KYBERNETIKA — VOLUME 23 (1987), NUMBER 2

ESTIMATING INTERACTIONS IN BINARY LATTICE DATA WITH NEAREST-NEIGHBOR PROPERTY

MARTIN JANŽURA

A method for estimating parameters (called interactions) of Gibbsian model for binary variables with the nearest-neighbor property on a lattice is developed. Consistency of the estimate is proved, an approximate way of calculation is given and demonstrated by a numerical example.

1. INTRODUCTION

A collection of binary data obtained from lattice points of a two-dimensional rectangular observation region is considered to be generated by some two state random field (r. f.) the distribution of which we are interested in.

As usual, homogeneity expressed by translation invariance of the r. f. will be assumed. Moreover, dependence structure of the r.f. will be supposed to be given only by interations between the neighboring variables.

Thanks to these assumptions we may restrict our considerations to the class of Markov r.f.'s. Unfortunately, the multi-dimensional Markov r.f.'s can not be treated as easily as one-dimensional Markov chains.

Therefore, since Markov r.f.'s represent special cases of Gibbs r.f.'s as defined in the frame of statistical mechanics, we shall follow the theory of Gibbs r.f.'s.

This approach brings another advantage consisting in the fact that the Gibbsian description of Markov r.f. involves quantitative expression of the mentioned above interactions as real-valued parameters.

Thus, the problem of finding the unknown distribution is transformed to a parameter estimation problem.

After the basic definitions and results in Section 2 (adopted mostly from [5])) the estimate is constructed and its consistency is proved in Section 3. The following Section 4 is devoted to the implementation of the proposed method and the approximate way of calculation based on the results valid for the one-dimensional case (cf. [3]). In the last Section 5 a numerical example is presented and the results are compared with those of Besag [1] and Strauss [7] who analysed the same data by different methods earlier.

2. PRELIMINARIES

Let $X = \{0, 1\}$ be the state space, $\mathscr{F} = \exp X$ the σ-algebra of all its subsets. We denote by $\mathscr{Z}$ the set of all integers, by $\mathscr{Z}^+$ the set of all positive integers, and by $T = \mathscr{Z}^2$ the two-dimensional lattice. We denote by $h = (1, 0)$, $v = (0, 1) \in T$ the base vectors of T. Let $(X^T, \mathscr{F}^T)$ be the infinite product measurable space.

For $A \subset T$ we denote by $Pr_A: X^T \to X^A$ the corresponding projection function. For the sake of brevity we shall write x_A instead of $Pr_A(x)$ for $x \in X^T$, and shortly x_t instead of $x_{\{t\}}$ for one-point subsets of T. For $x_A \in X^A$ we denote by $\bar{x}_A = Pr_A^{-1}(x_A) \in \mathscr{F}^T$ the corresponding measurable cylinder.

By *two state random field* (abbreviated to *r.f.*) we mean a probability measure μ defined on the space $(X^T, \mathscr{F}^T)$.

A r.f. μ is *stationary* if it is invariant under shifts in both the directions, i.e.

$$\mu = \theta_h^{-1}\mu\,, \quad \mu = \theta_v^{-1}\mu$$

where the shifts θ_h, θ_v: $X^T \to X^T$ are defined through $\theta_h(x)_t = x_{t+h}$, $\theta_v(x)_t = x_{t+v}$, respectively, for every $x \in X^T$, $t \in T$.

A stationary r.f. μ is *ergodic* if its restriction to the σ-algebra of invariant sets assumes only the values zero or one, i.e. for every $F \in \mathscr{S} = \{E \subset \mathscr{F}^T;\ \theta_h^{-1}E = E,\ \theta_v^{-1}E = E\}$ it holds: if $\mu(F) > 0$ then $\mu(F) = 1$.

For $t \in T$ let $\|t\|$ be the two-dimensional Euclidean norm. For $A \subset T$ we denote by $\partial A = \{t \in T;\ \inf_{s \in A} \|s - t\| = 1\}$ the neighborhood of the set A.

