Austrian Journal of Statistics
April 2017, Volume 46, 37–45.
http://www.ajs.or.at/
doi:10.17713/ajs.v46i3-4.669

High-order Vector Markov Chain with Partial Connections in Data Analysis

Yuriy Kharin
Belarusian State University,
Department of Mathematical
Modelling and Data Analysis

Michail Maltsew
Belarusian State University,
Research Institute
for Applied Problems
of Mathematics and Informatics

Abstract

A new mathematical model for discrete time series is proposed: homogenous vector Markov chain of the order s with partial connections. 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.

1. Introduction

Markov chain is a wide used mathematical model for discrete time series. It is applied in economics (Kemeny and Snell 1963), biology (Waterman 1999), sociology (Bonacich 2003) and in other fields. Markov chain of the order s (Doob 1953) is an adequate model for description of high-depth dependences in data. In practice data is often represented in time-indexed blocks, and it is reasonable to use vector Markov chains. The state space for such models consists of m-vectors for some finite value m. Unfortunately, it is difficult to use s-order Markov chain in practice, because the number of parameters D for this model increases exponentially when s grows. That is why small-parametric (parsimonious) models are used in applications (Kharin 2012). For such models D depends polynomially on s. Markov chain of order s with r partial connections ($\mathrm{MC}(s, r)$) is an example of a parsimonious model for the univariate case $m = 1$. It was developed in Belarusian state university (Kharin and Petlitskii 2007). Conditional distribution for this model doesn't depend on all s previous states but only on r selected states. In this paper we propose a generalization of the $\mathrm{MC}(s, r)$ for high-order vector Markov chain. Note in addition that parsimonious models are also useful in robust statistical analysis (Kharin 2016; Kharin and Kishylau 2015; Kharin and Shlyk 2009; Kharin and Zhuk 1998; Kharin 1997).

2. Vector Markov chain with partial connections

2.1. Mathematical model

Introduce the notation: $\mathbf{N}$ is the set of positive integers; $A = \{0, 1, \dots, N-1\}$ is the discrete set with N elements, $2 \le N < \infty$; $m \in \mathbf{N}$, $J_i = (j_{i1}, \dots, j_{im}) \in A^m$, $i = 1, 2, \dots$, is an m-dimensional vector; $J_a^b = (J_a, \dots, J_b)$, $a, b \in \mathbf{N}$, $a \le b$, is a sequence of $b - a + 1$ ordered m-dimensional vectors;

$$x_t = (x_{t1}, \dots, x_{tm}) \in A^m, t \in \mathbf{N}$$

is a homogeneous vector Markov chain of the order s $(2 \le s < \infty)$ with the state space A^m, with some initial probability distribution

$$\pi^{(0)}_{J_1,\dots,J_s} = \mathrm{P}\{x_1 = J_1, \dots, x_s = J_s\}, \tag{1}$$

and some $(s+1)$-dimensional matrix of transition probabilities:

$$P = (p_{J_1^s, J_{s+1}}),$$
$$p_{J_1^s, J_{s+1}} = \mathrm{P}\{x_t = J_{s+1} | x_{t-1} = J_s, \dots, x_{t-s} = J_1\}, t = s+1, s+2, \dots. \tag{2}$$

We will denote this Markov chain VMC(s) (Vector Markov Chain of the order s).

The number of independent parameters for the VMC(s) is determined by formula:

$$D_s = N^{ms}(N^m - 1). \tag{3}$$

In Table 1 we present the number of parameters for the binary VMC(s) when $m = 8$ for different values of s .

Table 1: The number of parameters for the binary VMC(s)

s	1	2	4	8	16
D_s	65 280	16 711 680	$\approx 1,095 \cdot 10^{12}$	$\approx 4,704 \cdot 10^{21}$	$\approx 8,677 \cdot 10^{40}$

Table 1 illustrates the "curse of dimensionality" for the s-order Markov chain. To overcome this difficulty we construct a modification of the VMC(s) similarly to the paper (Kharin and Petlitskii 2007).

