On Statistical Analysis of Markov Chains with Conditional Memory Depth

Yu.S. Kharin, M.V. Maltsew
Research Institute for Applied Problems of Mathematics and Informatics
Belarusian State University
Minsk, BELARUS
e-mail: `kharin@bsu.by, maltsew@mail.ru`

Abstract

A new mathematical model of the s-order Markov chain with conditional memory depth is proposed. Maximum likelihood estimators of parameters are constructed and their properties are analyzed. A statistical test on parameter values is constructed. Numerical results are presented.

Keywords: Markov Chain, Ergodicity, MLE, Statistical Test

1 Introduction

Markov chains [1] are used to solve many applied problems of statistical data analysis in genetics [2], economics and in many other fields. The Markov chain of the order s, $s \geq 1$ is a general model in these problems, but the number of parameters of this model increases exponentially when the order s grows. One needs to have the data set of huge size to estimate this general model. Thus development and research of high-order Markov chain models with a small number of parameters is a significant problem. The examples of such models are the Markov chain of order s with r partial connections [3], the Raftery model [4], the variable length Markov chain [5] and the Markov chain with conditional memory depth that is proposed in this paper.

2 Mathematical model

Introduce the notation: $2 \leq N < \infty$; $A = \{0, 1, \dots, N-1\}$ is the state space of N elements; $x_t \in A, t \in \mathbf{N}$, is a homogeneous Markov chain of the order s $(2 \leq s < \infty)$ defined at the probability space (Ω, F, P); $J_1^{s+1} = (j_1, \dots, j_{s+1}) \in A^{s+1}$ is the multiindex; $P = (p_{J_1^{s+1}})$ is the one-step probability transition matrix, $p_{J_1^{s+1}} = \mathrm{P}\{x_{t+s} = j_{s+1} | x_{t+s-1} = j_s, \dots, x_t = j_1\}, \forall t \in \mathbf{N}$; $1 \leq B_* \leq s-1$; $K = N^{B_*} - 1$; $1 \leq M \leq K+1$; $Q^{(1)}, \dots, Q^{(M)}$ are M different stochastic matrices of the order N: $Q^{(m)} = (q_{i,j}^{(m)}), i, j \in A,\ 1 \leq m \leq M$; $< J_n^m >= \sum\limits_{k=n}^{m} N^{k-1} j_k,\ 1 \leq n \leq m \leq s$; $\delta_{J_n^m, I_n^m} = \prod\limits_{k=n}^{m} \delta_{j_k, i_k}$ is the Kronecker symbol for multiindices J_n^m, I_n^m.

Define the Markov chain $x_t \in A$ of the order s as the Markov chain with conditional memory depth if its one-step transition probabilities have the following form:

$$p_{J_1^{s+1}} = \sum_{k=0}^{K} \delta_{<J_{s-B_*+1}^{s}>,k} \cdot q_{j_{b_k}, j_{s+1}}^{(m_k)}, \tag{1}$$

where $1 \leq m_k \leq M$, $1 \leq b_k \leq s - B_*$, $0 \leq k \leq K$, $\min\limits_{0 \leq k \leq K} b_k = 1$.

The sequence of elements $J_{s-B_*+1}^{s}$, which determines conditions in the formula (1), is called the base memory fragment (BMF). We can see that the state of the model x_t at time t doesn't depend on all previous states, but depends only on $B_* + 1$ states $(j_{b_k}, J_{s-B_*+1}^{s})$; the value of BMF $J_{s-B_*+1}^{s}$ determines not only the state j_{b_k}, but it also determines the transition matrix. The transition matrix $P = (p_{J_1^{s+1}})$ is defined by $D = 2(N^{B_*} + 1) + MN(N-1)$ independent parameters.

3 Statistical Analysis

Find ergodicity conditions for the Markov chain with conditional memory depth.

