AUSTRIAN JOURNAL OF STATISTICS
Volume 40 (2011), Number 1 & 2, 75–84

Statistical Analysis of Discrete Time Series Based on the MC(s, r)–Model

Yuriy Kharin and Andrei Piatlitski
Belarusian State University, Minsk

Abstract: A new model for discrete-valued time series is proposed: Markov chain of the order s with r partial connections MC(s, r). Statistical estimators for the parameters of the MC(s, r)-model are constructed. The asymptotic properties of these estimators are proved. Statistical tests on the parameters of this model are proposed and their performance is analyzed. The theory is illustrated on some real statistical data.

Keywords: High-order Markov Chain, Estimators, Hypotheses Testing.

1 Introduction

Mathematical modeling of complex systems and processes in genetics, meteorology, social sciences, medicine, economics usually leads to the problem of statistical analysis of discrete-valued data (see Avery and Henderson, 1999). If we analyze the data in dynamics, we come to the model of discrete-valued time series, i. e. a random process $x_t \in A$ on the probability space $(\Omega, F, \mathbf{P})$ with discrete time $t \in \mathbb{N} = \{1, 2, \dots\}$ and a finite set of N $(2 \leq N < \infty)$ states $A = \{0, 1, \dots, N-1\}$. An universal model for long-memory discrete time series is the high-order Markov chain (see Billingsley, 1961). Unfortunately, the number of parameters for the s-order Markov chain $D = N^s(N-1)$ increases exponentially with respect to the order s, and to identify this model is a computationally hard problem; in addition, we need to have data sets of huge size $n > D$. This situation generates topical problems of construction and statistical analysis of small-parametric models for high-order Markov chains, i. e. for the models determined by small number of parameters of the transition matrix. Some of these models are known: the Jacobs-Lewis model (see Jacobs and Lewis, 1978), the Mixture Transition Distribution model (see Raftery, 1985), the hidden Markov model (see Rabiner, 1989), the variable length Markov chain (see Buhlmann and Wyner, 1999).

A new model called "Markov chain with partial connections" was proposed by Kharin in 2004. In this paper, using the results from Kharin and Piatlitski (2007) we establish new properties for this model, consider methods for estimating the model parameters, analyze properties of the estimators, apply this model to real statistical data.

2 The MC(s, r)-Model and its Properties

Introduce the notation: $J_i^k = (j_i, j_{i+1}, \dots, j_k) \in A^{k-i+1}$ is a subsequence of $k - i + 1$ indices, $k \geq i$; $\{x_t\}$ is a homogeneous Markov chain of the s-th order with the state space A, the one-step transition probabilities $p_{J_1^{s+1}} = \mathbf{P}\{x_{t+s} = j_{s+1} | x_{t+s-1} = j_s, \dots, x_t = j_1\}$, and the initial probabilities $\pi_{J_1^s} = \mathbf{P}\{X_1^s = J_1^s\}$, $J_1^{s+1} \in A^{s+1}$.

Definition 1. *The Markov chain x_t is called the Markov chain of the s-th order with r partial connections (see Kharin and Piatlitski, 2007) if*

$$p_{J_1^{s+1}} = p_{j_1,\dots,j_s,j_{s+1}} = q_{j_{m_1^0},\dots,j_{m_r^0},j_{s+1}}\,, \qquad J_1^{s+1} \in A^{s+1}\,, \tag{1}$$

where $r \in \{1, 2, \dots, s\}$ is the number of connections; $M_r^0 = (m_1^0, m_2^0, \dots, m_r^0) \in \mathrm{M}$ is the integer-valued vector with r ordered components $1 = m_1^0 < \dots < m_r^0 \le s$, called the pattern of connections, M is the set of all admissible patterns; $Q = (q_{J_1^{r+1}})_{J_1^{r+1}\in A^{r+1}}$ is the stochastic matrix, that is, $q_{J_1^{r+1}} \ge 0$, $J_1^{r+1} \in A^{r+1}$, and $\sum_{j_{r+1}\in A} q_{J_1^{r+1}} = 1$, $J_1^r \in A^r$.

The equation (1) means that the transition probability to the state j_{s+1} depends not on all s preceding states $j_1, \dots, j_s$ but on r selected states $j_{m_1^0}, \dots, j_{m_r^0}$ only. The transition matrix for the MC(s, r) is completely determined by $N^r(N - 1)$ parameters, instead of D parameters. If $r = s$, then the MC(s, r) is the Markov chain of the s-th order.

Theorem 1. *Let d be the greatest common divisor of $m_2^0 - m_1^0, m_3^0 - m_2^0, \dots, m_r^0 - m_{r-1}^0, s + 1 - m_r^0$. If $d > 1$ and $x_1, x_2, \dots, x_s$ are independent, then the MC(s, r) can be represented as the sequence of d independent processes:*

$$x_{(t-1)d+1} = x_t^{(1)}, x_{(t-1)d+2} = x_t^{(2)}, \dots, x_{(t-1)d+d} = x_t^{(d)}\,, \qquad t \in \mathbb{N}\,,$$

where $\{x_t^{(i)}\}$ is the MC$(s/d, r)$ with the pattern $M_r^* = \big(m_1^0 + d^{-1}(m_i^0 - m_1^0)\big)_{1\le i\le r}$ and the transition matrix Q, $i = 1, 2, \dots, d$.

