Statistica Sinica **10**(2000), 1325-1344

ASYMPTOTIC PROPERTIES OF MAXIMUM (COMPOSITE) LIKELIHOOD ESTIMATORS FOR PARTIALLY ORDERED MARKOV MODELS

Hsin-Cheng Huang and Noel Cressie

Academia Sinica, Taipei and The Ohio State University

Abstract: Partially ordered Markov models (POMMs) are Markov random fields (MRFs) with neighborhood structures derivable from an associated partially ordered set. The most attractive feature of POMMs is that their joint distributions can be written in closed and product form. Therefore, simulation and maximum likelihood estimation for the models is quite straightforward, which is not the case in general for MRF models. In practice, one often has to modify the likelihood to account for edge components; the resulting composite likelihood for POMMs is similarly straightforward to maximize. In this article, we use a martingale approach to derive the asymptotic properties of maximum (composite) likelihood estimators for POMMs. One of our results establishes that under regularity conditions that are fairly easy to check, and Dobrushin's condition for spatial mixing, the maximum composite likelihood estimator is consistent, asymptotically normal, and also asymptotically efficient.

Key words and phrases: Acyclic directed graph, asymptotic efficiency, asymptotic normality, consistency, Dobrushin's condition, level set, Markov random field, martingale central limit theorem, strong mixing, triangular martingale array.

1. Introduction

Partially ordered Markov models (POMMs), introduced by Cressie and Davidson (1998), are a natural extension of one-dimensional Markov chains and a generalization of two-dimensional Markov mesh models (Abend, Harley and Kanal (1965)). The models are actually Markov random field (MRF) models with neighborhood structures derivable from an associated partially ordered set (poset). That poset determines the conditional distribution of a datum at any site $\boldsymbol{s}$, conditioned on data at sites that are less than $\boldsymbol{s}$ by the partial order; specifically, the conditional distribution depends only on the data at the adjacent lower neighborhood of $\boldsymbol{s}$. This development of POMMs in the spatial context has direct parallels to models on acyclic directed graphs and causal network models (e.g., Lauritzen and Spiegelhalter (1988); Lauritzen, Dawid, Larson and Leimer (1990)).

Generally, the maximum likelihood estimator (MLE) for a MRF model is analytically intractable and numerical solutions are usually computationally intensive. Therefore several alternative estimation procedures, which are computationally efficient but less efficient than maximum likelihood (ML), have been proposed. For example, Besag (1974, 1975) proposed the coding and the maximum pseudo-likelihood methods, and Possolo (1986) proposed the logit method for binary Markov random fields. However on a large sub-class of MRFs, namely the POMMs, the joint distributions can be written in closed and product form, and hence ML estimation is relatively straightforward. In practice one often has to modify the likelihood to account for edge components; the resulting composite likelihood can be similarly maximized in a straightforward manner.

Many authors have considered the consistency and the asymptotic normality of ML estimation. The literature for one-dimensional dependent processes in a multi-parameter framework includes Basawa, Feigin and Heyde (1976), Crowder (1976), Sweeting (1980), Heijmans and Magnus (1986a, b), and Sarma (1986). Unfortunately, only a small proportion of the literature on ML estimation is concerned with higher dimensional, dependent random fields. Gidas has proved the consistency (1988) and asymptotic normality (1993) of the MLE for a MRF model under certain conditions. However, his parameterization is somewhat restricted in the sense that the potential function of his MRF model is linear in the parameters. That is, his MRF model belongs to a certain exponential family (which is typically not the case for POMMs). In this article, we adapt a martingale approach used by Crowder (1976) to derive the asymptotic properties of maximum (composite) likelihood estimators for POMMs. For example, one of our results establishes that, under regularity conditions that are fairly easy to check and Dobrushin's condition (Dobrushin (1968)) for spatial mixing, the maximum composite likelihood estimator is consistent, asymptotically normal, and also asymptotically efficient. Although the approach by Sweeting (1980) may also be used to derive the asymptotic properties of the ML estimator, some stronger uniform-convergence condition has to be imposed, which may not be as easy to check in practice as the approach we are using.

In Section 2, we define a POMM on $\mathbb{Z}^d$, which is extended from Cressie and Davidson's (1998) definition for a POMM defined on a finite set of sites. The consistency and asymptotic normality of the maximum (composite) likelihood estimator for POMMs are established in Section 3. Section 4 contains an example in which the composite-likelihood results of Section 3 are applied to a specific spatial process on a two-dimensional square lattice. Proofs of the results stated in Section 3 are given in the Appendix.

2. Partially Ordered Markov Models

We give the basic definitions associated with a partially ordered Markov model (POMM); Cressie and Davidson (1998), Davidson and Cressie (1993), and Davidson, Cressie and Hua (1999) can be consulted for further details.

Consider first the notion of a partial order. Let D be a set of elements. Then $(D, \prec)$ is said to be a partially ordered set, or a poset, with partial order $\prec$, if for any $\boldsymbol{s}, \boldsymbol{s}', \boldsymbol{s}'' \in D$, the following three conditions are satisfied: (a) $\boldsymbol{s} \prec \boldsymbol{s}$; (b) $\boldsymbol{s} \prec \boldsymbol{s}'$ and $\boldsymbol{s}' \prec \boldsymbol{s}$ implies $\boldsymbol{s} = \boldsymbol{s}'$; (c) $\boldsymbol{s} \prec \boldsymbol{s}'$ and $\boldsymbol{s}' \prec \boldsymbol{s}''$ implies $\boldsymbol{s} \prec \boldsymbol{s}''$.

Next we introduce the notion of directed graphs and briefly discuss the relation between directed graphs and partial orders. A directed graph is a pair (V, F), where V denotes the set of vertices (or nodes) and F denotes the set (possibly empty) of directed edges between vertices. A directed edge is an ordered pair (v, v') that represents a directed connection from a vertex v to a different vertex v'. A directed path in a directed graph is a sequence of vertices $v_1, \ldots, v_k$, $k > 1$, such that (v_i, v_{i+1}) is a directed edge for each $i = 1, \ldots, k-1$. A cycle in a directed graph is a path $(v_1, \ldots, v_k)$ such that $v_1 = v_k$. An acyclic directed graph (ADG) is a directed graph that has no cycles in it. We shall use an acyclic directed graph to specify the spatial interdependencies between locations for a POMM and, as a consequence, the results given here in a spatial context might be adapted to the nonspatial models of Lauritzen, Dawid, Larson and Leimer (1990).

To construct a poset from an ADG (V, F), we define a binary relation $\prec$ on V such that $v \prec v'$ if either $v = v'$ or there exists a directed path from v to v'. Then it is straightforward to check that $(V, \prec)$ is a poset on V with the partial order $\prec$.

Before giving the definition of a POMM on $\mathbb{Z}^d$, the d-dimensional integer lattice, we introduce some notation and definitions. Consider a real-valued random field $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{Z}^d\}$. For any set $D \subset \mathbb{Z}^d$, let $Z(D) \equiv \{Z(\boldsymbol{s}) : \boldsymbol{s} \in D\}$. For any finite $D' \in D$, let $p(z(D'))$ denote the probability density (or probability mass function) of $Z(D')$. For any finite $D_1, D_2 \subset D$, let $p(z(D_1)|z(D_2))$ denote the conditional probability density (or conditional probability mass function) of $Z(D_1)$ conditioned on $Z(D_2) = z(D_2)$.

For $\boldsymbol{s}, \boldsymbol{s}' \in \mathbb{Z}^d$, let $d(\boldsymbol{s}, \boldsymbol{s}') \equiv \max_{1 \leq i \leq d} |s_i - s_i'|$, where $\boldsymbol{s} = (s_1, \ldots, s_d)$ and $\boldsymbol{s}' = (s_1', \ldots, s_d')$. For any $D_1, D_2 \subset \mathbb{Z}^d$, let $d(D_1, D_2) \equiv \inf\{d(\boldsymbol{s}_1, \boldsymbol{s}_2) : \boldsymbol{s}_1 \in D_1,\ \boldsymbol{s}_2 \in D_2\}$. For any $D \subset \mathbb{Z}^d$, let $diam\, D \equiv \sup\{d(\boldsymbol{s}, \boldsymbol{s}') : \boldsymbol{s}, \boldsymbol{s}' \in D\}$.

Consider now some definitions that are related to an ADG $(\mathbb{Z}^d, F)$ and its associated poset $(\mathbb{Z}^d, \prec)$. For any $\boldsymbol{s} \in \mathbb{Z}^d$, the *cone* of $\boldsymbol{s}$ is the set $cone\, \boldsymbol{s} \equiv \{\boldsymbol{s}' \in \mathbb{Z}^d \setminus \{\boldsymbol{s}\} : \boldsymbol{s}' \prec \boldsymbol{s}\}$, and the *adjacent lower neighborhood* of $\boldsymbol{s}$ is the set

$adjl\, \boldsymbol{s} \equiv \{\boldsymbol{s}' \in \mathbb{Z}^d : (\boldsymbol{s}', \boldsymbol{s}) \in F\}$. Throughout this article, we only consider ADGs that yield *finite* adjacent lower neighborhoods. For any $D \subset \mathbb{Z}^d$, the *cover* of D is the set $covr\, D \equiv \{\boldsymbol{s} \in \mathbb{Z}^d \setminus D : \ adjl\, \boldsymbol{s} \subset D\}$, and D is said to be a lower set if there exists an $\boldsymbol{s} \in \mathbb{Z}^d$ such that $cone\, \boldsymbol{s} \subset D$.

Definition 1. A set $D \subset \mathbb{Z}^d$ is said to be bounded below, or a b-set, if for any $\boldsymbol{s} \in D$ there exists an element $\boldsymbol{s}' \in L^0$ such that $\boldsymbol{s}' \prec \boldsymbol{s}$, where $L^0 \equiv \{\boldsymbol{s} \in D : \text{for any } \boldsymbol{s}' \in D, \text{ either } \boldsymbol{s} \prec \boldsymbol{s}' \text{ or } \boldsymbol{s}, \boldsymbol{s}' \text{ are unrelated}\}$ is the set of minimal elements of D.

It is not difficult to see that the complement of a lower set $D \subset \mathbb{Z}^d$, denoted by D^c, is bounded below.

Definition 2. The level sets of a b-set $D \subset \mathbb{Z}^d$ are a sequence of nonempty cover sets $\{L^n : n = 0, 1, \ldots\}$, defined recursively as $L^n \equiv covr\, \{\cup_{k=0}^{n-1} L^k \cup D^c\} \cap D$, $n \in \mathbb{N} \equiv \{1, 2, \ldots\}$, where L^0 is defined in Definition 1.

