The Annals of Applied Probability
1995, Vol. 5, No. 4, 1095–1125

A GENERALIZED MAXIMUM PSEUDO-LIKELIHOOD ESTIMATOR FOR NOISY MARKOV FIELDS

By David J. Barsky[1] and Alberto Gandolfi

University of California, Davis, and University of California, Berkeley

In this paper we present an asymptotic estimator, obtained by observing a noisy image, for the parameters of both a stationary Markov random field and an independent Bernoulli noise. We first estimate the parameter of the noise by solving a polynomial equation of moderate degree (about 6–7 in the one-dimensional Ising model and about 10–15 in the two-dimensional Ising model, for instance) and then apply the maximum pseudo-likelihood method after removing the noise. Our method requires no extra simulation and is likely to be applicable to any Markov random field, in any dimension. Here we present the general theory and some examples in one dimension; more interesting examples in two dimensions will be discussed at length in a companion paper.

1. Introduction. In recent applications, images as well as other processes have been modeled by a Markov random field, that is, a Gibbs state with a finite range interaction, sometimes degraded by noise [see Comets and Gidas (1992) and references contained therein].

We are interested here in statistical inference, that is, the estimation of parameters, for stationary Markov random fields. This section contains a brief discussion of the models together with some previous results and our findings; a more formal presentation is found in the rest of the paper.

We start with a single infinite black and white *image*, which is a specification of $+1$ (*black*) or -1 (*white*) at each vertex (*pixel*) of an infinite lattice. The lattice we consider is $\mathbb{Z}^d$ and typically $d = 2$. The statistical properties of the image are described by a stationary Markov random field (with stationary interaction), which depends on some parameters $\theta_0 = (\theta_0(1), \dots, \theta_0(s))$. A *noisy image* is obtained by independently flipping the sign (i.e., the color) of the image at each pixel with probability ε_0. The problem here is to estimate θ_0 and ε_0 by observing part of the noisy image, typically a finite rectangular array, with the sole a priori knowledge that θ_0 belongs to some subset $\Theta \subseteq \mathbb{R}^s$ and that ε_0 is small, typically $\varepsilon_0 < \frac{1}{2}$. [Here we assume sufficient knowledge of the structure of the Markov random field so that s is known. For questions related to the estimation of s in the case with no noise and no

Received Received April 1993; revised April 1994.
[1]Research supported in part by an NSF Postdoctoral Fellowship.
AMS 1991 *subject classifications*. Primary 62M40; secondary 62F12.
Key words and phrases. Maximum pseudo-likelihood estimator, consistent estimator, Markov random fields.

such knowledge, see Ji and Seymour (1991) and Denny and Wright (1991).] More precisely, the problem is to find a pair of functions (called estimators) $\hat{\theta}^{(\Lambda)}$ and $\hat{\varepsilon}^{(\Lambda)}$ of the noisy image in $\Lambda \subset \mathbb{Z}^d$, such that if $\{\Lambda_n\}_{n \in \mathbb{N}}$ is an increasing sequence of arrays whose union equals $\mathbb{Z}^d$, then $\hat{\theta}^{(\Lambda_n)} \to \theta_0$ and $\hat{\varepsilon}^{(\Lambda_n)} \to \varepsilon_0$ with probability 1 (with respect to the joint distribution, $P_{\varepsilon_0, \theta_0}$, of the Markov random field and the noise).

Various estimators have been proposed, both for specific Markov random fields and for more general models. Two of these are the maximum likelihood estimator [Dempster, Laird and Rubin (1977), Geman and McClure (1985) and Younes (1989)] and the maximum pseudo-likelihood estimator [Chalmond (1987) and Younes (1991)], both of which are based on the EM algorithm. Unfortunately, these estimators are obtained by iterative methods, requiring the simulation of a Markov random field at each iteration and resulting in a complex process. Other estimators are obtained by the methods of moments [Geman and McClure (1985) and Frigessi and Piccioni (1990)]. These methods do not require any extra simulation and are based on estimating various moments of the joint distribution $P_{\varepsilon_0, \theta_0}$ from the noisy image. Some combinations of these moments turn out to be functions of θ_0 independent of ε_0, so they can be used to estimate θ_0 provided they are invertible. Unfortunately, inverse functions cannot be easily produced even for the two-dimensional Ising model with zero external field, for which Onsager's exact solution of the model is available [one such inverse function was remarkably obtained in Frigessi and Piccioni (1990)], and they seem out of reach, if they exist at all, for all Markov random fields without an exact solution [i.e., most of them; see Baxter (1982)].

Our paper presents a new estimator for ε_0, whose computation requires no extra simulations, no iterations and which is (in principle) as easy and accurate as the solution of a moderate degree polynomial equation. An estimator of θ_0 is then obtained by a method analogous to the maximum pseudo-likelihood for the Markov random field alone [which is a very effective method; see Besag (1977)]. The advantages of our method lie in potentially very simple estimations. It is possible that it may also lend itself to new proofs of the identifiability of parameters. This method, however, is not without its own difficulties. The problem is now reduced to (1) producing the above-mentioned polynomial equation (whose form depends on the structure of the Markov random field and on Θ) from the noisy image and (2) determining a priori which one of the roots of this equation is an estimator for ε_0. In this paper, we describe how to produce suitable polynomial equations for any Markov random field (Sections 2 and 3). Determination of the correct root, however, is more difficult. We have made some progress in the general case, but have had enough ideas, patience or computer power to complete this programme only in some limited cases [described in Section 4 and in the companion paper, Barsky and Gandolfi (1995)]. We now briefly outline the main ideas in the paper; a rigorous treatment starts afresh in Section 2.

Our approach begins with construction of the polynomial equations. We list the probabilities of all possible specifications of colors in some fixed finite array, as given by the Markov random field alone. This listing uses a great number of parameters: θ_0 and many other probabilities whose functional dependences on θ_0 are known only from the exact solution of the model, rarely available and in any case not used in this paper. Then we describe how these probabilities are transformed under the noise, introducing the parameter ε_0. Next, we invert this transformation (which amounts to the inversion of a large matrix) and apply the inverse transformation (parametrized by a new variable, ε) to a corresponding list of probabilities of patterns of colors in the noisy image. [An analogous matrix can be found in Meloche and Ruben (1992).] If $\varepsilon = \varepsilon_0$, the inversion procedure described above returns us to the original list of probabilities, but for other values of ε the process only gives a list of functions of ε (and ε_0 and θ_0). Using the structure of the Markov random field, we can indicate some necessary conditions, in the form of polynomial relations, which must be satisfied in order for such a list of functions to be the list of probabilities for a Markov random field. [Some similar notions can be found in Newman (1987).] Each of these necessary conditions provides a polynomial equation in ε, and $\varepsilon = \varepsilon_0$ is always a root. The idea is, therefore, to estimate from the data (i.e., the noisy image) the list of probabilities already transformed by the noise, apply the inverse transformation with the parameter ε to this observed list of empirical probabilities and then solve one (or more) polynomial relations to determine for which value(s) of ε the inverse-transformed list satisfies some of the necessary condition(s) for being the list of probabilities for a Markov random field. Having found an estimator for ε_0, it is easy to "remove" the noise and use a maximum pseudo-likelihood method to estimate θ_0.

It is regrettable that we do not yet have a general method to indicate which real root of these polynomial equations is an estimator for ε. Some of the polynomial equations might even be identically zero, for some or for all $\theta \in \Theta$. Such equations are called *null relations*, and we shall discard them. However, the null relations depend on the specific models and need to be identified on a case by case basis. The nonnull relations, or *effective relations*, on the other hand, will generally have other roots besides $\varepsilon = \varepsilon_0$, and a priori identification of the root estimating ε_0 again is done case by case. Some restrictions on ε_0 are demanded; for example, if the Markov random field has a global spin–flip symmetry, then ε_0 cannot be distinguished from $1 - \varepsilon_0$ (which is reflected in the polynomial equations being invariant under the exchange $\varepsilon \mapsto 1 - \varepsilon$). Additionally, $\varepsilon_0 = \frac{1}{2}$ cannot generally be identified (which is reflected in the above-mentioned matrix being singular when $\varepsilon = \frac{1}{2}$), but even if $\varepsilon_0 \in [0, \frac{1}{2})$, some polynomial equations have multiple real roots.

One possible solution to the problem of multiple roots is the simultaneous use of two or more equations, looking for common roots. However, when estimated from empirical data, such a set of equations would typically not have any common roots and, at present, we have no good estimates on how far the roots of equations from the data can stray from their theoretical values. Additionally, it seems difficult to give a set of equations whose only common root is ε_0 for all $\theta_0 \in \Theta$. Nevertheless, it might be possible to obtain such a set of equations, and this issue could be the subject of further research.

In a different attempt to deal with the problem of multiple roots, we explicitly study our equations for the one-dimensional Ising models (i.e., Markov chains) in Section 4, and for the two-dimensional Ising model in a companion paper [Barsky and Gandolfi (1995)]. We have found that for several of these equations (1) $\varepsilon = \varepsilon_0$ is a single root and (2) it is the smallest real root. It might be the case for every Markov random field that there always are equations for which statements (1) and (2) both occur, so we have formulated a theorem of consistency for the estimation in the context of this case, hoping that this will be the only consistency result required by the present theory. The possibility remains that statements (1) and (2) hold (and thus our consistency theorem applies) even in situations which are too complicated for us to verify. If there is some physical intuition which suggests that these conditions are met, then our estimator could still be used for practical applications.

The reader can now either turn directly to Section 4 for a treatment (which we tried to make self-contained) of simple one-dimensional models or else first read Sections 2 and 3 for the abstract theory.

2. Definitions and the main result. Let $\mathbb{Z}^d$ be the *d-dimensional integer lattice* and let $\Lambda \subset \mathbb{Z}^d$ be any *box* of the form $\prod_{n=1}^{d}[i_n, i_n + j_n] \cap \mathbb{Z}^d$, for some $(i_1, \dots, i_d) \in \mathbb{Z}^d$ and $(j_1, \dots, j_d) \in \mathbb{Z}_+^d$. In the present paper our interest is focused on the *observed images* $y_\Lambda \in \{-1, 1\}^\Lambda$, which are the final result of some stochastic process. Our set of definitions is basically the description of this process.

Depending on the context, we indicate the *configuration space* $\{-1, 1\}^{\mathbb{Z}^d}$ by X, Y or Z; also, X_S, Y_S and Z_S will all indicate $\{-1, 1\}^S$, for $S \subseteq \mathbb{Z}^d$. We suppose that we are dealing with an *original image* $x \in X$, which has been corrupted by a *noise* $z \in Z$, resulting in an *observable image* $y \in Y$ given by $y_i = x_i \cdot z_i$ for all $i \in \mathbb{Z}^d$. Elements $i \in \mathbb{Z}^d$ are called *pixels*, and for any given pixel $i \in \mathbb{Z}^d$, x_i (resp., y_i) is called the *original* (resp., *observable*) *coloring* of i. The *observed image* y_Λ is the restriction of y to the box Λ. Apart from the distinction between observable and observed image, in the following we will often regard configurations in $\{-1, 1\}^{S_1}$ as the restrictions of configurations in $\{-1, 1\}^{S_2}$, if $S_1 \subseteq S_2 \subseteq \mathbb{Z}^d$.

The original image, the noise and the observable image are described by some elements of the sets $\mathscr{P}_X$, $\mathscr{P}_Z$ and $\mathscr{P}_Y$ of the probability measures on the Borel σ-algebra of X, Z and Y, respectively.

The original image. Let $\mathscr{C}$ be a *locally finite* (i.e., $|\{C \in \mathscr{C}: i \in C\}| < \infty$ for all $i \in \mathbb{Z}^d$, where $|A|$ denotes the cardinality of A) and *translation invariant* [i.e., $\tau_i(C) \in \mathscr{C}$ if $C \in \mathscr{C}$, where τ_i indicates the translation by the vector $i \in \mathbb{Z}^d$] collection of sets $C \subset \mathbb{Z}^d$ called *cliques*. Note that the local finiteness and translation invariance of $\mathscr{C}$ together imply that each clique is finite. An *interaction* ϕ based on $\mathscr{C}$ is a translation invariant real-valued function defined on $\bigcup_{C \in \mathscr{C}} X_C$. Let $\mathbf{0}$ indicate the origin of $\mathbb{Z}^d$. Then the *local interactions* of ϕ are the entries of the vector $\{\phi(\eta)\}_{\eta \in X_C, \mathbf{0} \in C \in \mathscr{C}}$. We use these interactions to define Markov random fields. Later on we shall see that these models can be reparametrized using fewer parameters than the total number of local interactions; it will then be advantageous to use a different, but equivalent, notation.

For now, let a set of cliques $\mathscr{C}$ and an interaction ϕ be fixed. For each finite $S \subset \mathbb{Z}^d$, the *energy function* U_S^ϕ: $X \to \mathbb{R}$ is defined by

$$U_S^\phi(x) = \sum_{\substack{C \in \mathscr{C} \\ C \cap S \text{ nonempty}}} \phi(x_C). \tag{2.1}$$

Note that the energy function can be thought of as a linear combination, with integer coefficients, of the local interactions.

For $\Lambda \subset \mathbb{Z}^d$ and $\bar{x} \in X$, a *finite volume Markov random field* for $\mathscr{C}$ and ϕ in Λ with boundary condition $\bar{x}$ is the probability measure

$$\mu_{\Lambda, \bar{x}}(x_\Lambda) = Z_{\Lambda, \bar{x}}^{-1} \exp(-U_\Lambda^\phi(x_\Lambda \vee \bar{x})), \tag{2.2}$$

where

$$Z_{\Lambda, \bar{x}} = \sum_{x_\Lambda \in X_\Lambda} \exp(-U_\Lambda^\phi(x_\Lambda \wedge \bar{x})).$$

Here $x_\Lambda \vee \bar{x} \in X$ is the configuration which agrees with x_Λ in Λ and with $\bar{x}$ in $\mathbb{Z}^d \setminus \Lambda$. A *Markov random field* for $\mathscr{C}$ and ϕ is any convex combination μ_ϕ of weak limits of $\mu_{\Lambda, \bar{x}}$ as $\Lambda \uparrow \mathbb{Z}^d$, that is, as Λ ranges over an increasing sequence of boxes which eventually covers the whole of $\mathbb{Z}^d$ [see Ruelle (1978), Chapter 1]. *Phase transition* occurs if there is more than one Markov random field for the given $\mathscr{C}$ and ϕ. In this paper, we only consider *translation invariant* Markov random fields; the set $\mathscr{M}_\phi$ of all such Markov random fields for the interaction ϕ is always nonempty [Ruelle (1978), Theorem 3.7]. Since our estimation scheme begins with a single infinite (noisy) image, we may assume that μ_ϕ is *ergodic* for the group of translations of $\mathbb{Z}^d$, as any original image is a typical configuration for some ergodic component of μ_ϕ.

