Identification of Markov Chains of Conditional Order

M. V. Maltsew [1], Yu. S. Kharin [2]

1) Belarusian State University, Minsk, BELARUS, maltsew@mail.ru

2) Belarusian State University, Minsk, BELARUS, kharin@bsu.by

Abstract: *A new special case of high-order Markov chains with a small number of parameters – Markov chain of conditional order – is considered. Statistical estimators for parameters of the model by observed time series are constructed; their asymptotic properties are analyzed. Results of computer experiments are presented.*

***Keywords*:** Markov Chain, Conditional Order, Statistical Estimation, Bayesian Information Criterion

1. INTRODUCTION

Discrete time series are widely used for modeling processes in applications. One usually needs to take into consideration dependence on the previous states of the process. Markov chain of the order s, $s \geq 1$, [1] is a well-known mathematical model adequate for these purposes. Markov chains are used in signal processing [2], genetics [3], economics [4], information security [5] and many other areas. However, it becomes difficult to use this model when the order s is large as the number of parameters increases exponentially as s grows. Therefore one needs to have the data set of huge size for fitting the model. Thus the problem of development and analysis of special kind of high-order Markov chains with a small number of parameters is rather important. Give some well-known examples of such models. For Markov chain of the order s with r partial connections [6] and for the variable length Markov chain [7] the transition probability depends on some selected states (and its number is quite small), but not on all s states. For Raftery model [8] one needs only one additional parameter for each order. Another mathematical model with a small numbers of parameters – Markov chain of conditional order [9] – is under consideration in this paper. The paper is devoted to identification of this model by observed data.

2. MATHEMATICAL MODEL

Introduce the notation: $\mathbf{N}$ is the set of natural numbers, $2 \leq N < \infty$; $A = \{0,1,\ldots,N-1\}$ is the finite state space with N elements; $J_m^n = (j_m,\ldots,j_n) \in A^{n-m+1}$, $n \geq m$, is the multiindex; $\{x_t \in A : t \in \mathbf{N}\}$ is a finite homogeneous Markov chain of the order s $(2 \leq s < \infty)$ defined at some probability space (Ω, F, P); $P = (p_{J_1^{s+1}})$ is the $(s+1)$-dimensional one-step transition probability matrix, $p_{J_1^{s+1}} = \mathrm{P}\{x_{t+s} = j_{s+1} \mid x_{t+s-1} = j_s, \ldots, x_t = j_1\}$, $t \in \mathbf{N}$; $L \in \{1, 2, \ldots, s-1\}$, $K = N^L - 1$ are natural numbers; $Q^{(1)},\ldots,Q^{(M)}$ are M $(1 \leq M \leq K+1)$ different square stochastic matrices of the order N: $Q^{(m)} = (q_{i,j}^{(m)})$, $0 \leq q_{i,j}^{(m)} \leq 1$, $\sum_{j\in A} q_{i,j}^{(m)} \equiv 1$, $i, j \in A$, $1 \leq m \leq M$;

$< J_n^m > = \sum_{k=n}^{m} N^{k-1} j_k \in \{0,1,\ldots, N^{m-n+1}-1\}$, $1 \leq n \leq m \leq s$, is the numeric representation of the multiindex J_n^m; $I\{C\}$ is the indicator of the event C.

Define the Markov chain $\{x_t \in A : t \in \mathbf{N}\}$ of conditional order if its one-step transition probabilities have the following form:

$$p_{J_1^{s+1}} = \sum_{k=0}^{K} I\{< J_{s-L+1}^{s} >= k\} q_{j_{b_k}, j_{s+1}}^{(m_k)} , \qquad (1)$$

where $1 \leq m_k \leq M$, $1 \leq b_k \leq s-L$, $0 \leq k \leq K$, $\min_{0 \leq k \leq K} b_k = 1$; and all elements of the set $\{1, 2,\ldots, M\}$ occur in the sequence $m_0, \ldots, m_K$. The sequence of elements J_{s-L+1}^{s}, that determines the condition in the formula (1), is called the base memory fragment (BMF) of the random sequence. According to the definition (1) the state of the model x_t at time t doesn't depend on all previous states, but depends only on L+1 states $(j_{b_k},\ J_{s-L+1}^{s})$; the value of BMF J_{s-L+1}^{s} determines not only the state j_{b_k}, but it also determines the transition matrix.