A r.f. μ is called *nearest-neighbor* if it satisfies

$$\mu(x_t \mid y_{T \setminus \{t\}}) = \mu(x_t \mid y_{\partial\{t\}})$$

for every $t \in T$, $x_t \in X$, and a.e. $y \in X^T[\mu]$.

We have used short notation here for the conditional distributions, i.e. $\mu(x_A \mid y_B)$ means $\mathsf{E}_\mu[\mathsf{I}_{\bar{x}_A} \mid Pr_B^{-1}(\mathscr{F}^B)]\,(y)$ for every $A, B \subset T$, $x_A \in X^A$, $y \in X^T$, where I denotes the indicator function.

A nearest-neighbor r.f. is *Gibbs r.f.* if

$$\mu(x_t \mid x_{\partial\{t\}}) = \frac{\exp\{x_t(U_0 + U_H(x_{t+h} + x_{t-h}) + U_V(x_{t+v} + x_{t-v}))\}}{1 + \exp\{U_0 + U_H(x_{t+h} + x_{t-h}) + U_V(x_{t+v} + x_{t-v})\}}$$

holds for every $t \in T$, $x \in X^T$, where $\boldsymbol{U} = (U_0, U_H, U_V) \in \mathscr{R}^3$ is a three-dimensional vector of real-valued parameters called *interactions*.

Let us note that a nearest-neighbor stationary r.f. is Gibbs r.f. whenever its conditional distributions are everywhere positive.

For every $U \in \mathscr{R}^3$ there exists at least one ergodic Gibbs r.f. (cf. Theorem 3·7 in [5]).

Providing the horizontal and the vertical interactions are sufficiently weak, e.g. $|U_H| + |U_V| < \frac{1}{2}$ (cf. [6]), the corresponding Gibbs r.f. is uniquely determined, and therefore ergodic.

For every $U \in \mathscr{R}^3$ let us denote the set of stationary, and ergodic Gibbs r.f.'s by $G_{St}(U)$, and $G_E(U)$, respectively.

Let $D \subset T$ be a finite rectangular lattice. For fixed $x_D = \{x_t; t \in D\} \in X^D$ let $Y_0(x_D) = \sum_{t \in D} x_t$ be the number of ones, and $Y_H(x_D) = \sum_{t \in D \cap (D-h)} x_t x_{t+h}$, resp. $Y_V(x_D) = \sum_{t \in D \cap (D-v)} x_t x_{t+v}$, be the number of adjacent horizontal, resp. vertical, pairs of ones.

For every $U = (U_0, U_H, U_V) \in \mathscr{R}^3$ we denote

$$Z_U(D) = \sum_{x_D \in X^D} \exp\{U_0 \,.\, Y_0(x_D) + U_H \,.\, Y_H(x_D) + U_V \,.\, Y_V(x_D)\}\,.$$

According to Theorem 3.4 in [5] the limit

$$p(U) = \lim_{m,n \to \infty} (m \,.\, n)^{-1} \,.\, \log Z_U(D(m, n))$$

exists, where $D(m, n) = \{(t_1, t_2) \in T; 0 \leqq t_1 < m, 0 \leqq t_2 < n\}$ for every $m, n \in \mathscr{Z}^+$, and $p: \mathscr{R}^3 \to \mathscr{R}$ is a strictly convex continuous function.

Now, let us define for every fixed $\beta = (\beta_0, \beta_H, \beta_V) \in \mathscr{R}^3$ the function $\Phi_\beta: \mathscr{R}^3 \to \mathscr{R}$ through the following formula

$$\Phi_\beta(U) = p(U) - U_0\beta_0 - U_H\beta_H - U_V\beta_V$$

for every $U \in \mathscr{R}^3$.

Let

$$\mathscr{M} = \{\beta \in \mathscr{R}^3; \exists\, U_\beta: \Phi_\beta(U_\beta) = \min_{U \in \mathscr{R}^3} \Phi_\beta(U)\}$$

be the set of parameters β for which the corresponding convex function attains its minimum, and let $\Psi: \mathscr{M} \to \mathscr{R}^3$ be the transform defined by $\Psi(\beta) = U_\beta$ for every $\beta \in \mathscr{M}$.