Markov random fields μ_ϕ satisfy the *Markov property*: for any finite $S \subset \mathbb{Z}^d$ there exists a finite $T \supset S$ such that $\mu_\phi(x_S \mid \bar{x}) = \mu_\phi(x_S \mid \bar{x}_{T \setminus S})$ for $x_S \in X_S$ and $\bar{x} \in X$. In particular, for $S = \{\mathbf{0}\}$ one may take $T = \bigcup_{C: \mathbf{0} \in C \in \mathscr{C}} C$; we denote this particular set by $\bar{N}_\mathbf{0}$ and we call it the *complete neighborhood* of the origin $\mathbf{0}$. Configurations $\xi \in X_{\bar{N}_\mathbf{0}}$ are called *complete local patterns*. For $i \in \mathbb{Z}^d$, $\bar{N}_i$ will be the translation $\tau_i(\bar{N}_\mathbf{0})$ of $\bar{N}_\mathbf{0}$ by the vector i. The *neighborhood* of i is $N_i = \bar{N}_i \setminus \{i\}$. It follows from the translation invariance of $\mathscr{C}$ that $i \in N_\mathbf{0}$ if and only if $-i \in N_\mathbf{0}$; thus $|N_\mathbf{0}|$ is even. *Local patterns* are configurations $\xi \in X_{N_\mathbf{0}}$, and each such configuration gives rise to a pair of *local characteristics*

$$\pi_\phi(x_\mathbf{0} \mid \xi) = \mu_\phi(x_\mathbf{0} \mid \xi) = Z_{\mathbf{0}, \xi}^{-1} \exp(-U_\mathbf{0}^\phi(x_\mathbf{0} \vee \xi)). \tag{2.3}$$

[Our notation, here and elsewhere, in writing $U_{\mathbf{0}}^{\phi}$ (and $Z_{\mathbf{0},\xi}$) is that when S is the singleton $\{\mathbf{0}\}$, we write $S = \mathbf{0}$ as an abbreviation.]

Note that local characteristics are functions of the $\sum_{C:\mathbf{0}\in C\in\mathscr{C}} 2^{|C|}$ local interactions. Moreover, the local characteristics are functions of ϕ independent of the specific $\mu_\phi \in \mathscr{M}_\phi$. Also, since the local characteristics are always strictly positive, the Markov property implies that $\mu_\phi(x_S) > 0$ for all $x_S \in X_S$, for every finite $S \subset \mathbb{Z}^d$.

The noise and the observable image. For some $\varepsilon \in [0,1]$, the statistical properties of the configurations $z \in Z$ are described by the Bernoulli probability measure $\nu_\varepsilon = \prod_{i\in\mathbb{Z}^d} \nu_{\varepsilon,i}$, defined on the Borel σ-algebra of Z, where $\nu_{\varepsilon,i}(z_i = -1) = \varepsilon = 1 - \nu_{\varepsilon,i}(z_i = 1)$. The action of the noise is given by setting $y_i = x_i \cdot z_i$. This amounts to flipping each pixel with probability ε, independently of the other pixels and of x. For any given interaction ϕ and noise level ε, the joint probability measure $P_{\phi,\varepsilon} = \mu_\phi \cdot \nu_\varepsilon$, defined on the Borel σ-algebra of Y, describes the statistical properties of the observable image. Eventually, the interactions ϕ will be parametrized by a vector θ. Then we will write $P_{\theta,\varepsilon}$ for the joint measure.

Estimation of parameters. Suppose now that the single infinite black and white observable image $y \in Y$ is fixed and we observe y_Λ as $\Lambda \uparrow \mathbb{Z}^d$. The statistical properties of y are described by P_{ϕ_0,ε_0} for a known $\mathscr{C}$, but with both ε_0 and ϕ_0 unknown. We want to estimate ε_0 and ϕ_0, but this is only feasible if both are identifiable. The restriction of ε_0 to $[0, \frac{1}{2})$ is sufficient for identification of ε_0, but ϕ_0 can only be identified modulo the following equivalence relation [see also Gidas (1991), Appendix, for a related discussion].

Two interactions ϕ_1 and ϕ_2 are *equivalent*, $\phi_1 \approx \phi_2$, if $\mathscr{M}_{\phi_1} = \mathscr{M}_{\phi_2}$ [or, equivalently, if $\mathscr{M}_{\phi_1}$ and $\mathscr{M}_{\phi_2}$ have nonempty intersection; see Gidas (1991)]. In our setting, the equivalence relation is better described by the following lemma. Note that, by translation invariance, an interaction ϕ is identified by the

$$r = \sum_{C:\mathbf{0}\in C\in\mathscr{C}} 2^{|C|}$$

local interactions, so it can be treated as a vector in $\mathbb{R}^r$. The energies corresponding to the choice of any s interactions $\overline{\phi}_1, \dots, \overline{\phi}_s \in \mathbb{R}^r$ can be written as a function $U_{\mathbf{0}} = (U_{\mathbf{0}}^{\overline{\phi}_1}, \dots, U_{\mathbf{0}}^{\overline{\phi}_s})$ defined on $X_{\overline{N}_0}$ and taking values in $\mathbb{R}^s$.

LEMMA 1. (a) *Two interactions ϕ_1 and ϕ_2 are equivalent iff*

$$U_{\mathbf{0}}^{\phi_1}(x_{\mathbf{0}} \vee \xi) - U_{\mathbf{0}}^{\phi_1}(-x_{\mathbf{0}} \vee \xi) = U_{\mathbf{0}}^{\phi_2}(x_{\mathbf{0}} \vee \xi) - U_{\mathbf{0}}^{\phi_2}(-x_{\mathbf{0}} \vee \xi), \tag{2.4}$$

for all $\xi \in X_{N_0}$. The two interactions are also equivalent iff

$$U_{\mathbf{0}}^{\phi}(x_{\mathbf{0}} \vee \xi) - U_{\mathbf{0}}^{\phi}(-x_{\mathbf{0}} \vee \xi) = 0, \tag{2.5}$$

for all $\xi \in X_{N_0}$, *with* $\phi = \phi_1 - \phi_2$. *In both* (2.4) *and* (2.5), x_0 *can be taken to be either* $+1$ *or* -1.

(b) *Define* $\mathscr{N} = \{\phi \in \mathbb{R}^r: \phi$ *satisfies* (2.5) *for all* $\xi \in X_{N_0}\}$ *and let* $\bar{\phi}_1, \dots, \bar{\phi}_s \in \mathbb{R}^r$ *be a basis of some maximal linear space* $\mathscr{S}$ *which is linearly independent of* $\mathscr{N}$. *If* $\theta = (\theta_1, \dots, \theta_s)$ *and* $\theta \cdot U_0$ *is the standard inner product of* θ *and* U_0 *in* $\mathbb{R}^s$, *then*

$$\theta \cdot U_0 = U_0^{(\sum_{i=1}^s \theta_i \bar{\phi}_i)}, \tag{2.6}$$

$$\theta \cdot [U_0(x_0 \vee \xi) - U_0(-x_0 \vee \xi)] = 0 \quad \text{for all } \xi \in X_{N_0} \text{ iff } \theta = 0, \tag{2.7}$$

and

$$s = \dim(\mathscr{S}) \le \min\{2^{|N_0|}, r\}. \tag{2.8}$$

PROOF. Two interactions ϕ_1 and ϕ_2 are equivalent iff all of the finite volume Markov random fields for ϕ_1 and ϕ_2 coincide, and this holds iff all ratios

$$\frac{\exp(-U_\Lambda^\phi(x_\Lambda \vee \bar{x}))}{\exp(-U_\Lambda^\phi(x'_\Lambda \vee \bar{x}))} \tag{2.9}$$

are the same for ϕ_1 and ϕ_2, whenever $\Lambda \subset \mathbb{Z}^d$, $\bar{x} \in X_{\mathbb{Z}^d \setminus \Lambda}$ and x_Λ, $x'_\Lambda \in X_\Lambda$. It may further be assumed in (2.9) that x_Λ and x'_Λ differ in exactly one pixel. Now all ratios (2.9) are of the form

$$\exp(-U_0^\phi(x_0 \vee \xi) + U_0^\phi(-x_0 \vee \xi)),$$

for some $\xi \in X_{N_0}$ and $x_0 = \pm 1$, so $\phi_1 \approx \phi_2$ iff (2.4) holds for all $\xi \in X_{N_0}$. Note that U_0^ϕ is linear in $\phi \in \mathbb{R}^r$, so that (2.5) holds (with $\phi = \phi_1 - \phi_2$) whenever (2.4) is satisfied. This linearity also yields (2.6), as $\theta \cdot U_0 = \sum_{i=1}^s \theta_i U_0^{\bar{\phi}_i} = U_0^{(\sum_{i=1}^s \theta_i \bar{\phi}_i)}$. Moreover, $\theta \cdot [U_0(x_0 \vee \xi) - U_0(-x_0 \vee \xi)] = 0$, for all $\xi \in X_{N_0}$ iff $U_0^{(\sum \theta_i \bar{\phi}_i)}(x_0 \vee \xi) - U_0^{(\sum \theta_i \bar{\phi}_i)}(-x_0 \vee \xi) = 0$, for all $\xi \in X_{N_0}$ iff $\sum_{i=1}^s \theta_i \bar{\phi}_i \approx 0$ (the interaction which is identically zero) iff $\theta = 0$ (the zero vector in $\mathbb{R}^s$), which proves (2.7). Finally, the nontrivial part of (2.8) (that $s \le 2^{|N_0|}$) follows from the facts that the dimension of $\mathscr{S}$ cannot exceed the number of linearly independent equations of type (2.5) and that there are only $2^{|N_0|}$ linear relations of this type—before checking for linear independence. □

Later on we show that actually a strict inequality holds in (2.8). It is now convenient to fix a basis $\bar{\phi}_1, \dots, \bar{\phi}_s$ of a maximal linear space $\mathscr{S} \subset \mathbb{R}^r$, which is independent of $\mathscr{N}$, and to replace the energy in (2.1)–(2.3) by $\theta \cdot U_\Lambda(x_\Lambda \vee \bar{x})$, with $\theta \in \mathbb{R}^s$ and $U_\Lambda = (U_\Lambda^{\bar{\phi}_1}, \dots, U_\Lambda^{\bar{\phi}_s})$. The parameters $\theta \in \mathbb{R}^s$ are now identifiable and they will replace ϕ in our various notations: μ_ϕ and $P_{\phi,\varepsilon}$ become μ_θ and $P_{\theta,\varepsilon}$, respectively. We may also assume, for simplicity, that each $\bar{\phi}_i$ has integer entries.

Fix $\Theta \subseteq \mathbb{R}^s$, $\theta_0 \in \Theta$ and $\varepsilon_0 \in [0, \frac{1}{2})$. We want to define functions $\hat{\varepsilon}^{(\Lambda)}(y)$ and $\hat{\theta}^{(\Lambda)}(y)$ such that

$$\hat{\varepsilon}^{(\Lambda)}(y) \to \varepsilon_0 \quad \text{as } \Lambda \uparrow \mathbb{Z}^d \tag{2.10}$$

and

$$\hat{\theta}^{(\Lambda)}(y) \to \theta_0 \quad \text{as } \Lambda \uparrow \mathbb{Z}^d \tag{2.11}$$

for $P_{\theta_0,\varepsilon_0}$-almost all $y \in Y$.

The reason for not necessarily taking $\Theta = \mathbb{R}^s$ is that the extra information provided by the knowledge of Θ can make the estimation of ε_0 easier, as will be seen in the consistency theorem.

The main quantity which we will be estimating from the data is most conveniently introduced as a function of the empirical process generated by an original (resp., observable) image, or more generally, as a function of probability measures in $\mathscr{P}_X$ (resp., $\mathscr{P}_Y$). For each box $\Lambda \subset \mathbb{Z}^d$ and image $x \in X$ (resp., $y \in Y$), define $x^{(\Lambda)} \in X$ (resp., $y^{(\Lambda)} \in Y$) to be the periodic extension of the restricted image x_Λ (resp., y_Λ). The *empirical process* $R_{\Lambda,x} \in \mathscr{P}_X$ is defined by

$$R_{\Lambda,x}(f) = \frac{1}{|\Lambda|} \sum_{i \in \Lambda} f\left(\tau_{-i} x^{(\Lambda)}\right)$$

for all continuous functions f: $X \to \mathbb{R}$. The empirical process $R_{\Lambda,y}$ is similarly defined. For $P \in \mathscr{P}_X$ (resp., $P \in \mathscr{P}_Y$), let M_P be the vector whose $2^{|\overline{N}_0|}$ entries, indexed by the complete local patterns $\bar{\xi} \in X_{\overline{N}_0}$ (resp., $\bar{\xi} \in Y_{\overline{N}_0}$), are given by

$$M_P(\bar{\xi}) = E^P\left(\mathbf{1}\left[x_{\overline{N}_0} = \bar{\xi}\right]\right),$$

where E^P indicates the expectation with respect to P and $\mathbf{1}$ is the indicator function of the event in the brackets. The components of M are thus the probabilities of the various local patterns. In the case of the empirical processes, the components are just the relative frequencies in some portion of the image. For simplicity we use the notation

$$\begin{aligned} M_{\Lambda,x} &= M_{R_{\Lambda,x}}, \\ M_{\Lambda,y} &= M_{R_{\Lambda,y}} \end{aligned} \tag{2.12}$$

and

$$M_\theta = M_{\mu_\theta}, \tag{2.13}$$

where, in (2.13), μ_θ is some Markov random field for the interaction θ. We suppress the dependence of M_θ on μ_θ as we shall eventually work (see Lemma 3) with properties of the vector which are independent of the particular choice of the measure $\mu_\theta \in \mathscr{M}_\theta$. Additionally,

$$M_{\theta,\varepsilon} = M_{P_{\theta,\varepsilon}} = M_\theta A_\varepsilon,$$

where the second equality makes reference to the $2^{|\overline{N}_0|} \times 2^{|\overline{N}_0|}$ matrix A_ε defined in (2.14).

