Math-Net.Ru

Общероссийский математический портал

Yu. S. Kharin, Экономные модели цепей Маркова высокого порядка для оценивания криптографических генераторов, *Матем. вопр. криптогр.*, 2016, том 7, выпуск 2, 131–142

DOI: <https://doi.org/10.4213/mvk191>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением

<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 54.224.187.69

26 апреля 2023 г., 14:56:40

Parsimonious models of high-order Markov chains for evaluation of cryptographic generators

Yu. S. Kharin

Belarusian State University, Minsk, Belarus

Получено 01.III.2015

Abstract. Parsimonious (small-parametric) high-order Markov chains determined by a small number of parameters may be used as models of output sequences in cryptographic generators and their blocks. The paper presents methods of statistical identification (parameter estimation and hypotheses testing) by the observed output sequence for Jacobs–Lewis model, Raftery MTD model, Markov chain with partial connections, Markov chain of conditional order.

Keywords: cryptograhic generator, output sequence, high-order Markov chain, parsimonious model, statistical identification

Экономные модели цепей Маркова высокого порядка для оценивания криптографических генераторов

Ю. С. Харин

Белорусский государственный университет, Минск, Белоруссия

Аннотация. Экономные (малопараметрические) сложные цепи Маркова, определяющиеся небольшим числом параметров, могут использоваться как модели выходных последовательностей криптографических генераторов и их блоков. Представлены методы статистической идентификации (оценивания параметров и проверки гипотез) по наблюдаемой выходной последовательности для модели Якобса–Льюиса, модели Рафтери, цепей Маркова с частичными связями, цепей Маркова условного порядка.

Ключевые слова: криптографический генератор, выходная последовательность, цепь Маркова высокого порядка, экономная модель, статистическая идентификация

1. Introduction

Cryptographic generators are necessary elements for cryptographic systems of information protection [1]. Classification of cryptographic generators based on their construction and implementation may be found in [9]. To evaluate cryptographic security of generators two approaches are known: 1) algebraic approach based on construction of some algebraic model for generator [1]; 2) stochastic approach based on the construction of some probabilistic model for the generator [15–17]. Modern cryptographic generators have a complex structure, and algebraic approach often appears to be not applicable. In these situations stochastic approach considered in this paper may be used. Following this approach the probabilistic model for the output sequence of the considered generator (or its block) is constructed and used to predict future elements of the output sequence [7].

Under stochastic approach the general model of the observed output sequence (of the cryptographic generator or its block) is the discrete time series ($\equiv$ discrete valued random sequence) $x_t = x_t(\omega) : \mathbb{N} \times \Omega \rightarrow \mathcal{A}$ defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$, where $t \in \mathbb{N} = \{1, 2, \dots\}$ is discrete time, $\mathcal{A} = \{0, 1, \dots, N-1\}$ is a finite set of $N \geq 2$ states.

The following probabilistic models are used for evaluation of cryptographic security: $x_t = \xi_t$ is a scheme of independent Bernoulli trials, $\mathbf{P}\{\xi_t = 1\} = 1 - \mathbf{P}\{\xi_t = 0\} = p$, $\mathcal{A} = V = \{0, 1\}$, $p \in [0, 1]$; x_t is a “corrupted periodic sequence”: $x_t = a_t \oplus \xi_t$, where a_t is some “deterministic” periodic sequence (e. g. a sequence generated by some LFSR or NLFSR); modification of the above-mentioned models using the Markovian model for ξ_t . A review of models is given in [16].

Modern cryptographic generators guarantee the uniform probability distribution of s -tuples in the output sequence for sufficiently large values of s , and to construct a useful probabilistic model we need to exploit high-order dependencies in x_t . A universal model for long-memory discrete time series is the high-order Markov chain [3, 5]. 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 identification of this model is a computationally hard problem; in addition, we need to have data sets of huge size $T > D$. This situation generates a topical problem of construction and statistical analysis of small-parametric (parsimonious) models for high-order Markov chains, i. e. the models determined by a small number of parameters. Some of these models are: the Jacobs – Lewis model [6], the Mixture Transition Distribution model [14], the hidden Markov model [2], the variable length Markov chain [4], Markov chain of the order s with r partial connections [11], Markov chain of conditional order [10].

