The Annals of Statistics
1992, Vol. 20, No. 1, 455–468

ON CONSISTENCY OF A CLASS OF ESTIMATORS FOR EXPONENTIAL FAMILIES OF MARKOV RANDOM FIELDS ON THE LATTICE

By Francis Comets

Université Paris 7

We prove strong consistency of a class of maximum objective estimators for exponential parametric families of Markov random fields on $\mathbb{Z}^d$, including both maximum likelihood and pseudolikelihood estimators, using large deviation estimates. We also obtain the optimality property for the maximum likelihood estimator in the sense of Bahadur.

1. Introduction. Markov random fields (MRF) on the lattice and, more generally, Gibbs distributions (GD), are known to provide pertinent models for interacting particle systems in statistical mechanics. Recently, they have been used in image processing to describe the local information shared by classes of images; they yield a quantitative interpretation of some low-level tasks as formulated by Geman and Geman (1984, 1987) and Azencott (1987). In this application, parametric estimation for exponential families of MRF is a crucial step. Usual estimation procedures are the maximum likelihood method—with computational drawbacks— and the maximum pseudolikelihood method owing to Besag [(1974) for the original coding method and (1977)].

When the interaction between variables is translation invariant, consistency was proved in a general setup for the maximum likelihood estimator (MLE) [Gidas (1991); see also the pioneer work of Pickard (1987) and the references therein] and consistency of the maximum pseudolikelihood estimator (MPLE) [Geman and Graffigne (1987)] for finite state space [see also Guyon (1986) and Gidas (1986)]. Though the proofs are simple when the underlying (true) distribution is ergodic, complications arise from the rich behavior of Gibbs distributions: Phase transitions occur when there exist many GD for a single value of the parameter. Then some GD are not ergodic, but they have long-range dependence. There may also exist nonstationary GD, even though the interaction is stationary.

In this paper, we give simple proofs of consistency for these estimators in broad generality, with exponential rate of convergence. We will use large deviation estimates for GD. These estimates imply that the spatial average of a sample from any GD is no worse behaved than the same average from the "worse" ergodic GD, and consequently, they overcome the previous difficulties for general models and general estimators. This technique has been used in another collaboration [Comets and Gidas (1992)] to prove consistency of the

Received October 1989; revised February 1991.

AMS 1980 *subject classifications*. Primary 62F10, 62M05; secondary 82A25, 60G60.

Key words and phrases. Maximum likelihood estimator, pseudolikelihood, Markov random field, objective function, Bahadur efficiency, large deviation.

MLE in the context of incomplete observations from a MRF. Our results in this paper do not cover the case of a gaussian MRF, which is studied in great detail by Künsch (1981).

For objective functions which are nice spatial averages, we prove in Section 3 that the maximizers are consistent estimators; this result covers both MLE and MPLE.

Asymptotic normality may fail in the framework of MRF, at least for standard normalization, precisely as an effect of phase transition. Therefore, we formulate optimality in terms of exponential deficiency rate. We show in Section 4 that the MLE is optimal and we discuss the effect of phase transition on the optimal rate in Section 5.

2. Markov random fields and statistical models. Let $\mathscr{X}$ be a Polish space (i.e., a metric complete separable space) and let $\Omega = \mathscr{X}^{\mathbb{Z}^d}$. Then the set $\mathscr{P}(\Omega)$ of all probability measures on Ω, equipped with its weak topology, is a Polish space too. We will denote by $\mathscr{P}_s(\Omega)$ the subset of stationary fields,

$$\mathscr{P}_s(\Omega) = \{Q \in \mathscr{P}(\Omega); Q \circ \tau^i = Q, \forall\, i \in \mathbb{Z}^d\},$$

where τ is the shift operator, $\tau^i\colon \Omega \to \Omega$, $(\tau^i x)_j = x_{i+j}$ for $x \in \Omega$. Let ρ be a probability measure on $\mathscr{X}$ and let $Q_\rho = \rho^{\otimes \mathbb{Z}^d}$ be the corresponding product measure on Ω. We will consider GD parametrized by $\theta \in \Theta$, with $\Theta = \mathbb{R}^m$, and specified in terms of m (known) energy functions $U^{(1)}, \dots, U^{(m)}$.

For $l = 1, \dots, m$ and finite $V \subset \mathbb{Z}^d$, let $I_V^{(l)}$ be a bounded continuous function on Ω, which depends only on $x_V = (x_i,\ i \in V)$; $I_V^{(l)}$ represent the contribution of the variables X_i, located in V, to the lth energy. For these m interactions $\{I_V^{(l)}; V \subset \mathbb{Z}^d \text{ finite}\}$, we assume:

$$\text{(2.1)} \qquad \text{translation invariance:} \qquad I_V^{(l)} \circ \tau^i = I_{i+V}^{(l)} \quad \text{for all } i, V, l,$$

$$\text{(2.2)} \qquad \text{summability:} \qquad |||\, I\, ||| = \left\{ \sum_{l=1}^{m} \left[\sum_{V:\, o \in V} \| I_V^{(l)} \|_\infty \right]^2 \right\}^{1/2} < \infty,$$

where o denotes the origin in $\mathbb{Z}^d$. Notice that we do not assume finite range for the interaction.

For any $\Lambda \subset \mathbb{Z}^d$ finite and any boundary condition (b.c.) $y \in \mathscr{X}^{\Lambda^c}$ outside Λ, the energy of a configuration $x_\Lambda \in \mathscr{X}^\Lambda$, given the b.c. y, is, by definition,

$$\text{(2.3)} \qquad U_\Lambda^{(l)}(x_\Lambda / y) = \sum_{V:\, V \cap \Lambda \neq \varnothing} I_V^{(l)}(x_\Lambda \vee y),$$

where $x_\Lambda \vee y$ is the configuration equal to x_Λ on Λ and to y on Λ^c. The Gibbs specification corresponding to $\theta \in \mathbb{R}^m$ is

$$\text{(2.4)} \qquad \pi_{\Lambda,\theta}(x_\Lambda / y) = Z_{\Lambda,y}(\theta)^{-1} \exp(\theta \cdot U_\Lambda(x_\Lambda / y)),$$

where U_Λ is the m-dimensional energy vector and

$$\text{(2.5)} \quad Z_{\Lambda,y}(\theta) = \int \exp(\theta \cdot U_\Lambda(x_\Lambda / y)) Q_\rho(dx_\Lambda) = \mathbb{E}^{\rho^{\otimes\Lambda}}\{\exp(\theta \cdot U_\Lambda(x_\Lambda / y))\}.$$

We will use notations $\mathbb{E}^P$ for the expectation under P like above and ${}_\Lambda x$ (resp. ${}_i x$) for x_{Λ^c} (resp. $x_{\{i\}^c}$).