For example, consider a directed graph $(\mathbb{Z}^2, F)$ with

$$F = \cup_{(u,v) \in \mathbb{Z}^2} \Big\{ \Big((u, v-1), (u, v)\Big),\ \Big((u-1, v), (u, v)\Big),\ \Big((u-1, v-1), (u, v)\Big) \Big\}.$$

Then $cone\,(u, v) = \{(u', v') : u' \leq u \text{ or } v' \leq v\}$, which is also a lower set, and $adjl\,(u, v) = \{(u, v-1), (u-1, v), (u-1, v-1)\}$. Also, if $D = cone\,(0, 0)$, then $covr\, D = \{(0, 0)\}$, and if $D = \{(u, v) : u \geq 0 \text{ and } v \geq 0\}$, then the level sets of D for $n = 0, 1, \ldots$, are given by $L^n = \{(u, v) \in D : u + v = n\}$.

It is straightforward to see that, for D a b-set, $\cup_{n=0}^{\infty} L^n = D$ and $L^i \cap L^j = \emptyset$, $i \neq j$. Also note that the elements in each level set L^n, $n = 0, 1, \ldots$, are mutually unrelated by the partial order $\prec$, and an element in L^i cannot be larger than an element in L^j if $i < j$. Therefore, for D a finite b-set with $|D| = m$, we can write $D = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m\}$ such that $\boldsymbol{s}_i \prec \boldsymbol{s}_j$ only if $i < j$, $i, j = 1, \ldots, m$. This kind of ordering based on level sets is important for proving the asymptotic properties of maximum (composite) likelihood estimators.

Definition 3. Let $(\mathbb{Z}^d, F)$ be an ADG with its associated poset $(\mathbb{Z}^d, \prec)$, and let L^* be a nonempty lower set of $\mathbb{Z}^d$. Let $U_{\boldsymbol{s}}$ denote any finite set such that $adjl\, \boldsymbol{s} \subset U_{\boldsymbol{s}} \subset cone\, \boldsymbol{s}$, and let $V_{\boldsymbol{s}}$ denote any finite set of points not related to $\boldsymbol{s}$ by the partial order $\prec$. Then $\{Z(\boldsymbol{s}) :\ \boldsymbol{s} \in \mathbb{Z}^d\}$ is said to be a partially ordered Markov model (POMM) on $\mathbb{Z}^d$ if, for all $\boldsymbol{s} \in \mathbb{Z}^d \setminus L^*$ and for any $U_{\boldsymbol{s}}$ and $V_{\boldsymbol{s}}$, $p(z(\boldsymbol{s}) | z(U_{\boldsymbol{s}} \cup V_{\boldsymbol{s}})) = p(z(\boldsymbol{s}) | z(adjl\, \boldsymbol{s}))$.

Notice that the POMM defined by Cressie and Davidson (1998) is on a finite set of sites. The existence of a POMM on $\mathbb{Z}^d$ can be proved by using Kolmogoroff's Extension Theorem (see Kolmogoroff (1933); Durrett (1991)). A special case of a POMM is the Markov Mesh Model, until now defined on a finite subset of $\mathbb{Z}^2$. It was introduced by Abend, Harley and Kanal (1965) and has attracted some interest for fast generation of textures (e.g., Devijver (1988); Goutsias (1989); Gray, Kay and Titterington (1994)).

3. Limit Theorems for Maximum (Composite) Likelihood Estimation

From now on, let $Z(\mathbb{Z}^d)$ be a POMM with lower set L^* and level sets $\{L^0, L^1, \ldots\}$. Let $\{\Lambda_n : n = 1, 2, \ldots\}$ be a strictly increasing sequence of finite subsets of $\mathbb{Z}^d \setminus L^*$ that satisfy

$$\lim_{n\to\infty} |\Lambda_n^*|/|\Lambda_n| = 1, \tag{1}$$

where for each $n \in \mathbb{N}$, $\Lambda_n^* \equiv \{\boldsymbol{s} \in \Lambda_n : adjl\, \boldsymbol{s} \subset \Lambda_n\}$ is the set of the interior points of Λ_n. Then for each $n \in \mathbb{N}$, the set of the edge points of Λ_n is $E_n \equiv \Lambda_n \setminus \Lambda_n^*$. That is, we assume that the proportion of the edge points within Λ_n is asymptotically negligible as $n \to \infty$. Using the idea of level sets, we can assume without loss of generality that

$$\Lambda_n = \{\boldsymbol{s}_{n,1}, \ldots, \boldsymbol{s}_{n,|\Lambda_n|}\}, \quad n \in \mathbb{N}, \tag{2}$$

such that $\boldsymbol{s}_{n,i} \prec \boldsymbol{s}_{n,j}$ only if $i < j$, $i, j = 1, \ldots, |\Lambda_n|$. It has to be noticed that for $m, n \in \mathbb{N}$ and a given $i \in \{1, \ldots, |\Lambda_m|\}$, $\boldsymbol{s}_{m,i}$ may not be the same as $\boldsymbol{s}_{n,i}$ if $m < n$, since the collection of the level sets of Λ_m is typically not a subcollection of the level sets of Λ_n. In the spatial context of POMMs, Cressie and Davidson (1998) give the joint probability density (or mass) function of $Z(\Lambda_n)$ as

$$\begin{aligned} p\left(z(\Lambda_n)\right) &= \prod_{k=1}^{|\Lambda_n|} p\left(z(\boldsymbol{s}_{n,k})|z(\boldsymbol{s}_{n,1}), \ldots, z(\boldsymbol{s}_{n,k-1})\right) \\ &= \prod_{\boldsymbol{s}_{n,k}\in E_n} p(z(\boldsymbol{s}_{n,k})|z(\boldsymbol{s}_{n,1}), \ldots, z(\boldsymbol{s}_{n,k-1})) \prod_{\boldsymbol{s}\in\Lambda_n^*} p\left(z(\boldsymbol{s})|z(adjl\, \boldsymbol{s})\right), \ n \in \mathbb{N}. \end{aligned}$$

In the context of graphical models, this result can be found in, *inter alia*, Kiiveri, Speed and Carlin (1984) and Lauritzen and Spiegelhalter (1988), although it should be noted that their results do not deal with the edge effects, arising from $z(E_n)$, in as much generality. The asymptotic properties of maximum likelihood estimators developed below are given in the spatial setting.

We assume that, for $\boldsymbol{s} \in \mathbb{Z}^d \setminus L^*$, the conditional probability density $p(z(\boldsymbol{s})| z(adjl\, \boldsymbol{s}); \boldsymbol{\theta})$ depends on the vector of parameters $\boldsymbol{\theta} \in \Theta$, where Θ is an open connected subset of $\mathbb{R}^p$. Hence, for each $n \in \mathbb{N}$, the log-likelihood function of

$Z(\Lambda_n)$ is

$$l_n(\boldsymbol{\theta}) = \sum_{\boldsymbol{s}_{n,k}\in E_n} \log\left\{p\left(Z(\boldsymbol{s}_{n,k})|Z(\boldsymbol{s}_{n,1}),\ldots,Z(\boldsymbol{s}_{n,k-1});\boldsymbol{\theta}\right)\right\} + \sum_{\boldsymbol{s}\in\Lambda_n^*} \log\left\{p\left(Z(\boldsymbol{s})|Z(adjl\,\boldsymbol{s});\boldsymbol{\theta}\right)\right\}. \quad (3)$$

In practice, with only the parametric form of $p\left(z(\boldsymbol{s})|z(adjl\,\boldsymbol{s});\boldsymbol{\theta}\right)$ given, we may not know the parametric structure of $p\left(z(\boldsymbol{s}_{n,k})|z(\boldsymbol{s}_{n,1}),\ldots,z(\boldsymbol{s}_{n,k-1});\boldsymbol{\theta}\right)$ for $\boldsymbol{s}_{n,k}\in E_n$, $n\in I\!N$. So we consider the second part of the log-likelihood function, without edge components, given by

$$l_n^*(\boldsymbol{\theta}) = \sum_{\boldsymbol{s}\in\Lambda_n^*} \log\left\{p\left(Z(\boldsymbol{s})|Z(adjl\,\boldsymbol{s});\boldsymbol{\theta}\right)\right\}. \quad (4)$$

We call (4) a composite likelihood after Lindsay (1988). Note that, by (1), $|E_n|/|\Lambda_n| \to 0$ as $n\to\infty$; hence the first part of the log-likelihood function contains comparatively little information, as is borne out by our efficiency results for maximum composite likelihood estimators.

Suppose that $\boldsymbol{\theta}_0$ is the vector of true values of the parameters. For each $n\in I\!N$, let $\hat{\boldsymbol{\theta}}_n$ be a solution of the likelihood equation, $\partial l_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = \mathbf{0}$. We call $\hat{\boldsymbol{\theta}}_n$ the maximum likelihood estimator (MLE) if $\hat{\boldsymbol{\theta}}_n$ is the global maximum of $l_n(\boldsymbol{\theta})$. Further, let $\hat{\boldsymbol{\theta}}_n^*$ be a solution of the composite-likelihood equation, $\partial l_n^*(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = \mathbf{0}$. We call $\hat{\boldsymbol{\theta}}_n^*$ the maximum composite likelihood estimator (MCLE) if $\hat{\boldsymbol{\theta}}_n^*$ is the global maximum of $l_n^*(\boldsymbol{\theta})$.