Lemma 2.1. It holds

(i) $\mathscr{M}$ is a open convex subset of $\mathscr{R}^3$;

(ii) Ψ is a continuous function;

(iii) $\Psi^{-1}(U)$ is a compact convex subset of $\mathscr{M}$ for every $U \in \mathscr{R}^3$.

The proof follows from properties of the function p by elementary considerations. □

Let us denote $C_0 = \{x \in X^T; x_0 = 1\}$, $C_H = \{x \in X^T; x_0 \,.\, x_h = 1\}$, and $C_V = \{x \in X^T; x_0 \,.\, x_v = 1\}$. For every r.f. μ let us denote $\beta(\mu) = (\mu(C_0), \mu(C_H), \mu(C_V)) \in [0, 1]^3$.

Proposition 2.2. For every $\mu \in G_{St}(U^0)$, $U^0 \in \mathscr{R}^3$, it holds

$$\Phi_{\beta(\mu)}(U^0) = \min_{U \in \mathscr{R}^3} \Phi_{\beta(\mu)}(U)\,.$$

Proof. Cf. Theorem 3.12 in [5]. □

3. ESTIMATING INTERACTIONS

For fixed $m, n \in \mathscr{Z}^+$ a collection $x_{D(m,n)} \in X^{D(m,n)}$ of binary data is now supposed to be generated by an unknown Gibbs r.f. $\mu \in G_{St}(U^0)$, $U^0 = (U_0^0, U_H^0, U_V^0) \in \mathscr{R}^3$. On the base of the given collection of observations we try to find the interactions. Considering the interactions as a vector parameter, we obtain a parameter estimation problem.

A solution of that problem proceeds from Proposition 2.2 of the preceding section. If we knew the vector $\boldsymbol{\beta}(\mu)$ we could obtain the unknown interactions $\boldsymbol{U}^0$ by minimizing the function $\Phi_{\boldsymbol{\beta}(\mu)}$. Thus, minimizing the function $\Phi_{\hat{\beta}(m,n)}$ where $\hat{\beta}(m, n) \in \mathscr{R}^3$ is a vector of some empirical values, we obtain an estimate $\hat{\boldsymbol{U}}_{m,n}$.

Naturally, the empirical values will be given by relative frequences, i.e.

$$\hat{\beta}(m, n) = ((m \, . \, n)^{-1} \, Y_0(x_{D(m,n)}), (n(m - 1))^{-1} \, Y_H(x_{D(m,n)}),$$
$$(m(n - 1))^{-1} \, Y_V(x_{D(m,n)})) \, .$$

Theorem 3.1. The function $\Phi_{\hat{\beta}(m,n)}$ attains its minimum with probability tending to one, the root $\hat{\boldsymbol{U}}_{m,n}$ being a consistent estimate of the interactions $\boldsymbol{U}^0$.

Proof. According to d-dimensional ergodic theorem (cf. Theorem VIII. 6.9 in [2]) there exist $\mathscr{S}$-measurable functions $g_0, g_H, g_V \colon X^T \to \mathscr{R}$ such that

$$\hat{\beta}(m, n) \to \boldsymbol{g} = (g_0, g_H, g_V) \quad \text{a.s. } [\mu] \quad \text{as} \quad m, n \to \infty$$

for every stationary μ.

From the representation theory for stationary Gibbs r.f.'s it follows (see e.g. Proposition 3.6 and Corollary 3.14 in [5]) that there exists a unique probability measure Q, defined on the set of stationary r.f.'s with a suitable σ-algebra such that

$$\mu(F) = \int_{G_E(\boldsymbol{U}^0)} \nu(F) \, \mathrm{d}Q(\nu)$$

holds for every $F \in \mathscr{F}^T$. Therefore

$$\mu\{x \in X^T; \boldsymbol{g}(x) \in \Psi^{-1}(U^0)\} = \int_{G_E(\boldsymbol{U}^0)} \nu\{x \in X^T; \boldsymbol{g}(x) \in \Psi^{-1}(\boldsymbol{U}^0)\} \, \mathrm{d}Q(\nu) = 1$$

as $\boldsymbol{g} = \mathsf{E}_\nu[\boldsymbol{g}] = (\mathsf{E}_\nu[g_0], \mathsf{E}_\nu[g_H], \mathsf{E}_\nu[g_V])$ a.s. $[\nu]$ for every ergodic ν, and $\mathsf{E}_\nu[\boldsymbol{g}] = \boldsymbol{\beta}(\nu) \in \Psi^{-1}(\boldsymbol{U}^0)$ for $\nu \in G_E(\boldsymbol{U}^0)$ by Proposition 2.2.