The set $\mathscr{G}(\theta)$ of infinite volume *Gibbs distributions* (GD) is the set of all $P_\theta \in \mathscr{P}(\Omega)$ such that, for all finite $\Lambda \subset \mathbb{Z}^d$, the conditional distribution of P_θ given ${}_\Lambda x$ satisfies

$$(2.6) \qquad P_\theta(dx_\Lambda / {}_\Lambda x) = \pi_{\Lambda,\theta}(x_\Lambda / {}_\Lambda x)\rho^{\otimes\Lambda}(dx_\Lambda)$$

for P_θ-a.e. ${}_\Lambda x \in \mathscr{X}^{\Lambda^c}$. This equation is known as the Dobrushin–Lanford–Ruelle equation, and existence of solutions P_θ is well known under the assumptions (2.1) and (2.2). A *Markov random field* (MRF) is a GD with finite range interaction, that is, $I_V = 0$ if the diameter of V is more than some positive constant. MRF have the Markov property in space.

We will consider a sequence of finite windows Λ_n of observation, increasing to $\mathbb{Z}^d$. For simplicity, we assume $\Lambda_n = [-n, n]^d$ and we write x_n for x_{Λ_n} and $\pi_{n,\theta}$ for $\pi_{\Lambda_n,\theta}, \dots$ with a slight abuse of notation. The statistical problem may be formulated as follows: We want to estimate θ from larger and larger windows Λ_n and observations $X_n = (X_i,\ i \in \Lambda_n)$ from finite volume GD,

$$(2.7) \qquad P_{n,\theta}(dx_n) = \left[\int \pi_{n,\theta}(x_n/y)\mu_n(dy)\right]\rho^{\otimes\Lambda_n}(dx_n),$$

where μ_n is an arbitrary distribution on $\mathscr{X}^{\Lambda_n^c}$. Hence, the sample may come from some infinite volume GD P_θ—when $\mu_n = P_\theta(dx_n)$—as well as from different finite volumes ones with a whole family of limit point as $n \to \infty$. Notice that we may add an isolated point to the space $\mathscr{X}$, in order to cover GD with free b.c. with formula (2.7). Note also that we do not consider the estimation problem of the true underlying measure P_θ in the first case, which cannot be solved in general from increasing pieces X_n of a single realization of P_θ.

The maximum likelihood estimator (MLE) $\hat{\theta}_{n,z} = \hat{\theta}_{n,z}(X_n)$ is chosen to maximize the (conditional) likelihood function

$$(2.8) \qquad l_{n,z}(\theta, X_n) = |\Lambda_n|^{-1} \log \pi_{n,\theta}(X_n/z)$$

for some b.c. z. Here, $|\Lambda|$ is the cardinality of Λ. Due to heavy integrals as in (2.5), brute force computations are intractable here, but the MLE may be approximated using stochastic algorithms, as proposed in Younes (1988). Besag (1974, 1977) proposed to replace the expectation in (2.5) with a single integral: The maximum pseudolikelihood estimator (MPLE) $\tilde{\theta}_{n,z}$ is, by definition, any measurable maximizer (possibly at infinity) of

$$(2.9) \qquad pl_{n,z}(\theta, X_n) = |\Lambda_n|^{-1} \log \prod_{i \in \Lambda_n} \pi_{\{i\},\theta}(X_i / {}_i[X_n \vee z])$$

for some b.c. z. When possible, the b.c. in (2.8) and (2.9) are chosen to be observed data, shrinking the window Λ_n somewhat; this situation is covered by our results below, with minor modifications.

We will assume the following: The model is *identifiable* if $\mathscr{G}(\theta) \cap \mathscr{G}(\theta_o) = \varnothing$ for all θ, θ_o with $\theta \neq \theta_o$. This condition is equivalent to $\mathscr{G}(\theta) \neq \mathscr{G}(\theta_o)$; see

Georgii (1988). In the appendix of Gidas (1991) this condition is shown to be related to the statistical mechanics concept of physical equivalence of the interactions.

Now we state our first result: Let $\|\cdot\|$ be a norm on $\mathbb{R}^m$.

THEOREM 2.1. *Assume that the model is identifiable. Independently of z in* (2.8) *and* (2.9), *we have* $\forall\ \varepsilon > 0$, $\exists\ c > 0$ *and* $\delta > 0$ *such that*

$$P_{n,\theta_o}\left(\|\hat{\theta}_{n,z} - \theta_o\| > \varepsilon\right) \le c\exp(-|\Lambda_n|\delta)$$

and

$$P_{n,\theta_o}\left(\|\tilde{\theta}_{n,z} - \theta_o\| > \varepsilon\right) \le c\exp(-|\Lambda_n|\delta).$$

Note that exponential consistency holds for any underlying GD P_{n,θ_o}, regardless of lack of spatial stationarity, and is not affected by phase transition phenomena, unlike the rate of convergence in the general ergodic theorem for GD. If $P_{n,\theta_o} = P_{\theta_o}$, a.s. convergence is a straightforward consequence of the theorem and of the Borel–Cantelli lemma. Our result for the MPLE is more general than those mentioned before that require finite state space $\mathscr{X}$ or stationarity of P_{θ_o}. The theorem will be proved as a consequence of Theorem 3.1 in the next section.

We end this section with some well-known properties of MRF. We start with a definition. Let $x_n \in \mathscr{X}^{\Lambda_n}$ and extend it by periodization outside Λ_n into a (periodic) configuration $x^{(n)} \in \Omega$. Define $R_{n,x}$ by

$$R_{n,x} = |\Lambda_n|^{-1}\sum_{i\in\Lambda_n}\delta_{\tau^i x^{(n)}} \in \mathscr{P}_s(\Omega) \tag{2.10}$$

where δ_y is the Dirac mass at point $y \in \Omega$; $R_{n,x}$ is stationary because $x^{(n)}$ is periodic. The empirical field of X on Λ_n is, by definition, the random distribution $R_{n,X}$. In this paper we will write spatial averages occurring in estimators in terms of $R_{n,X}$. For example, the marginal distribution on $\mathscr{X}$ of $R_{n,X}$ is the empirical measure. Furthermore, using (2.1) and (2.2), one can check that

$$|\Lambda_n|^{-1}U_n(X_n/y) = \mathbb{E}^{R_{n,X}}A_U + \varepsilon(n), \tag{2.11}$$

where

$$A_U(x) = \sum_{V:\, V\ni o}|V|^{-1}I_V(x)$$

and where the remainder term $\varepsilon(n)$ comes from the periodization, and is such that $\lim_{n\to\infty}\sup_y\|\varepsilon(n)\| = 0$. We now recall some facts.

FACT 1. $p_{n,y}(\theta) = |\Lambda_n|^{-1}\log Z_{n,y}(\theta)$ is a convex function of θ, $|p_{n,y}(\theta) - p_{n,y}(\theta')| \le \|\theta - \theta'\|_2 \, |||\, I\, |||$ and $p(\theta) = \lim_{n\to\infty} p_{n,y}(\theta)$ exists uniformly in y and is independent of y.

