The Annals of Applied Probability
1995, Vol. 5, No. 3, 603–612

CONSISTENCY OF THE MAXIMUM PSEUDO-LIKELIHOOD ESTIMATOR OF CONTINUOUS STATE SPACE GIBBSIAN PROCESSES

BY SHIGERU MASE

Hiroshima University

The purpose of this paper is to show the strong consistency of the maximum pseudo-likelihood estimator for continuous state space stationary Gibbsian processes under fairly general conditions. Besides the maximum pseudo-likelihood estimator of Besag, we consider its extension, the maximum pseudo-likelihood of second order. The framework of our study is Ruelle's theory of superstable potential functions.

1. Introduction. Although spatial statistics is a new branch of statistics, its progress is remarkable. This progress has been supported by advances of probabilistic tools which deal with spatial objects. One of the main subjects of spatial statistics is the analysis of mapped point patterns which arise in many application fields. The natural basis of analysis of mapped point patterns is the theory of Gibbsian distributions, which has a long history in statistical physics. Gibbsian distributions can yield a variety of realistic point patterns, but their statistical analysis has suffered from the complexity of Gibbsian distributions, which gives rise to cumbersome dependencies.

Statisticians are interested in the estimation problem for a parametric family of Gibbsian distributions. Several estimation methods have been proposed. Among them is the maximum pseudo-likelihood estimator (MPLE). Besag introduced the concept of pseudo-likelihood of a Gibbsian model on a lattice as the product of conditional distributions of sites given configurations of other sites in order to avoid the numerical complexity of calculating true likelihoods. Subsequently, the concept was extended to continuous models by taking a limit of pseudo-likelihoods of discretized models. Attractive features of the MPLE are that its computation is numerically easy and that it admits a fairly detailed theoretical analysis.

As to theoretical results for the MPLE, we refer readers to Jensen and Møller (1991), Comets (1992), Guyon and Künsch (1992) and Jensen and Künsch (1994). Comets proved the strong consistency of the MPLE for lattice models. Jensen and Møller proved the weak consistency of the MPLE for continuous models having finite range potentials. Guyon and Künsch proved the remarkable result that for stationary and ergodic Ising models the MPLE

Received January 1994; revised October 1994.

AMS 1991 *subject classifications.* Primary 62M30; secondary 62F12.

Key words and phrases. Consistency, Gibbsian process, superstable potential, pseudo-likelihood, ergodic decomposition, continuous model, parametric model.

is asymptotically normal regardless of parameter values, that is, even if a phase transition occurs. This result is extended in Jensen and Künsch (1994) to continuous models having finite range potentials without the assumption of ergodicity. In this paper we will prove the strong consistency of the MPLE for continuous models under general conditions. Our proof is based on the profound theory of superstable potentials due to Ruelle (1970), which was recently applied successfully to derive large deviation results for continuous Gibbsian models by Georgii (1994).

Beside the MPLE of Besag, we will propose and show the strong consistency of the MPLE of second order, which is an extension of the MPLE of first order, that is, the MPLE of Besag. The MPLE of second order is the maximizer of the pseudo-likelihood of second order, which is, roughly speaking, a limit of products of conditional distributions of every pair of sites given other configurations of discretized models. Although the practical usefulness of the MPLE of second order may be doubtful, its construction shows clearly that the MPLE of first order is an example of a series of moment-type estimators and how Ruelle's theory can be used to prove the consistency of such estimators.