Since the compact (by Lemma 2.1) set $\Psi^{-1}(\boldsymbol{U}^0)$ is contained in the open (by Lemma 2.1) set $\mathscr{M}$ with some ε-neighborhood $\Psi^{-1}(\boldsymbol{U}^0)^\varepsilon$, and since a.s. convergence yields convergence in probability, we have

$$\mu\{\hat{\beta}(m, n) \notin \mathscr{M}\} \leqq \mu\{\hat{\beta}(m, n) \notin \Psi^{-1}(\boldsymbol{U}^0)^\varepsilon\} \leqq \mu\{\|\hat{\beta}(m, n) - \boldsymbol{g}\| \geqq \varepsilon\} \to 0$$

as $m, n \to \infty$.

Hence the estimate $\hat{\boldsymbol{U}}_{m,n}$ exists with probability tending to one. The consistency follows immediately from the a.s. convergence of $\hat{\beta}(m, n)$ and the continuity of the transform Ψ (Lemma 2.1). $\square$

Remark. Let us realize that in case of non-uniqueness of the Gibbs r.f. the interactions do not determine the r.f. μ but the class $G_{St}(U^0)$ which the r.f. μ belongs to. Thus, from the point of view of the r.f.'s, the problem of interactions estimation is a discrimination problem with classes $G_{St}(U)$, $U \in \mathscr{R}^3$ (many of them consisting of one r.f. only, of course).

4. IMPLEMENTATION

Now, we will not be concerned with the properties of the estimate. Our main interest consists in the computational aspect of the problem. There is one substantial difficulty during minimization of the function Φ_β; the difficulty in evaluating the function p. Since direct computation from the definition is impossible, and no other method seems to be available, we shall substitute the function p by a function which approximates p and the values of which can be easily calculated.

For every $R \in \mathscr{Z}^+$ let us define the function $p^R\colon \mathscr{R}^3 \to \mathscr{R}$ through the formula

$$p^R(U) = \lim_{N\to\infty} (N \,.\, R)^{-1} \log Z_U^R(N) \quad \text{for every} \quad U = (U_0, U_H, U_V) \in \mathscr{R}^3 ,$$

where

$$Z_U^R(N) = \sum_{x_1,\ldots,x_{R.N}\in X} \exp\{U_0 \sum_{i=1}^{N.R} x_i + U_H \sum_{i=1}^{N.R-1} x_i x_{i+1} + U_V \sum_{i=1}^{N.R-R} x_i x_{i+R}\}$$

for every $N \in \mathscr{Z}^+$ ($p^R(U)$ exists again due to Theorem 3.4 in [5]).

Proposition 4.1. It holds $|p(U) - p^R(U)| \leqq 2 \,.\, R^{-1}|U_H|$ for every $U \in \mathscr{R}^3$ and $R \in \mathscr{Z}^+$, and therefore $p(U) = \lim_{R\to\infty} p^R(U)$.

Proof. Easy consideration shows that

$$|\log Z_U(D(R, N)) - (KL)^{-1} \log Z_U(D(L.\, R, K \,.\, N))| \leqq N|U_H| + R|U_V|$$

for every $R, N, K, L \in \mathscr{Z}^+$.

