Tatra Mt. Math. Publ. **39** (2008), 155–163

TATRA MOUNTAINS *Mathematical Publications*

A PARAMETRIC MODEL FOR DISCRETE-VALUED TIME SERIES

Martin Janžura — Lucie Fialová

ABSTRACT. A parametric model for statistical analysis of Markov chains type models is constructed. Within the proposed model we can estimate the unknown parameter by simultaneous optimization of a collection of short-range objective functions. The estimate is proved to be consistent and asymptotically normal. Under mild conditions the estimate agrees with the standard maximum likelihood one, and, therefore, it is also asymptotically efficient.

1. Introduction

A natural probabilistic model for time series with discrete (categorial) data is a Markov chain of some order, see, e.g., the classical references [2] or [1]. But whenever the order of the Markov property is considerably high, the general non-parametric model is extremely complex and hardly tractable, in particular from the statistical analysis point of view. Therefore a reasonable parametric model is needed. As standard parametric families we can consider the well-known logistic regression models, or, (see, e.g., [7]) the Gibbs-Markov distributions with pair-wise interactions which can be understood as two-sided logistic regression models. But all these models are numerically feasible only up to a certain (not very high) range.

In the present paper we introduce a parametric model which deals with small-dimensional sufficient statistics and which makes possible block-wise (separate) estimation of the multi-dimensional vector parameter. Thus, even a higher order model can be identified from a rather short data series. The proposed model corresponds, in fact, to Gibbs distribution with interactions that depend non-linearly on the unknown parameter. First, the problem of existence and uniqueness of such distributions is discussed. Then the estimator is constructed and its (mostly asymptotic) properties are investigated, namely the consistency and

2000 Mathematics Subject Classification: Primary 60J27; Secondary 62M05.

Keywords: Markov chains, Gibbs distributions, statistical estimation.

This research was supported by the GAČR grant No. 201/06/1323.

the asymptotic normality are proved. Moreover, it is shown that under some additional assumptions the estimate coincides with the standard maximum likelihood one which also yields to its asymptotic efficiency. For general results on Gibbs distribution we refer to [5], [6], and [10], for limit theorems to [4]. The statistical inference aspects are partly adapted from [7] and [9].

2. Preliminaries

2.1. Basic definitions

Let $\mathcal{X}$ denote some fixed finite state space, and Z the set of integers. For every $V \subset Z$ we denote by $\mathcal{F}_V = \sigma(\mathrm{Proj}_V)$ the σ-algebra of cylinder sets, i.e., the σ-algebra generated by the projection function $\mathrm{Proj}_V : \mathcal{X}^Z \to \mathcal{X}^V$, and by $\mathcal{L}_V$ the space of real-valued cylinder functions, i.e., every $f \in \mathcal{L}_V$ is $\mathcal{F}_V$-measurable or, equivalently, it depends only on the coordinates from the set $V \subset Z$.

Further, let $\mathcal{P}$ denote the set of all probability measures (quoted as random processes) on the space $\left(\mathcal{X}^Z, \mathcal{F}^Z\right)$, and $\mathcal{P}_S \subset \mathcal{P}$ the subset of shift-invariant probability measures (quoted as *stationary random processes*), i.e. $P \in \mathcal{P}_S$ iff $P = P \circ \tau^{-1}$, where $\tau : \mathcal{X}^Z \to \mathcal{X}^Z$ is the shift defined by $\left(\tau(x)\right)_s = x_{s+1}$ for every $s \in Z$, $x \in \mathcal{X}^Z$. Moreover, $P \in \mathcal{P}_S$ is *ergodic* if $P(E) \in \{0,1\}$ for every $E \in \mathcal{E} = \{E \in \mathcal{F}_Z; \tau(E) = E\}$. We write $P \in \mathcal{P}_E$.

For every $V \subset Z$ we denote by $P_V = P/\mathcal{F}_V$ the restriction of the random process to the σ-algebra of cylinder sets $\mathcal{F}_V$, i.e., the corresponding marginal distribution. Finally, we shall denote by $\mathcal{V} = \left\{W \subset Z; 0 < |W| < \infty\right\}$ the system of finite non-void subsets of Z.