Finally let us comment on applicability of the present result. The Gibbsian model was first applied in mathematical ecology in order to characterize mapped point patterns of plants and animal territories. Geman and Geman (1984) use the Gibbsian model as a prior of a Bayesian restoration of dirty images and gave rise to a new and wide interest in the Gibbsian model; see, for example, Guyon and Künsch (1992) and other papers in the same volume. Besag, York and Mollié (1991) described how this approach can be used in geographical epidemiology. The Gibbsian model is considered now to be a central model of spatial statistics. In all these applications, the model contains an extra parameter which should be estimated from data. Hence there occurs the usual task of constructing estimators and proving their properties such as consistency and asymptotic normality. Consistency is of first importance since it guarantees the possibility of statistical estimation of unknown parameters. Also it is important as a basis for discussing other statistical properties. For example, the consistency of MPLE proved in Jensen and Møller (1991) and the relevant formulation is the basis of the asymptotic normality result of Jensen and Künsch (1994). The author believes that the asymptotic normality result of Jensen and Künsch (1994) is valid generally and can be proved using the framework and the result of the present paper.

2. Preliminaries. Let $\mathscr{C}$ be the set of all configurations (locally finite subsets) of $\mathbb{R}^d$. Let $\mu_G = \mu \cap G$ for $\mu \in \mathscr{C}$ and $G \subset \mathbb{R}^d$. Denote the set of all bounded Borel sets by $\mathbb{B}_0$. The volume of G is denoted by $|G|$. We consider the smallest σ-algebra $\mathscr{F}$ of $\mathscr{C}$ which makes all number functions $N_G(\mu) = \#\mu_G$, $G \in \mathbb{B}_0$, measurable. The set of all configurations of $G \subset \mathbb{R}^d$ is denoted by $\mathscr{C}_G$ and $\mathscr{F}_G$ is the restriction of $\mathscr{F}$ to $\mathscr{C}_G$. Consider the partition of $\mathbb{R}^d$ by unit hypercubes $F(r)$ with center at $r = (r_1, \dots, r_d) \in \mathbb{Z}^d$ and let, in particular, $F = F(0)$. The number of points in $\mu \cap F(r)$ is denoted by $n(\mu, r)$. A sequence

$\{G_n\} \subset \mathbb{B}_0$ is called *regular* if there is an increasing sequence $\{G'_n\}$ of compact convex sets with $G_n \subset G'_n$, $\sup_n |G'_n|/|G_n| < +\infty$, and the supremum of radii of spheres lying in G'_n going to ∞.

Let $|r| = \max_i r_i$. A probability measure on $(\mathscr{C}, \mathscr{F})$ is *tempered* in the sense of Ruelle (1970) if it is supported by the set of configurations

$$S = \bigcup_{N \geq 1} S_N = \bigcup_{N \geq 1} \left\{ \mu \in \mathscr{C};\ \sum_{r:|r| \leq l} n(\mu, r)^2 \leq N^2 (2l+1)^d \text{ for all } l \right\}.$$

Let z be a *chemical potential*, an arbitrary real number, and $\Phi(x)$ be a *pair potential function* which is an even Borel measurable function from $\mathbb{R}^d$ to $(-\infty, +\infty]$. The constant e^{-z} is called the *activity*. The *interaction energy* of a finite configuration $(x)_n = (x_1, \dots, x_n)$ is defined by

$$U((x)_n) = \sum_{1 \leq i < j \leq n} \Phi(x_i - x_j). \tag{1}$$

If there is a constant $B > 0$ such that, for every finite configuration μ, $U(\mu) \geq -B\#\mu$, Φ is called *stable*. A potential is called *hard core* if it takes the value ∞ in a neighborhood of the origin. It is said to be *of finite range* if $\Phi(x) = 0$ except a neighborhood of the origin. For two configurations c and μ, the *mutual interaction energy* is defined (if meaningful) by $W(c, \mu) = \sum_{x \in c,\ y \in \mu} \Phi(x - y)$. Let us define $U(c, \mu) = U(c) + W(c, \mu \setminus c)$, the energy of c with the outer configuration μ. We also use the quantity $E_G(\mu) = \sum_{x \in \mu_G} W(\{x\}, \mu \setminus \{x\})$.

We will assume the following four conditions on the potential Φ.