We will use the notation:

$$M_r = \{(k_1, l_1), (k_2, l_2), \dots, (k_r, l_r)\} \subseteq M_* = \{(k, l) : 1 \le k \le s, 1 \le l \le m\} \tag{4}$$

is an ordered set of $1 \le r \le sm$ pairs of indices, $k_1 = 1$, which we will call a template-set; $\mathbf{M}_r$ is a set of all possible template-sets;

$$S_{M_r}(J_t, \dots, J_{t+s-1}) = (j_{t+k_1-1, l_1}, \dots, j_{t+k_r-1, l_r}),\ t = 1, 2, \dots$$

is a selector function, that associates s vectors with their r components: $S_{M_r} : A^{ms} \to A^r$; $Q = (q_{(i_1,\dots,i_r), I_{r+1}})$ is a stochastic $N^r \times N^m$ matrix, $i_1, \dots, i_r \in A$, $I_{r+1} \in A^m$.

The Markov chain $\{x_t \in A^m : t \in \mathbf{N}\}$ is called the vector Markov chain of the order s with r partial connections if its transition probabilities have the following form:

$$p_{J_1^s, J_{s+1}} = q_{S_{M_r}(J_1,\dots,J_s), J_{s+1}} = q_{(j_{k_1,l_1},\dots,j_{k_r,l_r}), J_{s+1}},\ J_1, \dots, J_s, J_{s+1} \in A^m. \tag{5}$$

We will denote this model VMC(s, r). The m-dimensional VMC(s, r) is determined by the following parameters:

- $s \ge 1$ is the order of the Markov chain;

- $r \in \{1, \ldots, sm\}$ is the number of connections;
- M_r is the template-set of connections;
- $Q = (q_{(i_1,\ldots,i_r),I_{r+1}})$ is a stochastic $N^r \times N^m$-matrix, $(i_1, \ldots, i_r) \in A^r$, $I_{r+1} \in A^m$.

The definition (5) of the VMC(s, r) means that the probability distribution of time series x_t at time point t depends not on all ms components of s previous states, but it depends only on r selected components determined by the template-set M_r. If $r = sm$, then $M_r = M_*$ and we have fully-connected s-order Markov chain: VMC$(s, ms) \equiv$ VMC(s). If $m = 1$, then the VMC(s, r) transforms into the Markov chain with partial connections (Kharin and Petlitskii 2007).

The number of independent parameters for the VMC(s, r) is determined by formula:

$$d = N^r(N^m - 1) + 2r - 1. \tag{6}$$

In Table 2 we present the number of parameters for the binary VMC(s, r) when $N = 2$, $m = 8$ for different values of s and r .

Table 2: The number of parameters for the binary VMC(s, r)

(s, r)	(1, 2)	(2, 4)	(4, 6)	(8, 8)	(16, 10)	(32, 16)
d	1 023	4 087	16 331	65 295	261 139	16 711 711

From comparison of Table 1 and Table 2 we can see sufficient gain in the number of parameters of the VMC(s, r)-model comparing to the VMC(s)-model.

2.2. Ergodicity conditions

Let us give now ergodicity conditions for the Markov chain of conditional order.

Theorem 1. *Homogenous vector Markov chain with partial connections VMC(s, r) is ergodic if and only if there exists a number $c \in \mathbf{N}$, such that the following inequality holds:*

$$\min_{J_1^s, J_{c+1}^{c+s} \in A^{ms}} \sum_{J_{s+1}^c \in A^{m(c-s)}} \prod_{t=1}^{c} q_{S_{M_r}}(J_t, \ldots, J_{t+s-1}), J_{t+s} > 0. \tag{7}$$

Proof. Consider the first-order Markov chain with extended state space, which is equivalent to x_t:

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

Transition matrix for $\bar{x}_t$ has the following form:

$$\bar{P} = (\bar{p}_{I_1^s, J_1^s}),\ \bar{p}_{I_1^s, J_1^s} = \mathbf{I}\{I_2^s = J_1^{s-1}\} p_{I_1^s, J_s}, \tag{8}$$

where $\mathbf{I}\{C\}$ is the indicator function of the event C, $p_{I_1^s, J_s}$ is determined by (5). Ergodicity of x_t is equivalent to ergodicity of $\bar{x}_t$. According to (Kemeny and Snell 1963) $\bar{x}_t$ is ergodic if and only if there exists a number $c \in \mathbf{N}$, such that the following inequality holds:

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

where $\bar{p}^{(c)}_{J_1^s, J_{c+1}^{c+s}}$ is the c-step transition probability from J_1^s to J_{c+1}^{c+s} for the Markov chain $\bar{x}_t$. In (Kharin and Maltsew 2011) the following representation for this probability was obtained:

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

Using this result and equation (5) we come to the criterion (7). Theorem is proved.