2.2. Entropy and entropy rate

For a pair of discrete probability measures μ, ν let $H(\mu) = \int [-\log \mu] \,\mathrm{d}\mu$ and $I(\mu|\nu) = \int \log \frac{\mu}{\nu} \,\mathrm{d}\mu$ denote (whenever the integrals exist) the *entropy* and the *information divergence* (relative entropy, Kullback–Leibler number), respectively.

Nevertheless, for the random processes we have to deal with the asymptotic versions of the quantities, i.e., with the *entropy rate*

$$\mathcal{H}(P) = \lim_{N\to\infty} (2N+1)^{-1} H\Big(P_{[-N,N]}\Big),$$

and the *asymptotic I-divergence* (relative entropy rate)

$$\mathcal{I}\big(P|Q\big) = \lim_{N\to\infty} (2N+1)^{-1} I\Big(P_{[-N,N]} \,|\, Q_{[-N,N]}\Big)$$

for P, $Q \in \mathcal{P}$, provided the integrals and limits exist. Let us note that the entropy rate $\mathcal{H}(P)$ exists for every $P \in \mathcal{P}_S$, while the relative entropy rate $\mathcal{I}(P|Q)$

exists only under some additional conditions (e.g., the Markov property) on the reference random process Q (see, e.g., [5]).

2.3. Gibbs distributions

The functions from $\mathcal{L} = \bigcup_{V \in \mathcal{V}} \mathcal{L}_V$ will be quoted as (finite range) *potentials*. For a potential $\Phi \in \mathcal{L}_V$, $V \in \mathcal{V}$, we define the *Gibbs specification* as the family of probability kernels $\Pi^\Phi = \{\Pi^\Phi_\Lambda\}_{\Lambda \in \mathcal{V}}$ where

$$\Pi^\Phi_\Lambda(x_\Lambda | x_{\Lambda^c}) = Z^\Phi_\Lambda(x_{\Lambda^c}) \exp\left\{ \sum_{j \in \Lambda - V} \Phi \circ \tau_j(x) \right\}$$

with the normalizing constant

$$Z^\Phi_\Lambda(x_{\Lambda^c}) = \sum_{y_\Lambda \in X_0^\Lambda} \exp\left\{ \sum_{j \in \Lambda - V} \Phi \circ \tau_j(y_\Lambda, x_{\Lambda^c}) \right\}$$

for every $\Lambda \in \mathcal{V}$. Here, $\Lambda - V = \{\lambda - v; v \in V, \lambda \in \Lambda\} = \big\{j \in Z; (j + V) \cap \Lambda \neq \emptyset\big\}$.

A random process $P \in \mathcal{P}$ is a *Gibbs distribution* with the potential $\Phi \in \mathcal{L}$ if

$$P_{\Lambda|\Lambda^c}(x_\Lambda | x_{\Lambda^c}) = \Pi^\Phi_\Lambda(x_\Lambda | x_{\Lambda^c}) \quad \text{a. s.} \quad [P]$$

for every $\Lambda \in \mathcal{V}$. The set of such P's will be denoted by $G(\Phi)$, and, in general, $G(\Phi) \neq \emptyset$. Thanks to the absence of phase transitions in the one-dimensional short-range systems (see, e.g., [10], Theorem 5.6.2) we have the Gibbs measure given *uniquely*, i.e., $G(\Phi) = \{P^\Phi\}$ for every $\Phi \in \mathcal{L}$. Moreover, $P^\Phi \in \mathcal{P}_E$, i.e., the respective Gibbs random process is *ergodic*.

