. C. Shields // The Annals of Statistics. – 2000. – Vol. 28, N 6. – P. 1601–1619. doi:10.1214/aos/1015957472.

References

1. Kemeni J. G., Snell J. L. *Finite Markov chains*. Moscow, Nauka, 1970. 272 p. (in Russian)
2. Woterman M. S. *Mathematical methods for analysis of DNA sequences*. Moscow, Mir, 1999. 350 p. (in Russian)
3. Bonacich P., Liggett T. M. Asymptotics of a matrix valued Markov chain arising in sociology. *Stochastic Processes and their Applications*, 2003, vol. 104, is. 1, pp. 155–171. doi:10.1016/S0304-4149(02)00231-4.
4. Kharin Yu. S., Agievich S. V., Vasiliev D. V., Matveev G. V. *Cryptology.* Minsk, Belarusian State University Publ., 2013. 512 p. (in Russian)
5. Zubkov A. M. Random data generators and their application. *Moskovskii universitet i razvitie kriptografii v Rossii: materialy konferentsii v Moskovskom gosudarstvennom universitete* [Moscow State University and the cryptography development in Russia: Proceedings of the Conference at the Moscow State University]. Moscow, Moscow Center for Continuous Mathematical Education Publ., 2003, pp. 200–206. (in Russian)
6. Dub J. L. *Probability processes*. Moscow, Foreign Literature Publishing House, 1956. 605 p. (in Russian)
7. Kharin Yu. Parsimonious models for high-order Markov chains and their statistical analysis. *VIII World Congress on Probability and Statistics.* Istanbul, Publ. House of Koc. Univ., 2012, pp. 168–169.
8. Kharin Yu. S. Markov chains with the *r*-partial connections and their statistical estimation. *Doklady Nacional'noj akademii nauk Belarusi* [Doklady of the National Academy of Sciences of Belarus], 2004, vol. 48, no. 1, pp. 40–44. (in Russian)
9. Shiryaev A. N. *Probability*. 2nd ed. Moscow, Nauka, 1989. 640 p. (in Russian)
10. Kharin Yu. S., Maltsev M. V. Algorithms of statistical analysis of the Markov chains with conditional memory depth. *Informatika* [Informatics], 2011, no. 1, pp. 34–43. (in Russian)
11. Borovkov A. A. *The theory of probabilities*. Moscow, Nauka, 1986. 432 p. (in Russian)
12. Basawa I. V., Prakasa Rao B. L. S. *Statistical inference for stochastic processes*. London, Academic Press, 1980. 438 p.
13. Kramer G. *Mathematical methods of statistics*. Moscow, Mir, 1975. 648 p. (in Russian)
14. Cawley G. C., Talbot N. L. C. On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, 2010, vol. 11, pp. 2079–2107.
15. Csiszar I., Shields P. C. The consistency of the BIC Markov order estimator. *The Annals of Statistics*, 2000, vol. 28, no. 6, pp. 1601–1619. doi:10.1214/aos/1015957472.

Информация об авторах

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

Мальцев Михаил Владимирович – ст. науч. сотрудник, заведующий лабораторией, НИИ прикладных проблем математики и информатики Белорусского государственного университета (пр. Независимости, 4, 220030, Минск, Республика Беларусь). E-mail: maltsew@bsu.by.

Сологуб Надежда Сергеевна – лаборант, НИИ прикладных проблем математики и информатики Белорусского государственного университета (пр. Независимости, 4, 220030, Минск, Республика Беларусь). E-mail: nssologub@gmail.com.

Для цитирования

Харин, Ю. С. Векторная цепь Маркова с частичными связями и статистические выводы о ее параметрах / Ю. С. Харин, М. В. Мальцев, Н. С. Сологуб // Докл. НАН Беларуси. – 2016. – Т. 60, № 6. – С. 14–21.

Information about the authors

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

Maltsew Michael Vladimirovich – Senior researcher, Head of the Research Laboratory, Research Institute for Applied Problems of Mathematics and Informatics of the Belarusian State University (4, Nezavisimosti Ave., 220030, Minsk, Republic of Belarus). E-mail: maltsew@bsu.by.

Sologub Nadezhda Sergeevna – Assistant, Research Institute for Applied Problems of Mathematics and Informatics of the Belarusian State University (4, Nezavisimosti Ave., 220030, Minsk, Republic of Belarus). E-mail: nssologub@gmail.com.

For citation

Kharin Yu. S., Maltsew M. V., Sologub N. S. Vector Markov chain with partial connections and statistical inferences on its parameters. *Doklady Natsional'noi akademii nauk Belarusi* [Doklady of the National Academy of Sciences of Belarus], 2016, vol. 60, no. 6, pp. 14–21. (in Russian)