FACT 2. The relative entropy of $P \in \mathscr{P}_s(\Omega)$ with respect to Q_ρ, given by

$$H(P) = \lim_{n \to \infty} |\Lambda_n|^{-1} \mathbb{E}^{P_{\Lambda_n}}\big[\log\{dP_{\Lambda_n}/dQ_{\rho,\Lambda_n}\}\big]$$

exists in $[0, +\infty]$. Moreover, H is a linear, lower semicontinuous function on $\mathscr{P}_s(\Omega)$, with compact level sets $\{R \in \mathscr{P}_s(\Omega);\ H(R) \le a\}$ for all $a \ge 0$.

FACT 3. $p(\theta)$ is given by the Gibbs variational formula

$$p(\theta) = \max\{\mathbb{E}^Q(\theta \cdot A_U) - H(Q);\ Q \in \mathscr{P}_s(\Omega)\} \tag{2.12}$$

and the set $\mathscr{G}_s(\theta)$ of stationary GD is exactly the set of maximizers Q of (2.12) (this is the Gibbs variational principle).

FACT 4. For all closed set $\mathscr{F}$ in $\mathscr{P}_s(\Omega)$ and all θ_o,

$$\begin{aligned} \limsup_{n\to\infty} |\Lambda_n|^{-1} \log P_{n,\theta_o}(R_{n,X} \in \mathscr{F}) &\le -\inf\{H_{\theta_o}(R);\ R \in \mathscr{F}\} \\ \text{with } H_{\theta_o}(R) &= -\mathbb{E}^R(\theta_o \cdot A_U) + H(R) + p(\theta_o). \end{aligned} \tag{2.13}$$

For a proof of Facts 1–3, see Georgii (1988); for Fact 4 (which is a large deviation upper bound) see Comets (1986), Föllmer and Orey (1988) and Olla (1988) for complete generality. A_U is a bounded and continuous function and, consequently, H_θ is itself linear, lower semicontinuous (l.s.c.) and has compact level sets.

REMARK 1. From Facts 2 and 3, we see that the infimum in Fact 4 is zero if and only if $\mathscr{F} \cap \mathscr{G}_s(\theta_o) \neq \varnothing$; otherwise, the P_{n,θ_o} probability goes to zero. In many statistical problems, the set $\mathscr{F}$ under consideration is often closed half-spaces; then the infimum is zero if and only if $\mathscr{F}$ contains ergodic GD. Therefore, the behavior of an arbitrary sequence of finite volume GDs P_{n,θ_o} can be controlled with the behavior of the ergodic GD.

REMARK 2. The function $p(\theta)$ is called the pressure. By definition, p is convex—as a limit of convex functions. In the appendix of Gidas (1987), it is shown that the model is identifiable if the pressure is strictly convex.

3. Consistency of maximum objective estimators. In this section, we prove the general consistency theorem (Theorem 3.1) and then show that it implies Theorem 2.1.

We will say that a function $k_n(\theta, x_n)$: $\Theta \times \mathscr{X}^{\Lambda_n} \to \mathbb{R}$ is an objective function if there exists a continuous function K on $\Theta \times \mathscr{P}_s(\Omega)$ such that

$$k_n(\theta, x_n) = K(\theta, R_{n,x}) + \varepsilon(n), \tag{3.1}$$

with $\lim_{n\to\infty} \varepsilon(n) = 0$ uniformly for $x_n \in \mathscr{X}^{\Lambda_n}$ and for θ in compact sets of Θ, and

$$\forall\, \theta_o, \forall\, P \in \mathscr{G}_s(\theta_o),\ \theta_o \text{ is the unique maximum of } K(\cdot, P). \tag{3.2}$$

If, in addition, $k_n(\theta, x_n)$ is concave in θ for all x_n, k_n will be called a concave objective function. A maximum objective estimator is a random variable $\bar{\theta}_n = \bar{\theta}_n(X_n)$ maximizing $k_n(\cdot, X_n)$. The reader may refer to Dacunha-Castelle and Duflo (1982), Definition 3.2.7, Sections 3.2.3 and 3.3.4, for a general theory of maximum objective estimators.

THEOREM 3.1. *Every maximum concave objective estimator $\bar{\theta}_n$ on $\Theta = \mathbb{R}^m$ is consistent. For all $\varepsilon > 0$ and all θ_o, we have*

$$\limsup_{n\to\infty} |\Lambda_n|^{-1} \log P_{n,\theta_o}\left(\left\|\bar{\theta}_n - \theta_o\right\| \geq \varepsilon\right) < 0.$$

REMARK. The concavity assumption is made here because the set Θ of parameters is the whole space $\mathbb{R}^m$. When Θ is compact, this assumption can be removed, using (3.1) to control the oscillation of k_n in θ, uniformly with respect to P.

PROOF. Let $\varepsilon > 0$, $\delta > 0$ and $\mathscr{S}(\theta_o, \varepsilon) = \{\theta;\ \|\theta - \theta_o\| = \varepsilon\}$. Since k_n is concave,

$$\begin{aligned} \left\{\left\|\bar{\theta}_n - \theta_o\right\| \geq \varepsilon\right\} &\subset \{\exists\ \theta \in \mathscr{S}(\theta_o, \varepsilon) : k_n(\theta, X_n) \geq k_n(\theta_o, X_n)\} \\ &\subset \left\{\max_{\theta \in \mathscr{S}(\theta_o, \varepsilon)} K(\theta, R_{n,X}) \geq K(\theta_o, R_{n,X}) - \delta\right\} \end{aligned} \tag{3.3}$$

for large n. In (3.3), we have used (3.1), the continuity of K and the compactness of $\mathscr{S}$. Again from continuity and compactness, the function $\max_{\theta \in \mathscr{S}(\theta_o, \varepsilon)} K(\theta, R)$ is continuous in R and the set

$$\mathscr{F}(\theta_o, \varepsilon, \delta) = \left\{R \in \mathscr{P}_s(\Omega);\ \max_{\theta \in \mathscr{S}(\theta_o, \varepsilon)} K(\theta, R) \geq K(\theta_o, R) - \delta\right\}$$

is closed in $\mathscr{P}_s(\Omega)$. Then, Fact 4 yields

$$\begin{aligned} &\limsup_{n\to\infty} |\Lambda_n|^{-1} \log P_{n,\theta_o}\{R_{n,X} \in \mathscr{F}(\theta_o, \varepsilon, \delta)\} \\ &\quad\leq -\inf\left\{H_{\theta_o}(R);\ R \in \mathscr{F}(\theta_o, \varepsilon, \delta)\right\}. \end{aligned} \tag{3.4}$$