Remark.

i) From the above definition we observe that the Gibbs distributions represent the infinite-dimensional counterparts to the exponential distributions.

ii) If $\Phi \in \mathcal{L}_A$, we have $\mathcal{P}^\Phi_{V|V^c} \in \mathcal{L}_{\overline{V}}$ with $\overline{V} = V - A + A$ and therefore $P^\Phi_{V|V^c} = P^\Phi_{V|\partial V}$ a.s., where $\partial V = \overline{V} \setminus V$, i.e., P^Φ obeys the "bilateral" Markov property, which is equivalent (see, e.g., [6], Chapter 10) to the usual "unilateral" one.

iii) The Gibbs distribution is determined equivalently by the *variational principle*

$$P^\Phi = \operatorname{argmax}_{Q \in \mathcal{P}_S} \left[\mathcal{H}(Q) - \int \Phi \, \mathrm{d}Q \right].$$

For details see, e.g., [10]. □

3. Problem

3.1. Statistical estimation

Suppose we are given a sequence of *data* $\hat{x}_{[1,n]} = (\hat{x}_1, \ldots, \hat{x}_n) \in \mathcal{X}^{[1,n]}$ and a *parametric family* of stationary random processes

$$\mathcal{P}_\Theta = \left\{P^\theta\right\}_{\theta\in\Theta} \subset \mathcal{P}_S$$

with $\Theta \subset R^K$, where K is a finite set of indices. In addition, we usually assume that each $P^\theta \in \mathcal{P}_\Theta$ is also r-Markov with some fixed order $r \geq 1$, i.e.

$$P^\theta\big(x_0|x_{-1}, \ldots, x_{-r-\ell}\big) = P^\theta\big(x_0|x_{-1}, \ldots, x_{-r}\big)$$

for every $\ell = 0, 1, \ldots$. Then, the standard *statistical estimator* assumes the form

$$\hat{\theta}^n = \operatorname{argmin}_{\theta\in\Theta} \int F^\theta \,\mathrm{d}\hat{P}^n,$$

where $\hat{P}^n$ is the empirical random process given for every $f \in \mathcal{L}_A$, $A \subset Z$ finite, by

$$\int f \,\mathrm{d}\hat{P}^n = |n_A|^{-1} \sum_{i\in n_A} f(\hat{x}_{A+i})$$

whenever $n_A \neq \emptyset$, where $n_A = \big\{i; A+i \subset [1,n]\big\}$, and F^θ is a suitable function, e.g.,

$$F^\theta\left(x_{[-r,0]}\right) = -\log P^\theta\left(x_0 \,|\, x_{-1}, \ldots, x_{-r}\right)$$

for the maximum likelihood estimator.

Apparently, with such an approach there might be troubles, in particular, when the range r of the Markov property is rather large. Then, consequently, the number of parameter $|K|$ is also large, and we can have, first of all, computational problems with evaluating the function F^θ and with the stability of the optimization procedure. If, moreover, the data size n is rather small, we have also a statistical problem with the "over-parametrization".

3.2. Goal

We would like to introduce a parametric model that will enable us to avoid, at least to some extend, the problems indicated above.

Suppose there exists a partition $\{K_j\}_{j\in J}$ of the index set K (i.e., $K = \bigcup_{j\in J} K_j$, $K_j \cap K_{j'} \neq \emptyset$ for $j \neq j'$). Then we may write $\theta = (\theta^j)_{j\in J}$, where $\theta^j \in R^{K_j}$ is the block corresponding to the index set $K_j \subset K$ for every $j \in J$.

Further, for every $j \in J$ let us suppose a function

$$F_j^{\theta^j} \in \mathcal{L}_{A_j}$$

with $A_j \subset Z$ finite, where $|A_j|$ is small enough to be compared with the range r of the Markov property.

We intend to estimate the parameter $\theta \in \Theta$ *block-wise separately*, i.e., $\hat{\theta}^n = (\hat{\theta}^{n,j})_{j\in J}$, where

$$\hat{\theta}^{n,j} = \operatorname{argmin}_{\theta^j \in R^{K_j}} \int F_j^{\theta^j} \, \mathrm{d}\hat{P}^n$$

for each $j \in J$ simultaneously. In the following section we show that such an approach can make sense. In particular, under what relations between the parametric distribution P^θ and the collection of objective functions $\{F_j^{\theta^j}\}_{j\in J}$ such an approach really works.