For each $n\in I\!N$ and for $\boldsymbol{s}_{n,k}\in\Lambda_n$, $k=1,\ldots,|\Lambda_n|$, let

$$\begin{aligned}
\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,k};\boldsymbol{\theta}) &\equiv \left(b_1^{(n)}(\boldsymbol{s}_{n,k};\boldsymbol{\theta}),\ldots,b_p^{(n)}(\boldsymbol{s}_{n,k};\boldsymbol{\theta})\right)^T \\
&\equiv \frac{\partial}{\partial\boldsymbol{\theta}} \log\left\{p\left(Z(\boldsymbol{s}_{n,k})|Z(\boldsymbol{s}_{n,1}),\ldots,Z(\boldsymbol{s}_{n,k-1});\boldsymbol{\theta}\right)\right\}, \\
\boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,k};\boldsymbol{\theta}) &\equiv \left(A_{j,k}^{(n)}(\boldsymbol{s}_{n,k};\boldsymbol{\theta})\right)_{p\times p} \\
&\equiv \frac{\partial^2}{\partial\boldsymbol{\theta}^2} \log\left\{p\left(Z(\boldsymbol{s}_{n,k})|Z(\boldsymbol{s}_{n,1}),\ldots,Z(\boldsymbol{s}_{n,k-1});\boldsymbol{\theta}\right)\right\}.
\end{aligned}$$

Then, by the Mean Value Theorem, the likelihood equation can be written as

$$\frac{\partial l_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \sum_{\boldsymbol{s}\in\Lambda_n} \boldsymbol{b}^{(n)}(\boldsymbol{s};\boldsymbol{\theta}_0) + \sum_{\boldsymbol{s}\in\Lambda_n} \tilde{\boldsymbol{A}}^{(n)}(\boldsymbol{s};\boldsymbol{\theta}_0,\boldsymbol{\theta})(\boldsymbol{\theta}-\boldsymbol{\theta}_0) = \mathbf{0}, \quad n\in I\!N, \quad (5)$$

where $\tilde{\boldsymbol{A}}^{(n)}(\boldsymbol{s};\boldsymbol{\theta}_0,\boldsymbol{\theta})$ denotes $\boldsymbol{A}^{(n)}(\boldsymbol{s};\boldsymbol{\theta})$ with rows evaluated at possibly different points on the line segment between $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}$.

In the following subsections, asymptotic results are established for the M(C)LE. Proofs are given in the Appendix.

3.1. Consistency of the maximum (composite) likelihood estimator

Several approaches have been taken in proving the consistency and asymptotic normality of the generic MLE. Here, a martingale approach used by Crowder (1976) is adapted to the spatial models we are considering. For each $n \in \mathbb{N}$, let $\mathcal{F}_{n,0} \equiv \{\Omega, \emptyset\}$, and let $\mathcal{F}_{n,k} \equiv \sigma\{\omega : Z(\boldsymbol{s}_{n,1}), \ldots, Z(\boldsymbol{s}_{n,k})\}$ be the σ-algebra generated by $Z(\boldsymbol{s}_{n,1}), \ldots, Z(\boldsymbol{s}_{n,k})$, for $k = 1, \ldots, |\Lambda_n|$. Throughout the paper we assume that for all $\boldsymbol{\theta} \in \Theta$, the distributions $P_{\boldsymbol{\theta}}$ of $Z(\mathbb{Z}^d)$ have common support, and for $\boldsymbol{\theta} \in \Theta$, $\boldsymbol{s}_{n,k} \in \Lambda_n$, $n \in \mathbb{N}$, and $j = 1, 2$,

$$\begin{aligned} &\int \frac{\partial^j}{\partial \boldsymbol{\theta}^j} p\left(z(\boldsymbol{s}_{n,k}) | z(\boldsymbol{s}_{n,1}), \ldots, z(\boldsymbol{s}_{n,k-1}); \boldsymbol{\theta}\right) dz(\boldsymbol{s}_{n,k}) \\ &= \frac{\partial^j}{\partial \boldsymbol{\theta}^j} \int p\left(z(\boldsymbol{s}_{n,k}) | z(\boldsymbol{s}_{n,1}), \ldots, z(\boldsymbol{s}_{n,k-1}); \boldsymbol{\theta}\right) dz(\boldsymbol{s}_{n,k}) = 0. \end{aligned} \tag{6}$$

Note that (6) implies that the following conditions hold for all $\boldsymbol{s}_{n,k} \in \Lambda_n$, $n \in \mathbb{N}$:

$$\begin{aligned} E\left\{\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,k}; \boldsymbol{\theta}_0) \middle| \mathcal{F}_{n,k-1}\right\} &= \boldsymbol{0}, \\ E\left\{-\boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,k}; \boldsymbol{\theta}_0) \middle| \mathcal{F}_{n,k-1}\right\} &= \mathrm{var}\left\{\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,k}; \boldsymbol{\theta}_0) \middle| \mathcal{F}_{n,k-1}\right\}. \end{aligned}$$

From (6) we have the following result.

Proposition 1. *Consider the log-likelihood function given by* (3)*, and suppose the true value of the parameter is* $\boldsymbol{\theta}_0$. *For each* $n \in \mathbb{N}$, $k = 1, 2, \ldots, |\Lambda_n|$, *let* $\boldsymbol{M}_{n,k} \equiv \sum_{i=1}^{k} \boldsymbol{b}(\boldsymbol{s}_{n,i}; \boldsymbol{\theta}_0)$ *and* $\boldsymbol{W}_{n,k} \equiv \sum_{i=1}^{k} \{\boldsymbol{A}(\boldsymbol{s}_{n,i}; \boldsymbol{\theta}_0) - E(\boldsymbol{A}(\boldsymbol{s}_{n,i}; \boldsymbol{\theta}_0) | \mathcal{F}_{n,i-1})\}$, *where the* $\{\boldsymbol{s}_{n,i}\}$ *are given by* (2). *Then* $\{\boldsymbol{M}_{n,k}, \ \mathcal{F}_{n,k}\}$ *and* $\{\boldsymbol{W}_{n,k}, \ \mathcal{F}_{n,k}\}$ *are triangular martingale arrays.*

The Weak Law of Large Number for a martingale array is given next.

Proposition 2. *Assume that* $\{(S_{n,i}, \ \mathcal{F}_{n,i}) : n \in \mathbb{N}, \ i = 1, \ldots, k_n\}$ *is a triangular martingale array, where* $k_n \to \infty$ *as* $n \to \infty$. *Let* $X_{n,1} = S_{n,1}$, *and* $X_{n,i} = S_{n,i} - S_{n,i-1}$, $n \in \mathbb{N}$, $i = 2, \ldots, k_n$. *Suppose that*

(i) $\displaystyle\sum_{i=1}^{k_n} P\left(|X_{n,i}| > k_n\right) \to 0, \text{ as } n \to \infty;$

(ii) $\displaystyle\frac{1}{k_n} \sum_{i=1}^{k_n} E\Big(X_{n,i} I(|X_{n,i}| \le k_n) \Big| \mathcal{F}_{n,i-1}\Big) \xrightarrow{p} 0, \text{ as } n \to \infty;$

(iii) $\frac{1}{k_n^2}\sum_{i=1}^{k_n}\left\{E\left(X_{n,i}I(|X_{n,i}|\le k_n)\right)^2 - E\Big(E\left(X_{n,i}I(|X_{n,i}|\le k_n)|\mathcal{F}_{n,i-1}\right)\Big)^2\right\}\to 0$, as $n\to\infty$.

Then as $n\to\infty$, $S_{n,k_n}/k_n \xrightarrow{p} 0$.

Corollary 1. *Assume that $\{(S_{n,i},\ \mathcal{F}_{n,i}) : n\in I\!N,\ i=1,\ldots,k_n\}$ is a triangular martingale array, where $k_n\to\infty$ as $n\to\infty$. Let $X_{n,1}=S_{n,1}$ and $X_{n,i}=S_{n,i}-S_{n,i-1}$, $n\in I\!N$, $i=2,\ldots,k_n$. If $\sup_{n,i}E\,|X_{n,i}|^{1+\delta}<\infty$ for some $\delta>0$, then as $n\to\infty$, $S_{n,k_n}/k_n\xrightarrow{p}0$.*

We are now able to establish the consistency of the MLE for POMMs.

Theorem 1. *Consider the log-likelihood function given by* (3) *and suppose the true value of the parameter is $\boldsymbol{\theta}_0$. Assume*

(A.1) $$\sup_{|\Lambda_n|\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le 1}\frac{1}{|\Lambda_n|}\sum_{i=1}^{|\Lambda_n|}\left\{\boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta})-\boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,i},\boldsymbol{\theta}_0)\right\}\xrightarrow{p}\boldsymbol{0},\ \text{as } n\to\infty;$$

(A.2) $\sup_{n,i,j}E\left(\left|b_j^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)\right|^q\right)<\infty$, *for some* $q>1$;

(A.3) $P\left\{\frac{-1}{|\Lambda_n|}\sum_{i=1}^{|\Lambda_n|}\boldsymbol{c}^T\boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)\boldsymbol{c}>\varepsilon\right\}\to 1$, *as $n\to\infty$ for all $\|\boldsymbol{c}\|=1$ and for some $\varepsilon>0$.*

Then there exists a solution $\hat{\boldsymbol{\theta}}_n$ of the likelihood equation (5) *such that, as $n\to\infty$, $\hat{\boldsymbol{\theta}}_n\xrightarrow{p}\boldsymbol{\theta}_0$. If in addition*

(A.4) $$P\left\{\frac{-1}{|\Lambda_n|}\sum_{i=1}^{|\Lambda_n|}\boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta})\ \textit{is positive-definite for all}\ \boldsymbol{\theta}\in\Theta\right\}\to 1,\ \text{as } n\to\infty,$$

then with probability tending to 1, *$\hat{\boldsymbol{\theta}}_n$ is the MLE.*

In practice, the second part of the log-likelihood function, given by (4), is all that we can use for inference. For each $n\in I\!N$, let $\{\boldsymbol{s}^*_{n,k}=\boldsymbol{s}_{n,m_k}:k=1\ldots,|\Lambda^*_n|\}$ be the subsequence of $\{\boldsymbol{s}_{n,1},\ldots,\boldsymbol{s}_{n,|\Lambda_n|}\}$ such that $\Lambda^*_n=\{\boldsymbol{s}^*_{n,1},\boldsymbol{s}^*_{n,2},\ldots,\boldsymbol{s}^*_{n,|\Lambda^*_n|}\}$, and define $\mathcal{F}^*_{n,0}\equiv\{\Omega,\emptyset\}$ and $\mathcal{F}^*_{n,k}\equiv\mathcal{F}_{n,m_k}$, $k=1,\ldots,|\Lambda^*_n|$. Then $p(z(\boldsymbol{s}^*_{n,k})|\mathcal{F}^*_{n,k-1})=p(z(\boldsymbol{s}^*_{n,k})|z(adjl\ \boldsymbol{s}^*_{n,k}))$, $\boldsymbol{s}^*_{n,k}\in\Lambda^*_n$, $k=1,\ldots,|\Lambda^*_n|$. Similar to Theorem 1, we have the following result for the MCLEs.