Corollary 1. If all elements of matrix Q are positive, then the VMC(s,r) is ergodic.

If the VMC(s,r) is ergodic one, then the stationary probability distribution exists (Doob 1953). We will denote it $(\pi_{J_1^s})$, $J_1^s \in A^{ms}$, and its marginal distributions will be denoted as

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

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

3. Statistical estimators for VMC(s,r) parameters

3.1. Likelihood function

Let us construct now statistical estimators for VMC(s,r) parameters. At first, we need to construct the likelihood function.

Introduce the notation: $X^{(n)} \in A^{mn}$ is the observed vector time series of length n;

$$\nu_{s+1}^{M_r}(i_1,\dots,i_r,I_{r+1}) = \sum_{t=1}^{n-s} \mathbf{I}\{S_{M_r}(x_t,\dots,x_{t+s-1}) = (i_1,\dots,i_r), x_{t+s} = I_{r+1}\}, \quad (9)$$

$$\nu_s^{M_r}(i_1,\dots,i_r) = \sum_{I_{r+1}\in A^m} \nu_{s+1}^{M_r}(i_1,\dots,i_r,I_{r+1}),\ (i_1,\dots,i_r) \in A^r,\ I_{r+1} \in A^m, \quad (10)$$

are frequency statistics for the VMC(s,r) based on $X^{(n)}$.

Lemma 1. If the true values s, r and M_r, are known, then the likelihood function for the VMC(s,r) has the following form:

$$L_n(X^{(n)},Q) = \pi^{(0)}_{x_1,\dots,x_s} \prod_{t=s}^{n-1} q_{S_{M_r}(x_{t-s+1},\dots,x_t),x_{t+1}}. \quad (11)$$

Proof. Equation (11) follows from the expression for the n-dimensional probability distribution that we get using theorem on compound probabilities, the Markov property and definition of the VMC(s,r):

$$\mathrm{P}\{x_1 = j_1, x_2 = j_2, \dots, x_n = j_n\} =$$
$$= \mathrm{P}\{X_1^s = J_1^s\} \prod_{t=s}^{n-1} \mathrm{P}\{x_{t+1} = J_{t+1} | X_1^t = J_1^t\} = \pi^{(0)}_{J_1,\dots,J_s} \prod_{t=s}^{n-1} q_{S_{M_r}(J_{t-s+1},\dots,J_t),J_{t+1}}.$$

Lemma is proved.

Corollary 2. The loglikelihood function for the VMC(s,r) has the following form:

$$l_n(X^{(n)},Q) = \ln \pi^{(0)}_{J_1,\dots,J_s} + \sum_{i_1,\dots,i_r\in A} \sum_{I_{r+1}\in A^m} \nu_{s+1}^{M_r}(i_1,\dots,i_r,I_{r+1}) \ln q_{(i_1,\dots,i_r),I_{r+1}}. \quad (12)$$

3.2. Estimators for transition probabilities

Construct now maximum likelihood estimators (MLE) for the matrix Q of one-step transition probabilities.

Theorem 2. If the true values s, r and the template-set M_r are known, then the maximum likelihood estimators for the one-step transition probabilities (5) are

$$\hat{q}_{(i_1,\dots,i_r),I_{r+1}} = \begin{cases} \dfrac{\nu_{s+1}^{M_r}(i_1,\dots,i_r,I_{r+1})}{\nu_s^{M_r}(i_1,\dots,i_r)}, \ if \ \nu_s^{M_r}(i_1,\dots,i_r) > 0, \\ \\ \dfrac{1}{N^m}, \ if \ \nu_s^{M_r}(i_1,\dots,i_r) = 0. \end{cases} \tag{13}$$

Proof. In order to construct the MLE we need to solve the following conditional extremum problem:

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

This problem splits into N^r subproblems for each set $(i_1,\dots,i_r)$:

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

Using the Lagrange multipliers method for solving these subproblems we get estimators (13). Theorem is proved.