This paper is devoted to identification (estimation of parameters and hypotheses testing) of the parsimonious models by the observed output sequence. The paper has the following structure. After Introduction (Section 1) we present our results on identification of some known parsimonious models of high-order Markov chains: for the Jacobs–Lewis model JL (s) in Section 2 and for the Raftery MTD model in Section 3. Section 4 and Section 5 are devoted to new parsimonious models: Markov chain MC (s, r) of order s with r partial connections and Markov chain of conditional order respectively.

2. Identification of the Jacobs–Lewis model JL (s)

Let us remind [5] that a homogeneous Markov chain of order $s \in \mathbb{N}$ is determined by the generalized Markov property ($t > s$; $i_1, \dots, i_t \in \mathcal{A}$):

$$\begin{aligned} \mathbf{P}\{x_t = i_t \mid x_{t-1} = i_{t-1}, \dots, x_1 = i_1\} = \\ = \mathbf{P}\{x_t = i_t \mid x_{t-1} = i_{t-1}, \dots, x_{t-s} = i_{t-s}\} = p_{(i_{t-s}, \dots, i_{t-1}), i_t}, \end{aligned} \quad (1)$$

where $P = (p_{i_1, \dots, i_s, i_{s+1}})$ is an $(s+1)$ -dimensional matrix of one-step transition probabilities satisfying the normalization condition:

$$\sum_{i_{s+1} \in \mathcal{A}} p_{i_1, \dots, i_s, i_{s+1}} \equiv 1, \quad i_1, \dots, i_s \in \mathcal{A}.$$

Jacobs–Lewis model JL (s) is defined [6] by the following stochastic difference equation of order $s \geq 2$ with random delay:

$$x_t = \mu_t x_{t-\eta_t} + (1 - \mu_t) \xi_t, \quad t > s, \quad (2)$$

where $\{\xi_t, \eta_t, \mu_t\}$ are jointly independent random variables with probability distributions:

$$\begin{aligned} \mathbf{P}\{\mu_t = 1\} &= 1 - \mathbf{P}\{\mu_t = 0\} = \rho; \\ \mathbf{P}\{\eta_t = j\} &= \lambda_j, \quad j \in \{1, 2, \dots, s\}, \quad \lambda_1 + \lambda_2 + \dots + \lambda_s = 1, \quad \lambda_s \neq 0; \\ \mathbf{P}\{\xi_t = k\} &= \pi_k, \quad k \in \mathcal{A}, \quad \sum_{k \in \mathcal{A}} \pi_k = 1; \end{aligned} \quad (3)$$

initial random values $x_1, x_2, \dots, x_s$ are independent with probability distribution $\mathbf{P}\{x_1 = k\} = \mathbf{P}\{x_2 = k\} = \dots = \mathbf{P}\{x_s = k\} = \pi_k$, $k \in \mathcal{A}$. The number of parameters $\pi = (\pi_k)$, $\lambda = (\lambda_i)$, ρ depends linearly on s : $D_{\text{JL}} = N + s - 1$.

The probabilistic model (2), (3) of generation of random sequence x_t is illustrated by Fig. 1. The generator consists of three elementary generators (G_1 for generation of ξ_t , G_2 for generation of η_t and G_3 for generation of binary sequence μ_t), a shift register and a selector that selects one of its input signals $x_{t-\eta_t}$, ξ_t depending on the value μ_t .

Fig. 1. Generation of x_t by JL(s) model

Theorem 1. *The discrete time series x_t determined by (2) and (3) is a homogeneous Markov chain of the order s with the initial probability distribution $\pi_{i_1, i_2, \dots, i_s} = \pi_{i_1} \pi_{i_2} \dots \pi_{i_s}$ and the $(s+1)$ -dimensional matrix of transition probabilities $P(\pi, \lambda, \rho) = (p_{i_1, \dots, i_s, i_{s+1}})$, $i_1, \dots, i_s, i_{s+1} \in \mathcal{A}$:*

$$p_{i_1, \dots, i_s, i_{s+1}} = (1-\rho) \pi_{i_{s+1}} + \rho \sum_{j=1}^s \lambda_j \mathbf{I}\{i_{s+1} = i_{s-j+1}\}, \quad (4)$$

where $\mathbf{I}\{B\}$ is the indicator of event B .