We comment here on our vector notation. We generally do not distinguish between row vectors and column vectors and use whichever notation seems to be most natural for the purpose at hand, as it will usually be clear from the context which is meant. For example, in writing $M_\theta A_\varepsilon$ above, we are using the usual probabilistic notation and regarding M_θ as a row vector. Later, in

Section 4 [above (4.4)] it will be equally evident that the vectors ϕ and $\mathbf{U}(\phi)$ are column vectors and that we are using the usual linear-algebraic notation in writing $\mathbf{U}(\phi) = \mathbf{U}\phi$.

The entries of A_ε are indexed by $\bar{\xi}^{(1)} \in X_{\bar{N}_0}$ and $\bar{\xi}^{(2)} \in Y_{\bar{N}_0}$ and given by

$$A_\varepsilon\big(\bar{\xi}^{(1)}, \bar{\xi}^{(2)}\big) = \varepsilon^D(1-\varepsilon)^{(|\bar{N}_0|-D)}, \tag{2.14}$$

where $D = D(\bar{\xi}^{(1)}, \bar{\xi}^{(2)}) = \sum_{i \in \bar{N}_0} \frac{1}{2}|\bar{\xi}_i^{(1)} - \bar{\xi}_i^{(2)}|$ is the Hamming distance between $\bar{\xi}^{(1)}$ and $\bar{\xi}^{(2)}$. As shown in Lemma 5, A_ε is invertible for $\varepsilon \neq \frac{1}{2}$ and its inverse A_ε^{-1} has components

$$A_\varepsilon^{-1}\big(\bar{\xi}^{(1)}, \bar{\xi}^{(2)}\big) = (2\varepsilon - 1)^{-|\bar{N}_0|}\varepsilon^D(\varepsilon - 1)^{(|\bar{N}_0|-D)}.$$

[Properties of related matrices appear in Barsky (1995) and Meloche and Ruben (1992).]

The estimator $\hat{\varepsilon}^{(\Lambda)}$ in (2.10) is one of the roots of a polynomial equation in ε constructed by relating the entries of $M_{\Lambda,y}A_\varepsilon^{-1}$ to the probabilities of complete local patterns in μ_{θ_0}. After this, $\hat{\varepsilon}^{(\Lambda)}$ is used to remove the noise from the data so that an estimator $\hat{\theta}^{(\Lambda)}$ satisfying (2.11) can be determined.

Before we can give the exact form of the polynomial equations, it is necessary to study Markov random fields in greater detail.

Structure of Markov random fields. We shall eventually see (in Lemma 3) that it is possible to produce $2^{|N_0|} - s$ polynomial equations (although several of these may be null). To be certain that we have any equations at all, we must first show that the upper bound for s given in (2.8) can be improved.

LEMMA 2. *Let $\mathscr{C}$ be a locally finite, translation invariant set of cliques, not all of size* 1. *Then if $\mathscr{S}$ is as in Lemma* 1,

$$s = \dim(\mathscr{S}) \leq \sum_{\substack{C:\mathbf{0}\in C\in\mathscr{C}\\ |C|\geq 2}} \frac{1}{|C|}(2^{|C|} - 1) < 2^{|N_0|}. \tag{2.15}$$

PROOF. From (2.8) we have that $s \leq r = \sum_{C:\mathbf{0}\in C\in\mathscr{C}} 2^{|C|}$, but s is in fact smaller for three reasons. We present these reasons as linear conditions which can (actually, the first condition must) be satisfied by the vectors of $\mathscr{S}$, thereby giving successive upper bounds to s.

In the first place, interactions are translation invariant. If τ_i stands both for the translation by the vector $i \in \mathbb{Z}^d$ and for the map induced by this translation on the configurations of X, then interactions satisfy $\phi(\tau_i\eta_C) = \phi(\eta_C)$ for all $i \in C$, $\eta_C \in X_C$ and $C \in \mathscr{C}$. Roughly speaking, the translation invariance implies that a clique C and all of its translates can contribute at most $2^{|C|}$ parameters to the sum which is the dimension of $\mathscr{S}$. More formally,

$$s \leq \sum_{C:\mathbf{0}\in C\in\mathscr{C}} \frac{1}{|C|}2^{|C|}.$$

In the second place, for each $C \in \mathscr{C}$ and for any interaction ϕ based on C, we may assume that $\phi(\overline{\eta}) = 0$ for at least one $\overline{\eta} \in X_C$. In fact, for any fixed $\overline{\eta} \in X_C$ (with $C \in \mathscr{C}$), we can define

$$\overline{\phi}(\eta) = \begin{cases} \phi(\eta) - \phi(\overline{\eta}), & \text{for } \eta \in X_C, \\ \phi(\eta), & \text{for } \eta \in X_{C'}, C' \neq C. \end{cases}$$

Then $\overline{\phi} \approx \phi$ by (2.4), which shows that

$$s \leq \sum_{C:\mathbf{0} \in C \in \mathscr{C}} \frac{1}{|C|}(2^{|C|} - 1).$$

Finally, we can also assume that $\phi(\eta) = 0$ for all $\eta \in X_C$ whenever $|C| = 1$. Indeed, we have just shown above that if $|C| = 1$, then it may be assumed that (acting on the color at that single pixel) $\phi(-1) = 0$. Assuming that $\phi(+1) \neq 0$ and that there is a clique $\tilde{C} \in \mathscr{C}$ with $|\tilde{C}| \geq 2$, define

$$\tilde{\phi}(\eta) = \begin{cases} \phi(\eta) + \phi(+1)\left|\left\{i \in \tilde{C}: \eta_i = +1\right\}\right|/|\tilde{C}|, & \text{for } \eta \in X_{\tilde{C}}, \\ 0, & \text{if } \eta \in X_C, |C| = 1, \\ \phi(\eta), & \text{otherwise.} \end{cases}$$

(In order to not upset the argument of the preceding paragraph, we must take the $\overline{\eta}$ for $X_{\tilde{C}}$ to be identically -1.) By (2.4), $\tilde{\phi} \approx \phi$, which concludes the proof of the first inequality in (2.15).

We next prove the second inequality in (2.15) for $d = 1$. Afterwards we will show how to reduce higher-dimensional cases to this setting. Given $\mathscr{C}$ and its associated complete neighborhood of $\mathbf{0}$, $\overline{N}_\mathbf{0} = \overline{N}_\mathbf{0}(\mathscr{C}) = \{i_{-|N_0|/2}, i_{-|N_0|/2+1}, \dots, i_{|N_0|/2}\}$ (where it may be assumed that i_n is increasing in n), define $\psi: \overline{N}_\mathbf{0} \to \overline{N} = \{-|N_\mathbf{0}|/2, -|N_\mathbf{0}|/2 + 1, \dots, |N_\mathbf{0}|/2\}$ by $\psi(i_n) = n$. In particular, note that $\psi(i_0) = \mathbf{0}$ since $i_0 = \mathbf{0}$. Now define $\mathscr{C}' = \{k + \psi(C): k \in \mathbb{Z}, \mathbf{0} \in C \in \mathscr{C}\}$. It is readily seen that $\mathscr{C}'$ is a locally finite, translation invariant set of cliques and that ψ induces a cardinality-preserving bijection between $\{C \in \mathscr{C}: \mathbf{0} \in C\}$ and $\{C' \in \mathscr{C}': \mathbf{0} \in C'\}$. Therefore,

$$\sum_{\substack{C:\mathbf{0} \in C \in \mathscr{C} \\ |C| \geq 2}} \frac{1}{|C|}(2^{|C|} - 1) = \sum_{\substack{C':\mathbf{0} \in C' \in \mathscr{C}' \\ |C'| \geq 2}} \frac{1}{|C'|}(2^{|C'|} - 1).$$

Since $\overline{N}_\mathbf{0}(\mathscr{C}') = \overline{N}$, it now suffices to prove that the second inequality in (2.15) holds for all $\mathscr{C}$ with $\overline{N}_\mathbf{0}(\mathscr{C}) = \overline{N}$. For such a $\mathscr{C}$, observe that the subset of those cliques containing $\mathbf{0}$ can be partitioned into equivalence classes by declaring a pair of cliques to be equivalent if they are translates of one another. Each equivalence class has a "least" representative: a clique for which the origin is the maximal pixel. Let $\mathscr{D}$ denote the set of least representatives for the equivalence classes of those cliques C having $\mathbf{0} \in C$ and $|C| \geq 2$. Then

$$\sum_{\substack{C:\mathbf{0}\in C\in\mathscr{C}\\|C|\geq 2}} \frac{1}{|C|}(2^{|C|}-1) = \sum_{C\in\mathscr{D}} (2^{|C|}-1).$$

The number of least representatives in $\mathscr{D}$ having cardinality n cannot exceed $\binom{|N_0|/2}{n-1}$ since any such representative must contain the origin and exactly $n-1$ pixels in $\{-|N_0|/2, -|N_0|/2+1,\dots,-1\}$. Thus

$$\begin{aligned}\sum_{C\in\mathscr{D}} (2^{|C|}-1) &= \sum_{n=2}^{N_0/2+1} \binom{|N_0|/2}{n-1}(2^n-1)\\ &= 2\cdot 3^{|N_0|/2} - 2^{|N_0|/2} - 1\\ &< 2^{|N_0|},\end{aligned}$$

for $|N_0| = 2, 4, \dots$.

For the higher-dimensional cases, we construct a map $\psi\colon \mathbb{Z}^d \to \mathbb{Z}$ which induces a cardinality-preserving bijection between $\{C \in \mathscr{C}\colon \mathbf{0} \in C\}$ and $\{C' \in \mathscr{C}'\colon \mathbf{0} \in C'\}$, where $\mathscr{C}'$ is some locally finite, translation invariant collection of cliques in $\mathbb{Z}$. Given $\mathscr{C}$, we note that it is always possible to choose integers $n_1,\dots,n_d$ so that $\psi(i_1,\dots,i_d) = \sum_{r=1}^{d} n_r i_r$ is injective on $\overline{N}_0(\mathscr{C})$. The collection $\mathscr{C}' = \{k + \psi(C)\colon k \in \mathbb{Z},\ \mathbf{0} \in C \in \mathscr{C}\}$ is a translation invariant, locally finite set of cliques in $\mathbb{Z}$ with $N_0(\mathscr{C}') = \psi(N_0(\mathscr{C}))$, and hence $|N_0(\mathscr{C}')| = |N_0(\mathscr{C})|$. Therefore,

$$\begin{aligned}\sum_{\substack{C:\mathbf{0}\in C\in\mathscr{C}\\|C|\geq 2}} \frac{1}{|C|}(2^{|C|}-1) &\leq \sum_{\substack{C':\mathbf{0}\in C'\in\mathscr{C}'\\|C'|\geq 2}} \frac{1}{|C'|}(2^{|C'|}-1)\\ &< 2^{|N_0(\mathscr{C}')|}\\ &= 2^{|N_0(\mathscr{C})|}\end{aligned}$$

as desired. □

It is clear that $\dim(\mathscr{S})$ might even be smaller than proven in Lemma 2. For instance, the first inequality in (2.15) yields $s \leq 3$ for the one-dimensional nearest-neighbor Ising model, but it is easy to verify that actually $s = 2$, with the two parameters corresponding to the two parameters of the 2×2 stochastic matrix defining the one-dimensional two state Markov chain. Also, the choice of the parameters made in this proof might not be the most convenient for a physical description of the model. In the Ising model, for example, one often uses the external field $h = \frac{1}{2}(\phi(+1) + \phi(-1))$ as a parameter.

We now fix $\mathscr{C}$ and Θ and turn to the issue of showing that the probabilities of the local patterns in a Markov random field satisfy certain polynomial equations. A *polynomial relation* for the vector $M = (M(\bar{\xi}))_{\bar{\xi}\in X_{N_0}}$ is any homogeneous polynomial $Q = Q(M)$ of the 2^{N_0} entries of M. Some particularly relevant polynomial relations are of the form

$$Q_\alpha(M) = \prod_{\xi \in X_{N_0}} [M(+1 \vee \xi)]^{\alpha_+(\xi)}[M(-1 \vee \xi)]^{-\alpha_-(\xi)} - \prod_{\xi \in X_{N_0}} [M(-1 \vee \xi)]^{\alpha_+(\xi)}[M(+1 \vee \xi)]^{-\alpha_-(\xi)}, \tag{2.16}$$

where $\alpha = (\alpha(\xi))_{\xi \in X_{N_0}}$ is a vector with integer entries, $\alpha_+(\xi) = \max\{\alpha(\xi), 0\}$ and $\alpha_-(\xi) = \min\{\alpha(\xi), 0\}$. We are especially interested in polynomial relations for the vector M_θ defined in (2.13). Some of these are described in the next lemma.

LEMMA 3. *Given $\mathscr{C}$ and Θ, there exist linearly independent integer vectors $\alpha^{(n)} \in \mathbb{Z}^{2^{|N_0|}}$, $n = 1, \dots, 2^{|N_0|} - s$, each with entries indexed by $\xi \in X_{N_0}$, such that*

$$Q_{\alpha^{(n)}}(M_\theta) = 0, \tag{2.17}$$

for all $\theta \in \Theta$, and for all Markov random fields $\mu_\theta \in \mathscr{M}_\theta$.