As δ tends to zero, the closed set $\mathscr{F}(\theta_o, \varepsilon, \delta\}$ shrinks to $\mathscr{F}(\theta_o, \varepsilon, 0)$, and since H_{θ_o} is lower semicontinuous, the last infimum increases to $\inf\{H_{\theta_o}(R);\ R \in \mathscr{F}(\theta_o, \varepsilon, 0)\}$. Combining this with (3.3) and (3.4), we obtain

$$\limsup_{n\to\infty} |\Lambda_n|^{-1} \log P_{n,\theta_o}\left\{\left\|\bar{\theta} - \theta_o\right\| \geq \varepsilon\right\} \leq -\inf\left\{H_{\theta_o}(R);\ R \in \mathscr{F}(\theta_o, \varepsilon, 0)\right\}.$$

Moreover, the closed set $\mathscr{F}(\theta_o, \varepsilon, 0)$ contains $\mathscr{G}_s(\theta)$ for $\|\theta - \theta_o\| \geq \varepsilon$ and then is not empty. Since H_{θ_o} is lower semicontinuous (l.s.c.) and has compact level sets, the infimum of H_{θ_o} on this set is achieved at some point R_o. According to Fact 3, $H_{\theta_o}(R_o) = 0$ if and only if $R_o \in \mathscr{G}_s(\theta_o)$, which does not hold since (3.2) implies $\mathscr{G}_s(\theta_o) \cap \mathscr{F}(\theta_o, \varepsilon, 0) = \varnothing$. Hence, $\inf\{H_{\theta_o}(R);\ R \in \mathscr{F}(\theta_o, \varepsilon, 0)\} = H_{\theta_o}(R_o)$ is strictly positive, which ends the proof. □

PROOF OF THEOREM 2.1. We now prove Theorem 2.1 by checking the assumptions of Theorem 3.1.

(a) The MLE $\hat{\theta}_{n,z}$ maximizes the concave function

$$
\begin{aligned}
l_{n,z}(\theta, X_n) &= |\Lambda_n|^{-1} \log \pi_{n,\theta}(X_n/z) \\
&= |\Lambda_n|^{-1} \theta \cdot U_n(X_n/z) - p_{n,z}(\theta) \qquad (3.5)\\
&= \mathbb{E}^{R_{n,X}} \theta \cdot A_U - p(\theta) + \varepsilon(n),
\end{aligned}
$$

where $\varepsilon(n)$ goes to 0 uniformly in z and for θ in compact sets, since we have (2.11) and uniform convergence of $p_{n,z}$ to p (from Ascoli's theorem and Fact 1). Hence, K is here the continuous function $K(\theta, R) = \mathbb{E}^R \theta \cdot A_U - p(\theta)$. Let us check (3.2). If $P \in \mathscr{G}_s(\theta_0)$, the variational principle in Fact 3 implies that

$$K(\theta, P) - K(\theta_o, P) = \mathbb{E}^P \theta \cdot A_U - p(\theta) - H(P) \le 0$$

from the variation formula, with equality if and only if $P \in \mathscr{G}_s(\theta)$. Therefore, condition (3.2) holds if and only if the model is identifiable and $l_{n,z}$ is a concave objective function.

(b) The MPLE $\tilde{\theta}_{n,z}$ maximizes the concave function

$$pl_{n,z}(\theta, X_n) = |\Lambda_n|^{-1} \sum_{i \in \Lambda_n} \Big(\theta \cdot U_{\{i\}}[(X_i/_i(X_n \vee z))] - p_{\{i\},_i(X_n \vee z)}(\theta)\Big).$$

Let us denote the continuous function $p_{\{i\},y}(\theta)$ of $(\theta, y) \in \Theta \times \mathscr{X}^{\mathbb{Z}^d - \{i\}}$ by $p_i(\theta, y)$. Using (2.2), one can check that

$$U_{\{i\}}(x_i/_i x) - U_{\{i\}}(y_i/_i y) = \varepsilon\big(\min\{\|j - i\|; y_j \ne x_j\}\big)$$

and that

$$p_i(\theta, y) - p_i(\theta, x) = \varepsilon_{\mathscr{C}}\big(\min\{\|j - i\|; y_j \ne x_j\}\big)$$

for all compact $\mathscr{C} \subset \Theta$. In addition to (2.1), this implies that

$$
\begin{aligned}
pl_{n,z}(\theta, X_n) &= \mathbb{E}^{R_{n,X}}\{\theta \cdot U_{\{o\}}(x_o/_o x) - p_o(\theta, {}_o x)\} + \varepsilon(n) \\
&= K(\theta, R_{n,X}) + \varepsilon(n),
\end{aligned}
$$

with ε as in (3.1). The function K is continuous and such that

$$
\begin{aligned}
&K(\theta, P) - K(\theta_o, P) \\
&\quad = \mathbb{E}^P\Big(\mathbb{E}^P\big[(\theta - \theta_o) \cdot U_{\{o\}}(x_o/_o x)/_o x\big] - p_o(\theta, {}_o x) + p_o(\theta_o, {}_o x)\Big).
\end{aligned}
$$

Let $P \in \mathscr{G}_s(\theta_o)$; the conditional distribution $P(dx_o/_o x)$ is P-a.s. equal to $\pi_{\{o\},\theta_o}(x_o/_o x)\rho(dx_o)$ given by (2.4). From Jensen inequality we obtain

$$
\begin{aligned}
\mathbb{E}^P\big[(\theta - \theta_o) \cdot U_{\{o\}}/_o x\big] &\le \log \mathbb{E}^P\big[\exp((\theta - \theta_o) \cdot U_{\{o\}})/_o x\big] \\
&= \log\Big\{\big[Z_{\{o\},_o x}(\theta_o)\big]^{-1} \int_{\mathscr{X}} \exp(\theta \cdot U_{\{o\}})\rho(dx_o)\Big\} \\
&= p_o(\theta, {}_o x) - p_o(\theta_o, {}_o x), \qquad P\text{-a.s.},
\end{aligned}
$$

with equality if and only if $(\theta - \theta_o) \cdot U_{\{o\}}(\cdot/_o x) = 0$ ρ-a.s. Hence, $K(\theta, P) -$

$K(\theta_o, P) \leq 0$ with equality if and only if $P(dx_o/_o x) = \pi_{\{o\},\theta}(x_o/_o x)\rho(dx_o)$ for P-a.e. $_o x$. Since P is stationary, each conditional distribution $P(dx_\Lambda/_\Lambda x)$ can be written in terms of $P(dx_o/_o x)$, and then

$$K(\theta, P) - K(\theta_o, P) = 0 \Leftrightarrow P \in \mathscr{G}_s(\theta).$$

Here again, condition (3.2) is equivalent to identifiability. □

4. Optimality. For those parameter values where phase transition does not occur, the MLE is asymptotically normal and efficient in the Fisher sense [Janzura (1988); Gidas (1991)], but at phase transition points θ_o, asymptotic normality of MLE does not always hold, at least for standard deterministic normalization (see the complete discussion in Section 5A). The Fisher information matrix $p''_{n,y}(\theta_o)$ may go to infinity as $n \to \infty$. On the other hand, some other estimators, like MPLE, behave nicely for all parameter values [see Guyon and Künsch (1990)] under ergodicity assumptions. It seems interesting to develop a general approach to optimality.