Thus the Markov chain of conditional order is determined by the following parameters: unconditional order s of the Markov chain; the length of BMF (L); K+1 parameters $\{b_k\}$ determining conditional orders of the Markov chain $s_k = s - b_k + 1 \in \{L+1, L+2, \ldots, s\}$; K+1 parameters $\{m_k\}$ determining transition matrix; M stochastic matrices of the order N, which are described by $MN(N–1)$ independent parameters. Hence the transition matrix $P = (p_{J_1^{s+1}})$ of the Markov chain of conditional order is defined by $D = 2(N^L + 1) + MN(N-1)$ independent parameters.

Note that if $L = s-1$, $b_0 = \ldots = b_K = 1$, we have fully connected Markov chain of the order s; similarly if $b_0 = \ldots = b_K = s-L$, we have fully connected Markov chain of the order L+1. If $M = K$+1, then all the parameters $\{m_k\}$ are different and each value k of BMF has its own transition matrix $Q^{(k)}$.

3. STATISTICAL ESTIMATION OF PARAMETERS

At first, let us give ergodicity conditions for the Markov chain of conditional order.

Theorem 1. Markov chain of conditional order is ergodic if and only if there exists a natural number $m \in \mathbf{N}$, $s \leq m < \infty$, such that the following inequality holds:

$$\min_{J_1^s, J_{1+m}^{s+m} \in A^s} \sum_{k=0}^{K} \sum_{J_{s+1}^{m} \in A^{m-s}} \prod_{i=1}^{m} I\{J_{i+s-L}^{i+s-1} = k\} q_{j_{b_k+i-1}, j_{i+s}}^{(m_k)} > 0 . \qquad (2)$$

In the sequel, we will consider ergodic Markov chains. Denote the probability distribution of the J_0^l by $\pi(r,\ J_0^l)$

$= \mathrm{P}\{x_t = j_0,\ X_{t+r+1}^{t+r+l} = J_1^l\}$, $J_0^l \in A^{l+1}$, $l = 0, 1,\ldots$, $\pi(0,\ J_0^l) = \pi(J_0^l)$.

Construct now estimators of parameters of the model. At first, let us obtain the maximum likelihood estimators (MLE) of the matrices $Q^{(1)},\ldots,Q^{(M)}$ using an observed realization X_1^n of length n. All other parameters are assumed to be known. We need the following notation:

$$1 \le l \le s,\ \ 0 \le l_0 \le s-l;$$

$$A^{s+1}(J_1^l) = \{I_1^{s+1} \in A^{s+1} : I_1^l = J_1^l\},$$

$$A^{1+l_0+l}(j_0^{l_0}, J_1^l) = \{I_1^{1+l_0+l} \in A^{1+l_0+l} : i_1 = j_0,\ I_{2+l_0}^{1+l_0+l} = J_1^l\};$$

$$\nu_{J_1^{s+1}}(n) = \sum_{t=1}^{n-s} \delta_{X_t^{t+s}, J_1^{s+1}},\ \ \nu_{J_1^l}(n) = \sum_{I_1^{s+1} \in A^{s+1}(J_1^l)} \nu_{I_1^{s+1}}(n),$$

$$\nu_{j_0, J_1^l}^{(l_0)}(n) = \sum_{I_1^{1+l_0+l} \in A^{1+l_0+l}(j_0^{l_0}, J_1^l)} \nu_{I_1^{1+l_0+l}}(n),$$

$\bar{P} = \left(\bar{p}_{<I_1^s>,<J_1^s>}\right)$ is the one-step transition probability matrix for the s-dimensional Markov chain of the first order $X^{(t)} = \left(x_t, x_{t+1},\ldots,x_{t+s-1}\right)$, $t \in \mathbf{N}$, with extended state space, $I_1^s, J_1^s \in A^s$, $\bar{p}_{<I_1^s>,<J_1^s>} = \delta_{I_2^s, J_1^{s-1}} p_{I_1^s, j_s}$. E_{N^s} is the identity matrix of the order N^s; $P^* = \lim\limits_{n\to\infty} \bar{P}^n$ is the limit matrix; $Z = (E_{N^s} - \bar{P} + P^*)^{-1} = (z_{I_1^s, J_1^s})$, $I_1^s, J_1^s \in A^s$, $h(J_0^{L+1}, I_0^L) = \sum\limits_{E_1^{s+1} \in A_j} \sum\limits_{F_1^{s+1} \in A_i} \pi_{E_1^s} z_{E_2^{s+1}, F_1^s}$ (we denote $A_j = A^{1+l_0+l}(j_0^{l_0}, J_1^l)$ for shot).