PROOF. A simple computation using (2.16), (2.13) and (2.3) shows that, for any $\theta \in \mathbb{R}^s$ (and any $\mu_\theta \in \mathscr{M}_\theta$), $Q_\alpha(M_\theta) = 0$ iff

$$\sum_{\xi \in X_{N_0}} \alpha(\xi) \log \frac{M_\theta(+1 \vee \xi)}{M_\theta(-1 \vee \xi)} = 0$$

iff

$$\sum_{\xi \in X_{N_0}} \alpha(\xi)\big[U_0^\theta(+1 \vee \xi) - U_0^\theta(-1 \vee \xi)\big] = 0. \tag{2.18}$$

Asking that (2.18) be satisfied for *all* $\theta \in \mathbb{R}^s$ is the same as seeking solutions to

$$\alpha \mathbf{U} = 0,$$

where $\mathbf{U}$ is the $2^{|N_0|} \times s$ matrix having entries

$$\mathbf{U}(\xi, i) = U_0^{\bar{\phi}_i}(+1 \vee \xi) - U_0^{\bar{\phi}_i}(-1 \vee \xi), \tag{2.19}$$

for $\xi \in X_{N_0}$ and $i \in \{1, \dots, s\}$. By the assumption that $\bar{\phi}_1, \dots, \bar{\phi}_s$ is a basis for a subspace of $\mathbb{R}^r$ which is linearly independent of $\mathscr{N}$, we know that the columns of $\mathbf{U}$ must be linearly independent. Since rank$(\mathbf{U}) = s$, we have that the nullity of the mapping $\alpha \mapsto \alpha\mathbf{U}$ is $2^{|N_0|} - s \geq 1$, where the inequality follows from Lemma 2. Finally, as the entries of $\mathbf{U}$ are all integers, it is possible to find a basis $\{\alpha^{(n)}\}_{n=1}^{2^{|N_0|}-s}$ for the null space of this mapping with each $\alpha^{(n)}$ having only integer entries. □

Definition of an estimator. In the estimation of ε_0, we will use some polynomial equations of the form $Q(M) = 0$, provided they satisfy some suitable conditions. To better understand these conditions, let us consider the application of the relations found in Lemma 3 to the vector

$$M_{\theta,\varepsilon',\varepsilon} = M_{\theta,\varepsilon'} A_{\varepsilon}^{-1} = M_{\theta} A_{\varepsilon'} A_{\varepsilon}^{-1}. \tag{2.20}$$

The case when $\varepsilon' = \varepsilon_0$ is of special interest. Also setting $\varepsilon = \varepsilon_0$ shows that $M_{\theta,\varepsilon_0,\varepsilon_0} = M_\theta$ and hence $\varepsilon = \varepsilon_0$ is a solution of

$$Q(M_{\theta,\varepsilon',\varepsilon}) = 0 \tag{2.21}$$

when $\varepsilon' = \varepsilon_0$.

As already mentioned in the Introduction, for any $\varepsilon' = \varepsilon_0 \in [0, \frac{1}{2})$, there will typically be several values of ε satisfying (2.21). The examples presented in Section 4 suggest that perhaps it is always possible to find equations (2.21) for which ε_0 is the smallest real root in ε when $\varepsilon' = \varepsilon_0$. Motivated by these examples, we will take our "suitable conditions" on the relations (2.17) to be some variant of requiring that ε_0 is the smallest real root in ε of (2.21) when $\varepsilon' = \varepsilon_0$. Actually, it will turn out to be sufficient to require that 0 is a single root in ε, and the only root in $(-\infty, 0]$, of (2.21) when $\varepsilon' = 0$. We mention here, primarily to establish terminology which will be used in the treatment of the examples (see Section 4), that one way for a relation Q to fail to satisfy this condition is for $Q(M_{\theta,\varepsilon_0,\varepsilon})$ to be identically zero in ε for some $\theta \in \Theta$ and $\varepsilon_0 \in [0, \frac{1}{2})$. Whenever this occurs, we say that Q is a *null relation*.

If we already knew the interaction θ_0, then we could further specialize (2.21) to the polynomial equation $Q(M_{\theta_0,\varepsilon',\varepsilon}) = 0$ for $\varepsilon' = \varepsilon_0$ (or for $\varepsilon' = 0$). However, it is because we suppose ourselves to be initially ignorant of the choice of $\theta_0 \in \Theta$ that we use the relations of Lemma 3 which are valid for *all* $\theta \in \Theta$. Actually, an examination of the proof of Lemma 3 shows that the relations (2.17) are satisfied for all $\theta \in \mathbb{R}^s$. The reason for keeping track of Θ is that it is simpler to verify the above-mentioned suitable conditions if we are allowed to use the a priori knowledge that the interaction belongs to the region Θ in the interaction space $\mathbb{R}^s$. In fact, when we subsequently turn to the estimation of θ_0, we will often make that estimation in the context of the larger space $\mathbb{R}^s$.

CONSISTENCY THEOREM. *Let $\mathscr{C}$ and Θ be given, let $\varepsilon_0 \in [0, \frac{1}{2})$ and $\theta_0 \in \Theta$ and let Q be a polynomial relation. Suppose that $Q(M_\theta) = 0$, for all $\theta \in \Theta$, and define*

$$M_{\theta,\varepsilon',\varepsilon} = M_\theta A_{\varepsilon'} A_{\varepsilon}^{-1}.$$

Further suppose that, for all $\theta \in \Theta$, the equation

$$Q(M_{\theta,0,\varepsilon}) = 0 \tag{2.22}$$

admits $\varepsilon = 0$ as a single root and has no real roots in $(-\infty, 0)$. Then for any $\gamma > 0$:

(i) *The smallest real root in $[-\gamma, 1]$, $\hat{\varepsilon}^{(\Lambda)}(y)$ say, of*

$$Q(M_{\Lambda,y,\varepsilon}) = 0, \tag{2.23}$$

where

$$M_{\Lambda,y,\varepsilon} = M_{\Lambda,y} A_\varepsilon^{-1},$$

is such that

$$\lim_{\Lambda \uparrow \mathbb{Z}^d} \hat{\varepsilon}^{(\Lambda)}(y) = \varepsilon_0,$$

(ii) *The vector* $\hat{\theta}^{(\Lambda)}(y)$ *which maximizes (in* $\mathbb{R}^s$*)*

$$\text{GMPL}(\theta, \Lambda, y) = \prod_{(x_0 \vee \xi) \in X_{\overline{N}_0}} [\pi_\theta(x_0 \mid \xi)]^{M_{\Lambda,y,\hat{\varepsilon}^{(\Lambda)}(y)}(x_0 \vee \xi)} \tag{2.24}$$

converges pointwise to θ_0 [*i.e.*, $\lim_{\Lambda \uparrow \mathbb{Z}^d} \hat{\theta}^{(\Lambda)}(y)(\eta) = \theta_0(\eta)$, *for all* $\eta \in X_C$ *and for all* $C \in \mathscr{C}$] *with* $P_{\theta_0,\varepsilon_0}$-*probability* 1.

3. Consistency. The consistency theorem is proven in this section after three introductory lemmas concerning the matrix A_ε. The first of these lemmas indicates that the noise transformation is well described by A_{ε_0}. We choose a collection of cliques $\mathscr{C}$ and a set of interaction parameters $\Theta \subset \mathbb{R}^s$, which are fixed throughout this section.

For $x \in X$, $z \in Z$ and $\Lambda \subset \mathbb{Z}^d$, define the matrix

$$A_{\Lambda,x,z}(\bar{\xi}^{(1)}, \bar{\xi}^{(2)}) = \begin{cases} \left[|\Lambda| M_{\Lambda,x}(\bar{\xi}^{(1)})\right]^{-1} \sum_{i \in A} \mathbf{1}\left[x_{\overline{N}_i}^{(\Lambda)} = \bar{\xi}^{(1)}, (x \cdot z)_{\overline{N}_i}^{(\Lambda)} = \bar{\xi}^{(2)}\right], & \text{if } M_{\Lambda,x}(\bar{\xi}^{(1)}) \neq 0, \\ 0, & \text{if } M_{\Lambda,x}(\bar{\xi}^{(1)}) = 0. \end{cases} \tag{3.1}$$

Note that if $y = x \cdot z$, then

$$M_{\Lambda,y} = M_{\Lambda,x} \cdot A_{\Lambda,x,z}. \tag{3.2}$$

LEMMA 4. *For all* $\theta_0 \in \Theta$, $\varepsilon_0 \in [0,1]$ *and* $\mu_{\theta_0} \in \mathscr{M}_{\theta_0}$,

$$\lim_{\Lambda \uparrow \mathbb{Z}^d} A_{\Lambda,x,z} = A_{\varepsilon_0} \tag{3.3}$$

with $P_{\theta_0,\varepsilon_0}$-*probability* 1.

PROOF. Let $x \in X$, $\bar{\xi}^{(1)} \in X_{\overline{N}_0}$ and define $S_x(\bar{\xi}^{(1)}) = \{i \in \mathbb{Z}^d\colon x_{\overline{N}_i} = \bar{\xi}^{(1)}\}$. As ν_{ε_0} is a Bernoulli measure and $\mu_{\theta_0}(\bar{\xi}^{(1)}) > 0$, for any $\bar{\xi}^{(3)} \in Z_{\overline{N}_0}$, we have

$$\lim_{\Lambda \uparrow \mathbb{Z}^d} |S_x(\bar{\xi}^{(1)}) \cap \Lambda|^{-1} \sum_{i \in S_x(\bar{\xi}^{(1)}) \cap \Lambda} \mathbf{1}\left[z_{\overline{N}_i} = \bar{\xi}^{(3)}\right] = \nu_{\varepsilon_0}(\bar{\xi}^{(3)}) \tag{3.4}$$

with ν_{ε_0}-probability 1. Since we assumed that μ_{θ_0} is ergodic,

$$\lim_{\Lambda \uparrow \mathbb{Z}^d} |\Lambda|^{-1} |S_x(\bar{\xi}^{(1)}) \cap \Lambda| = \lim_{\Lambda \uparrow \mathbb{Z}^d} M_{\Lambda,x}(\bar{\xi}^{(1)}) = M_{\theta_0}(\bar{\xi}^{(1)})$$

with μ_{θ_0}-probability 1. Therefore, writing $\bar{\xi}^{(2)} = \bar{\xi}^{(1)} \cdot \bar{\xi}^{(3)}$ and letting $y_{\bar{N}_i}^{(\Lambda)} = (x \cdot z)_{\bar{N}_i}^{(\Lambda)}$ for some $z \in Z$, we have

$$\begin{aligned}
&\lim_{\Lambda \uparrow \mathbb{Z}^d} \frac{1}{|\Lambda|} \sum_{i \in \Lambda} \mathbf{1}\Big[x_{\bar{N}_i}^{(\Lambda)} = \bar{\xi}^{(1)}, y_{\bar{N}_i}^{(\Lambda)} = \bar{\xi}^{(2)} \Big] \\
&\quad = \lim_{\Lambda \uparrow \mathbb{Z}^d} \frac{1}{|\Lambda|} \sum_{i \in S_x(\bar{\xi}^{(1)}) \cap \Lambda} \mathbf{1}\big[z_{\bar{N}_i} = \bar{\xi}^{(3)} \big] \\
&\quad = \nu_{\varepsilon_0}\big(\bar{\xi}^{(3)}\big) M_{\theta_0}\big(\bar{\xi}^{(1)}\big)
\end{aligned} \tag{3.5}$$

with $P_{\theta_0,\varepsilon_0}$-probability 1. It then follows that $\lim_{\Lambda \uparrow \mathbb{Z}^d} A_{\Lambda,x,z}(\bar{\xi}^{(1)}, \bar{\xi}^{(2)}) = \nu_{\varepsilon_0}(\bar{\xi}^{(3)})$, again with $P_{\theta_0,\varepsilon_0}$-probability 1, and this concludes the proof since $\nu_\varepsilon(\bar{\xi}^{(3)}) = \varepsilon^{D(\bar{\xi}^{(1)},\bar{\xi}^{(2)})}(1-\varepsilon)^{|\bar{N}_0| - D(\bar{\xi}^{(1)},\bar{\xi}^{(2)})}$. □

Lemma 4 was formulated for $\varepsilon_0 \in [0,1]$. However, A_ε is defined for any $\varepsilon \in \mathbb{R}$ and we now proceed to describe properties of A_ε for generic values of ε.

LEMMA 5. *For $a, b \in \mathbb{R}$, let $A_{a,b}$ be the matrix defined by*

$$A_{a,b}\big(\bar{\xi}^{(1)}, \bar{\xi}^{(2)}\big) = a^{D(\bar{\xi}^{(1)},\bar{\xi}^{(2)})} b^{|\bar{N}_0| - D(\bar{\xi}^{(1)},\bar{\xi}^{(2)})},$$

for any $\bar{\xi}^{(1)}, \bar{\xi}^{(2)} \in X_{\bar{N}_0}$. Then

$$A_{\varepsilon,1-\varepsilon} = A_\varepsilon, \tag{3.6}$$

$$A_{0,1} = I \quad (\textit{the identity matrix}), \tag{3.7}$$

$$A_{ab} \cdot A_{cd} = A_{ad+bc,\, ac+bd} \tag{3.8}$$

and

$$\begin{aligned}
A_\varepsilon^{-1} &= A_{\varepsilon,1-\varepsilon}^{-1} \\
&= A_{\varepsilon/(2\varepsilon-1),(\varepsilon-1)/(2\varepsilon-1)} = (2\varepsilon - 1)^{-|\bar{N}_0|} A_{\varepsilon,\varepsilon-1} \quad \textit{for } \varepsilon \neq \tfrac{1}{2}.
\end{aligned} \tag{3.9}$$

PROOF. Properties (3.6) and (3.7) are immediate and (3.9) follows from (3.8) and the homogeneity of $A_{a,b}$ in (a,b). Writing D_{ij} as shorthand for $D(\bar{\xi}^{(i)}, \bar{\xi}^{(j)})$, we verify property (3.8) as follows:

$$\begin{aligned}
&\sum_{\bar{\xi}^{(3)} \in X_{\bar{N}_0}} A_{ab}\big(\bar{\xi}^{(1)}, \bar{\xi}^{(3)}\big) A_{cd}\big(\bar{\xi}^{(3)}, \bar{\xi}^{(2)}\big) \\
&\quad = \sum_{\bar{\xi}^{(3)} \in X_{\bar{N}_0}} a^{D_{1,3}} b^{|\bar{N}_0| - D_{1,3}} c^{D_{3,2}} d^{|\bar{N}_0| - D_{3,2}} \\
&\quad = \sum_{l=0}^{|\bar{N}_0| - D_{1,2}} \binom{|\bar{N}_0| - D_{1,2}}{l} (ac)^l (bd)^{|\bar{N}_0| - D_{1,2} - l} \sum_{m=0}^{D_{1,2}} \binom{D_{1,2}}{m} (ad)^m (bc)^{D_{1,2} - m} \\
&\quad = (ac + bd)^{|\bar{N}_0| - D_{1,2}} (ad + bc)^{D_{1,2}} = A_{ad+bc,\, ac+bd}\big(\bar{\xi}^{(1)}, \bar{\xi}^{(2)}\big). \qquad \square
\end{aligned}$$

The next lemma describes how to use the condition on the roots of (2.22) to obtain a related condition on the roots of (2.21) when $\varepsilon' = \varepsilon_0$.