Proof. Denote $k_i = (m_i^0 - m_1^0)/d$, where $i = 1, \dots, r + 1$, $m_{r+1}^0 = s + 1$. The transition probability at the time $t + k_{r+1}d$ depends on the states at the times $t + k_1d, \dots, t + k_rd$, where $t \in \mathbb{N}$, $t - 1 = kd + l$, $0 \le l \le d - 1$, $k = 0, 1, \dots$. It is easy to see that the sets $\{(k + k_i)d + l + 1 : 1 \le i \le r + 1, k \ge 0\}$ are mutually disjoint for different l. □

Introduce the notation: $\delta_{I_1^k,\,J_1^k} = \prod_{l=1}^k \delta_{i_l,\,j_l}$ is the Kronecker symbol for $I_1^k, J_1^k \in A^k$; $p^*_{J_1^s,K_1^s} = \delta_{J_2^s,K_1^{s-1}}\, q_{j_{m_1^0},\dots,j_{m_r^0},k_s}$, $J_1^s, K_1^s \in A^s$; $p^{*\,(i)}_{J_1^s,K_1^s}$ are the i-steps transition probabilities.

Definition 2. *Let us call the MC(s, r) the ergodic Markov chain if there exists the limit $\lim_{i\to\infty} p^{*\,(i)}_{J_1^s,K_1^s} = \pi^*_{K_1^s}$, $K_1^s \in A^s$, where $\pi^*_{K_1^s} > 0$, $\sum_{K_1^s\in A^s} \pi^*_{K_1^s} = 1$.*

Theorem 2. *The MC(s, r) is ergodic if and only if there exists $i \in \mathbb{N}$ such that*

$$\min_{J_1^s, J_{s+i+1}^{2s+i}\in A^s} \sum_{J_{s+1}^{s+i}\in A^i} \prod_{k=1}^{s+i} q_{j_{k+m_1^0-1},\dots,j_{k+m_r^0-1},j_{k+s}} > 0\,.$$

Stationary probability distribution $(\pi^*_{J_1^s})_{J_1^s\in A^s}$ is the unique solution of the equations

$$\pi^*_{J_2^{s+1}} = \sum_{j_1\in A} \pi^*_{J_1^s}\, q_{j_{m_1^0},\dots,j_{m_r^0},j_{s+1}}\,,\ \ J_1^{s+1} \in A^s\,.$$

Proof. The MC(s, r) can be reduced to the first order vector Markov chain, then using the ergodic theorem from Kemeny and Snell (1960) we obtain the necessary statement. □

Corollary 1. *If $\min_{J_1^{r+1}\in A^{r+1}} q_{J_1^{r+1}} > 0$, then the MC$(s, r)$ is ergodic.*

Corollary 2. *Let the MC(s,r) be an ergodic Markov chain. The stationary probability distribution is uniform ($\pi^*_{J_1^s} \equiv N^{-s}$) if and only if the matrix Q is doubly stochastic:* $\sum_{j_1\in A} q_{J_1^{r+1}} \equiv 1$, $\sum_{j_{r+1}\in A} q_{J_1^{r+1}} \equiv 1$.

Corollary 3. *Assume that the MC(s,r) is an ergodic Markov chain. The stationary probability distribution has the multiplicative form $\pi^*_{J_1^s} = \prod_{i=1}^s \pi^*_{j_i}$, $J_1^s \in A^s$, if and only if* $\pi^*_{j_{r+1}} = \sum_{j_1\in A} \pi^*_{j_1} q_{J_1^{r+1}}$, $J_2^{r+1} \in A^r$; $\sum_{j\in A} \pi^*_j = 1$.

Definition 3. *The MC(s,r) is stationary if it is ergodic and its initial probabilities $(\pi_{J_1^s})_{J_1^s\in A^s}$ equal the stationary distribution $(\pi^*_{J_1^s})_{J_1^s\in A^s}$.*

Let the MC(s,r) is stationary, $\mu_{J_1^{r+1}}(M_r) = \mathbf{P}\{x_{t+m_1-1} = j_1, \ldots, x_{t+m_r-1} = j_r, x_{t+s} = j_{r+1}\} = \sum_{K_1^s\in A^s} \delta_{k_{m_1},j_1} \cdot \ldots \cdot \delta_{k_{m_r},j_r} \pi^*_{K_1^s} p_{K_1^s,j_{r+1}}$, $J_1^{r+1} \in A^{r+1}$, is the probability distribution of the $(r+1)$-tuple for some pattern $M_r \in \mathrm{M}$; the point instead of any index means summation on all possible values of this index: $\mu_{J_1^r\cdot}(M_r) = \sum_{j_{r+1}\in A} \mu_{J_1^{r+1}}(M_r)$.

3 Statistical Estimators for Parameters of the MC(s,r) and Their Performance

3.1 Statistical Estimation of the Matrix Q

Introduce the notation: $X_1^n = (x_1, x_2, \ldots, x_n) \in A^n$ is the realization of the MC(s,r) of the length $n > s$; $F\big(J_i^{i+s-1}; M_r\big) = (j_{i+m_1-1}, j_{i+m_2-1}, \ldots, j_{i+m_r-1})$ is the selector-function of the r-th order for some pattern $M_r \in \mathrm{M}$, $J_i^{i+s-1} \in A^s$, $i \in \mathbb{N}$;

$$\nu_{J_1^{r+1}}(X_1^n; M_r) = \sum_{t=1}^{n-s} \delta_{F\left(X_t^{t+s-1};M_r\right),J_1^r} \delta_{x_{t+s},j_{r+1}}, \quad J_1^{r+1} \in A^{r+1}, \tag{2}$$

are the frequency statistics of the MC(s,r) for some pattern $M_r \in \mathrm{M}$.