Corollary 2. *Consider the log-likelihood function given by* (3) *and suppose the true value of the parameter is $\boldsymbol{\theta}_0$. Assume*

(A.1′) $$\sup_{|\Lambda^*_n|\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le 1}\frac{1}{|\Lambda^*_n|}\sum_{i=1}^{|\Lambda^*_n|}\left\{\boldsymbol{A}^{(n)}(\boldsymbol{s}^*_{n,i};\boldsymbol{\theta})-\boldsymbol{A}^{(n)}(\boldsymbol{s}^*_{n,i},\boldsymbol{\theta}_0)\right\}\xrightarrow{p}\boldsymbol{0},\ \text{as } n\to\infty;$$

(A.2′) $\sup_{n,i,j} E\left(\left|b_j^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}_0)\right|^q\right) < \infty,$ for some $q > 1$;

(A.3′) $P\left\{\frac{-1}{|\Lambda_n^*|}\sum_{i=1}^{|\Lambda_n^*|}\boldsymbol{c}^T\boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}_0)\boldsymbol{c} > \varepsilon\right\} \to 1$ as $n \to \infty$, for all $\|\boldsymbol{c}\| = 1$ and for some $\varepsilon > 0$.

Then there exists a solution $\hat{\boldsymbol{\theta}}_n^*$ *of the composite-likelihood equation* $\partial l_n^*(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = \boldsymbol{0}$ *such that, as* $n \to \infty$, $\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0$. If in addition

(A.4′) $P\left\{\frac{-1}{|\Lambda_n^*|}\sum_{i=1}^{|\Lambda_n^*|}\boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}) \text{ is positive-definite for all } \boldsymbol{\theta}\in\Theta\right\}\to 1,$ as $n\to\infty$,

then with probability tending to 1, $\hat{\boldsymbol{\theta}}_n^*$ *is the maximum composite likelihood estimator.*

3.2. Asymptotic normality of the maximum (composite) likelihood estimator

The asymptotic normality of MLEs for POMMs is established in the following theorem.

Theorem 2. *Assume the conditions of Theorem* 1 *hold except that* (A.2) *holds for some* $q > 2$. *Define* $\boldsymbol{B}_n \equiv var(\sum_{i=1}^{|\Lambda_n|}\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0))$, $n \in I\!N$. *Assume*

(A.5) $\frac{1}{\boldsymbol{c}'\boldsymbol{B}_n\boldsymbol{c}}\sum_{i=1}^{|\Lambda_n|}E\left(\left(\boldsymbol{c}^T\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)\right)^2\middle|\mathcal{F}_{n,i-1}\right)\xrightarrow{p} 1,$ as $n\to\infty$, for all $\|\boldsymbol{c}\|=1$.

Then as $n \to \infty$, $\boldsymbol{B}_n^{1/2}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{I}_p)$.

Note that $\boldsymbol{B}_n$ in the statement of Theorem 2 is actually the Fisher information matrix of $\boldsymbol{\theta}_0$ contained in $Z(\Lambda_n)$. Further, we have as $n \to \infty$,

$$\begin{aligned}
&\boldsymbol{B}_n^{1/2}\left\{\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 - \boldsymbol{B}_n^{-1}\sum_{i=1}^{|\Lambda_n|}\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)\right\}\\
&=\boldsymbol{B}_n^{1/2}\left\{-\sum_{i=1}^{|\Lambda_n|}\tilde{\boldsymbol{A}}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0,\hat{\boldsymbol{\theta}}_n)\right\}^{-1}\left\{\sum_{i=1}^{|\Lambda_n|}\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)\right\}-\boldsymbol{B}_n^{-1/2}\sum_{i=1}^{|\Lambda_n|}\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)\\
&=\left\{\boldsymbol{B}_n^{1/2}\left(-\sum_{i=1}^{|\Lambda_n|}\tilde{\boldsymbol{A}}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0,\hat{\boldsymbol{\theta}}_n)\right)^{-1}\boldsymbol{B}_n^{1/2}-\boldsymbol{I}_p\right\}\left\{\boldsymbol{B}_n^{-1/2}\sum_{i=1}^{|\Lambda_n|}\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)\right\}\xrightarrow{p}\boldsymbol{0}.
\end{aligned}$$

Therefore, $\hat{\boldsymbol{\theta}}_n$ is also asymptotically efficient (see Basawa and Rao (1980, Section 7.2.4); Rao (1973, Section 5c.2)).

Corollary 3. *Assume the conditions of Theorem* 2. *Let* $\hat{\boldsymbol{B}}_n \equiv -\sum_{i=1}^{|\Lambda_n|} \boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,i}; \hat{\boldsymbol{\theta}}_n)$ *for all* $n \in I\!N$. *Then as* $n \to \infty$, $\hat{\boldsymbol{B}}_n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{I}_p)$, *and a* $100(1-\alpha)\%$ *asymptotic confidence region of* $\boldsymbol{\theta}_0$ *is given by* $\{\boldsymbol{\theta} : (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n)^T \hat{\boldsymbol{B}}_n(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n) \le \chi^2_{p;1-\alpha}\}$, *where* $\chi^2_{p;1-\alpha}$ *is the* $100(1-\alpha)$ *percentile of a* χ^2_p *distribution.*

Similar to Theorem 2, we have the following results for the MCLEs.

Corollary 4. *Assume that the conditions of Corollary* 2 *hold except that* (A.2′) *holds for some* $q > 2$. *Define* $\boldsymbol{B}_n^* \equiv var(\sum_{i=1}^{|\Lambda_n^*|} \boldsymbol{b}^{(n)}(\boldsymbol{s}^*_{n,i}; \boldsymbol{\theta}_0))$, $n \in I\!N$. *Assume*

$$\text{(A.5}'\text{)}\quad \frac{1}{\boldsymbol{c}'\boldsymbol{B}_n^*\boldsymbol{c}} \sum_{i=1}^{|\Lambda_n^*|} E\Big(\Big(\boldsymbol{c}^T\boldsymbol{b}^{(n)}(\boldsymbol{s}^*_{n,i}; \boldsymbol{\theta}_0)\Big)^2 \Big| \mathcal{F}^*_{n,i-1}\Big) \xrightarrow{p} 1, \text{ as } n \to \infty, \text{ for all } \|\boldsymbol{c}\| = 1.$$

Then as $n \to \infty$, $(\boldsymbol{B}_n^*)^{1/2}(\hat{\boldsymbol{\theta}}^*_n - \boldsymbol{\theta}_0) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{I}_p)$. *Moreover,* $\hat{\boldsymbol{\theta}}_n$ *is asymptotically efficient.*

3.3. Mixing conditions sufficient for consistency and asymptotic normality of the maximum composite likelihood estimator

In this section, we show that under a certain mixing condition a solution $\hat{\boldsymbol{\theta}}^*_n$ of the composite-likelihood equation, $\partial l_n^*(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = \boldsymbol{0}$, is consistent and asymptotically normal. We also introduce Dobrushin's condition, which can be used to ensure this mixing condition.

Consider a random field $Z(D)$, $D \subset Z\!\!Z^d$. Let $\mathcal{F}^Z_B$ denote the σ-algebra generated by $Z(B)$ for $B \subset D$. The strong-mixing coefficients for $Z(D)$ are defined as

$$\alpha^Z_{m,l}(n) \equiv \sup\{\alpha_Z(B_1, B_2) :\ |B_1| \le m,\ |B_2| \le l,\ d(B_1, B_2) \ge n,\ B_1, B_2 \subset D\}, \tag{7}$$

where $n \in I\!N$, $m, l \in I\!N \cup \{\infty\}$, and $\alpha_Z(B_1, B_2) \equiv \sup\{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_1 \in \mathcal{F}^Z_{B_1},\ A_2 \in \mathcal{F}^Z_{B_2}\}$. For $m, l \in I\!N$, let $m \wedge l \equiv \min\{m, l\}$. We use the following Proposition (Doukhan (1994, Chapter 1, Theorem 1)).

Proposition 3. *Consider a random field* $Z(D)$, $D \subset Z\!\!Z^d$. *Assume there is a* $\delta > 0$ *such that* $E\Big(|Z(\boldsymbol{s})|^{2+\delta}\Big) < \infty$ *and* $E\left(Z(\boldsymbol{s})\right) = 0$, *for* $\boldsymbol{s} \in D$. *If* $\sum_{n=1}^{\infty}(n+1)^{2d-1}[\alpha^Z_{1,1}(n)]^{\delta/(2+\delta)} < \infty$, *then there is a constant* $C > 0$, *depending only on* $\{\alpha^Z_{1,1}(n) : n \in I\!N\}$, *such that* $E(|\sum_{\boldsymbol{s}\in B} Z(\boldsymbol{s})|^2) \le C \sum_{\boldsymbol{s}\in B}\{E(|Z(\boldsymbol{s})|^{2+\delta})\}^{2/(2+\delta)}$ *for any finite set* $B \subset D$.

Applying the Markov inequality, we have the following.

Corollary 5. *Suppose the assumptions of Proposition* 3 *hold. Let* $\{B_n : n \in I\!N\}$ *be an increasing sequence of finite subsets of* D *such that* $|B_n| \to \infty$*, as* $n \to \infty$*. If* $\sup_{\boldsymbol{s} \in \cup B_n} E(|Z(\boldsymbol{s})|^{2+\delta}) < \infty$*, then as* $n \to \infty$*,* $\frac{1}{|B_n|} \sum_{\boldsymbol{s} \in B_n} Z(\boldsymbol{s}) \xrightarrow{p} 0$.

Dobrushin's condition for spatial mixing is given as follows.

Definition 4. Consider a random field $Z(D)$, $D \subset \mathbb{Z}^d$. Let $\pi^Z_{\boldsymbol{s}}(\cdot|\boldsymbol{x})$ denote the conditional probability measure of $Z(\boldsymbol{s})$ given that $Z(D \setminus \{\boldsymbol{s}\}) = \boldsymbol{x}$, and let $\|\nu\|_{\text{var}}$ denote the total variation norm of a signed measure ν. For $\boldsymbol{s} \in D$, $\boldsymbol{t} \in D \setminus \{\boldsymbol{s}\}$, let $\gamma^Z_{\boldsymbol{s},\boldsymbol{t}} \equiv \frac{1}{2} \sup_{\boldsymbol{x},\tilde{\boldsymbol{x}}} \|\pi^Z_{\boldsymbol{s}}(\cdot|\boldsymbol{x}) - \pi^Z_{\boldsymbol{s}}(\cdot|\tilde{\boldsymbol{x}})\|_{\text{var}}$, where the sup is taken over all configurations $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$ identical except at site $\boldsymbol{t}$. Then a random field $Z(D)$ is said to satisfy Dobrushin's condition (Dobrushin (1968)) if $\Gamma \equiv \sup_{\boldsymbol{s} \in D} \sum_{\boldsymbol{t} \in D \setminus \{\boldsymbol{s}\}} \gamma^Z_{\boldsymbol{s},\boldsymbol{t}} < 1$.