Taking limit for $L, K \to \infty$, we obtain

$$|p(U) - (R \,.\, N)^{-1} \log Z_U(D(R, N))| \leqq R^{-1}|U_H| + N^{-1}|U_V| \,.$$

Further, it holds

$$|\log Z_U(D(R, N)) - \log Z_U^R(N)| \leqq N \,.\, |U_H| \quad \text{for every} \quad R, N \in \mathscr{Z}^+ \,.$$

Hence, we conclude

$$|p(U) - (N \,.\, R)^{-1} \log Z_U^R(N)| \leqq 2R^{-1}|U_H| + N^{-1}|U_V| ,$$

and taking limit for $N \to \infty$, we obtain the result. □

Following e.g. [4], Section I.2.1, we obtain

$$p^R(U) = R^{-1} \log \lambda_{\max}(M_U^R) ,$$

where M_U^R is a strictly positive-valued $2^R \times 2^R$ matrix with elements defined through the formula

$$M_U^R((x_1, \ldots, x_R), (y_1, \ldots, y_R)) =$$

$$= \exp\{U_0 \sum_{i=1}^{R} x_i + U_H(\sum_{i=1}^{R-1} x_i x_{i+1} + x_R y_1) + U_V \sum_{i=1}^{R} x_i y_i\}$$

$$\text{for every} \quad x_1, \ldots, x_R\,, \quad y_1, \ldots, y_R \in X\,,$$

and $\lambda_{\max}(M_U^R)$ is the uniquely defined strictly positive eigenvalue of the matrix M_U^R ($\lambda_{\max}(M_U^R)$ exists due to the well-known Perron-Frobenius theorem).

Remark. The bound given in Proposition 4.1, ensuring the convergence, does not seem to be satisfactory from the approximation point of view. Nevertheless, computational experiments show that, in fact, the convergence is fast enough. Even p^5, which is quite easy to be calculated, gives a sufficient approximation.

4. EXAMPLE

In order to illustrate the proposed method we shall apply it to the data analysed by Besag [1] and later by Strauss [7].

We have

$$m = n = 24$$
$$Y_0 = 176$$
$$Y_H = 66$$
$$Y_V = 69\,.$$

Using the approximation with $R = 4, 5, 6$, respectively, we obtain the following estimates:

	$\hat{U}_0$	$\hat{U}_H$	$\hat{U}_V$
$R = 4$	$-1{\cdot}638$	$0{\cdot}553$	$0{\cdot}681$
$R = 5$	$-1{\cdot}619$	$0{\cdot}544$	$0{\cdot}665$
$R = 6$	$-1{\cdot}621$	$0{\cdot}545$	$0{\cdot}665$

Under the additional assumption $U_H = U_V$ (isotropy) we obtain with the aid of our method:

	$\hat{U}_0$	$\hat{U}_H = \hat{U}_V$
$R = 4$	$-1{\cdot}631$	$0{\cdot}612$
$R = 5$	$-1{\cdot}615$	$0{\cdot}602$
$R = 6$	$-1{\cdot}617$	$0{\cdot}604$

We may compare these results with those of Strauss [7] who obtained the estimates

$$\hat{U}_H = \hat{U}_V = 0{\cdot}594 \quad 0{\cdot}592 \quad 0{\cdot}611$$

by three different methods of expansion, and those of Besag [1] who used his "coding method" to obtain two estimates

$$\hat{U}_H = \hat{U}_V = 0{\cdot}589 \quad 0{\cdot}481 \, .$$

Clearly, our results agree quite well especially with the "more exact" estimates of Strauss.

(Received November 4, 1985.)

REFERENCES

[1] J. E. Besag: On the statistical analysis of nearest-neighbor systems. Proceedings 9. European meeting of statisticians, Budapest 1972.

[2] N. Dunford and J. T. Schwartz: Linear Operators, I. Interscience, New York 1958.

[3] M. Janžura: Estimating interactions in binary data sequences. Kybernetika *22* (1986), 5, 377—384.

[4] D. H. Mayer: The Ruelle-Araki Transfer Operator in Classical Mechanics. (Lecture Notes in Physics 123.) Springer-Verlag, Berlin—Heidelberg—New York 1980.

[5] D. Ruelle: Thermodynamic Formalism. Addison Wesley, Reading, Mass. 1978.

[6] D. Simon: A remark on Dobrushin's uniqueness theorem. Comm. Math. Phys. *68* (1979), 183—185.

[7] D. J. Strauss: Analysing binary lattice data with the nearest-neighbor property. J. Appl. Probab. *12* (1975), 702—712.

RNDr. Martin Janžura, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.