Theorem 2. If true values L, $\{b_k\}$ and $\{m_k = k\}$ are known, then the MLE for the one-step transition probabilities $q_{u,\mathrm{v}}^{(m_k)}$, $u, \mathrm{v} \in A$, are

$$\begin{cases} \hat{q}_{u,\mathrm{v}}^{(m_k)} = \sum\limits_{w \in A^{B_*}} I\{<w>= k\} \dfrac{\nu_{u,wv}^{(l_k)}(n)}{\nu_{u,w}^{(l_k)}(n)}, \text{ if } \nu_{u,w}^{(l_k)}(n) > 0, \\ 1/N, \text{ if } \nu_{u,w}^{(l_k)}(n) = 0. \end{cases} \tag{3}$$

where $l_k = s - b_k - L$.

Remark. If some parameters $\{m_k\}$, $k = 0, 1,\ldots,K$, are equal, i.e. one transition matrix corresponds to different base memory fragments, then the MLE have the following form:

$$\hat{q}_{u,\mathrm{v}}^{(m_k)} = \begin{cases} \dfrac{\sum\limits_{w \in M_{m_k}} \nu_{u,wv}^{b_k}(n)}{\sum\limits_{w \in M_{m_k}} \nu_{u,w}^{b_k}(n)}, \text{ if } \sum\limits_{w \in M_{m_k}} \nu_{u,w}^{b_k}(n) > 0, \\ 1/N, \text{ if } \sum\limits_{w \in M_{m_k}} \nu_{u,w}^{b_k}(n) = 0, \end{cases} \tag{4}$$

where $M_i = \{w \in A^{B_*} : m_{<w>} = i\}$, $i = 1,\ldots,M$, $\bigcup\limits_{i=1}^{M} M_i = A^{B_*}$.

Construct now estimators for the parameters b_k, which determine conditional order of the chain s_k, $k = 0,\ldots,K$.

Theorem 3. If the true values L, $\{m_k\}$ are known, then the MLE of $\{b_k\}$ are

$$\hat{b}_k = \underset{1 \le b \le s-L}{\arg\max} \sum_{i,j \in A} \nu_{i,wj}^{s-b-L}(n) \ln(\hat{q}_{i,j}^{m_k}),\ \ k = 0,\ldots,K. \tag{5}$$

Finally, estimate the length of BMF L and the order of the chain s. These estimators are constructed using Bayesian Informational Criterion [10]:

$$(\hat{s}, \hat{L}) = \underset{2 \le s \le S_+,\, 1 \le L \le L_+}{\arg\min}\ BIC(s, L), \tag{6}$$

$$BIC(s, L) = -\Big(\sum_{\substack{u,v \in A,\\ w \in A^L}} \sum_{k=0}^{K} \delta_{<w>,k} \nu_{u,wv}^{(s-\hat{b}_k-L)}(n) \ln \hat{q}_{u,v}^{(k)}\Big) + 2N^L \log n,$$

where $S_+ \ge 2$, $1 \le L_+ \le S_+ - 1$ are maximal admissible values of the parameters s and L respectively; estimators $\hat{Q}^{(i)}$, $i = 1,\ldots,M$, and $\hat{b}_k$, $k = 0,\ldots,K$, can be found using (3) and (5) respectively.

3. ASYMPTOTIC PROPERTIES OF THE ESTIMATORS (3), (5)

We'll use the notation:

$$q_{J_0^{L+1}} = \sum_{k=0}^{K} I\{< J_1^L = k >\} q_{j_0, j_{L+1}}^{(k)},\ \ \bar{q}_{J_0^{L+1}} = \sqrt{n}\,(\hat{q}_{J_0^{L+1}} - q_{J_0^{L+1}});$$

$$l_k(b) = \sum_{j_0, j_{L+1} \in A} \pi(r_b,\ J_0^{L+1}) \ln \frac{\pi(r_b,\ J_0^{L+1})}{\pi(r_b,\ J_0^L)},$$

where k is a integer number, such that $< J_1^L > = k$, $r_b = s - b - L$, $b = 1,\ \ldots,\ s\text{-}L$; $I_k(b) = I_k(b,\ j_{L+1}, J_1^L) =$

$$= \sum_{j_0, j_{L+1} \in A} \pi(r_b,\ J_0^{L+1}) \ln \frac{\pi(r_b,\ J_0^{L+1})}{\pi(r_b,\ J_0^L)\pi(j_{L+1})}$$