Theorem 3. *Let the pattern of connections M_r^0 is known. The maximum likelihood estimator (MLE) for the matrix Q is*

$$\hat{Q} = (\hat{q}_{J_1^{r+1}})_{J_1^{r+1}\in A^{r+1}}, \quad \hat{q}_{J_1^{r+1}} = \begin{cases} \hat{\mu}_{J_1^{r+1}}(M_r^0)/\hat{\mu}_{J_1^r\cdot}(M_r^0), \ if \ \hat{\mu}_{J_1^r\cdot}(M_r^0) > 0, \\ 1/N, \ if \ \hat{\mu}_{J_1^r\cdot}(M_r^0) = 0, \end{cases} \tag{3}$$

$$\hat{\mu}_{J_1^{r+1}}(M_r) = \nu_{J_1^{r+1}}(X_1^n; M_r)/(n-s) \tag{4}$$

is the frequency estimator for the probability $\mu_{J_1^{r+1}}(M_r)$, $J_1^{r+1} \in A^{r+1}$, $M_r \in \mathrm{M}$.

The proof can be found in Kharin and Piatlitski (2007).

Lemma 1. *Assume that the MC(s,r) is a stationary Markov chain. Then the statistics $\hat{\mu}_{J_1^{r+1}}(M_r)$, $J_1^{r+1} \in A^{r+1}$, defined by (4), are unbiased and consistent estimators with covariances* $\mathrm{Cov}\{\hat{\mu}_{J_1^{r+1}}(M_r), \hat{\mu}_{K_1^{r+1}}(M_r)\} = \sigma^{\hat{\mu}}_{J_1^{r+1},K_1^{r+1}}/(n-s) + \mathcal{O}(1/n^2)$, *where*

$$\sigma^{\hat{\mu}}_{J_1^{r+1},K_1^{r+1}} = \mu_{J_1^{r+1}}(M_r)(\delta_{J_1^{r+1},K_1^{r+1}} - \mu_{K_1^{r+1}}(M_r)) + \beta_{J_1^{r+1},K_1^{r+1}}(M_r)\,,$$

$\beta_{J_1^{r+1},K_1^{r+1}}(M_r) = \sum_{I_1^s,L_1^s\in A^s} \delta_{F(I_1^s;M_r),J_1^r}\,\delta_{F(L_1^s;M_r),K_1^r}\,p_{I_1^s,j_{r+1}}\,p_{L_1^s,k_{r+1}}\big(\pi^*_{I_1^s}\,c_{(I_2^s,j_{r+1}),L_1^s} + \pi^*_{L_1^s}\,c_{(L_2^s,k_{r+1}),I_1^s}\big)$, $c_{I_1^s,L_1^s} = \sum_{k=1}^{\infty}(p^{*\,(k)}_{I_1^s,L_1^s} - \pi^*_{L_1^s})$. *Moreover, the probability distribution* $\big(\sqrt{n-s}(\hat\mu_{J_1^{r+1}}(M_r) - \mu_{J_1^{r+1}}(M_r))\big)_{J_1^{r+1}\in A^{r+1}}$ *at* $n\to\infty$ *converges to the normal distribution with zero mean and the covariance matrix* $\Sigma^{\hat\mu} = (\sigma^{\hat\mu}_{J_1^{r+1},K_1^{r+1}})_{J_1^{r+1},K_1^{r+1}\in A^{r+1}}$.

Proof. The frequency statistics of the MC(s,r) defined by (2) are linear functions of the frequency statistics of the Markov chain of the s-th order (see Kharin and Piatlitski, 2007). Using the relationship (4) and generalization of Lemma 3.2, Theorem 3.3 from Billingsley (1961), we come to the necessary statement. □

Remark 1. *The matrix* $C = (c_{J_1^s,I_1^s})_{J_1^s,I_1^s\in A^s}$ *can be expressed linearly in terms of the fundamental matrix* $Z = (E - P^* + \Pi^*)^{-1}$ *of the Markov chain:* $C = Z - E$, *where* $E = (\delta_{J_1^s,I_1^s})_{J_1^s,I_1^s\in A^s}$ *is the identity matrix,* $P^* = (\delta_{J_2^s,I_1^{s-1}}p_{J_1^s,i_s})_{J_1^s,I_1^s\in A^s}$, $\Pi^* = (\pi^*_{I_1^s})_{J_1^s,I_1^s\in A^s}$. *The matrix* Z *can be efficiently calculated (see Kemeny and Snell, 1960).*

Theorem 4. *For the stationary MC*(s,r) *the statistics* $\hat q_{J_1^{r+1}}$, $J_1^{r+1}\in A^{r+1}$, *defined by (3), at* $n\to\infty$ *are asymptotically unbiased and consistent estimators with covariances* $\mathrm{Cov}\{\hat q_{J_1^{r+1}},\hat q_{K_1^{r+1}}\} = \sigma^{\hat q}_{J_1^{r+1},K_1^{r+1}}/(n-s) + \mathcal{O}(1/n^2)$, *where*

$$\sigma^{\hat q}_{J_1^{r+1},K_1^{r+1}} = \delta_{J_1^r,K_1^r}\,q_{J_1^{r+1}}\,(\delta_{j_{r+1},k_{r+1}} - q_{K_1^{r+1}})/\mu_{J_1^r\cdot}(M_r^0),\ \ J_1^{r+1},\ K_1^{r+1}\in A^{r+1}.$$

Moreover, the probability distribution of the N^{r+1}*-dimensional random vector* $\big(\sqrt{n-s}(\hat q_{J_1^{r+1}} - q_{J_1^{r+1}})\big)_{J_1^{r+1}\in A^{r+1}}$ *at* $n\to\infty$ *converges to the normal distribution with zero mean and the covariance matrix* $\Sigma^{\hat q} = (\sigma^{\hat q}_{J_1^{r+1},K_1^{r+1}})_{J_1^{r+1},K_1^{r+1}\in A^{r+1}}$.