Therefore, in this section we study efficiency of estimators, in the sense of Bahadur [see Kester (1985) for a review]. We start with a bound, which is somewhat analogous to the Cramér–Rao bound:

PROPOSITION 4.1. *Let $\bar{\theta}_n$ be a consistent estimator of θ, depending only on X_n. Then,*

$$\liminf_{n \to \infty} |\Lambda_n|^{-1} \log P_{n,\theta_o}\left(\|\bar{\theta}_n - \theta_o\| > \varepsilon\right) \geq -b(\theta_o, \varepsilon),$$

where

$$b(\theta_o, \varepsilon) = \inf\left\{H_{\theta_o}(R);\ R \in \mathscr{G}_s(\theta), \|\theta - \theta_o\| > \varepsilon\right\}.$$

PROOF. The Shannon–McMillan theorem [Föllmer (1988), Theorems 3.37 and 3.38] states that

$$\lim_{n \to \infty} |\Lambda_n|^{-1} \log\left(dR_{\Lambda_n}/dP_{\theta_o,\Lambda_n}\right) = H_{\theta_o}(R),$$

R-a.s. for ergodic random fields R. From consistency of $\bar{\theta}_n$, we have $\lim_{n\to\infty} R(\|\bar{\theta}_n - \theta_o\| > \varepsilon) = 1$ for all $R \in \mathscr{G}_s(\theta)$ and all θ with $\|\theta - \theta_o\| > \varepsilon$. Then, Theorem 2.1 in Bahadur, Zabell and Gupta (1980) yields

$$\liminf_{n \to \infty} |\Lambda_n|^{-1} \log P_{n,\theta_o}\left(\|\bar{\theta}_n - \theta_o\| > \varepsilon\right) \geq -H_{\theta_o}(R) \tag{4.1}$$

for all such R and θ. Moreover, any $R \in \mathscr{G}_s(\theta)$ is a convex combination of ergodic elements of $\mathscr{G}_s(\theta)$ [Georgii (1988)]. Since H_{θ_o} is linear, this implies that the infimum defining $b(\theta_o, \varepsilon)$ may be taken on ergodic fields R only. Hence (4.1) proves the result.

The bound $b(\theta_o, \varepsilon)$ is nondecreasing and positive for positive ε, but may have discontinuities in ε due to phase transition. Letting $b(\theta_o, \varepsilon^-) =$

$\lim_{\varepsilon' \nearrow \varepsilon} b(\theta_o, \varepsilon')$, we can derive

$$b(\theta_o, \varepsilon^-) = \inf\{H_{\theta_o}(R);\ R \in \mathscr{G}_s(\theta),\ \|\theta - \theta_o\| \geq \varepsilon\} \tag{4.2}$$

from the lower semicontinuity of H_{θ_o} and from (2.6).

The quantity

$$c(\theta_o, \varepsilon) = -\limsup_{n\to\infty} |\Lambda_n|^{-1} \log P_{n,\theta_o}\big(\|\bar{\theta}_n - \theta_o\| \geq \varepsilon\big),$$

sometimes called the inaccuracy rate, measures the deficiency of the estimator. It may be used, in theory, to compare estimators. But from the practical point of view, this rate involves some pressure function in our setup, and explicit formulas or the equivalent for small ε as well as for the limit variance when asymptotic normality holds are not available. We illustrate this with the MPLE in the next section. We now state optimality of the MLE:

THEOREM 4.2. *The MLE is inaccuracy rate optimal in the sense that*

$$\limsup_{n\to\infty} |\Lambda_n|^{-1} \log P_{n,\theta_o}\big(\|\hat{\theta}_{n,z} - \theta_o\| \geq \varepsilon\big) \leq -b(\theta_o, \varepsilon^-)$$

for all ε *such that* $b(\theta_o, \varepsilon^-) < b(\theta_o, \infty) = \lim_{\varepsilon\to\infty} b(\theta_o, \varepsilon)$.

Optimality for the MLE is connected with optimality of the likelihood ratio test in the sense of exact slopes [Bahadur (1971)]. Optimality for the MLE is believed to hold in general exponential models, but does not hold when the statistical curvature is nonzero. In Bahadur (1983), the MLE for parameters of the finite state space Markov chain is shown to be locally optimal. In Section 5, we discuss the influence of phase transition on the optimal rate b.

PROOF OF THEOREM 4.2. Let $\varepsilon > \varepsilon' \geq 0$. Since $l_{n,z}$ is concave,

$$\{\|\hat{\theta}_{n,z} - \theta_o\| \geq \varepsilon\} \subset \Big\{\sup_{\theta\in\mathscr{S}(\theta_o,\varepsilon)} l_{n,z}(\theta, X_n) \geq \sup_{\theta\in\mathscr{S}(\theta_o,\varepsilon')} l_{n,z}(\theta, X_n)\Big\}.$$

We can repeat the same arguments as between (3.3) and the first inequality after (3.4). Using here (3.5) and $K(\theta, R) = \mathbb{E}^R\theta \cdot A_U - p(\theta)$, we obtain

$$\limsup_{n\to\infty} |\Lambda_n|^{-1} \log P_{n,\theta_o}\big(\|\hat{\theta}_{n,z} - \theta_o\| \geq \varepsilon\big) \leq -\inf\{H_{\theta_o}(R);\ R \in \mathscr{F}(0)\}, \tag{4.3}$$

where $\mathscr{F}(\delta) = \mathscr{F}(\theta_o, \varepsilon, \varepsilon', \delta)$ is defined this time by

$$\mathscr{F}(\delta) = \Big\{R \in \mathscr{P}_s(\Omega);\ \max_{\theta\in\mathscr{S}(\theta_o,\varepsilon)} K(\theta, R) \geq \max_{\theta\in\mathscr{S}(\theta_o,\varepsilon')} K(\theta, R) - \delta\Big\}.$$

At this point, we need the fact that the stationary random fields with minimum entropy under linear constraints are Gibbs distributions. Though well known for finite volume, this fact is—curiously enough—not emphasized in the infinite volume case. Let us start with some definitions: The affine hull of a subset B of $\mathbb{R}^m$ is the smallest affine space containing B, and the relative interior $\mathrm{ri}(B)$ of B is the interior of B in its affine hull.