LEMMA 6. *Let $\varepsilon_0 \in [0, \frac{1}{2})$ and let Q be a polynomial relation satisfying $Q(M_\theta) = 0$. Suppose that, for all $\theta \in \Theta$, the equation $Q(M_{\theta,0,\varepsilon}) = 0$ admits $\varepsilon = 0$ as a single root and has no real roots in $(-\infty, 0)$. Then, for all $\theta \in \Theta$, the equation*

$$Q(M_{\theta,\varepsilon_0,\varepsilon}) = 0 \tag{3.10}$$

admits $\varepsilon = \varepsilon_0$ as a single root and has no real roots in $(-\infty, \varepsilon_0)$.

PROOF. Fix any $\theta \in \Theta$. It follows from Lemma 5 that

$$M_{\theta,0,\varepsilon'} = (2\varepsilon_0 - 1)^{-|\overline{N}_0|} M_\theta A_{\varepsilon', \varepsilon' - 1}.$$

Similarly,

$$\begin{aligned} M_{\theta,\varepsilon_0,\varepsilon''} &= (2\varepsilon_0 - 1)^{-|\overline{N}_0|} M_\theta A_{\varepsilon'' - \varepsilon_0, \varepsilon'' + \varepsilon_0 - 1} \\ &= -M_\theta A_{(\varepsilon'' - \varepsilon_0)/(1 - 2\varepsilon_0), (\varepsilon'' - \varepsilon_0)/(1 - 2\varepsilon_0) - 1}, \end{aligned}$$

where the last step uses the homogeneity of $A_{a,b}$ in (a, b). Since Q is a homogeneous polynomial, $Q(M_{\theta,0,\varepsilon'})$ has a root of order n at $\varepsilon' = \tilde{\varepsilon}$ iff $Q(M_{\theta,\varepsilon_0,\varepsilon''})$ has a root of order n at $\varepsilon'' = (1 - 2\varepsilon_0)\tilde{\varepsilon} + \varepsilon_0$. The hypotheses of the lemma now guarantee that (3.10) has a single root at $\varepsilon = \varepsilon_0$ and it cannot have any roots in $(-\infty, \varepsilon_0)$. □

We are now ready to prove the consistency of the estimators.

PROOF OF THE CONSISTENCY THEOREM. For $\varepsilon \in \mathbb{R}$, $x \in X$, $z \in Z$ and $\Lambda \subset \mathbb{Z}^d$, let $M_{\Lambda,x,z,\varepsilon} = M_{\Lambda,x} A_{\Lambda,x,z} A_\varepsilon^{-1}$, and for $\gamma, a, b > 0$, let $R_{\gamma,a,b} = \{t \in \mathbf{C}: -\gamma \le \mathrm{Re}(t) \le a, |\mathrm{Im}(t)| \le b\}$. For any $\gamma, b > 0$, $a \in (\varepsilon_0, \frac{1}{2})$ and $\bar{\xi} \in X_{\overline{N}_0}$, we claim that

$$\lim_{\Lambda \uparrow \mathbb{Z}^d} \sup_{\varepsilon \in R_{\gamma,a,b}} |M_{\Lambda,x,z,\varepsilon}(\bar{\xi}) - M_{\theta_0,\varepsilon_0,\varepsilon}(\bar{\xi})| = 0 \tag{3.11}$$

with $P_{\theta_0,\varepsilon_0}$-probability 1. In fact, for any $\bar{\xi} \in X_{\overline{N}_0}$,

$$\begin{aligned} &\sup_{\varepsilon \in R_{\gamma,a,b}} |M_{\Lambda,x,z,\varepsilon}(\bar{\xi}) - M_{\theta_0,\varepsilon_0,\varepsilon}(\bar{\xi})| \\ &\quad \le 2^{|\overline{N}_0|} \max_{\bar{\xi}^{(1)} \in X_{\overline{N}_0}} |M_{\Lambda,x,z}(\bar{\xi}^{(1)}) - M_{\theta_0,\varepsilon_0}(\bar{\xi}^{(1)})| \cdot \sup_{\substack{\varepsilon \in R_{\gamma,a,b} \\ \bar{\xi}^{(1)} \in X_{\overline{N}_0}}} |A_\varepsilon^{-1}(\bar{\xi}^{(1)}, \bar{\xi})| \\ &\quad \le 2^{2|\overline{N}_0|} \biggl\{ \max_{\bar{\xi}^{(1)}, \bar{\xi}^{(2)} \in X_{\overline{N}_0}} |A_{\Lambda,x,z}(\bar{\xi}^{(2)}, \bar{\xi}^{(1)}) - A_{\varepsilon_0}(\bar{\xi}^{(2)}, \bar{\xi}^{(1)})| \\ &\qquad + \max_{\bar{\xi}^{(2)} \in X_{\overline{N}_0}} |M_{\Lambda,x}(\bar{\xi}^{(2)}) - M_{\theta_0}(\bar{\xi}^{(2)})| \biggr\} \cdot \sup_{\substack{\varepsilon \in R_{\gamma,a,b} \\ \bar{\xi}^{(1)} \in X_{\overline{N}_0}}} |A_\varepsilon^{-1}(\bar{\xi}^{(1)}, \bar{\xi})|, \end{aligned}$$

so that (3.11) follows from the ergodic theorem, Lemma 4 and the continuity of the mapping $\varepsilon \mapsto A_{\varepsilon}^{-1}$.

For all $\varepsilon_0 \in [0, \frac{1}{2})$, it follows from the assumptions about the roots of (2.22) and from Lemma 6 that $\varepsilon = \varepsilon_0$ is a single root of $Q(M_{\theta_0,\varepsilon_0,\varepsilon}) = 0$ and that this equation has no other real root in $(-\infty, \varepsilon_0)$. Since $Q(M_{\theta_0,\varepsilon_0,\varepsilon})$ is a polynomial in ε, for every sufficiently small $\delta > 0$,

$$\operatorname{sgn}\left[Q\left(M_{\theta_0,\varepsilon_0,\varepsilon_0-\delta}\right)\right] = -\operatorname{sgn}\left[Q\left(M_{\theta_0,\varepsilon_0,\varepsilon_0+\delta}\right)\right] \tag{3.12}$$

and there exists $c = c(\delta) > 0$ such that

$$|Q\left(M_{\theta_0,\varepsilon_0,\varepsilon_0-\delta}\right)|, |Q\left(M_{\theta_0,\varepsilon_0,\varepsilon_0+\delta}\right)| \geq c \tag{3.13}$$

and

$$\inf_{\varepsilon \in R_{\gamma,\varepsilon_0-\delta,c}} |Q\left(M_{\theta_0,\varepsilon_0,\varepsilon}\right)| \geq c. \tag{3.14}$$

It follows from (3.11) and the continuity of Q in its arguments that, with $P_{\theta_0,\varepsilon_0}$-probability 1, there is a Λ large enough that $Q(M_{\Lambda,x,z,\varepsilon})$ satisfies (3.12)–(3.14) with c replaced by $c/2$ in (3.13) and (3.14). As $Q(M_{\Lambda,x,z,\varepsilon})$ also is a polynomial in ε, this implies that $Q(M_{\Lambda,x,z,\varepsilon}) = 0$ has at least one root in $[\varepsilon_0 - \delta, \varepsilon_0 + \delta]$, and no other roots in $R_{\gamma,\varepsilon_0-\delta,c/2}$. Therefore, $\hat{\varepsilon}^{(\Lambda)}(y)$ can be taken to be the smallest (real) root in $[-\gamma, \varepsilon_0 + \delta]$ of $Q(M_{\Lambda,y,\varepsilon}) = 0$, and letting $\delta \downarrow 0$ shows that

$$\hat{\varepsilon}^{(\Lambda)}(y) \to \varepsilon_0 \quad \text{as } \Lambda \uparrow \mathbb{Z}^d, \tag{3.15}$$

with $P_{\theta_0,\varepsilon_0}$-probability 1. This estimator for ε_0 allows us to "clean," that is, reconstruct, the probabilities of complete local patterns in μ_{θ_0}. For all $\bar{\xi} \in X_{\overline{N}_0}$,

$$\begin{aligned} &|M_{\Lambda,y,\hat{\varepsilon}^{(\Lambda)}(y)}(\bar{\xi}) - M_{\theta_0}(\bar{\xi})| \\ &\quad \leq 2^{|\overline{N}_0|}\left\{\max_{\bar{\xi}^{(1)} \in X_{\overline{N}_0}} |A_{\hat{\varepsilon}^{(\Lambda)}(y)}^{-1}(\bar{\xi}^{(1)}, \bar{\xi}) - A_{\varepsilon_0}^{-1}(\bar{\xi}^{(1)}, \bar{\xi})|\right. \\ &\qquad \left. + \max_{\bar{\xi}^{(1)} \in X_{\overline{N}_0}} |M_{\Lambda,y}(\bar{\xi}^{(1)}) - M_{\theta_0,\varepsilon_0}(\bar{\xi}^{(1)})| \cdot \max_{\bar{\xi}^{(1)} \in X_{\overline{N}_0}} |A_{\varepsilon_0}^{-1}(\bar{\xi}^{(1)}, \bar{\xi})|\right\}, \end{aligned}$$

so that

$$M_{\Lambda,y,\hat{\varepsilon}^{(\Lambda)}(y)} \to M_{\theta_0} \tag{3.16}$$

with $P_{\theta_0,\varepsilon_0}$-probability 1, by the continuity of the mapping $\varepsilon \mapsto A_{\varepsilon}^{-1}$, (3.15) and (3.11).

We now will use a maximum pseudo-likelihood method to estimate θ_0. In the case of no noise [Besag (1977)], the method uses the function

$$\begin{aligned} \operatorname{MPL}(\theta, \Lambda, x) &= \prod_{i \in \Lambda} \pi_\theta\left(x_i \mid x_{N_i}^{(\Lambda)}\right) \\ &= \prod_{(x_0 \vee \xi) \in X_{\overline{N}_0}} \pi_\theta(x_0 \mid \xi)^{M_{\Lambda,x}(x_0 \vee \xi)}, \end{aligned} \tag{3.17}$$

which converges to

$$\text{MPL}(\theta, \theta_0) = \prod_{(x_0 \vee \xi) \in X_{\bar{N}_0}} \pi_\theta(x_0 \mid \xi)^{M_{\theta_0}(x_0 \vee \xi)} \tag{3.18}$$

with μ_{θ_0}-probability 1, as $\Lambda \uparrow \mathbb{Z}^d$. Our method uses (2.24) instead of (3.17). We first prove the convergence of GMPL(θ, Λ, y) to MPL(θ, θ_0) when θ belongs to some compact set $\Theta_c \subset \mathbb{R}^s$. Note that here θ is not restricted to Θ. The functions $\theta \mapsto \pi_\theta(x_0 \mid \xi)$ are continuous and strictly positive. Therefore,

$$\begin{aligned}
&\sup_{\theta \in \Theta_c} |\log \text{GMPL}(\theta, \Lambda, y) - \log \text{MPL}(\theta, \theta_0)| \\
&\quad \le 2^{|\bar{N}_0|} \sup_{\bar{\xi} \in X_{\bar{N}_0}} |M_{\Lambda, y, \hat{\varepsilon}^{(\Lambda)}(y)}(\bar{\xi}) - M_{\theta_0}(\bar{\xi})| \sup_{\substack{\theta \in \Theta_c \\ (x_0 \vee \xi) \in X_{\bar{N}_0}}} |\log(\pi_\theta(x_0 \mid \xi))| \\
&\quad \le K(\Theta_c) \sup_{\bar{\xi} \in X_{\bar{N}_0}} |M_{\Lambda, y, \hat{\varepsilon}^{(\Lambda)}(y)}(\bar{\xi}) - M_{\theta_0}(\bar{\xi})|,
\end{aligned}$$

with $K(\Theta_c)$ a suitable constant depending on Θ_c, so that

$$\sup_{\theta \in \Theta_c} |\log \text{GMPL}(\theta, \Lambda, y) - \log \text{MPL}(\theta, \theta_0)| \to 0 \tag{3.19}$$

by (3.16), with $P_{\theta_0, \varepsilon_0}$-probability 1.