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

$$\min_{J_1^s, J_{1+m}^{s+m} \in A^s} \sum_{J_{s+1}^m \in A^{m-s}} \prod_{i=1}^{m} \sum_{k=0}^{K} \delta_{<J_{i+s-B_*}^{i+s-1}>,k} \cdot q_{j_{b_k+i-1}, j_{i+s}}^{(m_k)} > 0.$$

Define the following frequency statistics

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

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

$$A^{s+1}(J_1^l) = \{I_1^{s+1} \in A^{s+1} | I_1^l = J_1^l\},\ 1 \leq l \leq s,\ 0 \leq l_0 \leq s - l,$$
$$A^{1+l_0+l}(j_0^{l_0}, J_1^l) = \{I_1^{1+l_0+l} \in A^{1+l_0+l} | i_1 = j_0, I_{2+l_0}^{1+l_0+l} = J_1^l\}.$$

Theorem 2. If true values of $B_*, \{b_k\}$, and $\{m_k\}$ are known, then the maximum likelihood estimators (MLE) of the one-step transition probabilities $q_{u,v}^{(m_k)}$, $1 \leq m_k \leq M, u, v \in A$, are defined as follows:

$$\hat{q}_{u,v}^{(m_k)} = \{ \sum_{w \in A^{B_*}} \delta_{<w>,k} \frac{\nu_{u,wv}^{(l_k)}(n)}{\nu_{u,w}^{(l_k)}(n)},\ if\ \nu_{u,w}^{(l_k)}(n) > 0,\ 1/N,\ if\ \nu_{u,w}^{(l_k)}(n) = 0\}. \tag{2}$$

Theorem 3. If the Markov chain with conditional memory depth is stationary, then (2) are consistent estimators: $\hat{q}_{u,v}^{(m)} \xrightarrow{\mathrm{P}} q_{u,v}^{(m)},\ 1 \leq m \leq M,\ n \to \infty$.

Theorem 4. If the true values of B_*, $\{m_k\}$ are known, then the MLE of $\{b_k\}$ are

$$\hat{b}_k = arg \max_{1 \leq b \leq s - B_*} \sum_{i,j \in A} \nu_{i,wj}^{s-b-B_*}(n) \ln(\hat{q}_{i,wj}^{(m_k)}),\ k = 1, 2 \ldots, K. \tag{3}$$

Construct now the statistical test for the hypotheses $H_0 = \{x_t \in A$ is a "purely" random sequence: $q_{i,j}^{(m)} = 1/N,\ \forall i, j \in A,\ m = 1, 2, \ldots, M\}$ and $H_1 = \{x_t \in A$ is the Markov chain with conditional memory depth with one-step transition probabilities $q_{i,j}^{(m)} = q_{i,j}^{(m)}(n) = \frac{1}{N}(1 + \frac{\omega_{i,j}^{(m)}(n)}{\sqrt{n}}) > 0$, $\omega_{i,j}^{(m)}(n) \to \omega_{i,j}^{(m)}$ as $n \to \infty$, $\sum\limits_{j\in A} \omega_{i,j}^{(m)} = 0$, $\sum\limits_{m=1}^{M} \sum\limits_{i,j\in A} |\omega_{i,j}^{(m)}| > 0\}$. Define the following statistic: $\rho(n) = \sum\limits_{w\in B_*,u,v\in A} \sum\limits_{k=0}^{K} \delta_{<w>,k}(\xi_{u,wv}^{(l_k)})^2 - \frac{1}{N} \sum\limits_{w\in B_*,u\in A} (\sum\limits_{v\in A} \sum\limits_{k=0}^{K} \delta_{<w>,k} \cdot \xi_{u,wv}^{(l_k)})^2$.

Theorem 5. If the hypothesis H_0 is true, then as $n \to \infty$ the probability distribution of the statistic $\rho(n)$ converges to the χ^2- distribution with $U = N^{B_*+1}(N-1)$ degrees of freedom.

Using this theorem we can construct the statistical test:

$$\text{decide } \{H_0 : \rho(n) \leq \Delta;\ H_1 : \rho(n) > \Delta\}.$$

where $\Delta = G_U^{-1}(1-\alpha)$ is the $(1-\alpha)$-quantile of the standard χ^2- distribution with U degrees of freedom, α is the given significance level.

4 Numerical results

Evaluate performance of the proposed statistical procedures by computer experiments.