Proof. The generalized Markov property (1) follows from the definition (2), (3) of the JL(s) model: $x_t = f(x_{t-1}, \dots, x_{t-s}; \xi_t, \eta_t, \mu_t)$, where the function $f(\cdot)$ is determined by (2). By (3) and the total probability formula we have:

$$\begin{aligned} p_{i_1, \dots, i_s, i_{s+1}} &= (1-\rho) \pi_{i_{s+1}} + \\ &+ \rho \sum_{j=1}^s \mathbf{P}\{\eta_{s+1} = j\} \mathbf{P}\{x_{s+1} = i_{s+1} \mid x_s = i_s, \dots, x_1 = i_1, \mu_{s+1} = 1, \eta_{s+1} = j\} = \\ &= (1-\rho) \pi_{i_{s+1}} + \rho \sum_{j=1}^s \lambda_j \mathbf{I}\{i_{s+1} = i_{s-j+1}\}. \end{aligned}$$

Initial probability distribution follows from the model assumption. $\square$

Corollary 1. *Maximum likelihood estimators (MLEs) $\hat{\pi}$, $\hat{\lambda}$, $\hat{\rho}$ for the parameters π , λ , ρ of the JL(s) model by the observations $X_T = (x_1, \dots, x_T)' \in \mathcal{A}^T$ are determined by maximization of the following loglikelihood function $l = l(\pi, \lambda, \rho)$:*

$$l = \sum_{t=1}^s \ln \pi_{x_t} + \sum_{t=s+1}^T \ln \left((1-\rho) \pi_{x_t} + \rho \sum_{j=1}^s \lambda_j \mathbf{I}\{x_t = x_{t-j}\} \right) \rightarrow \max_{\pi, \lambda, \rho}. \quad (5)$$

In [12] consistent estimators $\tilde{\pi}$, $\tilde{\lambda}$, $\tilde{\rho}$ are found and used as the initial approximation for the MLEs $\hat{\pi}$, $\hat{\lambda}$, $\hat{\rho}$ in the iterative solution of the maximization problem (5).

Define the hypotheses H_0 , H_1 on the values of the parameters: $H_0 = \{\pi = \pi^0, \lambda = \lambda^0, \rho = \rho^0\}$, where π^0 , λ^0 , ρ^0 are some fixed hypothetical values for parameters (e. g., if $\rho^0 = 0$, $\pi_k^0 \equiv N^{-1}$ hypothesis H_0 means that x_t is uniformly distributed random sequence); $H_1 = \overline{H_0}$. Using the asymptotic normality property of the MLEs $\hat{\pi}$, $\hat{\lambda}$, $\hat{\rho}$, we get [12] the generalized probability ratio test of the asymptotic size $\varepsilon \in (0, 1)$:

$$d = d(X_T) = \mathbf{I}\{\Lambda_T \geq \Delta_\varepsilon\}, \quad \Lambda_T = 2 \left(l(\hat{\pi}, \hat{\lambda}, \hat{\rho}) - l(\pi^0, \lambda^0, \rho^0) \right), \quad (6)$$

where Δ_ε is ε -quantile of the χ_{N+s-1}^2 -distribution.

3. Identification of the Raftery MTD model

Raftery *Mixture Transition Distribution* (MTD) model [14] is defined by a special small-parametric (parsimonious) representation of the matrix P :

$$p_{i_1, \dots, i_{s+1}} = \sum_{j=1}^s \lambda_j q_{i_j, i_{s+1}}, \quad i_1, \dots, i_{s+1} \in \mathcal{A}, \quad (7)$$

where $Q = (q_{ik})$ is a stochastic $(N \times N)$ -matrix, $i, k \in \mathcal{A}$, and $\lambda = (\lambda_1, \dots, \lambda_s)'$ is an s -column vector such that $\lambda_1 > 0$, $\lambda_2, \dots, \lambda_s \geq 0$, $\lambda_1 + \lambda_2 + \dots + \lambda_s = 1$. This model has $D_{\text{MTD}} = N(N-1)/2 + s - 1$ parameters. The MTD model (7) may be generalized to obtain the MTDg model:

$$p_{i_1, \dots, i_{s+1}} = \sum_{j=1}^s \lambda_j q_{i_j, i_{s+1}}^{(j)}, \quad i_1, \dots, i_{s+1} \in \mathcal{A}, \quad (8)$$

where $Q^{(j)} = (q_{ik}^{(j)})$ is the j -th stochastic matrix corresponding to the time lag $s-j$. The number of parameters in the MTDg model equals

$$D_{\text{MTDg}} = s \left(\frac{N(N-1)}{2} + 1 \right) - 1.$$

Let us introduce the notation: the distribution $\Pi^* = (\pi_{i_1, \dots, i_s}^*), i_1, \dots, i_s \in \mathcal{A}$ is an s -variate stationary probability distribution of the ergodic Markov chain; $\pi^* = (\pi_0^*, \dots, \pi_{N-1}^*)'$ is an univariate stationary probability distribution; δ_{ij} is the Kronecker symbol.