is the Shannon information [5] on the symbol j_{L+1} contained in the sequence J_0^l under fixed BMF J_1^L. The position of the symbol j_0 is determined by b; similarly

$$I_k = I_k(j_{L+1},\ H_1^{s-L} J_1^L) =$$

$$= \sum_{H_1^{s-L} \in A^{s-L}} \sum_{j_{L+1} \in A} \pi(\ H_1^{s-L} J_1^{L+1}) \ln \frac{\pi(\ H_1^{s-L} J_1^{L+1})}{\pi(\ H_1^{s-L} J_1^L)\pi(j_{L+1})}. \tag{7}$$

$$\hat{l}_k(b) = \sum_{j_0, j_{L+1} \in A} \hat{\pi}(r_b,\ J_0^{L+1}) \ln \frac{\hat{\pi}(r_b,\ J_0^{L+1})}{\hat{\pi}(r_b,\ J_0^L)},$$

$$\hat{I}_k(b) = \sum_{j_0, j_{L+1} \in A} \hat{\pi}(r_b,\ J_0^{L+1}) \ln \frac{\hat{\pi}(r_b,\ J_0^{L+1})}{\hat{\pi}(r_b,\ J_0^L)\hat{\pi}(j_{L+1})}$$

are statistical estimators of the parameters $l_k(b)$ and $I_k(b)$ respectively; $\bar{\psi}_H(r) = \sqrt{n}\left(\hat{I}_H(r) - I_H(r)\right)$.

Theorem 4. If the Markov chain of conditional order (1) is stationary, then (2) gives consistent estimators at $n \to \infty$:

$$\hat{q}_{u,\mathrm{v}}^{(m)} \xrightarrow{\mathrm{P}} q_{u,\mathrm{v}}^{(m)},\ 1 \le m \le M. \tag{7}$$

Theorem 5. Under the conditions of Theorem 4 at $n \to \infty$ the random variables $\{\bar{q}_{J_0^{L+1}} : J_0^{L+1} \in A^{L+2}\}$ are jointly asymptotically normal with zero means and covariance matrix $\Sigma_q = \Sigma_q(J_0^{L+1}, I_0^{L+1})$:

$$\Sigma_q(J_0^{L+1},\ I_0^{L+1}) = I\{J_0^L = I_0^L\} q_{J_0^{L+1}} \frac{I\{j_{L+1} = i_{L+1}\} - q_{J_0^L i_{L+1}}}{\pi(J_0^L)}. \tag{8}$$

Prove now the consistency property for the estimators (5).

Lemma 1. Under the conditions of Theorem 4, if J_1^L, $< J_1^L > = k$, is fixed BMF, then

$$\arg\max_{1\le b\le s-L} l_k(b) = \arg\max_{1\le b\le s-L} I_k(b).$$

Lemma 2. Under the conditions of Lemma 1

$$I_k(b) = I_k.$$

Theorem 6. Under the conditions of Lemma 1at $n\to\infty$ the statistic $\hat{b}_k$ is a consistent estimator:

$$\hat{b}_k \xrightarrow{\text{P}} b_k .$$

Theorem 7. Under the conditions of Theorem 4 at $n\to\infty$ the random variables $\{\overline{\psi}_H(r)\}$ are jointly asymptotically normal with zero means and covariance matrix $\Sigma_\psi = \Sigma_\psi(r,\ k)$, $0 \le r,k \le s\text{-}L\text{-}1$,

$$\Sigma_\psi(r,k) = \sum_{f_0,\, f_{L+1}} \sum_{g_0,\, g_{L+1}} t_1 t_2 t_3 t_4 - t_5$$

$$t_1 = \ln\frac{\pi(r,\, f_0 H_1^L f_{L+1})}{\pi(r,\, f_0 H_1^L)\pi(f_{L+1})} - \frac{\pi(H_1^L f_{L+1})}{\pi(f_{L+1})},$$

$$t_2 = \ln\frac{\pi(k, g_0 H_1^L g_{L+1})}{\pi(k, g_0 H_1^L)\pi(g_{L+1})} - \frac{\pi(H_1^L g_{L+1})}{\pi(g_{L+1})},$$

$$t_3 = \sum_{E_1^{s+1}\in A^{s+1}} \pi(E_1^{s+1}) + \frac{\pi(r,\, f_0 H_1^L f_{L+1})}{\pi(r,\, f_0 H_1^L)} \cdot \frac{\pi(k, g_0 H_1^{L+1})}{\pi(k, g_0 H_1^L)},$$

$$t_4 = h(f_0 H_1^L f_{L+1}, g_0 H_1^L) + h(g_0 H_1^L g_{L+1},\, f_0 H_1^L),$$

$$t_5 = 3\pi(r,\, f_0 H_1^L f_{L+1})\pi(k, g_0 H_1^L g_{L+1}).$$

4. NUMERICAL EXPERIMENTS

Experiment 1. Illustrate consistency of the estimators $\hat{\mathbf{b}} = (\hat{b}_0,\dots,\hat{b}_K)$, calculated according to (5).