We need one more definition before we establish the lemmas that are used to prove Theorem 3 below.

Definition 5. A random field $Z(D)$, $D \subset \mathbb{Z}^d$, is said to be k-Markovian, where $k > 0$ is an integer, if $p\,(z(V)|z(U)) = p\,(z(V)|z((V)_k))$ for any finite set $V \subset D$, and for any finite set U such that $(V)_k \subset U \subset D \setminus V$, where $(V)_k \equiv \{\boldsymbol{s} \in D \setminus V : d(\boldsymbol{s}, V) \leq k\}$.

Under Dobrushin's condition, we can prove the following two results for strong-mixing coefficients. The first one can be found in Guyon (1995, Theorem 2.1.3) and Doukhan (1994, Section 2.2.2), and the second follows directly from the first (see the Appendix).

Lemma 1. *Assume a random field* $Z(D)$*,* $D \subset \mathbb{Z}^d$*, satisfies Dobrushin's condition. If* $Z(D)$ *is* k*-Markovian, then there exist positive constants* C_1 *and* C_2 *such that for any* $n \in I\!N$ *and* $m, l \in I\!N \cup \{\infty\}$*,* $\alpha^Z_{m,l}(n) \leq C_1(m \wedge l)\exp(-C_2 n)$*, where* $\alpha^Z_{m,l}(n)$ *is given by* (7).

Lemma 2. *Assume a random field* $Z(D)$*,* $D \subset \mathbb{Z}^d$*, satisfies Dobrushin's condition and is* k*-Markovian. For each* $\boldsymbol{s} \in D$*, let* $W(\boldsymbol{s}) \equiv f_{\boldsymbol{s}}\,(Z(V_{\boldsymbol{s}}))$ *for some measurable functions* $f_{\boldsymbol{s}} : I\!R^{|V|} \to I\!R$*, where* $V_{\boldsymbol{s}} \equiv (V + \boldsymbol{s}) \cap D$*. Then there exist positive constants* C_3 *and* C_4 *such that for any* $n \in I\!N$ *and* $m, l \in I\!N \cup \{\infty\}$*,* $\alpha^W_{m,l}(n) \leq C_3(m \wedge l)\exp(-C_4 n)$*, where* $\alpha^W_{m,l}(n)$ *is defined by* (7).

The asymptotic normality of MCLEs for POMMs can be established under some regularity conditions, including Dobrushin's condition, which are relatively easy to check.

Theorem 3. *Let* $Z(\mathbb{Z}^d)$ *be a POMM with lower set* L^**. Assume* (A.1′) *holds and* (A.2′) *holds for some* $q > 4$*. Suppose the random field* $Z(\mathbb{Z}^d \setminus L^*)$ *satisfies*

Dobrushin's condition, $\sup\{diam\,(adjl\,\boldsymbol{s} \cup \{\boldsymbol{s}\}) : \boldsymbol{s} \in \mathbb{Z}^d \setminus L^*\} \le k < \infty$, *and*

$$\liminf_{n\to\infty} \frac{1}{|\Lambda_n^*|} \boldsymbol{c}^T \boldsymbol{B}_n^* \boldsymbol{c} > 0, \tag{8}$$

for all $\|\boldsymbol{c}\| = 1$, *where* $\boldsymbol{B}_n^* \equiv var(\sum_{i=1}^{|\Lambda_n^*|} \boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i}^*; \boldsymbol{\theta}_0))$, $n \in \mathbb{N}$. *Then*

(i) *There exists a solution* $\hat{\boldsymbol{\theta}}_n^*$ *of the composite-likelihood equation,* $\partial l_n^*(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = \boldsymbol{0}$, *such that as* $n \to \infty$, $\hat{\boldsymbol{\theta}}_n^* \xrightarrow{p} \boldsymbol{\theta}_0$.

(ii) *As* $n \to \infty$, $(\boldsymbol{B}_n^*)^{1/2}\left(\hat{\boldsymbol{\theta}}_n^* - \boldsymbol{\theta}_0\right) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{I}_p)$, *and* $\hat{\boldsymbol{\theta}}_n$ *is asymptotically efficient.*

(iii) *If, in addition,* (A.4′) *holds, then with probability tending to* 1, $\hat{\boldsymbol{\theta}}_n^*$ *is the maximum composite likelihood estimator.*

4. Example: Conditional Binomial Distributions

Consider the poset $(\mathbb{Z}^2, \prec)$ with partial order defined by the adjacent lower neighbors as follows:

$$adjl\,(u,v) = \{(u-1,v),(u-1,v-1),(u,v-1),(u+1,v-1)\}, \quad (u,v)^T \in \mathbb{Z}^2.$$

This type of dependence was used by Cressie and Davidson (1998) and Davidson *et al.* (1999) for the analysis of textures on a finite, rectangular array of pixels. Suppose that the POMM is defined by

$$Z(u,v)|Z(adjl\,(u,v)) \sim Bin\left(G-1,\ \frac{\exp(T_{u,v}(\boldsymbol{\theta}))}{1+\exp(T_{u,v}(\boldsymbol{\theta}))}\right),$$

where $G \in \{2,3,\ldots\}$ is the number of grey levels, and $T_{u,v}(\boldsymbol{\theta}) \equiv \boldsymbol{\theta}^T\boldsymbol{H}(u,v)$ with the vector of parameters $\boldsymbol{\theta} \equiv (\beta_0,\beta_1,\beta_2,\beta_3,\beta_4)^T$ and $\boldsymbol{H}(u,v) \equiv (1, Z(u-1,v), Z(u-1,v-1), Z(u,v-1), Z(u+1,v-1))^T$. Let $\Lambda_n = \{(u,v) : 0 \le u \le n,\ 0 \le v \le n,\ u,v \in \mathbb{Z}\}$, $n \in \mathbb{N}$. Then $\{\Lambda_n : n \in \mathbb{N}\}$ is a sequence of finite subsets of $\mathbb{Z}^2$ that satisfy (1) with $\Lambda_n^* \equiv \{\boldsymbol{s} \in \Lambda_n : adjl\,\boldsymbol{s} \subset \Lambda_n\} = \{(u,v) : 1 \le u \le n-1,\ 1 \le v \le n,\ u,v \in \mathbb{Z}\}$, $n \in \mathbb{N}$. Hence, (4) can be written as

$$l_n^*(\boldsymbol{\theta}) = \sum_{(u,v)\in\Lambda_n^*} \left\{\log\left(C_{Z(u,v)}^{G-1}\right) + Z(u,v)T_{u,v}(\boldsymbol{\theta}) - (G-1)\log\left(1+\exp(T_{u,v}(\boldsymbol{\theta}))\right)\right\},$$

where $C_{Z(u,v)}^{G-1} \equiv (G-1)!/\{(Z(u,v))!(G-1-Z(u,v))!\}$. The first and second partial derivatives of $p\,(z(u,v)|z(adjl\,(u,v));\boldsymbol{\theta})$ can be calculated as

$$\boldsymbol{b}^{(n)}(u,v;\boldsymbol{\theta}) = \left(Z(u,v) - (G-1)\frac{\exp(T_{u,v}(\boldsymbol{\theta}))}{1+\exp(T_{u,v}(\boldsymbol{\theta}))}\right)\boldsymbol{H}(u,v), \quad (u,v)^T \in \mathbb{Z}^2,$$

$$\boldsymbol{A}^{(n)}(u,v;\boldsymbol{\theta}) = \frac{-(G-1)\exp(T_{u,v}(\boldsymbol{\theta}))}{(1+\exp(T_{u,v}(\boldsymbol{\theta})))^2}\boldsymbol{H}(u,v)\boldsymbol{H}(u,v)^T, \quad (u,v)^T \in \mathbb{Z}^2.$$

To prove the consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}_n^*$, we use Theorem 3. It is easy to check that all the assumptions are satisfied if Dobrushin's condition holds. Not surprisingly, further restrictions are needed on the spatial-dependence parameters for the model to satisfy Dobrushin's condition. An example for $G = 2$ is displayed in Figure 1, which shows the values of $\Gamma_{\boldsymbol{\theta}}$ for different values of $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)^T$, where $\Gamma_{\boldsymbol{\theta}}$ is defined in Definition 4. Recall that Dobrushin's condition is satisfied if and only if $\Gamma_{\boldsymbol{\theta}} < 1$ and notice how this region changes as certain parameters are varied while others are held fixed.

As a result of Theorem 3, we may conclude that the maximum composite likelihood estimator $\hat{\boldsymbol{\theta}}_n^*$ is consistent, asymptotically normal, and asymptotically efficient, provided Dobrushin's condition is satisfied.

Acknowledgements

This research was supported by the Office of Naval Research under grants N00014-93-1-0001 and N00014-99-1-0214, and was partially carried out at Iowa State University.

Appendix

This appendix contains proofs of the results stated in Section 3.

Proof of Proposition 2. Let $S'_{n,k_n} = \sum_{i=1}^{k_n} X_{n,i} I(|X_{n,i}| \leq k_n)$, $n \in I\!N$. Since $P(S_{n,k_n} \neq S'_{n,k_n}) \leq \sum_{i=1}^{k_n} P(|X_{n,i}| > k_n) \to 0$ as $n \to \infty$, we only have to show that $S'_{n,k_n}/k_n \xrightarrow{p} 0$ as $n \to \infty$. On account of (ii), it suffices to prove that as $n \to \infty$,

$$\frac{1}{k_n} \sum_{i=1}^{k_n} \Big\{ X_{n,i} I(|X_{n,i}| \leq k_n) - E\Big(X_{n,i} I(|X_{n,i}| \leq k_n) | \mathcal{F}_{n,i-1} \Big) \Big\} \xrightarrow{p} 0.$$

This follows from (iii) by using Chebyshev's inequality.