LEMMA 4.3. *Let A be a bounded continuous function from Ω to $\mathbb{R}^m$,*

$$B = \{\mathbb{E}^R A;\ R \in \mathscr{P}_s(\Omega),\ H(R) < \infty\} \subset \mathbb{R}^m$$

and let $\not{p}$ be the convex function on $\mathbb{R}^m$ given by

$$\not{p}(t) = \lim_{n\to\infty} |\Lambda_n|^{-1} \log \mathbb{E}^{Q_\rho}\Big\{\exp \sum_{i\in\Lambda_n} t\cdot A\circ\tau^i\Big\}.$$

Then, for $a \in \mathrm{ri}(B)$, the infimum $\inf\{H(R);\ R \in \mathscr{P}_s(\Omega),\ \mathbb{E}^R A = a\}$ is achieved, exactly at all $R \in \mathscr{G}_s(t\cdot A)$ with $\mathbb{E}^R A = a$ and t such that the subdifferential $\partial\not{p}(t)$ of $\not{p}$ at point t contains a. Moreover, the conclusion is true for a in the range $\bigcup_t \partial\not{p}(t)$ of $\partial\not{p}$.

The set $\mathscr{G}_s(t\cdot A)$ above is the set of stationary Gibbs distributions given in (2.4) and (2.6) with energy $t\cdot V$, where the energy function V is such that $A_V = A$ [with A_V given by (2.11)].

PROOF OF LEMMA 4.3. In order to prove the lemma, we write the variational formula in Fact 3 as $\not{p}(t) = \max_a\{t\cdot a - \lambda(a)\}$ with $\lambda(a) = \inf\{H(R);\ \mathbb{E}^R A = a\}$. Since H is affine and the constraint $\mathbb{E}^R A = a$ is linear in R, λ is convex; the (convex, proper, l.s.c.) functions $\not{p}$ and λ are Legendre conjugate. Therefore, we have $\mathrm{ri}(\mathrm{dom}\,\lambda) \subset \bigcup_t \partial\not{p}(t) \subset \mathrm{dom}\,\lambda$, with $\mathrm{dom}\,\lambda = \{a \in \mathbb{R}^m;\ \lambda(a) < \infty\}$; see the beginning of Section 24 in Rockafellar (1970).

By definition, B is the domain $\mathrm{dom}\,\lambda$. Then, for a as in the lemma, there exists some $t \in \mathbb{R}^m$ with $a \in \partial\not{p}(t)$; in particular, $\not{p}(t) = t\cdot a - \lambda(a)$. Since $a \in B$ and H is l.s.c., the infimum defining $\lambda(a)$ is achieved at some point R, and of course $\mathbb{E}^R t\cdot A - H(R) = \not{p}(t)$. The variational principle in Fact 3 implies that $R \in \mathscr{G}_s(t\cdot A)$ and that we have, for all $Q \in \mathscr{P}_s(\Omega)$, $\mathbb{E}^Q t\cdot A - H(Q) \le \mathbb{E}^R t\cdot A - H(R)$, with equality if and only if $Q \in \mathscr{G}_s(t\cdot A)$. Hence, for all Q such that $\mathbb{E}^Q A = a$, we have $H(Q) \ge H(R)$ with equality iff $Q \in \mathscr{G}_s(t\cdot A)$. This proves the lemma. □

We now end the proof of the theorem. Using the above notation with $A = A_U$, inequality (4.3) for $\delta = 0$ reads

$$\begin{aligned}
&\limsup_{n\to\infty} |\Lambda_n|^{-1} \log P_{n,\theta_o}\big(\|\hat\theta_{n,z} - \theta_o\| \ge \varepsilon\big)\\
&\quad\le -\inf\{H_{\theta_o}(R);\ R \in \mathscr{F}(0)\}\\
&\quad= -\inf_a\Big\{-\theta_o\cdot a + \lambda(a) - p(\theta_o);\\
&\qquad\qquad \max_{\theta\in\mathscr{S}(\theta_o,\varepsilon)}[\theta\cdot a - p(\theta)] \ge \max_{\theta\in\mathscr{S}(\theta_o,\varepsilon')}[\theta\cdot a - p(\theta)]\Big\}.
\end{aligned} \tag{4.4}$$

Clearly, the infima are achieved at some points $R \in \mathscr{F}(0)$ and a with $\mathbb{E}^R A_U = a$.

We claim that R is a GD; this will follow from the lemma, if we check that the point a lies in the range of ∂p (we have $\not{p} = p$). Since a belongs to the convex set dom λ, there exists a sequence a_q in ri(dom λ) with $\lim_{q\to\infty} a_q = a$ and $\lim_{q\to\infty} \lambda(a_q) = \lambda(a)$. Using the lemma, we see that there exists θ_q and R_q, with $a_q \in \partial p(\theta_q)$, $R_q \in \mathscr{G}_s(\theta_q)$ and $\mathbb{E}^{R_q} A_U = a_q$. From the variational principle in Fact 3, it follows that $\mathscr{G}_s(\theta) \subset \mathscr{F}(0)$ for $\theta \in \mathscr{S}(\theta_o, \varepsilon)$, and then $H_{\theta_o}(R) \le b(\theta_o, \varepsilon^-) < b(\theta_o, \infty)$ by assumption. This implies that $H_{\theta_o}(R_q) < b(\theta_o, \infty)$ for large q, which shows boundedness of the sequence θ_q and then existence of a limit point $\theta_\infty \in \mathbb{R}^m$. Since the graph of ∂p is closed [Theorem 24.4 in Rockafellar (1970)], we obtain that $a \in \partial p(\theta_\infty)$, and then the desired claim: R is a GD [in fact, $R \in \mathscr{G}_s(\theta_\infty)$].

Let θ_1 with $\|\theta_1 - \theta_o\| \le \varepsilon'$, $\theta \in \mathscr{S}(\theta_o, \varepsilon)$ and $\theta' \in \mathscr{S}(\theta_o, \varepsilon')$ with $\theta' \in [\theta, \theta_1]$. The variational principle in Fact 3 shows that $\mathbb{E}^Q A_U$ is tangent to the convex function p at point θ_1, when $Q \in \mathscr{G}_s(\theta_1)$. The strict convexity of p (cf. Remark 2 at the end of Section 2) implies that $K(\theta, Q) < K(\theta', Q)$, and then $Q \notin \mathscr{F}(0)$. Therefore,

$$\inf\{H_{\theta_o}(Q); Q \in \mathscr{F}(0)\} = H_{\theta_o}(R) \ge \inf\{H_{\theta_o}(Q); Q \in \mathscr{G}_s(\theta), \|\theta - \theta_o\| > \varepsilon'\}$$
$$= b(\theta_o, \varepsilon')$$

from convexity again. As $\varepsilon' \nearrow \varepsilon$, $b(\theta_o, \varepsilon') \nearrow b(\theta_o, \varepsilon^-)$, and the first inequality in (4.4) shows the theorem. □

5. About the inaccuracy rate. In this section, we discuss the inaccuracy rate and the effects of phase transition.

A. From (4.2), (2.13) and from convexity, we can rewrite the optimal inaccuracy rate

$$b(\theta_o, \varepsilon^-) = \inf\{(\theta - \theta_o) \cdot \mathbb{E}^R A_U - p(\theta) + p(\theta_o); R \in \mathscr{G}_s(\theta), \|\theta - \theta_o\| = \varepsilon\}.$$

Recall that in the previous formula, $\mathbb{E}^R A_U$ belongs to the subdifferential $\partial p(\theta)$ of p at point θ. We give now approximations of $b(\theta_o, \varepsilon^-)$ for small ε, in various situations.