Consistency of estimators (13) follows from the next theorem.

Theorem 3. If the VMC(s,r) is stationary Markov chain, then MLE (13) are consistent estimators as $n \to \infty$:

$$\hat{q}_{(i_1,\dots,i_r),I_{r+1}} \xrightarrow{\text{P}} q_{(i_1,\dots,i_r),I_{r+1}}, \ i_1,\dots,i_r \in A, \ I_{r+1} \in A^m. \tag{14}$$

Proof. Normalized frequencies of the states for the s-order Markov chain tend to the stationary probability distribution as $n \to \infty$ (Kharin and Maltsew 2011):

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

Since frequencies $\nu_{s+1}^{M_r}(i_1,\dots,i_r,I_{r+1})$, $\nu_s^{M_r}(i_1,\dots,i_r)$ are sums of the frequencies of the $(s+1)$-tuples, we have convergence property for $n \to \infty$:

$$\hat{\pi}_s^{M_r}(i_1,\dots,i_r) = \frac{1}{n-s}\nu_s^{M_r}(i_1,\dots,i_r) \xrightarrow{\text{P}} \pi_s^{M_r}(i_1,\dots,i_r), \tag{15}$$

$$\hat{\pi}_{s+1}^{M_r}(i_1,\dots,i_r,I_{r+1}) = \frac{1}{n-s}\nu_{s+1}^{M_r}(i_1,\dots,i_r,I_{r+1}) \xrightarrow{\text{P}}$$

$$\xrightarrow{\text{P}} \pi_{s+1}^{M_r}(i_1,\dots,i_r,I_{r+1}) = \pi_s^{M_r}(i_1,\dots,i_r) q_{(i_1,\dots,i_r),I_{r+1}}. \tag{16}$$

Using (15), (16) and theorem on functional transformations of convergent random sequences from (Borovkov 1998), we come to (14). Theorem is proved.

3.3. Estimators for the template-set

To construct estimators for the template-set M_r we will also use the maximum likelihood method. Let

$$H(M_r) = \sum_{i_1,\dots,i_r\in A} \sum_{I_{r+1}\in A^m} \nu_{s+1}^{M_r}(i_1,\dots,i_r,I_{r+1}) \ln \frac{\nu_{s+1}^{M_r}(i_1,\dots,i_r,I_{r+1})}{\nu_s^{M_r}(i_1,\dots,i_r)} \tag{17}$$

be the plug-in statistical estimator for the Shannon conditional information on the future vector x_{t+1}.

Theorem 4. If the order s and the number of connections r are a priori known, then the MLE for the template-set M_r is

$$\hat{M}_r = arg \max_{M_r \in \mathbf{M}_r} H(M_r). \tag{18}$$

Proof. Formula (18) follows from (17) and from the representation of the loglikelihood function (12) for the VMC(s, r).

Computational complexity of the exhaustive search in the formula (18) is $\mathcal{O}(nmN^{r+m}(sm)^{r-1})$, that is why we can use it only for small values of m and r. Therefore we developed a special algorithm for calculation of estimators (18) to reduce computational complexity. This algorithm is based on step-by-step extension of the initial template-set.

Let $\mathbf{M}_r^+(M_{r-1})$ be the set of templates built by extension of the template M_{r-1} by one element from the set $M_* \setminus M_{r-1}$, $r = 2, 3, \dots$. At the first step of the algorithm we find the initial template $\hat{M}_{r_-}$, $r_- \geq 1$, using the exhaustive search in (18). Then we find sequentially the estimators:

$$\hat{M}_{r_-+1}, \hat{M}_{r_-+2}, \dots, \hat{M}_r. \tag{19}$$

Estimator $\hat{M}_{r'}$, $r' = r_- + 1, r_- + 2, \dots, r$, is constructed as follows:

$$\hat{M}_{r'} = arg \max_{M_{r'} \in \mathbf{M}_{r'}(M_{r'-1})} H(M_{r'}), \tag{20}$$

i. e. we extend the template-set $\hat{M}_{r'-1}$ by one additional element.