Next, we show that log GMPL(θ, Λ, y) is a strictly concave function of θ. Let θ, $\theta' \in \mathbb{R}^s$, with $\theta' \ne 0$. Then

$$\begin{aligned}
&\frac{d}{dt} \log \text{GMPL}(\theta + t\theta', \Lambda, y) \\
&\quad = \sum_{(x_0 \vee \xi) \in X_{\bar{N}_0}} M_{\Lambda, y, \hat{\varepsilon}^{(\Lambda)}(y)}(x_0 \vee \xi) \pi_{\theta + t\theta'}(-x_0 \mid \xi) \theta' \\
&\qquad \times [U_0(-x_0 \vee \xi) - U_0(x_0 \vee \xi)]
\end{aligned} \tag{3.20}$$

and

$$\begin{aligned}
&\frac{d^2}{dt^2} \log \text{GMPL}(\theta + t\theta', \Lambda, y) \\
&\quad = -\sum_{(x_0 \vee \xi) \in X_{\bar{N}_0}} M_{\Lambda, y, \hat{\varepsilon}^{(\Lambda)}(y)}(x_0 \vee \xi) \pi_{\theta + t\theta'}(-x_0 \mid \xi) \pi_{\theta + t\theta'}(x_0 \mid \xi) \\
&\qquad \times (\theta' \cdot [U_0(-x_0 \vee \xi) - U_0(x_0 \vee \xi)])^2.
\end{aligned} \tag{3.21}$$

Since $\theta' \ne 0$ and $M_{\theta_0}(\bar{\xi}) > 0$, it follows from (2.7), (3.16) and (3.21) that, for all $\theta \in \mathbb{R}^s$, log GMPL(θ, Λ, y) is strictly concave with $P_{\theta_0, \varepsilon_0}$-probability 1 if Λ is sufficiently large.

Finally, note that (3.20) and (3.21) also hold (with M_{θ_0} replacing $M_{\Lambda, y, \hat{\varepsilon}^{(\Lambda)}(y)}$) if MPL$(\theta, \theta_0)$ replaces GMPL(θ, Λ, y). In this case, we may set $\theta = \theta_0$ in (3.20) and use (2.7) to see that log MPL(θ, θ_0) has a unique maximum and that it achieves this maximum at $\theta = \theta_0$. Therefore, by taking Θ_c such that $\theta_0 \in \Theta_c$, it follows from (3.19) that, with $P_{\theta_0, \varepsilon_0}$-probability 1, if Λ is large enough, log GMPL(θ, Λ, y) has a unique maximum in $\mathbb{R}^s$ at $\hat{\theta}^{(\Lambda)}(y) \in \Theta_c$ and that $\hat{\theta}^{(\Lambda)}(y) \to \theta_0$ as $\Lambda \uparrow \mathbb{Z}^d$. □

The consistency theorem was formulated for the (homogeneous) polynomial relations. Lemma 3 indicates how to produce at least one polynomial equation of the type (2.17). However, it may happen that all of the polynomial relations which are obtained in this fashion are *null relations* in that each is identically zero in ε for certain values if $\theta \in \Theta$. In such circumstances, it may be possible to use further insight into the model to obtain additional nonnull (or *effective*) relations. An example is given in Section 4 (for the general nearest-neighbor Markov random field on $\mathbb{Z}$). Loosely speaking, the polynomial relations which we have shown how to construct in Lemma 3 can fail to be effective if either Θ is taken to be an unreasonably large subset of $\mathbb{R}^s$ or there are some special symmetries in the process. In the latter case, it is not unreasonable to expect that the presence of the many symmetries necessary to make all of these polynomial relations null should lead to (enough of) an exact solution of the model to enable one to construct some additional relations which are not null. It is also conceivable that some these new relations may not be polynomial relations in the sense defined above, in that they may be nonhomogeneous. Actually, in the example of Section 4 mentioned above, we are able to find two effective relations: one is homogeneous (and thus is covered by the consistency theorem); the other is not.

To be able to handle situations in which we wish to consider nonhomogeneous polynomials, we point out that if no use is made of Lemma 6 and the homogeneity of Q, then the proof of the consistency theorem yields the following result.

COROLLARY. *Let $\mathscr{C}$ and Θ be given and let $\varepsilon_0 \in [0, \frac{1}{2})$ and $\theta_0 \in \Theta$. Suppose that $Q = Q(M)$ is a polynomial in the entries of M such that, for every $\theta \in \Theta$, $Q(M_\theta) = 0$ and the only solution $\varepsilon \in (-\infty, \varepsilon_0]$ of*

$$Q(M_{\theta,\varepsilon_0,\varepsilon}) = 0$$

is a single root at ε_0. Then for any $\gamma > 0$:

(i) *The smallest real root in $[-\gamma, 1]$, call it $\hat{\varepsilon}^{(\Lambda)}(y)$, of*

$$Q(M_{\Lambda,y,\varepsilon}) = 0$$

is such that

$$\lim_{\Lambda \uparrow \mathbb{Z}^d} \hat{\varepsilon}^{(\Lambda)}(y) = \varepsilon_0.$$

(ii) *The vector $\hat{\theta}^{(\Lambda)}(y)$ which maximizes*

$$\mathrm{GMPL}(\theta, \Lambda, y) = \prod_{(x_0 \vee \xi) \in X_{\bar{N}_0}} [\pi_\theta(x_0 \mid \xi)]^{M_{\Lambda,y,\hat{\varepsilon}^{(\Lambda)}(y)}(x_0 \vee \xi)}$$

in $\mathbb{R}^s$ converges pointwise to θ_0 with $P_{\theta_0,\varepsilon_0}$-probability 1.

4. Example in one dimension. The reader should be able to read this section right after the Introduction (comments between brackets are for the benefit of those who have also read Sections 2 and 3), occasionally looking at some parts of Sections 2 and 3 when indicated.

In this section we consider the one-dimensional nearest-neighbor Markov random field on $\{-1, 1\}^{\mathbb{Z}}$. Stationary one-dimensional nearest-neighbor Markov random fields are in one-to-one correspondence with a collection of stationary one-step Markov chains, the latter defined by means of a two-parameter transition matrix $\begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$, with a, $b \in (0,1)$. The Markov chains with $a = 0$, $a = 1$, $b = 0$ or $b = 1$ do not correspond to Markov random fields because each of these chains prohibits some pattern of a pair of colors, which corresponds to having an interaction which is infinite on that pattern. We shall further assume that $a \neq 1 - b$, which is the same as requiring that the Markov chain not be a Bernoulli process, since in that case the interaction is not identifiable. We can use the correspondence between Markov random fields and Markov chains as a method for writing down the probabilities of the specification of colors in some finite array as explicit functions of the interaction ϕ_0. This explicit functional dependence is called an *exact solution* of the model.

The existence of an exact solution should mean that the estimation of the parameters θ_0 and ε_0 is not overly difficult, and this presentation serves mainly to illustrate our method in a simple case and also to offer an explicit estimator that the reader might want to compare with his/her preferred one. It would now be easy to obtain our polynomial relations from the exact solution (eventually we will do this, and the impatient reader can go directly to the subsection on the general nearest-neighbor Markov random fields below) but we choose instead to begin working from the general setup of Markov random fields, using interactions between pixel colors (along the lines of Section 2). In this way we shall also illustrate what one does when the exact solution is not available.

Before we go on, let us mention some possible alternative estimators which could be used in this situation. First, as the necessary probabilities can be easily obtained from the exact solution, it appears that one could try to use the maximum likelihood estimator to estimate a, b and ε simultaneously. In practice this may not work. In our investigation of symmetric Markov chains (i.e., $a = b$) we found that, for small images, the maximum likelihood estimator was afflicted with a strong "endpoint" problem: it always returned an estimate which had $a = 0$, $a = \frac{1}{2}$, $a = 1$ or $\varepsilon = \frac{1}{2}$ (all of which are ruled out by our assumptions on the model), and for larger images, even though one can explicitly write down the likelihood function, this is a polynomial of high degree in several variables, and it may not be easy to locate where it achieves its maximum. Second, in situations where maximum likelihood fails because the estimation is based on incomplete data, one could do the estimation via the EM algorithm, but as we remarked in the Introduction, this iterative method can be computationally quite intensive. Third, if one has the a priori information that the Markov chain is symmetric, then the method of moments, as in Frigessi and Piccioni (1990), turns out to be a computationally easier task. However, it is not clear how to extend the method of moments to the asymmetric case $a \neq b$, and we thus make this our starting point, providing a computationally fast method for estimating the parameter in an asymmetric one-step Markov chain with independent symmetric noise. The general case, with $a, b \in (0,1)$ and $a \neq 1 - b$ (but without the hypothesis that $a \neq b$) will be discussed afterward. At the end of this section we compare (in the symmetric case) our estimator to the moment estimator of Frigessi and Piccioni (1990).

Asymmetric nearest-neighbor Markov random fields. As mentioned above, we begin our treatment by studying the *interactions*, which are functions of the colors in cliques. It is suggested that the reader who has not already done so, should read the definitions in Section 2 up to (2.3). Here cliques are all the elements of $\mathbb{Z}$ and all the pairs (which we can take to be ordered, for simplicity) of nearest-neighbor elements of $\mathbb{Z}$. Modulo translations, there are two cliques in the present model, for a total of six specifications of colors:

$$(+,+),(+,-),(-,+),(-,-),(+),(-). \tag{4.1}$$

Here and in the following we abbreviate by dropping the 1 in $+1$ and -1. It is assumed that a noise level $\varepsilon_0 \in [0, \frac{1}{2})$ and an asymmetric interaction ϕ_0 are fixed but unknown. An interaction is a real-valued function of the six specifications above and it is asymmetric if it is not invariant under the exchange of $+$ and $-$ in the specification of colors above. Our goal is to estimate the parameters ϕ_0 and ε_0, which characterize the Markov random field and the noise process, respectively, by observing the specification of colors in finite portions of the observable image $y \in \{-1,1\}^{\mathbb{Z}}$. This image is distributed according to the product measure $P_{\phi_0,\varepsilon_0} = \mu_{\phi_0} \times \nu_{\varepsilon_0}$, where μ_{ϕ_0} is a Markov random field of ϕ_0 and ν_{ε_0} is the Bernoulli measure with density ε_0. The interaction can only be estimated up to an equivalence class, where two interactions are said to be equivalent if they generate the same Markov random field (see Lemma 1).

The neighborhood of the origin is $\{-1,1\} \subset \mathbb{Z}$, the complete neighborhood of the origin is $\{-1,\mathbf{0},1\} \subset \mathbb{Z}$ and interactions are vectors in $\mathbb{R}^6$. We list the $2^{|N_0|}$ local patterns

$$(++),(+-),(-+),(--) \tag{4.2}$$

and the $2^{|\bar{N}_0|}$ complete local patterns

$$\begin{gathered}(+++),(++-),(-++),(-+-),(+-+),\\(+--),(--+),(---),\end{gathered} \tag{4.3}$$

and we fix these orderings.

The energy function (at the origin) is

$$U^{\phi}\left(\bar{\xi}_{(-1,\mathbf{0},1)}\right) = \phi\left(\bar{\xi}_{(-1,\mathbf{0})}\right) + \phi\left(\bar{\xi}_{(\mathbf{0},1)}\right) + \phi\left(\bar{\xi}_{\mathbf{0}}\right)$$

for any interaction ϕ and any $\bar{\xi}_{(-1,0,1)} \in X_{\bar{N}_0}$. An examination of this energy function leads to the construction of some polynomial relations (following Lemmas 1 and 3).

1. Let $\mathbf{U}(\phi)$ be the vector indexed by $\xi = \xi_{(-1,1)} \in X_{N_0}$ and given by $\mathbf{U}(\phi)(\xi) = U^{\phi}(+ \vee \xi) - U^{\phi}(- \vee \xi)$, where we define $(\pm \vee \xi) = (\xi_{(-1)}, \pm, \xi_{(+1)}) \in X_{\bar{N}_0}$. [For special choices of ϕ, $\mathbf{U}(\phi)$ is a column of the related matrix $\mathbf{U}$ in (2.19).]

   This defines a map
$$\phi = (\phi(++), \phi(+-), \phi(-+), \phi(--), \phi(+), \phi(-)) \mapsto \mathbf{U}(\phi) = \mathbf{U}\phi$$
   from $\mathbb{R}^6$ to $\mathbb{R}^4$ described by the matrix
$$\mathbf{U} = \begin{pmatrix} 2 & -1 & -1 & 0 & 1 & -1 \\ 1 & 0 & 0 & -1 & 1 & -1 \\ 1 & 0 & 0 & -1 & 1 & -1 \\ 0 & 1 & 1 & -2 & 1 & -1 \end{pmatrix} \tag{4.4}$$
   when using the order given in (4.1) and (4.2).
2. The model can be parametrized (see Lemma 1) by first finding a basis of a maximal linear space $\mathscr{S}$ independent of the null space of the matrix $\mathbf{U}$ and then writing the interactions as linear combinations of these basis elements, which can be chosen to be vectors in $\mathbb{Z}^6$. A customary choice in the present case is to form the basis from $\phi_1 = (1, -1, -1, 1, 0, 0)$ and $\phi_2 = (0, 0, 0, 0, 1, -1)$, in which case the two parameters are called β and h. With this choice of parameters, the map $(a, b) \to (\beta, h)$ is somewhat unpleasant and not entirely necessary here. Just for the purpose of showing how to interpret our restrictions on a and b in the language of interactions, we mention [see, e.g., Denny and Wright (1991)] that
$$e^h = \sqrt{\frac{1-a}{1-b}} \quad \text{and} \quad e^{\beta} = \left(\frac{(1-b)(1-a)}{ab}\right)^{1/4}.$$
   Regardless of which basis we choose, we have that $\dim(\mathscr{S}) = 2 < 4 = 2^{|N_0|}$ (as predicted by Lemma 2).
3. We find polynomial relations (see Lemma 3) by solving the equation
$$\alpha\mathbf{U} = 0,$$
   that is, by finding vectors α which are orthogonal to the columns of $\mathbf{U}$. Since $\mathbf{U}$ is a rank 2 [$2 = \dim(\mathscr{S})$] linear transformation into $\mathbb{R}^4$ [$4 = 2^{|N_0|}$], it is possible to find two such independent vectors, for instance,
$$\alpha^{(1)} = (0, 1, -1, 0) \tag{4.5}$$
   and
$$\alpha^{(2)} = (1, -2, 0, 1). \tag{4.6}$$
4. Next, for any $\theta = (\theta(1), \theta(2)) \in \mathbb{R}^2$, let M_θ be the vector indexed by $\bar{\xi} \in X_{\bar{N}_0}$ of the probabilities $M_\theta(\bar{\xi}) = \mu_{\theta(1)\phi_1 + \theta(2)\phi_2}(\bar{\xi})$, where ϕ_1 and ϕ_2 are a basis for $\mathscr{S}$. A pair of polynomial equations can be obtained from (2.16), using

the α's in (4.5) and (4.6):

$$Q_{\alpha^{(1)}}(M) = M(++-)M(--+) - M(+--)M(-++) = 0 \tag{4.7}$$

and

$$Q_{\alpha^{(2)}}(M) = M(+++)(M(+--))^2 M(-+-) - M(+-+)(M(++-))^2 M(---) = 0. \tag{4.8}$$