Example 1 (simulated data). The state space $A = \{0, 1\}, N = 2,\ s = 4,\ M = 2,\ B_* = 2,\ b_0 = b_1 = 2,\ b_2 = b_3 = 1,\ m_0 = m_2 = 1,\ m_1 = m_3 = 2,\ Q^{(1)} = \begin{pmatrix} 0.18 & 0.82 \\ 0.41 & 0.59 \end{pmatrix},\ Q^{(2)} = \begin{pmatrix} 0.77 & 0.23 \\ 0.09 & 0.91 \end{pmatrix}$.

Numerical experiments were conducted by the following scheme. The Markov chain with conditional memory depth was simulated using the developed computer program. Then estimates $\hat{Q}^1$ and $\hat{Q}^2$ of matrices $Q^{(1)}$ and $Q^{(2)}$ were calculated using the formula (2). After that the estimate of the variance $\hat{v}_n^u = \sum\limits_{k=1}^{2} \sum\limits_{i,j=0}^{1} (\hat{q}_{ij}^{(k)} - q_{ij}^{(k)})^2$ was calculated, where n is the length of the u-th realization of the Markov chain; $n = 500, 750, 1000, \ldots, 10000$; $u = 1, 2, \ldots, U$. Estimates $\{\hat{v}_n^u\}$ were calculated for each n and for $U = 1000$ independent replications. The performance of the estimators (3) $\hat{v}_n = \frac{1}{U} \sum\limits_{u=1}^{U} \hat{v}_n^u$ was obtained; its dependence on n is plotted in Figure 1.

Figure 1: Performance of the estimators (2)

Similarly we illustrate performance of estimators (3). In order to do this, we calculate the frequency of the true decision $\varepsilon = \frac{1}{U}\sum_{u=1}^{U}\varepsilon_u$, $\varepsilon_u = \delta_{\hat{b},b}$, $b = (b_0, \dots, b_K)$. The results are given in Table 1.

Table 1: Performance of the estimators (3)

n	500	1000	1500	2000	2500	3000	3500	4000	≥ 4500
ε	0.322	0.695	0.896	0.936	0.982	0.986	0.992	0.995	1.000

Example 2 (real data). DNA sequence of drosophila from the genetic data bank [6] was used. The sequence length $n = 12500$, the state space $A = \{A, C, G, T\}$, $N = 4$, $B_* = 2$, $s \in \{3, 4, \dots, 11, 16, 24, 32, 48\}$. Estimates of the vector b were obtained for each value s. In addition the value $\Delta_s = \frac{1}{N^K}\sum_{k=1}^{K}\Delta_{sk}$ was calculated, where $\Delta_{sk} = \sum_{i,j\in A}(\hat{q}_{ij}^{(k)} - 1/N)^2$ characterizes the departure of the observed DNA sequence from the "purely" random sequence. Estimates Δ_s^0 were calculated for the "purely" random sequence simulated by the computer program. The value Δ_s^0 was varied from 0.12 to 0.17, and the value Δ_s was varied from 0.46 to 0.48 for different s. So we can see that the observed DNA sequence could not be considered as the "purely" random sequence. Is was detected in experiments, that for some values of the BMF the correspondent components of the vector b were unchanged for increasing values of s. Here are these BMF values and correspondent values of the vector b component: AA - 1, AC - 2, CG - 3, GT - 1.

References

[1] Kemeny J., Snell J. (1963). Finite Markov Chains. Princeton NJ: D. Van Nostrand Company.

[2] Waterman M. S. (1999). Mathematical Methods for DNA Sequences. Chapman and Hall/CRC: Boca Raton.

[3] Kharin Yu. S., Piatlitski A. I. (2007). Markov Chain with Partial Associations and statistical conclusions of its parameters. Discrete mathematics. Vol. 19, 2, pp. 109-130.

[4] Raftery A.E. (1985). A Model for High-Order Markov Chains. J. Royal Statistical Society. Vol. B-47, 3, pp. 528-539.

[5] Buhlmann P., Wyner A. (1999). Variable Length Markov Chains. The Annals of Statistics. Vol. 27, 2, pp. 480-513.

[6] http://www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html