Proof. Consistency of the estimators immediately follows from the equation (3), Lemma 1 and the theorem on functional transformations of random sequences (see Borovkov, 1999).

Since $\partial\hat q_{J_1^{r+1}}/\partial\hat\mu_{K_1^{r+1}}(M_r^0) = \delta_{J_1^r,K_1^r}(\delta_{j_{r+1},k_{r+1}} - \hat q_{J_1^{r+1}})/\hat\mu_{J_1^r\cdot}(M_r^0)$, then using the continuity theorem for random sequences (see Borovkov, 1999) and Lemma 1, we have that the probability distribution of the random vector $\big(\sqrt{n-s}(\hat q_{J_1^{r+1}} - q_{J_1^{r+1}})\big)_{J_1^{r+1}\in A^{r+1}}$ converges to the normal distribution with zero mean and the covariances

$$\sigma^{\hat q}_{J_1^{r+1},K_1^{r+1}} = \sum_{i_{r+1},l_{r+1}\in A}\frac{(\delta_{j_{r+1},i_{r+1}} - q_{J_1^{r+1}})(\delta_{k_{r+1},l_{r+1}} - q_{K_1^{r+1}})}{\mu_{J_1^r\cdot}(M_r^0)\mu_{K_1^r\cdot}(M_r^0)}\,\sigma^{\hat\mu}_{(J_1^r,i_{r+1}),(K_1^r,l_{r+1})}\,.$$

Asymptotic unbiasedness and expressions for asymptotic covariances of the estimators $\hat q_{J_1^{r+1}}$, $J_1^{r+1}\in A^{r+1}$, follow from the continuity theorem for moments (see Borovkov, 1999) and the boundedness of the estimators $\hat q_{J_1^{r+1}}\in[0,1]$, $J_1^{r+1}\in A^{r+1}$. □

Corollary 4. *Under the assumptions of Theorem 4, the statistic* $\hat Q$ *is the mean-square consistent estimator. Moreover, the mean-squared error of* $\hat Q$ *is*

$$\Delta_n(\hat Q) = \mathbf{E}\{\|\hat Q - Q\|^2\} = (n-s)^{-1}\sum_{J_1^{r+1}\in A^{r+1}} q_{J_1^{r+1}}(1 - q_{J_1^{r+1}})/\mu_{J_1^r\cdot}(M_r^0) + \mathcal{O}(1/n^2)\,,$$

where $\|\cdot\|$ *is the Euclidean matrix norm.*

Consider the problem of statistical testing of two hypotheses: $H_0 = \{Q = Q^0\}$, where $Q^0 = (q^0_{J_1^{r+1}})_{J_1^{r+1} \in A^{r+1}}$ is some fixed stochastic matrix; $H_1 = \overline{H}_0$ is the alternative.

The test statistic based on Theorem 4 is

$$\rho = \sum\nolimits_{J_1^r \in A^r} \sum\nolimits_{j_{r+1} \in A,\, q^0_{J_1^{r+1}} > 0} \nu_{J_1^r \cdot}(X_1^n; M_r^0)(\hat{q}_{J_1^{r+1}} - q^0_{J_1^{r+1}})^2 / q^0_{J_1^{r+1}}\,. \tag{5}$$

The parametric family of decision rules for testing the hypotheses H_0, H_1 has the following form: to accept $\{H_0$, if $\rho \leq \Lambda$; H_1, if $\rho > \Lambda\}$, where $\Lambda > 0$ is some parameter (threshold value).

Theorem 5. *Let the MC(s, r) is stationary. Then the asymptotic probability distribution of the statistic ρ defined by (5) under H_0 at $n \to \infty$ is the standard χ^2 distribution with $U = \sum_{J_1^r \in A^r} (|D_{J_1^r}| - 1)$ degrees of freedom, where $D_{J_1^r} = \{j_{r+1} \in A : \ q_{J_1^{r+1}} > 0\}$.*

Corollary 5. *Assume that the conditions of Theorem 5 be satisfied. If $\Lambda = G_U^{-1}(1-\varepsilon)$ is the $(1-\varepsilon)$-quantile for the standard χ^2 distribution with U degrees of freedom, then the probability of the type I error (the significance level of the test) tends to ε at $n \to \infty$.*

Thus, the statistical test for the hypotheses H_0, H_1 based on Theorem 5, Corollary 5 consists of the following four steps.

1. Computation of the statistics $\nu_{J_1^{r+1}}(X_1^n; M_r^0)$, $J_1^{r+1} \in A^{r+1}$, by (2).
2. Computation of the statistic ρ by (5).
3. Computation of the P-value: $P = 1 - G_U(\rho)$, where $G_U(\cdot)$ is the probability distribution function of the standard χ^2 distribution with U degrees of freedom.
4. The decision rule (at the significance level ε): If the P-value $\geq \varepsilon$, then to conclude that the hypothesis H_0 is true. Otherwise, to conclude that the alternative H_1 is true.