3.4. Estimators for the order and the number of connections

In order to estimate the order s and the number of connections r we use the Bayesian information criterion (BIC) (Csiszar and Shields 1999), that has the following expression for our model:

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

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

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

where $s_+ \geq 1$, $1 \leq r_+ \leq ms_+$ are maximal admissible values of s and r respectively, d is the number of independent parameters of the VMC(s, r) determined by formula (6).

4. Results of computer experiments

4.1. Simulated data

Estimation of the matrix Q

Let us illustrate the properties of the constructed statistical estimators by computer simulation. Binary vector Markov chain with the following values of parameters:

$$m = 4,\ s = 4,\ r = 6,\ M_6 = \{(1,1), (2,2), (2,4), (3,1), (3,2), (4,3)\} \tag{21}$$

was simulated. Matrix Q has dimension $2^6 \times 2^4$, its elements were generated as random variables with the uniform probability distribution in the interval [0, 1].

We generated 100 independent realizations of VMC(s, r), each realization consisted of n binary m-vectors, $n \in \{10^5, 2 \cdot 10^5, \ldots, 100 \cdot 10^5\}$. Statistical estimators for transition probabilities were constructed according to formula (13), and the mean square estimation error for estimation of Q was calculated:

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

Figure 1 represents dependence of Δ_n on the time series length n and illustrates the consistency property of statistical estimators (13).

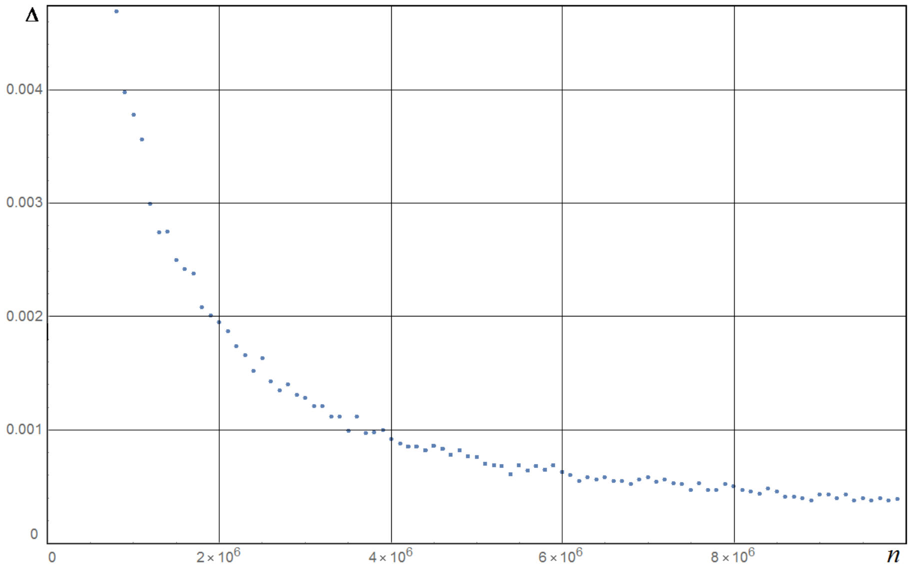

Figure 1: Estimation error of Q

Estimation of the template-set M_r

Let us analyze now properties of the estimators $\hat{M}_r$ built by the extension algorithm presented in the subsection 3.3. In computer experiments we simulated VMC(s, r) with parameters (21) for the time series length $n = 1000, 10000, 20000, \ldots, 150000$. For each n we simulated $U = 1000$ independent realizations. For each realization the template-set estimator $\hat{M}_r$ was computed with the extension algorithm. Then we calculate the frequency estimate of true decision:

$$\delta_n = \frac{1}{U} \sum_{u=1}^{U} \mathbf{I}\{\hat{M}_r^{(u)} = M_r\}, \tag{22}$$

where $\hat{M}_r^{(u)}$ is the estimator for the template-set obtained for the u-th realization.

Results presented in Table 3 illustrate the consistency property of the estimator $\hat{M}_r$.

Table 3: Error of estimating M_r

n	1000	10000	20000	30000	50000	60000	80000	100000	120000	150000
δ_n	0.001	0.341	0.609	0.762	0.914	0.943	0.979	0.993	0.998	1

4.2. Real data

In conclusion let us present results of experiments on real genetic data. We tested genetic DNA sequence LN589993 (https://www.ncbi.nlm.nih.gov/nuccore) which consists of four nucleotides (A, C, G, T). Detection of dependences in DNA sequences is an important problem in bioinformatics (Waterman 1999; Voloshko, Medved, and Kharin 2016), and we constructed stochastic model for the observed sequence using VMC(s, r).