Theorem 2. *If the model (8) is valid and for some $K \in \mathbb{N}$ all elements of the matrix $(Q^{(1)})^K$ are positive, then the stationary probability distribution Π^* satisfies the system of equations*

$$\pi_{i_1, \dots, i_s}^* = \prod_{l=0}^{s-1} \left(\pi_{i_{s-l}}^* + \sum_{j=l+1}^s \lambda_j \left(q_{i_{j-l}, i_{s-l}}^{(j)} - \sum_{r=0}^{N-1} q_{r, i_{s-l}}^{(j)} \pi_r^* \right) \right), \quad i_1, \dots, i_{s+1} \in \mathcal{A}.$$

Corollary 2. *If the model (7) is valid, then the stationary bivariate marginal probability distribution of the random vector $(x_{t-m}, x_t)'$ satisfies the relation $\pi_{ki}^*(m) = \pi_k^* \pi_i^* + \pi_k^* \lambda_{s-m+1} (q_{ki} - \pi_i^*)$, $1 \leq m \leq s$, $k \in \mathcal{A}$.*

The proof of Theorem 2 and of its corollary may be found in [12].

Let us construct estimators for the parameters of the MTD model by applying the property from Corollary 2. We define the following statistics as functions of the observed realization $X_1^T = (x_1, \dots, x_T)'$ for $i, k \in \mathcal{A}$, $j = 1, \dots, s$:

$$\begin{aligned} \tilde{\pi}_i &= \frac{1}{T-2s+1} \sum_{t=s+1}^{T-s+1} \delta_{x_t, i}; \\ \tilde{\pi}_{ki}(j) &= \frac{1}{T-2s+1} \sum_{t=s+j}^{T-s+j} \delta_{x_{t-j}, k} \delta_{x_t, i}; \\ \tilde{q}_{ki} &= \begin{cases} \sum_{j=1}^s \frac{\tilde{\pi}_{ki}(j)}{\tilde{\pi}_k} - (s-1) \tilde{\pi}_i, & \tilde{\pi}_k > 0, \\ \frac{1}{N} & \text{otherwise;} \end{cases} \\ z_{ki}(j) &= \frac{\tilde{\pi}_{ki}(s-j)}{\tilde{\pi}_k} - \tilde{\pi}_i; \\ d_{ki} &= \tilde{q}_{ki} - \tilde{\pi}_i; \\ \tilde{\lambda} &= \arg \min_{\lambda} \sum_{i, k \in \mathcal{A}} \sum_{j=1}^s (z_{ki}(j) - \lambda_j d_{ki})^2. \end{aligned} \tag{9}$$

Theorem 3. *Under conditions of the Corollary 2 the statistics (9) are asymptotically unbiased and consistent estimators for Q and λ as $T \rightarrow \infty$.*

Proof. It is easy to show that the definitions of consistency and asymptotic unbiasedness are satisfied. $\square$

The estimators $\tilde{Q}, \tilde{\lambda}$ defined by (9) give a good initial approximation for iterative maximization of the loglikelihood function, which yields the MLEs $\hat{Q}, \hat{\lambda}$:

$$l(Q, \lambda) = \sum_{t=s+1}^T \ln \sum_{j=1}^s \lambda_j q_{x_{t-s+j-1}, x_t} \rightarrow \max_{Q, \lambda}.$$

Generalized probability ratio test of the asymptotic (as $T \rightarrow \infty$) size $\varepsilon \in (0, 1)$ for testing hypotheses (on the values of parameters Q, λ): $H_0 = \{Q = Q^0, \lambda = \lambda^0\}$, $H_1 = \bar{H}_0$ is constructed as test (6) in the previous section for JL (s) model.

4. Identification of the MC(s, r) model

Let's introduce the notation: $r \in \{1, 2, \dots, s\}$ is the parameter called the *number of partial connections*; $M_r^0 = (m_1^0, \dots, m_r^0) \in M$ is an arbitrary integer r -vector with ordered components $1 = m_1^0 < m_2^0 < \dots < m_r^0 \leq s$ which is called the *connection template*; M is the set of cardinality $K = |M| = C_{s-1}^{r-1}$ which is composed of all possible connection templates with r partial connections; and $Q^0 = (q_{j_1, \dots, j_{r+1}}^0)$ is some $(r+1)$ -dimensional stochastic matrix, $j_1, \dots, j_{r+1} \in \mathcal{A}$.