4. Solution

4.1. Parametric model

From now, let us assume that

(A1) $F_j^{\theta^j}$ is a *smooth strongly convex function* of θ^j for every $j \in J$.

Then, we may denote

$$\boldsymbol{f}^\theta = \left(\boldsymbol{f}_j^{\theta^j}\right)_{j\in J} = \left(f_{j,i}^{\theta^j}\right)_{i\in K_j,\, j\in J},$$

where $f_{j,i}^{\theta_j} = \frac{\partial F_j^{\theta^j}}{\partial \theta_i^j}$ for every $i \in K_j$, $j \in J$.

Let $P^{\alpha,\theta}$ be the *Gibbs measure* with respect to the *potential* $\langle \alpha, \boldsymbol{f}^\theta \rangle$, i.e., (see Section 2.3),

$$P^{\alpha,\theta}_{V|V^c}(x_V|x_{V^c}) = \frac{\exp\left\{\sum_{t\in V-A} \left\langle \alpha, \boldsymbol{f}^\theta \circ \tau^t(x) \right\rangle\right\}}{\sum_{y_V\in\mathcal{X}^V} \exp\left\{\sum_{t\in V-A} \left\langle \alpha, \boldsymbol{f}^\theta \circ \tau^t(y_V, x^c_V) \right\rangle\right\}}$$

holds a. s. for every $V \in \mathcal{V}$, where $A = \bigcup_{j\in J} A_j$ and $V - A = \{v - a; v \in V, a \in A\} = \big\{s \in Z; (s + A) \cap V \neq \emptyset\big\}$. Or, equivalently,

$$P^{\alpha,\theta} = \operatorname{argmax}_{Q\in\mathcal{P}_S} \left[\mathcal{H}(Q) - \int \left\langle \alpha, \boldsymbol{f}^\theta \right\rangle \mathrm{d}Q\right].$$

Now, let us introduce the function $U : R^K \times R^K \to R^K$ given by

$$U(\alpha, \theta) = \int \boldsymbol{f}^\theta \mathrm{d}P^{\alpha,\theta}$$

for every α, $\theta \in R^K$, and define $\alpha(\theta)$ *implicitly* by

$$U\big(\alpha(\theta), \theta\big) = 0.$$

Let Θ denote the set where $\alpha(\theta)$ is defined. We have to assume the basic regularity (identifiability) condition

(A2) $\quad U(\alpha^1, \theta) = U(\alpha^2, \theta)$ iff $\alpha^1 = \alpha^2$.

The condition is closely related to the notion of *equivalence of potentials*. Potentials Φ_1, Φ_2 are equivalent, we write $\Phi_1 \approx \Phi_2$, if $G(\Phi_1) = G(\Phi_2)$. Potentials $\boldsymbol{\Phi} = (\Phi_1, \ldots, \Phi_K)$ are mutually non-equivalent if $\langle \alpha, \boldsymbol{\Phi} \rangle \approx 0$ yields $\alpha = 0$.

Thus, the condition (A2) is satisfied whenever potentials $\boldsymbol{f}^\theta$ are mutually non-equivalent, which is a bit stronger condition than the standard linear independence for finite systems. The mutual non-equivalence can be checked with the aid of some characteristics (see [8]).

Finally, we have to assume

(A3) $\quad \Theta \neq \emptyset$,

since otherwise our system would be empty. Again, for a finite system the assumption reads as: 0 belongs to the relative interior of the set of all possible values of the statistics. Here, it is more complicated, e.g., (A3) holds if $C_\theta(\cdot)$ assumes its minimum for some fixed $\theta \in R^K$, where

$$C_\theta(\alpha) = \max_{Q \in P_S} \left[\mathcal{H}(Q) + \int \langle \alpha, \boldsymbol{f}^\theta \rangle \, \mathrm{d}Q \right]$$

is a smooth convex function with $\nabla_\alpha C_\theta(\alpha) = U(\alpha, \theta)$ (see, e.g., [6], Chapter 16).