(i) The regular case occurs when p is twice continuously differentiable in a neighborhood of θ_o. In particular, we have $p'(\theta) = \mathbb{E}^R A_U$ for all $R \in \mathscr{G}_s(\theta)$. From the Taylor formula,

$$(5.1)\qquad b(\theta_o, \varepsilon^-) \sim \varepsilon^2/2 \inf\{t^* p''(\theta_o) t; t \in \mathbb{R}^m, \|t\| = 1\} \quad \text{as } \varepsilon \to 0.$$

It is known that the MLE is asymptotically normal and Fisher efficient in this case [Janzura (1988); Gidas (1991)], when $p''(\theta_o)$ is invertible:

$$(5.2)\qquad |\Lambda_n|^{1/2}(\hat{\theta}_{n,z} - \theta_o) \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}(0, p''(\theta_o)^{-1}).$$

(ii) *First-order phase transition* occurs in the nearest neighbor (n.n.) Ising model, for $\beta_o > \beta_c$ and $h = 0$. In this model, we have binary variables X_i and

$$U_\Lambda(x/y) = \beta/2 \Sigma^* x_i x_j + \beta \Sigma^{**} x_i y_i + h \sum_{i\in\Lambda} x_i,$$

where Σ^* ranges over i and j in Λ with $|i - j| = 1$, and Σ^{**} over $i \in \Lambda$,

$j \in \Lambda^c$ with $|i - j| = 1$. Hence, we set $\rho = (1/2)(\delta_{+1} + \delta_{-1})$, $\theta = (\beta, h)$, and $A_U = (x_o[x_u + x_{-u} + x_v + x_{-v}]/2, x_o)$ where $u = (1,0)$ and $v = (0,1)$. The pressure function p is not differentiable in the second direction at $\theta_o = (\beta_o, 0)$, but has an expansion at the neighborhood of θ_o: The one-sided limit $\lim_{\theta \to \theta_o, h > 0} p''(\theta)$ exists [Lebowitz (1972), and the exponential decay of the correlations under $P_{\theta,+}$ for $\theta = (\beta, 0)$ with $\beta > \beta_c$], is equal by symmetry to $\lim_{\theta \to \theta_o, h < 0} p''(\theta)$ and will be abusively denoted by $p''(\theta_o)$. Then, $p''(\theta)$ is continuous on $\mathbb{R}^+ \times \mathbb{R}$, and Taylor expansion leads to

$$p(\theta_o) = p(\theta) + (\theta_o - \theta) \cdot \mathbb{E}^{R_\theta} A_U + \tfrac{1}{2}(\theta_o - \theta)^* p''(\theta_o)(\theta_o - \theta) + o\left(\|\theta_o - \theta\|^2\right)$$

for arbitrary $R_\theta \in \mathscr{G}_s(\theta)$, since R_θ is unique for $h \neq 0$ [and $p'(\theta) = \mathbb{E}^{R_\theta} A_U$] and since only the second component of $\mathbb{E}^{R_\theta} A_U$ depends on the particular R_θ when $h = 0$. From this expansion, we still obtain the equivalent given in (5.1). The question of whether asymptotic normality holds for the MLE under some P_{θ_o} or not remains completely open in this case.

REMARK. The Curie–Weiss model is the infinite-dimensional approximation of the Ising model, and it shares the qualitative behavior of MRF. In this model, let $\hat{h}_n$ be the MLE of h assuming β is known; Comets and Gidas (1991) show that $|\Lambda_n|^{1/2}(\hat{h}_n - h)$ converges to the mixture of a gaussian distribution [with variance given by $p''(\theta_o)$] and of a Dirac mass at point 0, at first-order phase transition points. Note that the infinite volume GD is not extremal in this case.

(iii) *Second-order phase transition* occurs in the n.n. Ising model without external field ($h \equiv 0$), in $d = 2$ dimensions, at point $\beta_o = \beta_c$. From the celebrated Onsager formula for $p(\beta)$, we compute that $b(\beta_o, \varepsilon^-) \sim C\varepsilon^2 \log 1/\varepsilon$ for $\|\beta\| = |\beta|$. This should be compared with the result from Gidas (1987), that for periodic boundary conditions, $(|\Lambda_n| \log|\Lambda_n|)^{1/2}(\hat{\beta}_n - \beta)$ converges to a gaussian distribution under P_{β_c}.

B. In this setup, a major difficulty is to compute, explicitly enough, the inaccuracy rate of general estimators; this amounts to computing some pressure function (different from p). In the nongaussian framework, Onsager's formula is the only explicit formula, up to now, for the pressure. For instance, we can obtain, with computations similar to those in the proof of Theorem 4.2, the inaccuracy rate c of the MPLE:

$$\begin{aligned} -c(\theta_o, \varepsilon) &\le \liminf_{n\to\infty} |\Lambda_n|^{-1} \log P_{n,\theta_o}\left(\|\tilde{\theta}_{n,z} - \theta_o\| > \varepsilon\right) \\ &\le \limsup_{n\to\infty} |\Lambda_n|^{-1} \log P_{n,\theta_o}\left(\|\tilde{\theta}_{n,z} - \theta_o\| \ge \varepsilon\right) \le -c(\theta_o, \varepsilon^-) \end{aligned}$$

with

$$\begin{aligned} c(\theta_o, \epsilon^-) = \inf\Big\{ & H_{\theta_o}(R);\, R \in \mathscr{P}_s(\Omega) \colon \exists\, \theta \in \mathscr{S}(\theta_o, \varepsilon), \\ & \mathbb{E}^R\left(U_{\{o\}}(X_o/_oX) - \mathbb{E}^{\pi\{o\},\theta}[U_{\{o\}}/_oX]\right) = 0\Big\}. \end{aligned} \quad (5.3)$$

The above condition on R is that the normal equation for the MPLE should be satisfied in expectation. Note that we can write $b(\theta_o, \varepsilon)$ in analogy with (5.3), like

$$b(\theta_o, \varepsilon^-) = \inf\left\{H_{\theta_o}(R);\ R \in \mathscr{R}_s(\Omega) : \exists\ \theta \in \mathscr{S}(\theta_o, \varepsilon),\ \mathbb{E}^R A_U \in \partial p(\theta)\right\}$$

based on the integrated form of the normal equation for the MLE.