Numerical experiment was conducted as follows: U = 100 independent realizations of the Markov chain of conditional order of fixed length n was simulated for the parameters: N=2, A={0, 1}, s=2, L=4, M=K+1=16, $\mathbf{b}$ = (4, 10, 19, 18, 17, 1, 1, 1, 3, 25, 28, 13, 7, 23, 2, 1), $1 \le b_i \le 28$, $0 \le i \le K$=15. The length realization n varied from 10^3 to 10^7. Parameter $\mathbf{b}$ and matrices $Q^{(1)},\dots,Q^{(M)}$ are unknown.

Estimators $\hat{\mathbf{b}}$ were calculated for each u-th realization $u = 1, \dots, U$, and corresponding matrices $\hat{Q}^{(i)} = \hat{Q}^{(i)}(\hat{\mathbf{b}}), i = 1,\dots,M$ were obtained. We used estimate of the variance $\hat{\text{v}}_n^u = \sum_{k=0}^{K}\sum_{i,j=0}^{N}(\hat{q}_{ij}^{(k)} - q_{ij}^{(k)})^2$ as a measure of accuracy of the estimators for u-th realization of the length n. For each value of n estimators $\{\hat{\text{v}}_n^u\}$ were computed; then the total mean square error for the estimators $\hat{\text{v}}_n = \frac{1}{U}\sum_{u=1}^{U}\hat{\text{v}}_n^u$ was evaluated. The results are plotted on Figure 1 as little circles; large circles correspond to similar errors obtained using the correct values of $\mathbf{b}$.

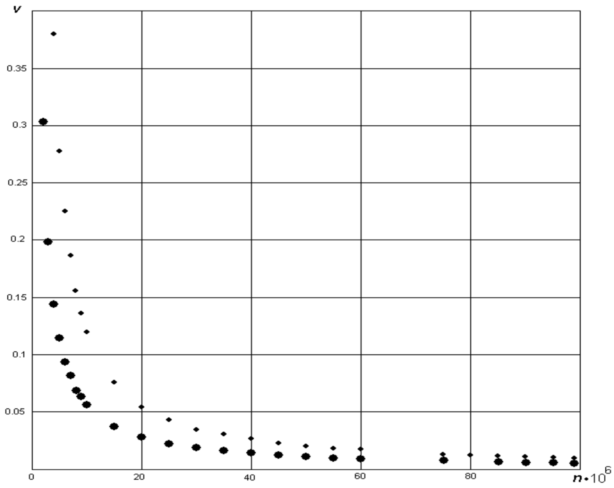

Fig.1 – Dependence v = v(*n*)

Experiment 2. Estimate BMF length L and the order of the Markov chain s using the Bayesian Information Criterion (BIC). Also we compare BIC with another well-known criterion – Akaike Information Criterion (AIC) [11]:

$$BIC(s,L) = -l_n(X_1^n,\, L,\, s) + \frac{D}{2}\ln(n),$$

$$AIC(s,L) = -l_n(X_1^n,\, L,\, s) + D,$$

where $D = 2(N^L + 1) + MN(N-1)$ is the number of independent parameters of the model,

$$l_n(X_1^n,\, L,\, s) = \sum_{\substack{u,v\in A,\\ w\in A^L}} \sum_{k=0}^{K} I(<w>=k)\nu^{l_k}(uwv)\ln q_{u,v}^{(m_k)}$$

is a log-likelihood function $(\hat{s}, \hat{L}) = \arg\min_{2\le s\le \bar{S},\, 1\le L\le \bar{L}} BIC(s, L)$ or $(\hat{s}, \hat{L}) = \arg\min_{2\le s\le \bar{S},\, 1\le L\le \bar{L}} AIC(s, L)$.