Proposition 1. *Let* (A1), (A2), *and* (A3) *hold. Then*

i) $\nabla_\alpha U(\alpha, \theta) = B_\alpha\big(\boldsymbol{f}^\theta\big) = \sum_{t \in Z} \mathrm{cov}_{P^{\alpha,\theta}} \left(\boldsymbol{f}^\theta, \boldsymbol{f}^\theta \circ \tau^t \right) > 0,$

ii) $\Theta \subset R^K$ *is open, and* $\alpha : \Theta \to R^K$ *is well-defined smooth function.*

Proof.

i) follows from the strong convexity of the function $C_\theta(\cdot)$ for the mutually non-equivalent $\boldsymbol{f}^\theta$ (see [3]).

ii) is a standard result of mathematical calculus (theorem on the "implicit function").

□

Thus, we shall deal with the *parametric family*

$$\left\{ \overline{P}^{\theta} \right\}_{\theta \in \Theta}, \quad \text{where} \quad \overline{P}^{\theta} = P^{\alpha(\theta),\theta} \quad \text{for every} \quad \theta \in \Theta.$$

4.2. Estimate

Since by definition we have $\int \boldsymbol{f}^{\overline{\theta}} \mathrm{d}\overline{P}^{\overline{\theta}} = 0$ for $\overline{\theta} \in \Theta$, or, equivalently

$$\overline{\theta}^{j} = \operatorname{argmin}_{\theta^j \in R^{K_j}} \int F_j^{\theta_j} \, \mathrm{d}\overline{P}^{\overline{\theta}} \quad \text{for every } j \in J \text{ and } \overline{\theta} \in \Theta,$$

we may follow our intention and define the *estimate* $\hat{\theta}^n = (\hat{\theta}^{n,j})_{j \in J}$ precisely as introduced in Section 3.2, i.e.,

$$\hat{\theta}^{n,j} = \operatorname{argmin}_{\theta^j \in R^K} \int F_j^{\theta_j} \mathrm{d}\hat{P}^n \quad \text{for all} \quad j \in J \quad \text{simultaneously.}$$

Theorem 2. *Under* (A1), (A2), *and* (A3), *the estimate* $\hat{\theta}^n = (\hat{\theta}^{n,j})$ *exists with probability tending to* 1, *it is consistent and asymptotically normal with the covariance matrix*

$$C_\theta = \left[\int \nabla f^\theta \, \mathrm{d}\overline{P}^{\theta} \right]^{-1} B_{\alpha(\theta)}(\boldsymbol{f}^\theta) \left[\int \nabla f^\theta \, \mathrm{d}\overline{P}^{\theta} \right]^{-1}.$$

If, in addition, ∇f^θ *is non-random, then* $\hat{\theta}^n$ *agrees with the maximum likelihood estimate, and it is also asymptotically efficient.*

Proof. By the ergodic theorem (cf., e.g., Theorem 14.A8 in [6]), for a.e. $[\overline{P}^{\theta^0}]$ $x \in \mathcal{X}^Z$ and every $j \in J$ the (strongly) convex functions $\int F_j^{\theta_j} \mathrm{d}\hat{P}^n$ tend to the (strongly) convex function $\int F_j^{\theta_j} \mathrm{d}\overline{P}^{\theta^0}$. Therefore the convergence is uniform on compact subsets of Θ, and, consequently, $\operatorname{argmin}_{\theta \in \Theta} \int F_j^{\theta_j} \mathrm{d}\hat{P}^n \to \operatorname{argmin}_{\theta \in \Theta} \int F_j^{\theta_j} \mathrm{d}\overline{P}^{\theta^0}$ a.s. (P^{θ^0}) for $n \to \infty$. That proves the existence and consistency. For the asymptotic normality let us observe

$$n^{\frac{1}{2}} \int f^{\theta^0} \mathrm{d}\hat{P}^n = n^{\frac{1}{2}} \left(\int f^{\theta^0} \mathrm{d}\hat{P}^n - \int f^{\hat{\theta}^n} \mathrm{d}\hat{P}^n \right) = - \int \nabla f^{\tilde{\theta}^n} \mathrm{d}\hat{P}^n \left[n^{\frac{1}{2}} \left(\hat{\theta}^n - \theta^0 \right) \right],$$