Proof of Corollary 1. It is sufficient to show that Conditions (i), (ii), and (iii) of Proposition 2 hold. First, note that $\sup_{n,i} E|X_{n,i}|^{1+\delta} < \infty$ for some $\delta > 0$ implies that $\{X_{n,i} : n \in I\!N,\ i = 1, \ldots, k_n\}$ is uniformly integrable. So as $n \to \infty$,

$$\sum_{i=1}^{k_n} P\left(|X_{n,i}| > k_n\right) \leq \sup_i k_n P\left(|X_{n,i}| > k_n\right) \leq \sup_i E\Big(|X_{n,i}| I\left(|X_{n,i}| > k_n\right) \Big) \to 0,$$

and

$$P\Big(\Big| \frac{1}{k_n} \sum_{i=1}^{k_n} E\Big(X_{n,i} I\left(|X_{n,i}| \leq k_n\right) \Big| \mathcal{F}_{n,i-1} \Big) \Big| > \varepsilon \Big) \leq \frac{1}{\varepsilon k_n} \sum_{i=1}^{k_n} E\Big(|X_{n,i}| I\left(|X_{n,i}| > k_n\right) \Big)$$

$$\leq \frac{1}{\varepsilon} \sup_i E\Big(|X_{n,i}| I\left(|X_{n,i}| > k_n\right) \Big) \to 0.$$

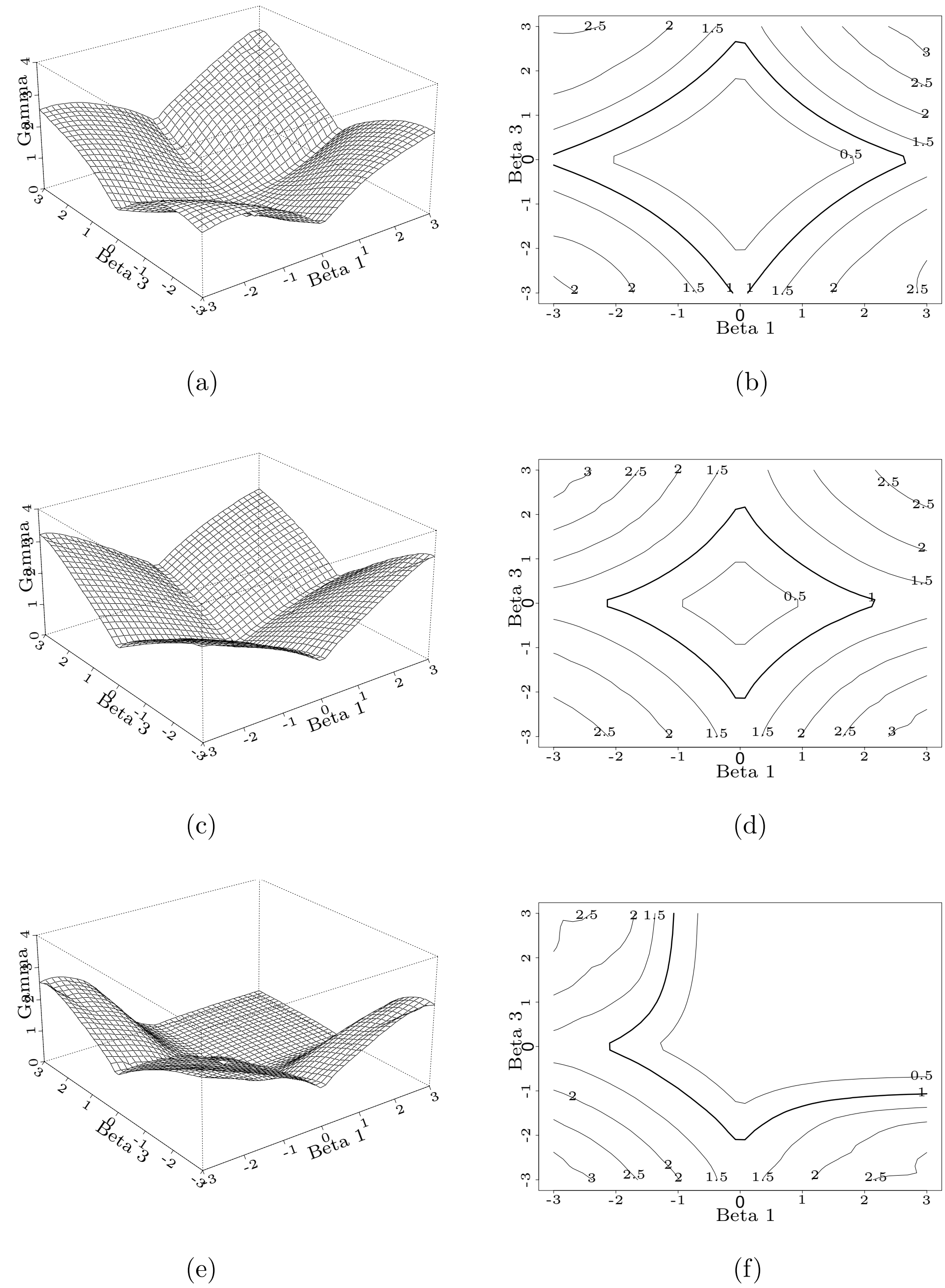

Figure 1. Three-dimensional plots and contour plots of $\Gamma_{\boldsymbol{\theta}}$ (where $\Gamma_{\boldsymbol{\theta}}$ is defined in Definition 4, except that now the parameter $\boldsymbol{\theta}$ is featured) versus β_1 and β_3. Dobrushin's condition is satisfied if and only if $\Gamma_{\boldsymbol{\theta}} < 1$; the contour $\Gamma_{\boldsymbol{\theta}} = 1$ is made bold in the plots. (a), (b): $\beta_0 = -3,\ \beta_2 = 0,\ \beta_4 = 0$; (c), (d): $\beta_0 = 0,\ \beta_2 = 0,\ \beta_4 = 0$; (e), (f): $\beta_0 = 3,\ \beta_2 = 0,\ \beta_4 = 0$.

This establishes Conditions (i) and (ii) of Proposition 2. Now

$$\begin{aligned}
&\frac{1}{k_n^2}\sum_{i=1}^{k_n}\Big\{E\Big(X_{n,i}I(|X_{n,i}|\le k_n)\Big)^2 - E\Big(E(X_{n,i}I(|X_{n,i}|\le k_n)|\mathcal{F}_{n,i-1}))\Big)\Big\}\\
&\le \frac{1}{k_n^2}\sum_{i=1}^{k_n}\Big\{E\Big(X_{n,i}I(|X_{n,i}|\le k_n)\Big)^2\Big\}\\
&\le \frac{1}{k_n^2}\sum_{i=1}^{k_n}\Big\{k_n^{1-\delta}E\Big(X_{n,i}I(|X_{n,i}|\le k_n)\Big)^{1+\delta}\Big\}\\
&\le \frac{1}{k_n^\delta}\sup_{n,i}E\,|X_{n,i}|^{1+\delta}\to 0, \quad \text{as } n\to\infty,
\end{aligned}$$

so Condition (iii) of Proposition 2 is satisfied.

Proof of Theorem 1. By Proposition 1 and Corollary 1, Assumption (A.2) implies that, as $n\to\infty$,

$$\frac{1}{|\Lambda_n|}\sum_{i=1}^{|\Lambda_n|}\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)\xrightarrow{p}\boldsymbol{0}, \tag{9}$$

where $\{\boldsymbol{s}_{n,i}\}$ is given by (2). Recall from (5) that for each $n\in I\!N$,

$$\begin{aligned}
\frac{\partial l_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} &= \sum_{i=1}^{|\Lambda_n|}\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)+\sum_{i=1}^{|\Lambda_n|}\Big\{\tilde{\boldsymbol{A}}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0,\boldsymbol{\theta})-\boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)\Big\}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\\
&\quad+\sum_{i=1}^{|\Lambda_n|}\boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0).
\end{aligned}$$

Pre-multiplying by $\frac{1}{|\Lambda_n|}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T$, then (9) and Assumptions (A.1) and (A.3) together imply there exists $\Delta>0$ such that, as $n\to\infty$, $P\{\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|=\delta}\frac{1}{|\Lambda_n|}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\frac{\partial l_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}<0\}\to 1$, for any $0<\delta<\Delta$. It follows from Lemma 2 of Aitchison and Silvey (1958) that as $n\to\infty$, $P\{\frac{\partial l_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$ has a zero at $\hat{\boldsymbol{\theta}}_n$ such that $\|\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\|<\delta\}\to 1$, for any $0<\delta<\Delta$. Therefore, $\hat{\boldsymbol{\theta}}_n\xrightarrow{p}\boldsymbol{\theta}_0$ as $n\to\infty$. Moreover, if Assumption (A.4) holds, then with probability tending to 1, $\hat{\boldsymbol{\theta}}_n$ is the point of the global maximum of $l_n(\boldsymbol{\theta})$, and hence the MLE.

Proof of Theorem 2. From Assumption (A.3), we have

$$\liminf_{n\to\infty}\left\{\inf_{\|\boldsymbol{c}\|=1}\boldsymbol{c}^T\left(\frac{1}{|\Lambda_n|}\boldsymbol{B}_n\right)\boldsymbol{c}\right\}\ge\varepsilon, \tag{10}$$

for some $\varepsilon > 0$. It follows that $\boldsymbol{B}_n$ is positive-definite for large n. Moreover, Assumptions (A.1) and (A.3) imply that the inverse of $\sum_{\Lambda_n} \tilde{\boldsymbol{A}}^{(n)}(\boldsymbol{s};\boldsymbol{\theta}_0,\hat{\boldsymbol{\theta}}_n)$ exists with probability tending to 1. Therefore, with probability tending to 1,

$$\begin{aligned}
&\boldsymbol{B}_n^{1/2}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)\\
&=\boldsymbol{B}_n^{1/2}\Big\{-\sum_{i=1}^{|\Lambda_n|}\tilde{\boldsymbol{A}}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0,\hat{\boldsymbol{\theta}}_n)\Big\}^{-1}\Big\{\sum_{i=1}^{|\Lambda_n|}\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)\Big\}\\
&=\Big\{\boldsymbol{B}_n^{-1/2}\Big(-\sum_{i=1}^{|\Lambda_n|}\tilde{\boldsymbol{A}}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0,\hat{\boldsymbol{\theta}}_n)\Big)\boldsymbol{B}_n^{-1/2}\Big\}^{-1}\Big\{\boldsymbol{B}_n^{-1/2}\sum_{i=1}^{|\Lambda_n|}\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)\Big\}.
\end{aligned}$$

Applying Slutsky's Theorem it is enough to show that, as $n\to\infty$,

$$\Big\{\boldsymbol{B}_n^{-1/2}\Big(-\sum_{i=1}^{|\Lambda_n|}\tilde{\boldsymbol{A}}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0,\hat{\boldsymbol{\theta}}_n)\Big)\boldsymbol{B}_n^{-1/2}\Big\}^{-1}\xrightarrow{p}\boldsymbol{I}_p, \tag{11}$$

$$\boldsymbol{B}_n^{-1/2}\sum_{i=1}^{|\Lambda_n|}\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)\xrightarrow{d}N(\boldsymbol{0},\boldsymbol{I}_p). \tag{12}$$