Corollary 6. *If the conditions of Theorem 5 hold and $H_1 = \{Q = Q^1\}$ is true, where*

$$Q^1 = (q^1_{J_1^{r+1}})_{J_1^{r+1} \in A^{r+1}}, \ q^1_{J_1^{r+1}} = q^0_{J_1^{r+1}}(1 + d_{J_1^{r+1}}/\sqrt{n-s}\,)\,, \tag{6}$$

$\sum_{j_{r+1} \in A} d_{J_1^{r+1}} q^0_{J_1^{r+1}} = 0$, $\sum_{J_1^{r+1} \in A^{r+1}} |d_{J_1^{r+1}}| > 0$, *then at $n \to \infty$ the power of the developed test $w \to 1 - G_{U,a}\big(G_U^{-1}(1-\varepsilon)\big)$, where $G_{U,a}(\cdot)$ is the probability distribution function of the noncentral χ^2 distribution with U degrees of freedom and the noncentrality parameter $a = \sum_{J_1^{r+1} \in A^{r+1}} \mu_{J_1^{r+1}}\,(M_r^0)\ d^2_{J_1^{r+1}}$.*

Remark 2. *The equation (6) means the contiguity property of the alternative H_1: the increase of the length n implies approaching of H_1 to H_0 with the rate $\mathcal{O}(1/\sqrt{n}\,)$.*

3.2 Statistical Estimation of the Pattern M_r^0

Introduce the notation:

$$H(M_r) = -\sum\nolimits_{J_1^{r+1} \in A^{r+1}} \mu_{J_1^{r+1}}(M_r) \log\Big(\mu_{J_1^{r+1}}(M_r)/\mu_{J_1^r \cdot}(M_r)\Big) \geq 0 \tag{7}$$

is the conditional entropy of the future symbol $x_{t+s} \in A$ relative to the past set by the selector $F\left(X_t^{t+s-1};\, M_r\right) \in A^r$, $M_r \in \mathrm{M}$; $\hat{H}(M_r)$ is the "plug-in" estimator of the conditional entropy that is generated by substitution of true probabilities $\mu_{J_1^{r+1}}(M_r)$ in (7) by their estimators $\hat{\mu}_{J_1^{r+1}}(M_r)$, $J_1^{r+1} \in A^{r+1}$.

Theorem 6. *If the order s and the number of connections r are known, then the maximum likelihood estimator for the pattern of connections M_r^0 is*

$$\hat{M}_r = \arg\min_{M_r \in \mathrm{M}} \hat{H}(M_r)\,. \tag{8}$$

Theorem 7. *If the MC(s,r) is a stationary Markov chain, then the estimator $\hat{M}_r$ defined by (8) at $n \to \infty$ is consistent: $\hat{M}_r \xrightarrow{\mathbf{P}} M_r^0$.*

Proof. Using equations (1), Lemma 1 and the theorem on functional transformations of random sequences (see Borovkov, 1999), we have

$$-\sum\nolimits_{K_1^{s+1}\in A^{s+1}} \pi_{K_1^s}\, p_{K_1^{s+1}} \log p_{K_1^{s+1}} = H(M_r^0),\ \ \hat{H}(M_r^0) \xrightarrow{\mathbf{P}} H(M_r^0)\,.$$

Since the matrix $P = (p_{K_1^{s+1}})_{K_1^{s+1}\in A^{s+1}}$ uniquely determines the pair $(Q,\ M_r^0)$ for the MC(s,r), we obtain the consistency of the estimator $\hat{M}_r$. □

Theorem 8. *If the conditions of Theorem 7 hold, then the probability distribution of the random vector $\left(\sqrt{n-s}(\hat{H}(M_r) - H(M_r))\right)_{M_r\in\mathrm{M}}$ at $n \to \infty$ converges to the normal distribution with zero mean and the covariance matrix $\Sigma^{\hat{H}} = \left(\sigma^{\hat{H}}_{M_r',\,M_r''}\right)_{M_r',\,M_r''\in\mathrm{M}}$:*

$$\sigma^{\hat{H}}_{M_r',\,M_r''} = \sum\nolimits_{I_1^{s+1}\in A^{s+1}} \pi^*_{I_1^s}\, p_{I_1^{s+1}} \log\frac{\mu_{F(I_1^s;M_r'),\,i_{s+1}}(M_r')}{\mu_{F(I_1^s;M_r')\cdot}(M_r')} \log\frac{\mu_{F(I_1^s;M_r''),\,i_{s+1}}(M_r'')}{\mu_{F(I_1^s;M_r'')\cdot}(M_r'')} -$$
$$-H(M_r')H(M_r'') + \sum\nolimits_{I_1^{s+1},\,L_1^{s+1}\in A^{s+1}} p_{I_1^{s+1}} p_{L_1^{s+1}} \left(\pi^*_{I_1^s} c_{I_2^{s+1},L_1^s} + \pi^*_{L_1^s} c_{L_2^{s+1},I_1^s}\right) \times$$
$$\times \log\left(\mu_{F(I_1^s;M_r'),\,i_{s+1}}(M_r') / \mu_{F(I_1^s;M_r')\cdot}(M_r')\right) \log\left(\mu_{F(L_1^s;M_r''),\,l_{s+1}}(M_r'') / \mu_{F(L_1^s;M_r'')\cdot}(M_r'')\right).$$

Proof. Since $\hat{H}(M_r)$ is the continuously differentiable functional transformation, then

$$\hat{H}(M_r) = \sum_{J_1^{r+1}\in A^{r+1}} \hat{\mu}_{J_1^{r+1}}(M_r) \log\frac{\hat{\mu}_{J_1^r\cdot}(M_r)}{\hat{\mu}_{J_1^{r+1}}(M_r)},\ \frac{\partial \hat{H}(M_r)}{\partial \hat{\mu}_{K_1^{r+1}}(M_r)} = \log\frac{\hat{\mu}_{J_1^r\cdot}(M_r)}{\hat{\mu}_{J_1^{r+1}}(M_r)}\,. \tag{9}$$

Then the statement of this theorem follows from the equation (9), the generalization of Lemma 1 and the continuity theorem for random sequences (see Borovkov, 1999). □

Theorem 8 can be used for estimation of the error probability $\mathbf{P}\{\hat{M}_r \neq M_r^0\}$.