where $\tilde{\theta}^n = \varepsilon_n \hat{\theta}^n + (1 - \varepsilon_n)\, \theta_0$ with some $\varepsilon_n \in [0,1]$. Since

$$n^{\frac{1}{2}} \int f^{\theta^0} \mathrm{d}\hat{P}^n \Rightarrow \mathcal{N}_K \left(0,\, B_{\alpha(\theta)}(\boldsymbol{f}^\theta) \right) \qquad \text{for } n \to \infty \text{ in distribution } [\overline{P}^{\theta^0}]$$

by the central limit theorem (see [4]), and

$$\int \nabla f^{\hat{\theta}^n} \mathrm{d}\hat{P}^n \to \int \nabla f^{\theta^0} \mathrm{d}\overline{P}^{\theta^0} \qquad \text{for } n \to \infty \text{ a.s. } [\overline{P}^{\theta^0}]$$

again by the ergodic theorem, we obtain the claimed asymptotic normality.

Since we have asymptotically $-n^{-1}\log \overline{P}^{\theta}_{[1,n]} \doteq C_\theta\big(\alpha(\theta)\big) - \int\big\langle \alpha(\theta), f^\theta\big\rangle \,\mathrm{d}\hat{P}^n$ (see, e.g., [9]), we shall define the maximum likelihood estimate in the form

$$\hat{\hat{\theta}}^n = \mathrm{argmin}_{\theta\in\Theta}\left\{C_\theta\big(\alpha(\theta)\big) - \int\big\langle \alpha(\theta), f^\theta\big\rangle \,\mathrm{d}\hat{P}^n\right\}.$$

Standardly, by differentiating we obtain the system of normal equations $\int\big[\nabla\alpha(\theta)\, f^\theta - \nabla f^\theta \alpha(\theta)\big]\,\big(\mathrm{d}\overline{P}^{\theta} - \mathrm{d}\hat{P}^n\big) = 0$, where, by definition

$$\int \nabla f^\theta \,\mathrm{d}\overline{P}^{\theta} + \sum_{t\in Z}\mathrm{cov}_{\overline{P}^\theta}\Big(f^\theta,\ \big[\nabla\alpha(\theta)\, f^\theta + \nabla f^\theta\alpha(\theta)\big]\circ\tau^t\Big) = 0.$$

Whenever ∇f^θ is a non-random function, the system of normal equations turns to $\nabla\alpha(\theta)\int f^\theta \mathrm{d}\hat{P}^n = 0$, where now

$$-\nabla\alpha(\theta) = B_{\alpha(\theta)}\big(\boldsymbol{f}^\theta\big)^{-1}\int \nabla f^\theta \mathrm{d}\overline{P}^{\theta} > 0.$$

Therefore the MLE $\hat{\hat{\theta}}^n$ is given implicitly by $\int f^{\hat{\hat{\theta}}^n}\,\mathrm{d}\hat{P}^n = 0$, and, thus, coincides with our estimate. The asymptotic Fisher's information matrix is given by

$$\begin{aligned}\mathcal{J}(\theta) &= \lim_{n\to\infty}\int \nabla^2_\theta\left(C_\theta\big(\alpha(\theta)\big) - \int\big\langle \alpha(\theta),\ f^\theta\big\rangle\,\mathrm{d}\hat{P}^n\right)\mathrm{d}\overline{P}^{\theta}\\ &= B_{\alpha(\theta)}\,\big(\nabla\alpha(\theta)\, f^\theta + \nabla f^\theta\alpha(\theta)\big)\end{aligned}$$

which for non-random ∇f^θ turns to

$$\nabla\alpha(\theta)\, B_{\alpha(\theta)}\big(\boldsymbol{f}^\theta\big),\,\theta)\,\nabla\alpha(\theta) = \left[\int \nabla f^{\,\theta}\mathrm{d}\overline{P}^{\theta}\right]\,\Big[B_{\alpha(\theta)}\big(\boldsymbol{f}^\theta\big)\Big]^{-1}\,\left[\int \nabla f^{\theta}\mathrm{d}\overline{P}^{\theta}\right].$$

