Consistent Estimation of the Basic Neighborhood of Markov Random Fields

Imre Csiszár[1], and Zsolt Talata[2]
A. Rényi Institute of Mathematics, Hungarian Academy of Sciences
POB 127, H-1364 Budapest, Hungary
{csiszar, zstalata}@renyi.hu

Abstract — **For Markov random fields with finite set of states, a modification of the Bayesian Information Criterion admits strongly consistent estimation of the smallest region that determines the conditional distributions. Phase transition or non-stationarity do not affect the result.**

I. Introduction

The basic neighborhood (BN) of a Markov Random Field (MRF) is the smallest region that determines conditional distributions at the sites. Statistical estimation of the BN is a problem similar to but more difficult than estimation of the order of a Markov chain. The latter was considered in [2], [3], and this work continues those. The Bayesian Information Criterion (BIC) used in [2] involves maximum likelihood and a penality term equal to one half the number of free parameters times the log of the sample size.

As no maximum likelihood formula is available for MRF's, here we use maximum pseudo-likelihood [1]; in the penalty term, a factor analogous but not equal to the number of free parameters will be used. The so obtained Pseudo-Bayesian Information Criterion (PIC) provides a strongly consistent estimator of the BN, not assuming any prior bound on its size, although a bound on the size of the hypothetical BN that grows with the sample size is required, unlike in [2].

II. Statement of the result

A *random field* on $\mathbb{Z}^d$ is a family of random variables $\{ X(i) : i \in \mathbb{Z}^d \}$; we assume they take values in a finite set A, with strictly positive finite-dimensional joint distributions. Write $X(\Delta) = \{ X(i) : i \in \Delta \}$ for finite $\Delta \subset \mathbb{Z}^d$, with possible values $a(\Delta) = \{ a(i) \in A : i \in \Delta \}$ called *blocks*. The translate of Δ when $0 \in \mathbb{Z}^d$ is translated to $i \in \mathbb{Z}^d$ is denoted by Δ^i. A *neighborhood* (of the origin) is any central symmetric finite set $\Gamma \subset \mathbb{Z}^d \backslash \{0\}$.

A random field is a *Markov random field* if there exists a neighborhood Γ such that

$$\begin{aligned}\mathrm{Prob}\{ X(i) = a \mid X(\Delta^i) = a(\Delta) \} \\ = \mathrm{Prob}\{ X(0) = a \mid X(\Gamma) = a(\Gamma) \} \quad (1)\end{aligned}$$

for each $\Delta \supseteq \Gamma$ with $0 \notin \Delta$, $a \in A$, $a(\Delta) \in A^{\Delta}$, and $i \in \mathbb{Z}^d$.

MRF's are the same as Gibbs fields with finite range interactions. Note that the conditional probabilities on the right of (1) need not uniquely determine the joint distribution of the $X(i)$'s, a phenomenon called *phase transition*. Moreover, the joint distribution is not necessarily stationary (translation invariant).

Standard results [4] imply that among the neighborhoods Γ satisfying (1) there is a minimal one, the *basic neighborhood* Γ_0. We are concerned with its statistical estimation from observing a realization of the MRF on an increasing sequence of finite regions $\Lambda_n \subset \mathbb{Z}^d$, $n \in \mathbb{N}$. We draw the statistical inference about a hypothetical BN Γ based on the blocks $a(\Gamma) \in A^\Gamma$ and $a(\Gamma, 0) \in A^{\Gamma \cup \{0\}}$ appearing in the sample $x(\Lambda_n)$, where $a(\Gamma, 0)$ is a shorthand for $a(\Gamma \cup \{0\})$.

Write $C(r) = \{ (j_1, \dots, j_d) \in \mathbb{Z}^d : \max |j_k| \le r \}$, and let $r_n \to \infty$, $r_n = o(\log^{1/(2d)} |\Lambda_n|)$. Let $\bar\Lambda_n$ be the set of those $i \in \Lambda_n$ for which the translate $C^i(r_n)$ is contained in Λ_n. For $\Gamma \subseteq C(r_n)$, denote

$$N_n(a(\Gamma)) = \left| \left\{ i \in \bar\Lambda_n : x(\Gamma^i) = a(\Gamma) \right\} \right|$$

$$N_n(a(\Gamma, 0)) = \left| \left\{ i \in \bar\Lambda_n : x(\Gamma^i \cup \{i\}) = a(\Gamma, 0) \right\} \right|.$$

The *maximum pseudo-likelihood* MPL_Γ, depending on the sample $x(\Lambda_n)$, for Γ as a hypothetical BN is defined by

$$\log \mathrm{MPL}_\Gamma(x(\Lambda_n)) = \sum_{a(\Gamma,0)} N_n(a(\Gamma, 0)) \log \frac{N_n(a(\Gamma, 0))}{N_n(a(\Gamma))}.$$

We define the *Pseudo-Bayesian Information Criterion* with (an arbitrary) constant $c > 0$ as

$$\mathrm{PIC}_\Gamma(x(\Lambda_n)) = -\log \mathrm{MPL}_\Gamma(x(\Lambda_n)) + c\,|A|^{|\Gamma|} \log |\Lambda_n|.$$

Theorem. *The PIC-estimator*

$$\widehat\Gamma_{\mathrm{PIC}}(x(\Lambda_n)) = \arg \min_{\Gamma \subseteq C(r_n)} \mathrm{PIC}_\Gamma(x(\Lambda_n))$$

of the BN is strongly consistent, that is, $\widehat\Gamma_{\mathrm{PIC}}(x(\Lambda_n)) = \Gamma_0$ *eventually almost surely as* $n \to \infty$*, provided that the sample regions satisfy* $|\bar\Lambda_n| / |\Lambda_n| \to 1$.

The inclusion $\widehat\Gamma_{\mathrm{PIC}} \supseteq \Gamma_0$ is proved by a familiar entropy argument. The proof of $\widehat\Gamma_{\mathrm{PIC}} \not\supset \Gamma_0$ relies upon the following strong typicality result:

Proposition. *Uniformly for all blocks* $a(\Gamma, 0) \in A^{\Gamma \cup \{0\}}$ *with* $\Gamma_0 \subseteq \Gamma \subseteq C(r_n)$*, the ratio* $N_n(a(\Gamma, 0)) / N_n(a(\Gamma))$ *differs by less than* $\varepsilon_n \cdot [\log N_n(a(\Gamma)) / N_n(a(\Gamma))]^{1/2}$ *from the right hand side of (1), eventually almost surely, where* $\varepsilon_n \to 0$.

References

[1] J. Besag, "Efficiency of pseudo-likelihood estimation for simple Gaussian fields," *Biometrika*, vol. 64, pp. 616–618, 1977.

[2] I. Csiszár, and P. C. Shields, "The consistency of the BIC Markov order estimator," *Ann. Statist.*, pp. 1601–1619, 2000.

[3] I. Csiszár, "Large-scale typicality of Markov sample paths and consistency of MDL order estimators," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1616–1629, 2002.

[4] H. O. Georgii, *Gibbs Measures and Phase Transitions*, de Gruyter, Berlin, 1988.

[1]Supported by the Hungarian National Foundation for Scientific Research, Grants T26041, T32323, TS40719, and T046376.

[2]Supported by the Hungarian National Foundation for Scientific Research, Grant T046376.