5. Let $M_{\theta,\varepsilon_0,\varepsilon} = M_\theta A_{\varepsilon_0} A_\varepsilon^{-1}$, where A_ε is the matrix defined in (2.14). It is clear that $\varepsilon = \varepsilon_0$ is a solution of $Q_{\alpha^{(i)}}(M_{\theta,\varepsilon_0,\varepsilon}) = 0$ for $i = 1, 2$, but it is not easy to tell whether the relations $Q_{\alpha^{(i)}}$ are null in the sense that, for some value of θ, $Q_{\alpha^{(i)}}(M_{\theta,\varepsilon_0,\varepsilon}) = 0$ for all ε. To decide whether these relations are null or effective, we now turn to the exact solution, because proceeding without it would prove quite arduous.
6. It is most convenient to change from the original parameters $\theta_0 = (\theta_1, \theta_2)$ to the parameters a and b appearing in the transition matrix for the Markov chain corresponding to the Markov random field. Let us indicate this change of parameters by the transformation $(a, b) \mapsto \theta(a, b)$. We then have

$$\begin{aligned} M_{a,b} &= M_{\theta(a,b)} \\ &= (a+b)^{-1}\big[b(1-a)^2, ab(1-a), ab(1-a), a^2b, ab^2, \\ &\qquad ab(1-b), ab(1-b), a(1-b)^2\big], \end{aligned} \tag{4.9}$$

where we use the ordering of the complete local patterns given in (4.3). A direct calculation (left to the reader) shows that $Q_{\alpha^{(1)}}$ is a null relation. On the other hand, temporarily assuming that $\varepsilon_0 = 0$, we note that $Q_{\alpha^{(2)}}(M_{\alpha(a,b),0,\varepsilon}) = 0$ reduces to

$$\begin{aligned} &\varepsilon(\varepsilon-1)ab(b-a)[a-(1-b)]^2(\varepsilon^2-\varepsilon+ab) \\ &\quad\times\big[(a+b)\varepsilon^2-(a+b)\varepsilon+ab(2-a-b)\big] = 0, \end{aligned} \tag{4.10}$$

where we have ignored various factors of $a + b$ and $2\varepsilon - 1$. Our assumptions on the parameters a and b ($a, b \neq 0$, $a \neq 1 - b$ and $a \neq b$) guarantee that (4.10) is not identically zero. Therefore, we only need to study the roots of the last two factors in order to verify that they cannot be confused with the root at $\varepsilon = \varepsilon_0$ $(= 0)$ when estimating ε_0 from the data. It is easily seen that the real roots (in ε) of last two factors in (4.10) are in (0,1) for all $a, b \in (0,1)$. The homogeneity of the polynomial $Q_{\alpha^{(2)}}$ allows us to conclude from the observation that $\varepsilon = 0$ is (a single root and) the smallest real root of (4.10), that for any $\varepsilon_0 \in [0, \frac{1}{2})$ the root at $\varepsilon = \varepsilon_0$ of $Q_{\alpha^{(2)}}(M_{\theta(a,b),\varepsilon_0,\varepsilon}) = 0$ is (a single root and) the smallest real root (see Lemma 6).
7. We are now ready to define our estimators. Recall that for any given interaction $\phi_0 \in \mathbb{R}^6$ there is a vector $\theta_0 \in \mathbb{R}^2$ with $\theta_0(1)\phi_1 + \phi_0(2)\phi_2 = \phi_0$,

where ϕ_1 and ϕ_2 are the elements of some basis for $\mathscr{S}$, as described in step 2 above. We will estimate ε_0 and θ_0. First, let $y^{(\Lambda)}$ be the periodic extension to $\mathbb{Z}$ of the observed image y_Λ (which is the restriction of the observable image y to Λ) and define $M_{\Lambda,y}$ to be the vector, indexed by $\bar{\xi} \in X_{\overline{N}_0}$, of the empirical frequencies of the specifications of colors $\bar{\xi}$ in $y^{(\Lambda)}$ in Λ [see (2.12)]. Then, form the equation in ε,

$$Q_{\alpha^{(2)}}\left(M_{\Lambda,y} A_\varepsilon^{-1}\right) = 0 \tag{4.11}$$

[as in (2.23)]. Next, find the smallest real root $\hat{\varepsilon}^{(\Lambda)}(y)$ of (4.11) (as we proved in the consistency theorem):

$$\hat{\varepsilon}^{(\Lambda)}(y) \to \varepsilon_0$$

with P_{ϕ_0,ε_0}-probability 1 as $\Lambda \uparrow \mathbb{Z}^d$. Finally, consider the function GMPL: $\mathbb{R}^2 \to \mathbb{R}$ given [as in (2.24)] by

$$\begin{aligned} \theta &\mapsto \mathrm{GMPL}(\theta, \Lambda, y) \\ &= \prod_{(x_0 \vee \xi) \in X_{\overline{N}_0}} \left[\pi_\theta(x_0 \mid \xi)\right]^{M_{\Lambda,y,\hat{\varepsilon}^{(\Lambda)}(y)}(x_0 \vee \xi)}, \end{aligned} \tag{4.12}$$

where $\pi_\theta(x_0 \mid \xi) = M_\theta(x_0 \vee \xi)/(M_\theta(+ \vee \xi) + M_\theta(- \vee \xi))$, for $\xi \in X_{N_0}$. The vector $\hat{\theta}^{(\Lambda)}(y) \in \mathbb{R}^2$ which maximizes GMPL(θ, Λ, y) satisfies [see the consistency theorem (and its proof)]

$$\hat{\theta}^{(\Lambda)}(y) \to \theta_0$$

with P_{ϕ_0,ε_0}-probability 1 as $\Lambda \uparrow \mathbb{Z}^d$, and this concludes the discussion of the asymmetric case.

General nearest-neighbor Markov random fields. We retain our "finite energy" (a, $b \neq 0$) and non-Bernoulli ($a \neq 1 - b$) assumptions on the parameters a and b, but now relax the asymmetry ($a \neq b$) hypothesis. Note that it was essential in (4.10) that $a \neq b$. If we are not given this a priori information, then $Q_{\alpha^{(2)}}$ is also a null relation and unfortunately no effective relations are derived from the most abstract version of our theory. However, one can now go further and add more relations by using the exact solution (4.9) explicitly. An examination of (4.9) yields a pair of equations which are independent of the two null relations (4.7) and (4.8):

$$Q_3(M) = M(++-) - M(-++) = 0 \tag{4.13}$$

and

$$Q_4(M) = M(+++)M(-+-) - (M(++-))^2 = 0. \tag{4.14}$$

Another direct calculation (left to the reader) shows that Q_3 is a null relation. However, it is also readily seen that $Q_4(M_{\theta(a,b),0,\varepsilon}) = 0$ is equivalent (neglecting various factors of $a + b$ and $2\varepsilon - 1$) to

$$\varepsilon(\varepsilon - 1)ab[a - (1 - b)]^2 = 0. \tag{4.15}$$

Therefore, Q_4 is an effective relation, and as above, for any $\varepsilon_0 \in [0, \frac{1}{2})$ and $\phi_0 \in \mathbb{R}^6$, we can define the estimators $\hat{\varepsilon}^{(\Lambda)}$ and $\hat{\theta}^{(\Lambda)}$ to be the smallest (real) root in ε of $Q_4(M_y A_\varepsilon^{-1}) = 0$ and the vector in $\mathbb{R}^2$ maximizing (4.12) with this current value of $\hat{\varepsilon}^{(\Lambda)}$, respectively. These estimators converge (with $P_{\theta_0,\varepsilon_0}$-probability 1) to ε_0 and θ_0, respectively, where θ_0 is related to the interaction ϕ_0 by $\theta_0(1)\phi_1 + \theta_0(2)\phi_2 = \phi_0$. (Although Q_4 was not obtained via the construction described in Lemma 3, it is a polynomial relation, and so the convergence of the estimators is according to the consistency theorem.)

For the sake of completeness we point out that there are eight components to $M_{a,b}$. We know that these eight components are functions of two parameters, they sum to 1 and they satisfy $Q_{\alpha^{(1)}}(M_{a,b}) = Q_{\alpha^{(2)}}(M_{a,b}) = Q_3(M_{a,b}) = Q_4(M_{a,b}) = 0$. Thus we should be able to extract one more relation from the complete solution, and indeed we can. One choice for a fifth independent equation is

$$
\begin{aligned}
Q_5(M) = {} & M(-+-)[M(-+-) + M(+-+)] \\
& \times [M(+++) + M(++-)]^2 \\
& - M^2(++-)M(+-+) = 0.
\end{aligned}
\tag{4.16}
$$

This fifth equation is different from the first four in that it is not homogeneous, but it is still possible to base the estimation on this relation. As a demonstration of how to proceed in the case of a nonhomogeneous polynomial we next show that Q_5 is effective.

A somewhat lengthy calculation shows that $Q_5(M_{\theta(a,b),\varepsilon_0,\varepsilon}) = 0$ reduces to

$$
\begin{aligned}
& (\varepsilon - \varepsilon_0)[\varepsilon - (1-\varepsilon_0)](2\varepsilon - 1)^2 ab[a - (1-b)]^2 \\
& \quad \times [\varepsilon^2 - \varepsilon + ab + (1-4b)\varepsilon_0(1-\varepsilon_0)] \\
& \quad \times \{3(a+b)\varepsilon^3 - [2a + 7b + 5(a-b)\varepsilon_0]\varepsilon^2 \\
& \qquad + [b(5 - a^2 - ab) + (4a - 6b + 4a^2 b + 4ab^2)\varepsilon_0 \\
& \qquad\qquad + (a + b - 4a^2 b - 4ab^2)\varepsilon_0^3]\varepsilon \\
& \qquad + [b(a^2 - 1) + b(1 - 5a^2 + ab)\varepsilon_0 \\
& \qquad\quad + (b - 2a + 8a^2 b - 4ab^2)\varepsilon_0^2 \\
& \qquad\qquad + (a - b - 4a^2 b + 4ab^2)\varepsilon_0^3]\} = 0.
\end{aligned}
\tag{4.17}
$$

We emphasize that because of the nonhomogeneity of Q_5 it is no longer sufficient only to verify that the smallest root in ε of $Q_5(M_{\theta(a,b),0,\varepsilon}) = 0$ is $\varepsilon = 0$. We must show more generally that the smallest root in ε of $Q_5(M_{\theta(a,b),\varepsilon_0,\varepsilon}) = 0$ is $\varepsilon = \varepsilon_0$. Another feature of the lack of homogeneity is that it now becomes critical to remember the denominators $a + b$ and $2\varepsilon - 1$ in $M_{a,b}$ and A_ε^{-1} when computing (4.17), whereas they could have been neglected in the similar computations of (4.10) and (4.15). To see that the penultimate factor in (4.17) does not vanish for any $\varepsilon \le \varepsilon_0$, we note that at $\varepsilon = \varepsilon_0$, this term, its derivative with respect to ε and its second derivative with respect to ε are $ab(2\varepsilon_0 - 1)^2$ (positive), $(2\varepsilon_0 - 1)$ (negative) and 2 (positive), respectively. Hence this factor must be strictly positive for all $\varepsilon \le \varepsilon_0$. A similar line of reasoning shows that the last term in (4.17) is strictly negative for all $\varepsilon \le \varepsilon_0$. Therefore, Q_5 is an effective (albeit nonhomogeneous) polynomial. We can now produce the estimators $\hat{\varepsilon}^{(\Lambda)}$ and $\hat{\theta}^{(\Lambda)}$, as described above for the homogeneous cases, by using $Q_5(M_y A_\varepsilon^{-1}) = 0$ to find $\hat{\varepsilon}^{(\Lambda)}$ from the observed image y_Λ. (The convergence of the estimators is guaranteed by the corollary at the end of Section 3.)

Comparison with other estimators. In this section we present some calculations which compare our estimator with an estimator obtained by the method of moments. For simplicity, we focus on the estimation of the noise parameter ε. The comparison is in the context of a small string of n (≈ 10) pixels, and we will explicitly compute the estimators by analyzing all possible configurations of two allowed colors.

We now indicate the specific choices made for this comparison, but the results should not be too dependent on these details. The original image is an element $x \in X = \{-1, 1\}^n$. The measure μ_a describing the original image is taken to be symmetric, as this is the only case where the method of moments can be used. Note that our estimator suffers no such deficiency. The transition matrix for the original chain is $\begin{pmatrix} 1-a & a \\ a & 1-a \end{pmatrix}$ and the measure is given by

$$\mu_a(x) = a^{|\{i \in \{1,\dots,n-1\}:\, x_i \ne x_{i+1}\}|}(1-a)^{|\{i \in \{1,\dots,n-1\}:\, x_i = x_{i+1}\}|},$$

for $a \in (0,1)$, up to a factor of $\frac{1}{2}$ which we ignore. The noise is described by the Bernoulli measure with density $\varepsilon \in [0, 1/2)$ of -1's, and the product measure describing the observed image $y \in Y$ is indicated by $P_{a,\varepsilon}$.