□

Example. Suppose $\mathcal{X} = \{0,1\}$, $K = \{0,1,\dots,r\}$, J={1,...,r}, $K_1 = \{0,1\}$, and $K_j = \{j\}$ for $j = 2,\dots,r$. Let

$$\begin{aligned}F_1^{\theta_0,\theta_1}(x_0,x_1) &= -\log\frac{e^{\theta_0 x_0 + \theta_1 x_0 x_1}}{2 + e^{\theta_0} + e^{\theta_0+\theta_1}} \qquad \text{with}\quad A_1 = \{0,1\},\\ F_j^{\theta_j}(x_0,x_j) &= -\log\frac{e^{\theta_j x_0 x_j}}{3 + e^{\theta_j}} \qquad \text{with}\quad A_j = \{0,j\}\end{aligned}$$

for $j = 2,\dots,r$, then $\overline{P}^{\theta} \in G\big(\alpha_0(\theta)\,x_0 + \sum\limits_{k=1}^{r}\alpha_k(\theta)\,x_0x_k\big)$ is r-Markov Gibbs distribution with "pair-wise interactions" depending non-linearly on the unknown parameter, i.e., it is a"curved exponential distribution". Then all assumptions are satisfied, and the estimate now assumes the form:

$$\hat{\theta}_0^n = \log\left[2\,\frac{\hat{P}_{00}^n - \hat{P}_{01}^n}{1-\hat{P}_{00}^n}\right], \quad \hat{\theta}_1^n = \log\left[\frac{\hat{P}_{01}^n}{\hat{P}_{00}^n - \hat{P}_{01}^n}\right], \quad \text{and} \quad \hat{\theta}_j^n = \log\left[\frac{3\hat{P}_{0j}^n}{1-\hat{P}_{0j}^n}\right]$$

for $j = 2, \ldots, r$, where $\hat{P}_{0j}^n = \hat{P}^n(x_0 = 1, x_j = 1)$ for $j = 0, \ldots, r$.

REFERENCES

[1] BASAWA, I. V.—PRAKASA RAO, B. L. S.: *Statistical Inference for Stochastic Processes*. Academic Press Inc., London, 1980.

[2] BILLINGSLEY, P.: *Statistical Inference for Markov Processes*. University of Chicago Press, Chicago, 1961.

[3] DOBRUSHIN, R. L.—NAHAPETIAN, B. S.: *Strong convexity of the pressure for lattice systems of classical statistical physics*, Teoret. Mat. Fiz. **20** (1974), 223–234. (In Russian)

[4] DOBRUSHIN, R. L.—TIROZZI, B.: *The central limit theorem and the problem of equivalence of ensembles*, Comm. Math. Phys. **54** (1977), 174–192.

[5] FÖLLMER, H.: *On entropy and information gain in random fields*, Z. Wahrs. verw. Geb. **26** (1973), 207–217.

[6] GEORGII, H. O.: *Gibbs Measures and Phase Transitions*. De Gruyter, Berlin, 1988.

[7] JANŽURA, M.: *Statistical analysis of Gibbs-Markov binary random sequences*, Statist. Decisions **12** (1994), 367–384.

[8] JANŽURA, M.: *Asymptotic behaviour of the error probabilities in the pseudo-likelihood ratio test for Gibbs–Markov distributions*. In: Proc. 5th Prague Symposium on Asymp. Statistics (P. Mandl, M. Hušková, eds.), Physica–Verlag, Heidelberg, 1994, pp. 285–296.

[9] JANŽURA, M.: *Asymptotic results in parameter estimation for Gibbs random fields*, Kybernetika **33** (1997), 133–159.

[10] RUELLE, D.: *Statistical Mechanics, Rigorous Results*. Benjamin, New York, 1969.

Received October 12, 2006

Institute of Information Theory
and Automation
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 4
182 08 Prague
CZECH REPUBLIC
E-mail: janzura@utia.cas.cz