3.3 Statistical Estimation of r and s

Let $s \in [s_-, s_+]$, $r \in [r_-, r_+]$, $1 \le s_- < s_+ < \infty$, $1 \le r_- < r_+ < s_+$. For estimation of r and s let us use the Bayesian information criterion (see Csiszar and Shields, 2000), which in our case has the form: $BIC(s,r) = 2(n-s)\hat{H}(\hat{M}_r) + U\log(n-s)$, where $U = \sum_{J_1^r\in A^r}\left(|D_{J_1^r}| - 1 + \delta_{\hat{\mu}_{J_1^r\cdot}(\hat{M}_r),0}\right)$, $D_{J_1^r} = \{j_{r+1}\in A : \hat{\mu}_{J_1^{r+1}}(\hat{M}_r) > 0\}$.

Statistical estimators for r and s are defined by the minimization:

$$BIC(s,r) \to \min_{s_-\le s\le s_+,\ r_-\le r\le r_+}\ . \tag{10}$$

where the first s_+ elements of the sequence X_1^n are initial (for the MC(s,r) the first $(s_+ - s)$ elements does not take part in the computation of the $BIC(s,r)$).

Theorem 9. *If the MC(s,r) is stationary, then $\hat r$, $\hat s$ defined by (10) are consistent.*

Proof. For some s^*, r^* introduce the variable $\alpha = \hat H(\hat M_r) + \sum_{J_1^{r^*+1}\in A^{r^*+1}} \hat\mu_{J_1^{r^*+1}}(\hat M^*_{r^*})\times$ $\times \log\left(\hat\mu_{J_1^{r^*+1}}(\hat M^*_{r^*})/\hat\mu_{J_1^{r^*}\cdot}(\hat M^*_{r^*})\right)$. Let $s^* \in [s_-, s)$, $r^* \in [r_-, r_+]$, then using Theorems 4, 7 we have $\alpha \xrightarrow[n\to\infty]{\mathbf{P}} \sum_{J_1^{s+1}\in A^{s+1}} \pi_{J_1^{s+1}} \log\left(q^*_{F(J^s_{s-s^*+1};\, M^*_{r^*}),\, j_{s+1}} / q_{F(J_1^s;\, M_r^0),\, j_{s+1}}\right) = -\varepsilon$, where $q^*_{F(J^s_{s-s^*+1};\, M^*_{r^*}),\, j_{s+1}} = \mu_{F(J^s_{s-s^*+1};\, M^*_{r^*}),\, j_{s+1}}(M^*_{r^*}) / \mu_{F(J^s_{s-s^*+1};\, M^*_{r^*})\cdot}(M^*_{r^*})$. Applying equation (1) and the Jensen's inequality (see Borovkov, 1999) we obtain that $\varepsilon > 0$. Thus, $\mathbf{P}\left\{BIC(s^*, r^*) - BIC(s,r) \geq 2(n-s_+)\varepsilon + (U^* - U)\log(n - s_+)\right\} \to 1$.

Similarly, as in the proof of the previous case, we conduct the proof for $s^* \in [s, s_+]$, $r^* \in [r_-, r)$. For $s^* \in [s, s_+]$, $r^* \in [r, r_+]$ and $n \to \infty$, we have $\mathbf{P}\{\alpha \leq u\} \to 1$, where $u = \sum_{J_1^{r^*+1}\in A^{r^*+1}} (\nu_{J_1^{r^*+1}}(X_1^n; \hat M^*_{r^*}) - \nu_{J_1^{r^*}\cdot}(X_1^n; \hat M^*_{r^*}) q_{F^*(J_1^{r^*}; M_r^0),\, j_{r^*+1}})^2 / \nu_{J_1^{r^*+1}}(X_1^n; \hat M^*_{r^*})$, $F^*\left(J_1^{r^*};\, M_r^0\right) = (j_{l_1}, j_{l_2}, \ldots, j_{l_r})$, $l_k \in \{1, 2, \ldots, r^*\}$, $m^*_{l_k} = s^* - s + m_k$, $k = 1, 2, \ldots, r$. Using the proof of Theorem 1 from Dorea and Lopes (2006), we obtain the consistency property of $\hat r$, $\hat s$. □

4 Applications of the MC(s,r) to Real Data

In this section, we analyze several sets of real statistical data illustrating three fields of applications of the MC(s,r). Performance of the fitted models was evaluated in the computer experiments by the BIC as in Berchtold (2002), Raftery and Tavare (1994).