The Markov chain of conditional order is simulated with the following parameters: $A = \{0,\ 1\}$, $N = 2$, $s = 4$, $M = 2$, $L = 2$, $b_0 = 2$, $b_1 = 2$, $b_2 = 1$, $b_3 = 1$, $m_0 = 1$, $m_1 = 2$, $m_2 = 1$, $m_3 = 2$. The length of the chain $n = 20000$. Matrices $Q^{(1)}$ and $Q^{(2)}$ are:

$$Q^{(1)} = \begin{pmatrix} 0.18 & 0.82 \\ 0.41 & 0.59 \end{pmatrix},\ Q^{(2)} = \begin{pmatrix} 0.77 & 0.23 \\ 0.09 & 0.91 \end{pmatrix}.$$

Maximal admissible values of the order and the BMF are $S_+ = 8$, $L_+ = 4$. If M is unknown, let us assume its maximal admissible value $M = N^L$, $D = N^L(2+N(N\text{-}1))$ (each value of BMF has its own transition matrix). The values of s, B and the corresponding values of BIC and AIC are given in Table 1.

Table 1. Estimates of *s*, *L*

(*s*,*L*)	**BIC**	**AIC**	**(*s*,*L*)**	**BIC**	**AIC**
(2,1)	13477	13445	(6,2)	*9997*	9934
(3,1)	12359	12327	(6,3)	10072	9946
(3,2)	11964	11901	(6,4)	10217	9964
(4,1)	10701	10670	(7,1)	10701	10670
(4,2)	*9997*	*9934*	(7,2)	*9997*	9934

(4,2)	10073	9946	(7,4)	10071	9944
(5,1)	10701	10670	(7,3)	10211	9958
(5,2)	*9997*	*9934*	(8,1)	10701	10670
(5,3)	10072	9946	(8,2)	*9997*	9934
(5,4)	10221	9968	(8,3)	10069	9942
(6,1)	10701	10670	(8,4)	10207	9954

We can see from Table 1 that BIC(s, L) attains the minimum if L = 2 and s = 4, 5,…, 8. The true order of the chain is equal to s = 4; the following pairs (s, L): (4, 2), (5, 2), (6, 2), (7, 2), (8, 2) are equivalent because we have the same values of conditional orders s_k for all ones. AIC(s, L) reaches the minimum on equivalent pairs (4, 2) and (5, 2). So both BIC and AIC attains the minimum on true values of s and L.

6. CONCLUSION

In this paper we consider the Markov chain of conditional order which is a new model of the high-order Markov chain with a small number of parameters. Statistical estimators for parameters of the model are constructed. Asymptotic properties of transition probabilities and conditional orders are analyzed. Numerical experiments illustrate the theoretical results.

6. REFERENCES

[1] J. Kemeny, J. Snell. *Finite Markov Chains*. Princeton NJ: D. Van Nostrand Company, 1990.

[2] Y. Li, Y. Dong , H. Zhang , H. Zhao , H. Shi ,X. Zhao. Spectrum Usage Prediction Based on High-order Markov Model for Cognitive Radio Networks, *2010 10th IEEE International Conference on Computer and Information Technology (CIT 2010)*, p. 2784-2788.

[3] M.S. Waterman. *Mathematical Methods for DNA Sequences*. Chapman and Hall/CRC: Boca Raton, 1999.

[4] W. K.Ching, High-order Markov chain models for categorical data sequences *Wiley Periodicals, Inc. Naval Research Logistics*. 2004. Vol. 51. P. 557 – 574.

[5] Yu. S. Kharin, V. I. Bernik, G. V. Matveev, S. V. Agievich. *Mathematical and computer foundation of cryptology*. Novoe znanie, Minsk, 2003.

[6] Yu. S. Kharin, A. I. Piatlitski. Markov chain with partial connections MC(s, r) and statistical conclusions of its parameters, *Discrete mathematics*, Vol. 19, №2 (2007). p. 109-130.

[7] P. Buhlmann, A. Wyner. Variable Length Markov Chains, *The Annals of Statistics*, Vol. 27, № 2 (1999). p. 480-513

[8] A. E. Raftery. A Model for High-Order Markov Chains, *J. Royal Statistical Society*, Vol. B-47, № 3 (1985). p. 528–539.

[9] Yu. S. Kharin, M. V.Maltsew Algorithms for statistical analysis of Markov chains with conditional memory depth, *Informatics* №1 (2011). p. 34-43.

[10] I. Csiszar, P. C. Shields. Consistency of the BIC order estimator, *Electronic research announcements of the American mathematical society*, Vol. 5 (1999). P. 123-127.

[11] H. Akaike A A new look at the statistical model identification, *IEEE Transactions on Automatic Control* 19 (6) (1974). P. 716-723