A Markov chain of order s with r partial connections [9, 11], denoted as MC(s, r), is defined by specifying the one-step transition probabilities:

$$p_{i_1, \dots, i_s, i_{s+1}} = q_{i_{m_1^0}, \dots, i_{m_r^0}, i_{s+1}}^0, \quad i_1, \dots, i_{s+1} \in \mathcal{A}. \quad (10)$$

The relation (10) implies that the probability of the process entering a state i_{s+1} at time $t > s$ depend only on its r states at times $t-s-1+m_j^0$, $j = 1, \dots, r$. Thus, instead of $D = N^s(N-1)$ parameters, the model (10) is defined by $D_{\text{MC}(s,r)} = N^r(N-1) + r - 1$ independent parameters that determine the matrices Q^0, M_r^0 . The reduction in the number of parameters may be very significant: for instance, if $N = 2$, $s = 32$, $r = 3$, then we have $D \approx 4.1 \cdot 10^9$, and $D_{\text{MC}(32,3)} = 10$.

Note that if $s = r$, $M_r^0 = (1, 2, \dots, s)$, then $P = Q^0$, and MC(s, s) is a Markov chain of order s . A constructive example of MC(s, r) for modeling of output sequences generated by cryptographic devices and their blocks is a binary ($N = 2$) autoregression of order s with r nonzero coefficients, a special case of which is a linear recursive sequence defined in the ring $\mathbb{Z}_2$ and generated by a degree s polynomial with r nonzero coefficients [13].

Let's introduce the notation: $J_s = (j_1, \dots, j_s) = (J_{s-1}, j_s) \in \mathcal{A}^s$ is a multi-index of order s , the function

$$S_t : \mathcal{A}^T \times M \rightarrow \mathcal{A}^r, (X_T; M_r) \rightarrow (x_{t+m_1-1}, \dots, x_{t+m_r-1}) \in \mathcal{A}^r$$

is called a *selector of order r* with parameters $M_r \in M$ and $t \in \{1, \dots, T-s+1\}$, $\Pi_{K_s} = \mathbf{P}\{X_s = K_s\}$ is the initial s -variate probability distribution of the Markov chain $\text{MC}(s, r)$, $\Pi_{K_s}^*$ is its stationary distribution, $\nu_{r+1}(J_{r+1}; M_r) = \sum_{t=1}^{T-s} \mathbf{I}\{S_t(X_T; M_{r+1}) = J_{r+1}\}$ is the frequency of the $(r+1)$ -tuple $J_{r+1} \in \mathcal{A}^{r+1}$ corresponding to the connection template $M_{r+1} = (M_r, s+1)$ and satisfying the normalization condition $\sum_{J_{r+1} \in \mathcal{A}^{r+1}} \nu_{r+1}(J_{r+1}; M_r) = T-s$, an index replaced by a dot denotes summation over all values of this index: $\nu_{r+1}((J_r, \cdot); M_r) = \sum_{j_{r+1} \in \mathcal{A}} \nu_{r+1}(J_{r+1}; M_r)$, $\nu_{r+1}((\cdot, j_{r+1}); M_r) = \sum_{J_r \in \mathcal{A}^r} \nu_{r+1}(J_{r+1}; M_r)$.

Theorem 4. *The model $\text{MC}(s, r)$ defined by (10) is ergodic if and only if there exists an integer $l \geq 0$ such that*

$$\min_{J_s, J'_s \in \mathcal{A}^s} \sum_{K_l \in \mathcal{A}^l} \prod_{i=1}^{s+l} q_{S_i((J_s, K_l, J'_s); M_{r+1}^0)}^0 > 0.$$

The proof is based on transformation of $\text{MC}(s, r)$ into a special Markov chain of order one with s -dimensional state space.

Let us apply the plug-in principle to construct the information functional $\hat{I}_{r+1}(M_r)$ of the observed realization X_T . In other words let us construct a sample-based estimator for the Shannon information on the future symbol $x_{t+s} \in \mathcal{A}$ contained in the r -tuple $S_t(X_T; M_r)$.