4.1 Wind Modeling

Consider the time-series of the daily average wind speed y_t at Malin Head (North of Ireland) during the period 1961-1978 (see Raftery and Tavare, 1994), $n = 6574$. We want to model two extreme situations (see Berchtold, 2002): days with exceptional low and high wind speed. Accordingly, we classified the data into three categories (the wind speed is given in knots, 1 knot = 0.5148 m/s):

$$x_t = \begin{cases} 0, & \text{if the wind speed is low, i. e. } y_t < 5; \\ 1, & \text{if the wind speed is normal, i. e. } y_t \in [5, 20]; \\ 2, & \text{if the wind speed is high, i. e. } y_t > 20. \end{cases}$$

The data $x_1, x_2, \ldots, x_n$ were fitted by the MC(s,r)-model $\left(1 \leq s \leq 7,\ 1 \leq r \leq s\right)$ using the algorithms defined in Section 3. Table 1 presents the computer results. As it is seen from Table 1 the best fitted model is the Markov chain of the order $\hat s = 3$ with $\hat r = 2$ partial connections, $\hat M_r = (1, 3)$,

$$\begin{pmatrix} 0.267 & 0.081 & 0 & 0.219 & 0.038 & 0 & 0.211 & 0.017 & 0 \\ 0.733 & 0.861 & 0.625 & 0.775 & 0.819 & 0.525 & 0.790 & 0.720 & 0.432 \\ 0 & 0.058 & 0.375 & 0.006 & 0.143 & 0.475 & 0 & 0.263 & 0.568 \end{pmatrix}^T .$$

It means that the speed at the day $t+3$ depends mainly on the speed at the days $t+2$, t. If the speed at the days t, $t+2$ was normal, then the probability to stay in the state "normal" at the day $t+3$ is high (0.819).

Table 1: Different modelings of the wind speed data

Model	BIC	Model	BIC	Model	BIC	Model	BIC
MC(1,1)	8127.52	MC(4,2)	8139.12	MC(5,5)	8621.97	MC(7,1)	9041.43
MC(2,1)	8777.63	MC(4,3)	8164.79	MC(6,1)	9016.23	MC(7,2)	8163.07
MC(2,2)	8096.08	MC(4,4)	8332.77	MC(6,2)	8148.48	MC(7,3)	8197.91
MC(3,1)	8849.90	MC(5,1)	8984.10	MC(6,3)	8190.78	MC(7,4)	8323.19
MC(3,2)	**8079.81**	MC(5,2)	8129.83	MC(6,4)	8350.82	MC(7,5)	8599.09
MC(3,3)	8143.13	MC(5,3)	8177.92	MC(6,5)	8576.92	MC(7,6)	8973.15
MC(4,1)	8956.11	MC(5,4)	8349.62	MC(6,6)	8969.54	MC(7,7)	9575.64

4.2 Song of the Bird "Wood Pewee"

Consider one of the records of the morning twilight song with three possible values corresponding to the three distinct phrases of the "Wood Pewee" song: $N = 3$, $n = 1327$. These data were analyzed in Raftery and Tavare (1994) and Berchtold (2002) by different models indicated in Table 2: HMM 2 (1) is the two states first-order hidden Markov model (see Rabiner, 1989); "Pattern" is the best model found by Raftery and Tavare (1994); MTDg l is the order l MTDg model (see Raftery, 1985); DCMM K $(l;h)$ is the K states double chain Markov model, where l is the order of the hidden Markov chain, h is the order of the observed Markov chain (see Berchtold, 2002). We applied the MC(s,r) to the same data.

Table 2: Different modelings of the "Wood Pewee" song

Model	BIC	Model	BIC	Model	BIC
MC(1,1)	1424.2	MC(4,2)	975.2	DCMM 2 (2;MTD 2)	832.2
MC(2,1)	1179.1	**MC(4,3)**	**765.9**	DCMM 3 (1;2)	795.8
MC(2,2)	801.9	MC(4,4)	768.3	DCMM 3 (2;2)	774.5
MC(3,1)	1930.7	"Pattern"	827.0	DCMM 2 (1;2)	857.2
MC(3,2)	1171.1	HMM 2 (1)	2223.2	DCMM 2 (1;MTD 2)	818.4
MC(3,3)	808.6	MTDg 2	1037.4	DCMM 2 (1;MTDg 2)	843.2
MC(4,1)	1541.4	MTDg 3	1032.7	**DCMM 2 (2;2)**	**733.0**

As it is seen from Table 2 the best fitted models are the DCMM 2 (2;2) and the MC(4,3) with $\hat{M}_r = (1, 3, 4)$. The matrix $\hat{Q}$ is represented in Figure 1, where dark-grey color is the transition probability to the state "0", light-grey color — to "1", white color — to "2". One can see that the fitted model MC(4,3) reveals significant dependencies in the observed time series.

4.3 DNA Analysis

DNA sequences use the alphabet of 4 bases $\{A, C, G, T\}$. Here, we study a binary representation of this alphabet, the purine-pyrimidine alphabet (see Berchtold, 2002).