It follows from (10) that the largest eigenvalue of $(\frac{1}{|\Lambda_n|}\boldsymbol{B}_n)^{-1/2}$ is bounded above as n tends to ∞. Therefore by Assumptions (A.1) and (A.5), Proposition 1 and Corollary 1,

$$\begin{aligned}
&\boldsymbol{B}_n^{-1/2}\Big\{-\sum_{i=1}^{|\Lambda_n|}\tilde{\boldsymbol{A}}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0,\hat{\boldsymbol{\theta}}_n)\Big\}\boldsymbol{B}_n^{-1/2}-\boldsymbol{I}_p\\
&=\boldsymbol{B}_n^{-1/2}\Big\{-\sum_{i=1}^{|\Lambda_n|}\tilde{\boldsymbol{A}}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0,\hat{\boldsymbol{\theta}}_n)-\boldsymbol{B}_n\Big\}\boldsymbol{B}_n^{-1/2}\\
&=\Big(\frac{1}{|\Lambda_n|}\boldsymbol{B}_n\Big)^{-1/2}\Big\{\frac{-1}{|\Lambda_n|}\sum_{i=1}^{|\Lambda_n|}\Big(\tilde{\boldsymbol{A}}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0,\hat{\boldsymbol{\theta}}_n)-\boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)\Big)\Big\}\Big(\frac{1}{|\Lambda_n|}\boldsymbol{B}_n\Big)^{-1/2}\\
&\quad+\Big(\frac{1}{|\Lambda_n|}\boldsymbol{B}_n\Big)^{-1/2}\Big\{\frac{1}{|\Lambda_n|}\sum_{i=1}^{|\Lambda_n|}\Big(\boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)-E\Big(\boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)\Big|\mathcal{F}_{n,i-1}\Big)\Big)\Big\}\\
&\quad\cdot\Big(\frac{1}{|\Lambda_n|}\boldsymbol{B}_n\Big)^{-1/2}+\boldsymbol{B}_n^{-1/2}\Big\{-\sum_{i=1}^{|\Lambda_n|}E\Big(\boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)\Big|\mathcal{F}_{n,i-1}\Big)-\boldsymbol{B}_n\Big\}\boldsymbol{B}_n^{-1/2}\xrightarrow{p}\boldsymbol{0},\\
&\qquad\qquad\text{as } n\to\infty.
\end{aligned}$$

This gives (11). It remains to prove (12). It is not difficult to show that Assumption (A.2) with $q > 2$ implies the following Lindeberg condition:

$$\frac{1}{\boldsymbol{c}'\boldsymbol{B}_n\boldsymbol{c}}\sum_{i=1}^{|\Lambda_n|} E\left(\left|\boldsymbol{c}^T\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)\right|^2 I\left(\left|\boldsymbol{c}^T\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)\right|^2 \geq \varepsilon(\boldsymbol{c}'\boldsymbol{B}_n\boldsymbol{c})\right)\Bigg|\mathcal{F}_{n,i-1}\right) \xrightarrow{p} 0 \tag{13}$$

as $n \to \infty$, for all $\|\boldsymbol{c}\| = 1$ and any $\varepsilon > 0$. By Assumption (A.5), (13), and the Central Limit Theorem for a martingale array (e.g., Durrett and Resnick (1978, Theorem 2.3)), we have $(\boldsymbol{c}^T\boldsymbol{B}_n\boldsymbol{c})^{-1/2}\sum_{i=1}^{|\Lambda_n|}\boldsymbol{c}^T\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0) \xrightarrow{d} N(0,1)$ as $n \to \infty$, for all $\|\boldsymbol{c}\| = 1$. It follows that $\boldsymbol{B}_n^{-1/2}\sum_{i=1}^{|\Lambda_n|}\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0) \xrightarrow{d} N(\boldsymbol{0},\boldsymbol{I}_p)$ as $n \to \infty$, by the Cramer-Wold device (see Cramer and Wold (1936); Durrett (1991)).

Proof of Corollary 3. Using the same arguments as for (11) in the proof of Theorem 2, it can be shown that $\boldsymbol{B}_n^{-1/2}\hat{\boldsymbol{B}}_n\boldsymbol{B}_n^{-1/2} \xrightarrow{p} \boldsymbol{I}_p$ as $n \to \infty$. It follows that $\hat{\boldsymbol{B}}_n^{1/2}\boldsymbol{B}_n^{-1/2} \xrightarrow{p} \boldsymbol{I}_p$ as $n \to \infty$. By Slutsky's Theorem we have, as $n \to \infty$, $\hat{\boldsymbol{B}}_n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = (\hat{\boldsymbol{B}}_n^{1/2}\boldsymbol{B}_n^{-1/2})\boldsymbol{B}_n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N(\boldsymbol{0},\boldsymbol{I}_p), (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^T\hat{\boldsymbol{B}}_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} \chi_p^2$. This completes the proof.

Proof of Corollary 4. The asymptotic normality can be obtained using the same arguments as in the proof Theorem 2. We therefore give only a proof for the asymptotic efficiency of $\hat{\boldsymbol{\theta}}_n^*$. First, note that $\boldsymbol{B}_n \equiv \mathrm{var}(\sum_{\boldsymbol{s}\in\Lambda_n}\boldsymbol{b}^{(n)}(\boldsymbol{s};\boldsymbol{\theta}_0)) = -E(\sum_{\boldsymbol{s}\in\Lambda_n}\boldsymbol{A}^{(n)}(\boldsymbol{s};\boldsymbol{\theta}_0)) = -E(\sum_{\boldsymbol{s}\in E_n}\boldsymbol{A}^{(n)}(\boldsymbol{s};\boldsymbol{\theta}_0) + \sum_{\boldsymbol{s}\in\Lambda_n^*}\boldsymbol{A}^{(n)}(\boldsymbol{s};\boldsymbol{\theta}_0)) = -E(\sum_{\boldsymbol{s}\in E_n}\boldsymbol{A}^{(n)}(\boldsymbol{s};\boldsymbol{\theta}_0)) + \boldsymbol{B}_n^*$, $n \in I\!N$. From (10) in the proof of Theorem 2, the smallest eigenvalue of $\frac{1}{|\Lambda_n|}\boldsymbol{B}_n$ is bounded away from zero for large n and, since $\frac{1}{|\Lambda_n|}E(\sum_{\boldsymbol{s}\in E_n}\boldsymbol{A}^{(n)}(\boldsymbol{s};\boldsymbol{\theta}_0)) \leq \frac{1}{|\Lambda|}\sum_{\boldsymbol{s}\in E_n}E(\boldsymbol{b}^{(n)}(\boldsymbol{s};\boldsymbol{\theta}_0)(\boldsymbol{b}^{(n)}(\boldsymbol{s};\boldsymbol{\theta}_0))^T) \to \boldsymbol{0}$, as $n \to \infty$, we have $\boldsymbol{B}_n^{-1}\boldsymbol{B}_n^* \to \boldsymbol{I}_n$ as $n \to \infty$. So

$$\begin{aligned}
&\boldsymbol{B}_n^{1/2}\left\{\hat{\boldsymbol{\theta}}_n^* - \boldsymbol{\theta}_0 - \boldsymbol{B}_n^{-1}\sum_{i=1}^{|\Lambda_n|}\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)\right\}\\
&= \boldsymbol{B}_n^{1/2}\left\{-\sum_{i=1}^{|\Lambda_n^*|}\tilde{\boldsymbol{A}}^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}_0,\hat{\boldsymbol{\theta}}_n^*)\right\}^{-1}\left\{\sum_{i=1}^{|\Lambda_n^*|}\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}_0)\right\} - \boldsymbol{B}_n^{-1/2}\sum_{i=1}^{|\Lambda_n|}\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i};\boldsymbol{\theta}_0)\\
&= (\boldsymbol{B}_n^*)^{1/2}\left\{-\sum_{i=1}^{|\Lambda_n^*|}\tilde{\boldsymbol{A}}^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}_0,\hat{\boldsymbol{\theta}}_n^*)\right\}^{-1}\left\{\sum_{i=1}^{|\Lambda_n^*|}\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}_0)\right\}\\
&\quad -(\boldsymbol{B}_n^*)^{-1/2}\sum_{i=1}^{|\Lambda_n^*|}\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}_0) + o_p(1)\\
&= \left\{(\boldsymbol{B}_n^*)^{1/2}\left(-\sum_{i=1}^{|\Lambda_n|}\tilde{\boldsymbol{A}}^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}_0,\hat{\boldsymbol{\theta}}_n^*)\right)^{-1}(\boldsymbol{B}_n^*)^{1/2} - \boldsymbol{I}_p\right\}
\end{aligned}$$

$$\cdot\left\{(\boldsymbol{B}_n^*)^{-1/2}\sum_{i=1}^{|\Lambda_n^*|}\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}_0)\right\}+o_p(1)\xrightarrow{p}\mathbf{0}.$$

That is, $\hat{\boldsymbol{\theta}}_n$ is asymptotically efficient (see Basawa and Rao (1980, Section 7.2.4); Rao (1973, Section 5c.2)).

Proof of Lemma 2. Consider any finite set $B \subset D$. Let $B^* \equiv \cup_{\boldsymbol{s}\in B} V_{\boldsymbol{s}}$. Then $\mathcal{F}_B^W \subset \mathcal{F}_{B^*}^Z$, since $W(\boldsymbol{s}) \equiv f_{\boldsymbol{s}}(Z(V_{\boldsymbol{s}}))$, $\boldsymbol{s} \in D$. Assume that $diam\, V = v$. Suppose $B_1, B_2 \subset D$ with $|B_1| \le m$, $|B_2| \le l$, and $d(B_1, B_2) \ge n \ge v$. Let $B_i^* \equiv \cup_{\boldsymbol{s}\in B_i} V_{\boldsymbol{s}}$, $i = 1, 2$. Then $|B_1^*| \le mv$, $|B_2^*| \le lv$, and $d(B_1^*, B_2^*) \ge n - v + 1$. So, by Lemma 1, $\alpha_{m,l}^W(n) \le \alpha_{mv,lv}^Z(n-v+1) = C_1\{(mv)\wedge(lv)\}\exp\{-C_2(n-v+1)\} = C_3(m \wedge l)\exp(-C_4 n)$, for some positive constants C_1, C_2, C_3 and C_4. This completes the proof.