Theorem 5. *The maximum likelihood estimators $\widehat{M}_r, \widehat{Q} = (\widehat{q}_{J_{r+1}})$, where $J_{r+1} \in \mathcal{A}^{r+1}$, for the parameters M_r^0, Q^0 may be defined as*

$$\begin{aligned} \widehat{M}_r &= \arg \max_{M_r \in M} \widehat{I}_{r+1}(M_r), \\ \widehat{q}_{J_{r+1}}(M_r) &= \begin{cases} \frac{\nu_{r+1}(J_{r+1}; \widehat{M}_r)}{\nu_{r+1}((J_r, \cdot); \widehat{M}_r)}, & \text{if } \nu_{r+1}((J_r, \cdot); \widehat{M}_r) > 0, \\ 1/N, & \text{if } \nu_{r+1}((J_r, \cdot); \widehat{M}_r) = 0. \end{cases} \end{aligned} \quad (11)$$

Theorem 6. *If $\text{MC}(s, r)$ defined by (10) is stationary and the connection template $M_r^0 \in M$ satisfies the identifiability condition, then the maximum likelihood estimators $\widehat{M}_r, \widehat{Q}$ defined by (11) are consistent as $T \rightarrow \infty$:*

$$\widehat{M}_r \xrightarrow{\mathbf{P}} M_r^0, \quad \widehat{Q} \xrightarrow{\mathbf{L}_2} Q^0,$$

and the following asymptotic expansion holds for the mean square error of $\widehat{Q}$:

$$\Delta_T^2 = \mathbf{E}\{\|\widehat{Q} - Q^0\|^2\} = \frac{1}{T-s} \sum_{J_{r+1} \in \mathcal{A}^{r+1}} \frac{(1-q_{J_{r+1}}^0) q_{J_{r+1}}^0}{\mu_{r+1}((J_r, \cdot); M_r^0, M_r^0)} + o\left(\frac{1}{T}\right),$$

$$\mu_{r+1}(J_{r+1}; M_r, M_r^0) = \sum_{K_{s+1} \in \mathcal{A}^{s+1}} \mathbf{I}\{S_1(K_{s+1}; M_{r+1}) = J_{r+1}\} \Pi_{K_s}^* p_{K_{s+1}}. \quad (12)$$

Theorems 5 and 6 have been proved in [9].

The estimators (11) have been used to construct a statistical test for the null hypothesis $H_0 : Q^0 = Q_0$ against the alternative $H_1 = \overline{H}_0$, where $Q_0 = (q_{0J_{r+1}})$ is some given stochastic matrix. The decision rule of a given asymptotic size $\varepsilon \in (0, 1)$ may be written as follows [9]:

$$d(X_T) = \mathbf{I}\{\rho > \Delta\}, \quad \rho = \sum_{J_{r+1}: q_{0J_{r+1}} > 0} \nu_{r+1}((J_r, \cdot); \widehat{M}_r) \frac{(\widehat{q}_{J_{r+1}} - q_{0J_{r+1}})^2}{q_{0J_{r+1}}}, \quad (13)$$

where Δ is the $(1-\varepsilon)$ -quantile of the χ^2 distribution with L degrees of freedom.

Note that to estimate the orders $s \in [s_-, s_+]$, $r \in [r_-, r_+]$ ($1 \leq s_- < s_+ < \infty$, $1 \leq r_- < r_+ < s_+$) we use a modification of the Bayesian Information Criterion (BIC) presented in [8].

Performance of the statistical estimators (12) and the test (13) was evaluated by Monte-Carlo simulation experiments, where the model parameters were fixed: $N = 2$, $s = 256$, $r = 6$; the chosen values of Q^0 and M_r^0 are omitted due to space limitation. For each simulated observation length of T , 10^4 simulation rounds were performed. Fig. 2 illustrates the numerical results obtained for this MC (256, 6) model: the mean square error Δ_T^2 of the estimator $\widehat{Q}$ is plotted against the observation length T ; the curve has been computed theoretically from the leading term of the expansion (12), and the circles are the experimental values obtained by the simulations.

Fig. 2. Dependence of Δ_T^2 on T for $N = 2$, $s = 256$, $r = 6$

5. Markov chain of conditional order and its identification

Let's introduce the notation: $J_m^n = (j_m, j_{m+1}, \dots, j_n) \in \mathcal{A}^{n-m+1}$, $n \geq m$, is the multiindex (subsequence of indices from a sequence $j_1, j_2, \dots$),

$$\langle J_m^n \rangle = \sum_{k=n}^m N^{k-n} j_k \in \{0, 1, \dots, N^{m-n+1} - 1\}$$

is the numeric representation of the multiindex $J_m^n \in \mathcal{A}^{m-n+1}$, $L \in \{1, 2, \dots, s-1\}$, $K = N^L - 1$ are some positive integers, $Q^{(1)}, \dots, Q^{(M)}$ are M ($1 \leq M \leq K+1$) different square stochastic matrices of the order N :

$$Q^{(m)} = (q_{i,j}^{(m)}), \quad 0 \leq q_{i,j}^{(m)} \leq 1, \quad \sum_{j \in \mathcal{A}} q_{i,j}^{(m)} \equiv 1, \quad i, j \in \mathcal{A}, \quad 1 \leq m \leq M,$$

and $\pi_{J_1^s}^0 = \mathbf{P} \{X_1^s = J_1^s\}$ is the initial probability distribution.