We recoded the sequence to binary form: "A" corresponds to 00, "C" corresponds to 10, "G" corresponds to 01, "T" corresponds to 11. The sequence was splitted on $n = 5708$ nonoverlapping triplets and represented as 6-variate $(m = 6)$ time series. Using step-by-step extension algorithm we estimated the template-set, and using (13) we estimated the transition probabilities and the mean square deviation of matrix Q from uniform distribution

$$\Delta = \max_{i_1,\dots,i_r \in \{0,1\},\, I \in \{0,1\}^m} |\hat{q}_{(i_1,\dots,i_r),I} - 1/64|. \tag{23}$$

For $s = 3$, $r = 7$ we get the following results:

$$\hat{M} = \{(1,3),(1,4),(2,1),(2,2),(3,2),(3,3),(3,4)\},\ \Delta = 0.175.$$

As we can see deviation of matrix $\hat{Q}$ from the case of "pure randomness" is quite significant, and constructed VMC(3, 7) model detects stochastic dependences in the analyzed genetic sequence.

References

Bonacich P (2003). "Asymptotics of a Matrix Valued Markov Chain Arising in Sociology." *Stochastic Processes and their Applications*, **104**(1), 155–171.

Borovkov A (1998). *Mathematical Statistics.* Gordon and Breach, New York.

Csiszar I, Shields P (1999). "Consistency of the BIC Order Estimator." *Electronic research announcements of the American mathematical society*, **5**, 123–127.

Doob J (1953). *Stochastic Processes.* Wiley, New York.

Kemeny J, Snell J (1963). *Finite Markov Chains.* D. Van Nostrand Company, Princeton NJ.

Kharin A (2016). "Performance and Robustness Evaluation in Sequential Hypotheses Testing." *Communications in Statistics - Theory and Methods*, **45**(6), 1693–1709.

Kharin A, Kishylau D (2015). "Robust Sequential Test for Hypotheses about Discrete Distributions in the Presence of "Outliers"." *Journal of Mathematical Sciences*, **205**(1), 68–73.

Kharin A, Shlyk P (2009). "Robust Multivariate Bayesian Forecasting under Functional Distortions in the ξ^2-metric." *Journal of Statistical Planning and Inference*, **139**(11), 3842–3846.

Kharin Y (1997). "Robustness of Clustering under Outliers." *Lecture Notes in Computer Science*, **1280**, 501–512.

Kharin Y (2012). "Parsimonious Models for High-order Markov Chains and Their Statistical Analysis." *VIII World Congress on Probability and Statistics*, pp. 168–169.

Kharin Y, Maltsew M (2011). "Algorithms for Statistical Analysis of Markov Chain with Conditional Memory Depth." *Informatics*, **1**, 34–43(in Russian).

Kharin Y, Petlitskii A (2007). "A Markov Chain of Order s with r Partial Connections and Statistical Inference on Its Parameters." *Discrete Mathematics and Applications*, **17**(3), 295–317.

Kharin Y, Zhuk E (1998). "Filtering of Multivariate Samples Containing "Outliers" for Clustering." *Pattern Recognition Letters*, **19**(11), 1077–1085.

Voloshko V, Medved E, Kharin Y (2016). "Multiresolution Statistical Analysis of DNA." *Proceedings of the 13th international conference*, pp. 178–181.

Waterman M (1999). *Mathematical Methods for DNA Sequences*. Chapman and Hall/CRC, Boca Raton, Florida.

Affiliation:

Yuriy Kharin
Department of Mathematical Modelling and Data Analysis
Belarusian State University
Independence av. 4
220030 Minsk, Belarus
E-mail: kharin@bsu.by

Michail Maltsew
Research Institute for Applied Problems of Mathematics and Informatics
Belarusian State University
Independence av. 4
220030 Minsk, Belarus
E-mail: maltsew@bsu.by

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 46 *Submitted:* 2016-11-15
April 2017 *Accepted:* 2017-02-02