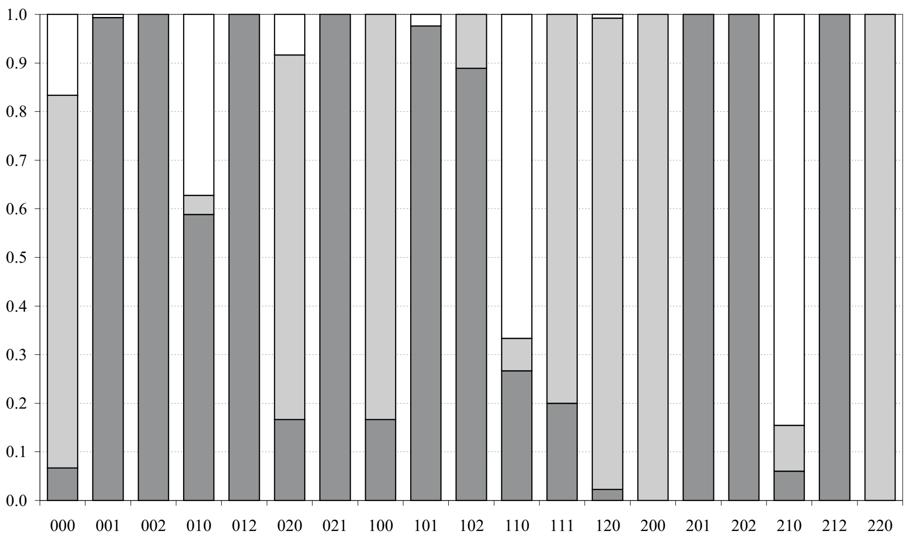

Figure 1: Illustration of the matrix $\hat{Q}$ for the “Wood Pewee” song

Each base is recoded as either purine $(\{A, G\})$ or pyrimidine $(\{C, T\})$. We considered methanococcus maripaludis C5 genomic DNA sequence from DNA Data Bank of Japan (`www.ddbj.nig.ac.jp`), its length is $n = 2^{17}$. We analyzed these data using the MC(s, r)-model $\big(1 \le s \le 32,\ 1 \le r \le \min\{s, 8\}\big)$.

The best fitted model is the Markov chain of the order $\hat{s} = 27$ with $\hat{r} = 5$ partial connections, the pattern $\hat{M}_r = (1, 7, 16, 19, 27)$ and the matrix $\hat{Q}$ (represented in Figure 2, where dark-grey color is the transition probability to the state “0”, light-grey color — to “1”). As it is seen from Figure 2 the MC(27,5) provides a good fitting of the examined DNA sequence.

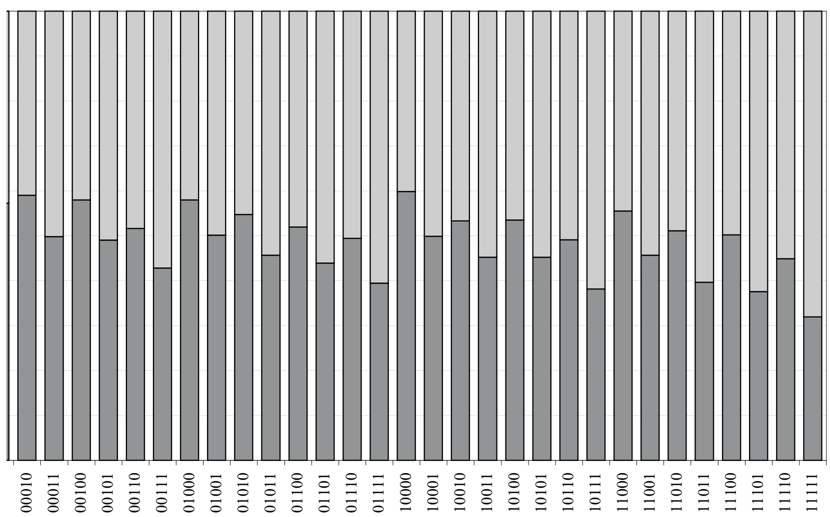

Figure 2: Illustration of the matrix $\hat{Q}$ for the DNA sequence

5 Conclusion

In this paper, we considered a new model for discrete-valued time series — Markov chain with partial connections MC(s, r). The properties of the MC(s, r)-model were analyzed. Consistent estimators of the parameters were constructed. A statistical test for the value of the matrix Q is proposed, its power is evaluated. The proposed model is illustrated in fitting of real statistical data.

References

Avery, P. J., and Henderson, D. A. (1999). Fitting Markov chain models to discrete state series such as DNA sequences. *Applied Statistics*, *48*, 53-61.

Berchtold, A. (2002). High-order extensions of the double chain Markov model. *Stochastic Models*, *18*, 193-227.

Billingsley, P. (1961). Statistical methods in Markov chains. *The Annals of Mathematical Statistics*, *32*, 12-40.

Borovkov, A. A. (1999). *Mathematical Statistics*. Amsterdam: Taylor and Francis.

Buhlmann, P., and Wyner, A. J. (1999). Variable length Markov chains. *The Annals of Statistics*, *27*, 480-513.

Csiszar, I., and Shields, P. C. (2000). The consistency of the BIC Markov order estimator. *The Annals of Statistics*, *28*, 1601-1619.

Dorea, C. C. Y., and Lopes, J. S. (2006). Convergence rates for Markov chain order estimates using EDC criterion. *Bulletin of the Brazilian Mathematical Society*, *37*, 561-570.

Jacobs, P. A., and Lewis, P. A. W. (1978). Discrete time series generated by mixtures I: Correlational and runs properties. *Journal of the Royal Statistical Society, Series B*, *40*, 94-105.

Kemeny, J. G., and Snell, J. L. (1960). *Finite Markov Chains*. New York: D. van Nostrand Com.

Kharin, Y. S., and Piatlitski, A. I. (2007). A Markov chain of order s with r partial connections and statistical inference on its parameters. *Discrete Mathematics and Applications*, *17*, 295-317.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*, 257-286.

Raftery, A. (1985). A model for high-order Markov chains. *Journal of the Royal Statistical Society, Series B*, *47*, 528-539.

Raftery, A., and Tavare, S. (1994). Estimation and modelling repeated patterns in high order Markov chains with the MTD model. *Applied Statistics*, *43*, 179-199.

Author's address:

Yurij Kharin and Andrei Piatlitski
Department of Mathematical Modeling and Data Analysis
Belarusian State University
Independence Avenue 4
220030 Minsk, Belarus
E-Mail: `Kharin@bsu.by`