There are various ways to generate estimators for ε by our method. For instance, by using (4.15) with $b = a$ we see that (4.14) is an effective estimator in the symmetric case. However, in order to avoid inappropriate asymmetries, it is better to use a slightly modified (symmetrized) version:

$$\begin{aligned} Q_{4'}(M) = {} & ((M(++-) + M(-++) + M(+--) + M(--+))/4)^2 \\ & -((M(-+-) + M(+-+))/2) \\ & \times((M(+++) + M(---))/2) = 0. \end{aligned} \tag{4.18}$$

This is also an effective relation because it reduces to (4.15) when evaluated at the vector of probabilities $M_{a,0,\varepsilon}$. Given an image y, we compute the empirical vector M_y by

$$M_y(uvw) = |\{i \in \{1,\dots,n-2\}:\, y_i = u,\, y_{i+1} = v,\, y_{i+2} = w\}|$$

and then solve

$$Q_{4'}(M_y A_\varepsilon^{-1}) = 0, \tag{4.19}$$

with the matrix A_ε as in (2.14). If (4.19) has real roots and the smallest real root is in $[0, \frac{1}{2})$, then we say that the estimator exists, we define it to be that smallest root and we denote it by $\hat{\varepsilon}_{\mathrm{GMPL}}(y)$. The roots of (4.19) which are complex, negative or greater than $\frac{1}{2}$ are eliminated because they do not make sense as possible values of ε, given our assumptions on the noise. When $\varepsilon = \frac{1}{2}$ the parameters are generally not identifiable, and for this reason $\frac{1}{2}$ is also excluded from the possible values taken by the estimator.

To implement an estimator based on the method of moments (MOM), one first computes the estimated moments

$$N_1(y) = \frac{1}{n-1} \sum_{i=1}^{n-1} y_i y_{i+1} \quad \text{and} \quad N_2(y) = \frac{1}{n-2} \sum_{i=1}^{n-2} y_i y_{i+2}.$$

It can be seen from Frigessi and Piccioni (1990) that

$$\hat{\varepsilon}_{\mathrm{MOM}}(y) = \left[1 - \operatorname{sgn}(N_1(y)) N_1(y) / \sqrt{N_2(y)}\,\right] / 2 \tag{4.20}$$

is an estimator for ε. As above (and for similar reasons), we adopt the convention of saying that this estimator exists only if $[N_2(y) > 0$ and$]$ $\hat{\varepsilon}_{\mathrm{MOM}}(y) \in [0, \frac{1}{2})$.

To compare the two estimators we generate all possible images $y \in Y$ and then compute $\hat{\varepsilon}_{\mathrm{GMPL}}(y)$ and $\hat{\varepsilon}_{\mathrm{MOM}}(y)$. The two estimators can be then compared by computing (for j = GMPL, MOM)

$$p_j(a, \varepsilon) = P_{a,\varepsilon}(\hat{\varepsilon}_j \text{ exists})$$

and

$$E_j(a, \varepsilon) = E_{a,\varepsilon}(\hat{\varepsilon}_j \mid \hat{\varepsilon}_j \text{ exists}),$$

where $E_{a,\varepsilon}$ is the expectation with respect to $P_{a,\varepsilon}$. The consistency theorem of Section 3 [resp., Frigessi and Piccioni (1990)] tells us that, for each $a \in (0, 1)$, $a \neq \frac{1}{2}$ and $\varepsilon \in [0, \frac{1}{2})$, $p_{\mathrm{GMPL}}(a, \varepsilon)$ [resp., $p_{\mathrm{MOM}}(a, \varepsilon)$] tends to 1 and $E_{\mathrm{GMPL}}(a, \varepsilon)$ [resp., $E_{\mathrm{MOM}}(a, \varepsilon)$] tends to ε as the number of pixels n diverges. We have computed the (polynomial) functions $p_j(a, \varepsilon)$ and the (rational) functions $E_j(a, \varepsilon)$ for j = GMPL, MOM and $n = 6, 7, 8, 9, 10$. Rather than transcribe the (long) formulas here, we first describe our findings in words and then provide some (crude) quantification.

The functions $p_j(a, \varepsilon)$ (j = GMPL, MOM) are both maximized at the points ($a = 0$, $\varepsilon = 0$) and ($a = 1$, $\varepsilon = 0$) (where these probabilities take the value 1), they tend to their respective infima (when $n = 9$, $35/128$ for both; when $n = 10$, $67/256$ for p_{GMPL} and $37/128$ for p_{MOM}) as either $a \to \frac{1}{2}$ or $\varepsilon \nearrow \frac{1}{2}$ and there are no other local extrema in $[0, 1] \times [0, \frac{1}{2}]$. We remark that these findings are in accordance with the intuition that the estimation should be simplest when the noise is small (ε near 0) and the model more "deterministic" (a near either 0 or 1), and the estimation should become more difficult as the image pixels become symmetric Bernoulli random variables (a or ε near $\frac{1}{2}$). As might be expected from the preceding two sentences, for both estimators, $E_j(a, \varepsilon)$ is close to ε in the vicinity of the points ($a = 0$, $\varepsilon = 0$) and ($a = 1$, $\varepsilon = 0$). For ε small and a near $\frac{1}{2}$, $E_j(a, \varepsilon)$ is much larger than ε

(positive bias), and for ε near $\frac{1}{2}$, $E_j(a, \varepsilon)$ is significantly smaller than ε (negative bias).

One way to quantify the results described above so that we may compare the two estimators is to compute the average values of the probabilities of the existence of the estimators and the deviation of the conditional expectation from ε (i.e., the bias) with respect to a uniform (prior) measure on the parameter space $\{(a, \varepsilon): 0 < a < 1,\ a \neq \frac{1}{2},\ 0 \leq \varepsilon < \frac{1}{2}\}$. We will therefore report in Table 1 (for j = GMPL, MOM)

$$\bar{p}_j = 2\int_{(0,1)\times[0,1/2)} p_j(a, \varepsilon)\, da\, d\varepsilon, \tag{4.21}$$

$$\bar{E}_j^{(1)} = 2\int_{(0,1)\times[0,1/2)} |E_j(a, \varepsilon) - \varepsilon|\, da\, d\varepsilon \tag{4.22}$$

and

$$\bar{E}_j^{(2)} = \left(2\int_{(0,1)\times[0,1/2)} \left[E_j(a, \varepsilon) - \varepsilon\right]^2 da\, d\varepsilon\right)^{1/2}. \tag{4.23}$$

The integrals of (4.21) give some indication of the utility of the estimators [the average probability of having a "feasible" estimate], while the integrals of (4.22) and (4.23) serve to quantify the bias.

The results shown in Table 1 give a rough indication that the estimator that we are presenting here performs more or less with the same accuracy as the method of moments. In terms of reducing the magnitude of the bias, our estimator appears a little better than the moment estimator. As far as the probability of existence of the estimators is concerned, the method of moments appears to be better than ours, but we will next report some evidence that indicates that this may not be the case when there are more pixels. First let us comment that the reason for restricting the investigation reported in Table 1 to images of 10 or less pixels is essentially a computing constraint imposed by working on small computers: the length of the calculations grows geometrically with the number of pixels. A computationally less intensive procedure is to simply count the number of observed images y for which either estimator exists; for j = GMPL, MOM we define the frequencies

$$f_j = \left|\left\{y \in Y = \{-1, 1\}^n : \hat{\varepsilon}_j \text{ exists}\right\}\right|. \tag{4.24}$$

TABLE 1

n	$\bar{p}_{\text{GMPL}}$	$\bar{p}_{\text{MOM}}$	$\bar{E}^{(1)}_{\text{GMPL}}$	$\bar{E}^{(1)}_{\text{MOM}}$	$\bar{E}^{(2)}_{\text{GMPL}}$	$\bar{E}^{(2)}_{\text{MOM}}$
6	0.256	0.393	0.110	0.127	0.136	0.159
7	0.315	0.369	0.104	0.147	0.129	0.183
8	0.244	0.350	0.107	0.104	0.133	0.129
9	0.332	0.350	0.118	0.123	0.146	0.152
10	0.321	0.364	0.102	0.097	0.124	0.118

The relative frequencies $f_j/2^n$ can be interpreted as the average of $p_j(a, \varepsilon)$ with respect to some prior distribution which gives rise to the uniform product measure on the space Y of observed images. Since the computation time is significantly shorter for this gauge of the feasibility of the estimators, it is possible for us to treat some cases of more pixels, as we have done in Table 2.

Note that although $\bar{p}_j$ differs in magnitude from $f_j/2^n$ for those strings for which we are able to obtain both numbers, the comparison between $\bar{p}_{\mathrm{GMPL}}$ and $\bar{p}_{\mathrm{MOM}}$ seems similar to that between $f_{\mathrm{GMPL}}/2^n$ and $f_{\mathrm{MOM}}/2^n$. If this correspondence holds also for $n > 10$, then the data at the bottom of Table 2 suggest that there are situations where our estimator has a significantly higher probability of existence than the moment estimator. Our overall conclusion is that the estimator that we are presenting here performs more or less with the same accuracy as the method of moments, and perhaps improves on it for larger images. Of course, this is only a first indication, and more detailed studies and simulations of the behavior of the various estimators for large n are needed.

It is important to mention that, while in this case the complexity of computing the estimators is about the same, in higher dimensions the complexity of the method of moments increases so as to become essentially untreatable. On the other hand, the complexity of our estimator does not substantially change with the dimension of the image: this is a key aspect of our method. What does change then is the difficulty in giving a rigorous proof that the root we indicated (the smallest one) is indeed the correct one, but this burden does not fall on the actual user of the algorithm and it needs to be done only once. However, those proofs do require more elaborate methods, and we present, in a separate paper [Barsky and Gandolfi (1995)], the rigorous verification that the smallest root is the correct one for a simple two-dimensional case (the Ising model). We actually verify the conjecture about the smallest root in the following cases:

1. no external field and the same vertical and horizontal positive interac-

TABLE 2

n	f_{GMPL}	f_{MOM}	$f_{\mathrm{GMPL}}/2^n$	$f_{\mathrm{MOM}}/2^n$
6	12	20	0.188	0.313
7	32	36	0.250	0.281
8	44	68	0.172	0.266
9	140	140	0.273	0.273
10	268	296	0.262	0.289
11	748	528	0.365	0.258
12	1,112	1136	0.271	0.277
13	2,564	2460	0.313	0.300
14	4,708	5832	0.287	0.356
15	12,488	9700	0.381	0.296

tions, both in the presence and absence of phase transition; in these cases the method of moments [see Frigessi and Piccioni (1990)] can also be used;
2. nonzero external field and the same vertical and horizontal positive interactions.

A general polynomial relation that is possibly valid under the mild assumption that the interactions are positive, but with no other constraint, is also presented, and if the conjecture were verified in that case, it would include the earlier mentioned cases 1 and 2 as well [see Barsky and Gandolfi (1995)].

At present we lack a general method to verify our conjecture, but no evidence so far has indicated that it might not hold in the general setup of the present paper. We then consider that if no contrary indication shows up, it might be possible to implement this method in applied problems and to take advantage of its simplicity without waiting for a rigorous proof. To lower the risk of a mistake, one could even use two or three of the estimators produced by our equations at the same time. There are several cases which can be studied next to test the conjecture about the smallest root; in particular, longer range one-dimensional Markov chains, examples in one or two dimensions with more allowed colors per pixel (where the matrix A becomes more complicated). There also remains the possibility that a simple convexity argument will yield a proof of the general conjecture, although if it exists, it has eluded us thus far. In the meantime, we hope that it will be possible to put these estimators to practical use.

Acknowledgment. We would like to thank A. Frigessi, B. Gidas and C. M. Newman for useful conversations.

REFERENCES

BARSKY, D. J. (1995). Some combinatorial aspects of the Ehrenfest diffusion process. Unpublished manuscript.

BARSKY, D. J. and GANDOLFI, A. (1995). A generalized maximum pseudo-likelihood estimator for Markov random fields with noise, II: two-dimensional examples. Unpublished manuscript.

BAXTER, R. J. (1982). *Exactly Solved Models in Statistical Mechanics*. Academic Press, New York.

BESAG, J. (1977). Efficiency of pseudo-likelihood estimation for simple Gaussian fields. *Biometrika* **64** 616–618.

CHALMOND, B. (1987). Image restoration using an estimated Markov model. *Signal Process.* **15** 115–129.

COMETS, F. and GIDAS, B. (1992). Parameter estimation for Gibbs distributions from partially observed data. *Ann. Appl. Probab.* **2** 142–170.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood for incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38.

DENNY, J. L. and WRIGHT, A. L. (1991). Inference about the shape of neighboring points in fields. In *Spatial Statistics and Imaging* (A. Possolo, ed.) 46–54. IMS, Hayward, CA.

FRIGESSI, A. and PICCIONI, M. (1990). Parameter estimation for the two-dimensional Ising fields corrupted by noise. *Stochastic Process. Appl.* **34** 297–311.

GEMAN, S. and MCCLURE, D. E. (1985). Bayesian image analysis: an application to single photon emission tomography. In *Proceedings of the Statistical Computing Section* 12–18. Amer. Statist. Assoc., Alexandria, VA.

GIDAS, B. (1991). Parameter estimation for Gibbs distributions, I: fully observed data. In *Markov Random Fields: Theory and Applications* (R. Chellappa and A. Jain, eds.). Academic Press, New York.

JI, C. and SEYMOUR, L. (1991). On the selection of Markov random fields texture models. Mimeo Series 2062, Dept. Statistics, Univ. North Carolina, Chapel Hill.

MELOCHE, J. and RUBEN, M. Z. (1992). Black and white image restoration. Technical Report 15, Dept. Statistics, Univ. British Columbia.

NEWMAN, C. M. (1987). Decomposition of binary random fields and zeros of partition functions. *Ann. Probab.* **15** 1126–1130.

RUELLE, D. (1978). *Thermodynamic Formalism*. Addison-Wesley, Reading, MA.

YOUNES, L. (1989). Parameter inference for imperfectly observed Gibbsian fields. *Probab. Theory Related Fields* **82** 625–645.

YOUNES, L. (1991). Parameter estimation for imperfectly observed Gibbs fields and some comments on Chalmond's EM Gibbsian algorithm. Preprint.

DEPARTMENT OF MATHEMATICS
CALIFORNIA STATE UNIVERSITY,
SAN MARCOS
SAN MARCOS, CALIFORNIA 92096

DEPARTIMENTO DI MATEMATICA
UNIVERSITA' DI ROMA TOR VERGATA
VIA DELLA RICERCA SCIENTIFICA
00133 ROMA
ITALY