Proof of Theorem 3. By Corollary 2 and Corollary 4, it suffices to prove (A.3′) and (A.5′). Since $\sup\{diam\,(adjl\ \boldsymbol{s} \cup \{\boldsymbol{s}\}) : \boldsymbol{s} \in \mathbb{Z}^d \setminus L^*\} \le k$, it follows that $Z(\mathbb{Z}^d \setminus L^*)$ is k-Markovian. For any $\|\boldsymbol{c}\| = 1$, define

$$F(\boldsymbol{s}) \equiv \begin{cases} \boldsymbol{c}^T\boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}_0)\boldsymbol{c} - E\left(\boldsymbol{c}^T\boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}_0)\boldsymbol{c}\right), & \text{if } \boldsymbol{s} = \boldsymbol{s}_{n,i}^* \in \cup\Lambda_n^*, \\ 0, & \text{otherwise.}\end{cases}$$

Then by Lemma 2 we have for all $n \in \mathbb{N}$,

$$\alpha_{1,1}^F(n) \le r_1\exp(-nr_2), \tag{14}$$

for some positive constants r_1 and r_2. Also by (A.2′) (with $q > 4$), for any $\|\boldsymbol{c}\| = 1$ we have

$$\sup_{\boldsymbol{s}\in\cup\Lambda_n^*} E\left(|F(\boldsymbol{s})|^{q/2}\right) < \infty. \tag{15}$$

Applying Corollary 5 to $\{F(\boldsymbol{s}) : \boldsymbol{s} \in \cup\Lambda_n^*\}$, (14) and (15) imply that as $n \to \infty$,

$$\frac{1}{|\Lambda_n^*|}\sum_{i=1}^{|\Lambda_n^*|}\left\{\boldsymbol{c}^T\boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}_0)\boldsymbol{c} - E\left(\boldsymbol{c}^T\boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}_0)\boldsymbol{c}\right)\right\} \xrightarrow{p} 0 \tag{16}$$

for all $\|\boldsymbol{c}\| = 1$. Since $\liminf_{n\to\infty}\frac{1}{|\Lambda_n^*|}\boldsymbol{c}^T\boldsymbol{B}_n^*\boldsymbol{c}$ is a continuous function of $\boldsymbol{c}$, and $\{\boldsymbol{c} : \|\boldsymbol{c}\| = 1\}$ is a compact set of $\mathbb{R}^p$, by (8) we have $\inf_{\|\boldsymbol{c}\|=1}\liminf_{n\to\infty}\frac{1}{|\Lambda_n^*|}\sum_{i=1}^{|\Lambda_n^*|} E(-\boldsymbol{c}^T\boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}_0)\boldsymbol{c}) = \inf_{\|\boldsymbol{c}\|=1}\liminf_{n\to\infty}\frac{1}{|\Lambda_n^*|}\boldsymbol{c}^T\boldsymbol{B}_n^*\boldsymbol{c} = \delta > 0$. It follows that as $n \to \infty$, $P\{\frac{-1}{|\Lambda_n^*|}\sum_{i=1}^{|\Lambda_n^*|}(\boldsymbol{c}^T\boldsymbol{A}^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}_0)\boldsymbol{c}) > \frac{\delta}{2}\} \to 1$ for all $\|\boldsymbol{c}\| = 1$. This proves (A.3′). Similarly, for any $\|\boldsymbol{c}\| = 1$, define

$$G(\boldsymbol{s}) \equiv \begin{cases} E\left(\left(\boldsymbol{c}^T\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}_0)\right)^2\middle|\mathcal{F}_{n,i-1}^*\right) - E\left(\boldsymbol{c}^T\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}_0)\right)^2, & \text{if } \boldsymbol{s} = \boldsymbol{s}_{n,i}^* \in \cup\Lambda_n^*, \\ 0, & \text{otherwise.}\end{cases}$$

Applying Corollary 5 to $\{G(\boldsymbol{s}) : \boldsymbol{s} \in \cup\Lambda_n^*\}$ as in the proof for (16), we obtain as $n \to \infty$, $\frac{1}{|\Lambda_n^*|}\sum_{i=1}^{|\Lambda_n^*|}\{E(\boldsymbol{c}^T\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}_0)|\mathcal{F}_{n,i-1}^*) - \boldsymbol{c}^T\boldsymbol{B}_n^*\boldsymbol{c}\} \xrightarrow{p} 0$, for all $\|\boldsymbol{c}\| = 1$. It follows that as $n \to \infty$, $\frac{1}{\boldsymbol{c}^T\boldsymbol{B}_n^*\boldsymbol{c}}\sum_{i=1}^{|\Lambda_n^*|}\{E(\boldsymbol{c}^T\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}_0)|\mathcal{F}_{n,i-1}^*) - \boldsymbol{c}^T\boldsymbol{B}_n^*\boldsymbol{c}\} = \frac{|\Lambda_n^*|}{\boldsymbol{c}^T\boldsymbol{B}_n^*\boldsymbol{c}}\frac{1}{|\Lambda_n^*|}\sum_{i=1}^{|\Lambda_n^*|}\{E(\boldsymbol{c}^T\boldsymbol{b}^{(n)}(\boldsymbol{s}_{n,i}^*;\boldsymbol{\theta}_0)|\mathcal{F}_{n,i-1}^*) - \boldsymbol{c}^T\boldsymbol{B}_n^*\boldsymbol{c}\} \xrightarrow{p} 0$ for all $\|\boldsymbol{c}\| = 1$. This proves (A.5′) and completes the proof.

References

Abend, K., Harley, T. J. and Kanal, L. N. (1965). Classification of binary random patterns. *IEEE Trans. Inform. Theory* **IT-11**, 538-544.

Aitchison, J. and Silvey, S. D. (1958). Maximum-likelihood estimation of parameters subject to restraints. *Ann. Math. Statist.* **29**, 813-828.

Basawa, I. V., Feigin, P. D. and Heyde, C. C. (1976). Asymptotic properties of maximum likelihood estimators for stochastic processes. *Sankhyā, Ser. A* **38**, 259-270.

Basawa, I. V. and Rao, L. S. P. (1980). *Statistical Inference for Stochastic Processes.* Academic Press, London.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* **34**, 192-225.

Besag, J. (1975). Statistical analysis of non-lattice data. *Statistician* **24**, 179-195.

Cramer, H. and Wold, H. (1936). Some theorems on distribution functions. *J. London Math. Soc.* **11**, 290-295.

Cressie, N. and Davidson, J. L. (1998). Image analysis with partially ordered Markov models. *Comput. Statist. Data Anal.* **29**, 1-26.

Crowder, M. J. (1976). Maximum likelihood estimation for dependent observations. *J. London Math. Soc.* **38**, 45-53.

Davidson, J. L. and Cressie, N. (1993). Markov pyramid models in images analysis. In *Image Algebra and Morphological Image Processing IV* (Edited by E. R. Dougherty, P. D. Gader and J. C. Serra), Society of Photo-Optical Instrumentation Engineers (SPIE) Proceedings **2030**, 179-190. SPIE, Bellingham, WA.

Davidson, J. L., Cressie, N. and Hua, X. (1999). Texture synthesis and pattern recognition for partially ordered Markov models. *Pattern Recognition* **32**, 1475-1505.

Devijver, P. A. (1988). Image segmentation using causal Markov random field models. In *Pattern Recognition* (Edited by J. Kittler), Lecture Notes in Computer Science **301**, 131-143. Springer, New York.

Dobrushin, P. L. (1968). The description of a random field by means of conditional probabilities and conditions of its regularity. *Theory Probab. Appl.* **13**, 197-224.

Doukhan, P. (1994). *Mixing: Properties and Examples.* Springer-Verlag, New York.

Durrett, R. and Resnick, S. I. (1978). Functional limit theorems for dependent variables. *Ann. Probab.* **6**, 829-846.

Durrett, R. (1991). *Probability: Theory and Examples.* Wadsworth, Belmont, CA.

Gidas, B. (1988). Consistency of maximum likelihood and pseudolikelihood estimators for Gibbs distributions. In *Stochastic Differential System, Stochastic Control Theory and Applications* (Edited by W. Fleming and P. L. Lions), 129-145. Springer, New York.

Gidas, B. (1993). Parameter estimation for Gibbs distributions from fully observed data. In *Markov Random Fields* (Edited by R. Chellappa and A. Jain), 471-498. Academic Press, Boston.

Goutsias, J. K. (1989). Mutually compatible Gibbs random fields. *IEEE Trans. Inform. Theory* **35**, 1233-1249.

Gray, A. J., Kay, J. W. and Titterington, D. M. (1994). An empirical study of the simulation of various models used for images. *IEEE Trans. Pattern Anal. Machine Intelligence* **16**, 507-513.

Guyon, X. (1995). *Random Fields on a Network: Modeling, Statistics, and Applications.* Springer-Verlag, New York.

Heijmans, R. D. H. and Magnus, J. R. (1986a). On the first-order efficiency and asymptotic normality of maximum likelihood estimators obtained from dependent observations. *Statist. Neerlandica* **40**, 169-188.

Heijmans, R. D. H. and Magnus, J. R. (1986b). Consistent maximum-likelihood estimation with dependent observations: the general (non-normal) case and the normal case. *J. Econometrics* **32**, 253-285.

Kiiveri, H., Speed, T. P. and Carlin, J. B. (1984). Recursive causal models. *J. Austral. Math. Soc.* **36**, 30-52.

Kolmogoroff, A. (1933). *Grundbegriffe der Wahrscheinlichkeitsrochnung.* Springer-Verlag, Berlin.

Lauritzen, S. L. and Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *J. Roy. Statist. Soc. Ser. B* **50**, 157-224.

Lauritzen, S. L., Dawid, A. P., Larson, B. N. and Leimer, H. G. (1990). Independence properties of directed Markov fields. *Networks* **20**, 491-505.

Lindsay, B. G. (1988). Composite likelihood methods. In *Statistical Inference from Stochastic Processes* (Edited by N. U. Prabhu), Contemporary Mathematics **80**, 221-239. American Mathematical Society, Providence, RI.

Possolo, A. (1986). Estimation of binary Markov fields. Technical Report no. **77**, Department of Statistics, University of Washington, Seattle, WA.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications.* 2nd edition. Wiley, New York.

Sarma, Y. R. (1986). Asymptotic properties of maximum likelihood estimators from dependent observations. *Statist. Probab. Lett.* **4**, 309-311.

Sweeting, T. J. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *Ann. Statist.* **8**, 1375-1381.

Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan.

E-mail: hchuang@stat.sinica.edu.tw

Department of Statistics, The Ohio State University, Columbus, OH 43210-1247, U.S.A.

E-mail: ncressie@stat.ohio-state.edu

(Received March 1999; accepted February 2000)