The Markov chain $\{x_t \in \mathcal{A} : t \in \mathbb{N}\}$ is called the *Markov chain of conditional order* (MCCO (s)), if its one-step transition probabilities have the following parsimonious form [10]:

$$p_{J_1^{s+1}} = \sum_{k=0}^K \mathbf{I} \{ \langle J_{s-L+1}^s \rangle = k \} q_{j_{b_k}, j_{s+1}}^{(m_k)}, \quad (14)$$

where $1 \leq m_k \leq M$, $1 \leq b_k \leq s-L$, $0 \leq k \leq K$, $\min_{0 \leq k \leq K} b_k = 1$; it is assumed that all elements of the set $\{1, 2, \dots, M\}$ occur in the sequence $m_0, \dots, m_K$. The sequence of elements J_{s-L+1}^s is called the *base memory fragment* (BMF) of the random sequence, L is the length of BMF, the value $s_k = s - b_k + 1$ is called the *conditional order*. Thus the conditional probability distribution of the state x_t at time t depends not on all s previous states, but it depends only on $L+1$ selected states corresponding to (j_{b_k}, J_{s-L+1}^s) . Note that if $L = s-1$, $s_0 = s_1 = \dots = s_K = s$, we have the fully-connected Markov chain of the order s . If $M = K+1$, then each transition matrix corresponds to only one value of BMF, otherwise there exists a common matrix which corresponds to several values of BMF. The number of independent parameters equals

$$D_{\text{MCCO}(s)} = 2(N^L + 1) + MN(N-1).$$

As in previous section we construct the MLEs $\{\widehat{Q}^{(i)}\}$, $\widehat{L}$, $\{\widehat{s}_k\}$, $\{\widehat{m}_k\}$ by maximization the loglikelihood:

$$l_T(X_T, \{Q^{(i)}\}, L, \{s_k\}, \{m_k\}) = \quad (15)$$

$$= \ln \pi_{X_1^s}^0 + \sum_{J_0^{L+1} \in \mathcal{A}^{L+2}} \sum_{k=0}^K \mathbf{I}\{\langle J_1^L \rangle = k\} \nu_{L+2, s_k - L - 1}^s(J_0^{L+1}) \ln q_{j_0, j_{L+1}}^{(m_k)} \rightarrow \max,$$

where $\nu_{l,y}^s(J_1^l) = \sum_{t=1}^{n-s} \mathbf{I}\{x_{t+s-l-y+1} = j_1, X_{t+s-l+2}^{t+s} = J_2^l\}$, $l \geq 2, 0 \leq y \leq s-l+1$, is the frequency of the state $J_1^l \in \mathcal{A}^l$ with the time gap of length y between the elements j_1 and J_2^l ; $\nu_{s+1}(J_1^{s+1}) = \nu_{s+1,0}^s(J_1^{s+1})$ is the frequency of $(s+1)$ -tuple J_1^{s+1} .

From (15) we get the following expressions for the MLEs (for simplicity of expressions we give here results only for the case $M = K+1$) [10]:

$$\widehat{q}_{j_0, j_{L+1}}^{(k+1)} = \begin{cases} \sum_{J_1^L \in \mathcal{A}^L} \mathbf{I}\{\langle J_1^L \rangle = k\} \frac{\nu_{L+2, g(s_k, L)}^s(J_0^{L+1})}{\nu_{L+1, g(s_k, L)}^s(J_0^L)}, & \text{if } \nu_{L+1, g(s_k, L)}^s(J_0^L) > 0, \\ 1/N, & \text{if } \nu_{L+1, g(s_k, L)}^s(J_0^L) = 0, \end{cases}$$

$$\widehat{s}_k = \arg \max_{L+1 \leq y} \sum_{J_1^L \in \mathcal{A}^L} \mathbf{I}\{\langle J_1^L \rangle = k\} \sum_{j_0, j_{L+1} \in \mathcal{A}} \nu_{L+2, g(y, L)}^s(J_0^{L+1}) \ln(\widehat{q}_{j_0, j_{L+1}}^{(k+1)}).$$