Lemma 4.3 implies that the minimizers R in (5.3) are GD, $R \in \mathscr{G}_s(t \cdot A_U + s \cdot [U_{\{o\}} - \mathbb{E}^{\pi_{\{o\}},\theta}\{U_{\{o\}}/_o x\}])$, where $t, s \in \mathbb{R}^m$ are subject to the MPLE normal equation given in (5.3). Then, the rate $c(\theta_o, \varepsilon^-)$ can be expressed in terms of the corresponding pressure function (on $\mathbb{R}^{2m}$). This new pressure function is rather complex compared to p, and so c is compared to b. Hence, comparison of estimators through the inaccuracy rate is not realistic.

C. We end with a remark: First-order phase transition may cause discontinuities of $b(\theta_o, \varepsilon)$ in ε. To illustrate our purpose, let us consider the n.n. Ising model, with $\theta = (\beta, h)$. Our theorems do not require for $\|\cdot\|$ the full properties of a norm; they remain valid with $\|\theta\| = \max\{(2 - r)h^+, rh^-, |\beta|\}$ for all $r \in (0, 2)$. Let $\beta_o > \beta_c + 1$. For small enough positive h_o, $H_{\theta_o}(R)$ achieves its maximum for $R \in \mathscr{G}_s(\theta)$ with θ in the rectangle with boundary $\beta = \beta_o \pm 1$, $h = 0$ or $h = 2$, on the open segment line $(h = 0, |\beta - \beta_o| > 1)$ only. Then, if $r = 1/h_o$ and $\varepsilon = 1$, we have $b(\theta_o, \varepsilon^-) < b(\theta_o, \varepsilon)$ due to the jump in the magnetization $\mathbb{E}^R x_o$ through this segment line.

One can easily imagine such a jump in the rate, when $\|\cdot\|$ is the usual norm, provided that the critical set of parameters coincides with a piece of sphere.

REFERENCES

AZENCOTT, R. (1987). Image analysis and Markov fields. In *Proceedings of the First International Conference on Applied Mathematics, Paris* (J. McKenna and R. Temen, eds.) 53–61. SIAM, Philadelphia.

BAHADUR, R. R. (1971). *Some Limit Theorems in Statistics*. SIAM, Philadelphia.

BAHADUR, R. R. (1983). Large deviations of the maximum likelihood estimate in the Markov chain case. In *Recent Advances in Statistics* (M. H. Rizvi, J. S. Rustagi and D. Siegmund, eds.) 273–286. Academic, New York.

BAHADUR, R. R., ZABELL, S. L. and GUPTA, J. C. (1980). Large deviations, tests and estimates. In *Asymptotic Theory of Tests and Estimation* (I. M. Chakravarti, ed.) 33–64. Academic, New York.

BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* **36** 192–236.

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

COMETS, F. (1986). Grandes deviations pour des mesures de Gibbs sur $\mathbb{Z}^d$. *C. R. Acad. Sci. Paris Sér. I Math.* **303** 511–514.

COMETS, F. and GIDAS, B. (1991). Asymptotics of maximum likelihood estimators for the Curie–Weiss model. *Ann. Statist.* **19** 557–578.

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

DACUNHA-CASTELLE, D. and DUFLO, M. (1982). *Probabilitiés et Statistiques* **2**. Masson, Paris.

FÖLLMER, H. (1988). Random fields and diffusion processes. *Ecole d'Eté de Probabilité de St. Flour. Lecture Notes in Math.* **1362** 101–204. Springer, Berlin.

FÖLLMER, H. and OREY, S. (1988). Large deviation for the empirical field of a Gibbs measure. *Ann. Probab.* **16** 961–977.

GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions Pattern Analysis and Machine Intelligence* **PAMI-6** 721–741.

GEMAN, S. and GEMAN, D. (1987). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions Pattern Analysis and Machine Intelligence* **PAMI-9** 39–55.

GEMAN, S. and GRAFFIGNE, C. (1987). Markov random fields image models and their applications to computer vision. In *Proceedings of the International Congress Mathematicians* (A. M. Gleason, ed.) **2** 1496–1517. Amer. Math. Soc., Providence, R.I.

GEORGII, H. O. (1988). *Gibbs Measure and Phase Transitions.* de Gruyter, Berlin.

GIDAS, B. (1986). Consistency of maximum likelihood and pseudolikelihood estimators for Gibbs distributions. In *Stochastic Differential Systems with Application to Electronic/Computer Engineering, Control Theory and Operations Research* (W. Fleming and P. L. Lions, eds.) 1–17. Springer, Berlin.

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, New York. To appear.

GUYON, X. (1986). Estimation d'un champ de Gibbs. Preprint, Univ. Paris I.

GUYON, X. (1987). Estimation d'un champ par pseudo-vraisemblance conditionnelle: Etude asymptotique et application au cas markovien. In *Spatial Processes and Spatial Time Series Analysis* (F. Droesbeke, ed.) 15–62. Publ. Fac. Univ. St Louis, Bruxelles.

GUYON, X. and KUNSCH, H. R. (1990). Asymptotic comparison of estimators in the Ising model. In *Stochastic Models, Statistical Methods and Algorithms in Image Analysis* (P. Barone, A. Frigessi and M. Piccioni, eds.) Springer, Berlin. To appear.

JANZURA, M. (1988). Statistical analysis of Gibbs random fields. In *Information Theory, Statistical Decision Functions, Random Processes* (P. Maudl, ed.) 429–438. Reidel, Dordrecht.

KESTER, A. D. M. (1985). *Some Large Deviation Results in Statistics.* Math. Centrum, Amsterdam.

KÜNSCH, H. R. (1981). Thermodynamics and statistical analysis of gaussian random fields. *Z. Wahrsch. Verw. Gebiete* **58** 407–422.

LEBOWITZ, J. (1972). Bounds on the correlations and analyticity properties of ferromagnetic Ising spin systems. *Comm. Math. Phys.* **28** 313–321.

OLLA, S. (1988). Large deviation for Gibbs random fields. *Probab. Theory Related Fields* **77** 343–357.

PICKARD, D. K. (1987). Inference for discrete Markov field: The simplest non-trivial case. *J. Amer. Statist. Assoc.* **82** 90–96.

ROCKAFELLAR, R. T. (1970). *Convex Analysis.* Princeton Univ. Press.

YOUNES, T. (1988). Estimation and annealing for Gibbsian fields. *Ann. Inst. H. Poincaré Probab. Statist.* **24** 269–294.

UNIVERSITÉ PARIS 7
U.F.R. DE MATHÉMATIQUES
TOUR 45-55, 5ÈME ÉTAGE
U.R.A. CNRS 756 ET 1321
2, PLACE JUSSIEU
75251 PARIS CEDEX 05
FRANCE