To estimate the order s and the BMF length L we use Bayesian information criterion:

$$(\widehat{s}, \widehat{L}) = \arg \min_{\substack{1 \leq L' \leq L_+, \\ 2 \leq s' \leq S_+}} \text{BIC}(s', L'),$$

$$\text{BIC}(s', L') = D_{\text{MCCO}(s)} \ln(n - s') -$$

$$- 2 \sum_{J_0^{L'+1} \in \mathcal{A}^{L'+2}} \sum_{k=0}^K \mathbf{I}\{\langle J_1^{L'} \rangle = k\} \nu_{L'+2, \widehat{g}(s_k, L')}^{s'}(J_0^{L'+1}) \ln \widehat{q}_{j_0, j_{L'+1}}^{(k+1)},$$

where $S_+ \geq 2$, $1 \leq L_+ \leq S_+ - 1$, are maximal admissible values of s and L respectively.

Statistical decision rule for testing the hypotheses on the values of parameters for the MCCO(s) model is constructed [10] by MLEs $\{\widehat{Q}^{(k)}\}$, $\{\widehat{s}_k\}$ in the same way as in the previous section.

The author would like to thank A. M. Zubkov for fruitful discussions and comments.

References

- [1] Alferov A., Zubov A., Kuzmin A., Cheremushkin A., *Fundamentals of Cryptography*, Moscow: Gelios ARV, 2001 (in Russian), 480 pp.
- [2] Berchtold A., “High-order extensions of the double chain Markov model”, *Stoch. Models*, **18**:2 (2002), 193–227.
- [3] Billingsley P., “Statistical methods in Markov chains”, *Ann. Math. Statist.*, **32** (1961), 12–40.
- [4] Buhlmann P., Wyner A., “Variable length Markov chains”, *Ann. Statist.*, **27**:2 (1999), 480–513.
- [5] Doob J., *Stochastic Processes*: N.Y.: Wiley, 1953.
- [6] Jacobs P. A., Lewis P. A. W., “Discrete time series generated by mixtures. I: Correlational and runs properties”, *J. Royal Statist. Soc. Ser. B.*, **40**:1 (1978), 94–105.
- [7] Kharin Yu., *Robustness in Statistical Forecasting*, N. Y.: Springer, 2013, 356 pp.
- [8] Kharin Yu., “Statistical analysis of discrete time series based on the MC (s, r)-model”, *Austrian J. Statist.*, **40**:1&2 (2011), 75–84.
- [9] Kharin Yu., Agievich S., Vasiljev D., Matveev G., *Cryptology*: Minsk: BSU, 2013 (in Russian), 512 pp.
- [10] Kharin Yu., Maltsev M., “Markov chain of conditional order: properties and statistical analysis”, *Austrian J. Statist.*, **43**:3&4 (2014), 205–217.
- [11] Kharin Yu. S., Piatlitski A. I., “A Markov chain of order s with r partial connections and statistical inference on its parameters”, *Diskretnaja Matematika*, **19**:2 (2007), 109–130 (in Russian); English transl.: *Discrete Math. Appl.*, **17**:3 (2007), 295–317.
- [12] Kharin Yu., Yarmola A., “Statistical estimation of parameters for the MTD-model of discrete time series”, *Vesti of Nat. Acad. of Sci. of Belarus*, 2 (2006), 20–26. (in Russian)
- [13] Maksimov Yu. I., “On Markov chains in binary shift registers with random elements”. In: *Trudy po diskretnoi matematike*, **1**, Moscow: TVP Sci. Publ., 1997, 203–220. (in Russian)
- [14] Raftery A., “A model for high-order Markov chains”, *J. Royal Statist. Soc. Ser. B*, **47**:3 (1985), 528–539.
- [15] Sachkov V. N., *Introduction to Combinatorial Methods of Discrete Mathematics*, Moscow: Nauka, 1982 (in Russian), 489 pp.
- [16] Shoitov A. M., “Probabilistic models of pseudorandom sequences in cryptography”. In: *Problemy informacionnoi bezopasnosti*, Moscow: Moscow State Univ., 2007, 116–134 (in Russian)
- [17] Zubkov A. M., “Generators of pseudorandom numbers and their applications”. In: *Matematika i bezopasnost' informacionnyh tehnologij*, Moscow: Moscow State Univ., 2003, 200–206 (in Russian)