(C1) Φ is *superstable*; that is, there are two constants $A > 0$ and $B \geq 0$ such that, for each finite subset $R \subset \mathbb{Z}^d$ and configuration $c \subset \bigcup_{r \in R} F(r)$,

$$U(c) \geq \sum_{r \in R} \left[A n(c, r)^2 - B n(c, r) \right].$$

(C2) Φ is *lower regular*; that is, there exists a positive decreasing function Ψ on the positive integers such that $\sum_r \Psi(|r|) < \infty$ and, if R and S are two finite subsets of $\mathbb{Z}^d$ and μ (resp. ν) is a configuration in $\bigcup_{r \in R} F(r)$ [resp. $\bigcup_{s \in S} F(s)$], then

$$W(\mu, \nu) \geq -\tfrac{1}{2} \sum_{r \in R} \sum_{s \in S} \Psi(|r - s|) \left[n(\mu, r)^2 + n(\nu, r)^2 \right]. \tag{2}$$

(C3) $\Phi^- = \Phi - |\Phi|$ is lower regular.

(C4) the following integral exists for some (hence for every) $\beta > 0$:

$$\int_{\mathbb{R}^d} |1 - e^{-\beta \Phi(x)}|\, dx$$

and $\Phi(x) \to 0$ as $|x| \to \infty$.

REMARK 1. A nonnegative potential is stable and lower regular. If, in addition, it is bounded by 0 from below in a neighborhood of the origin, it is superstable. A finite range stable potential is lower regular. Simple and very effective criteria for stability, superstability and lower regularity are known; see Ruelle (1970, 1983).

We assume that the reader is familiar with the definition and basic properties of (both local and global) Gibbsian distributions; see Preston (1976) and Ruelle (1970, 1983). For a fixed potential Φ, let $P_{z,\beta}$ be a global Gibbsian distribution with activity e^{-z} and potential $\beta\Phi$. Ruelle (1970) showed that there exists a tempered and stationary Gibbsian distribution for every parameter (z, β) if Φ satisfies conditions C1, C2 and C4. If z is large enough, there exists a unique Gibbsian distribution. It is known that the set $\mathscr{G}_{z,\Phi}$ of Gibbsian distributions for given z and Φ may not be a singleton (phase transition). There exists at least one member of $\mathscr{G}_{z,\Phi}$ which is both stationary and ergodic. In particular, if it is a singleton, the unique member is stationary and ergodic.

In proofs of the next section it is crucial to show the integrability of several random variables with respect to global Gibbsian distributions. This is possible if we borrow the results of Ruelle (1970). We collect necessary results in the following lemma.

LEMMA 1. *Let Φ be a function satisfying conditions* C1, C2 *and* C4, *and let $\mathbb{P}$ be a tempered Gibbsian distribution with activity e^{-z} and potential $\beta\Phi$. Then there are constants $\gamma > 0$ and δ only depending on z and $\beta\Phi$ such that the following inequality holds*:

$$(3)\quad \mathbb{P}\left\{\sum_{|r|\le l} n(X,r)^2 \ge N^2(2L+1)^d\right\} \le \exp\left[-(\gamma N^2 - e^{\delta})(2l+1)^d\right].$$

Also there exist constants $g > 0$ and $q \ge 0$ such that if $\Delta \in \mathbb{B}_0$ has diameter $L \ge 1$,

$$(4)\qquad \mathbb{P}\{N_{\Delta}(X) \le m\} \le \exp\left[-g\frac{m^2}{L^d} + qm\right].$$

The same inequalities also hold for the local Gibbsian distribution $\mathbb{P}_G$ if $G \subset \bigcup_{|r|\le l} F(r)$ or $G \subset \Delta$, respectively.

The proof of these inequalities for local Gibbsian distributions is given in Ruelle [(1970), Corollaries 2.8 and 2.9]. The same proof also applies to the case of global Gibbsian distributions if we start from the inequality

$$(5)\qquad \sigma_G^m(\mu) \le \exp\sum_r\left[-\gamma n(\mu,r)^2 + \delta n(\mu,r)\right]$$

[see Ruelle (1970), Corollary 5.3], where σ_G^m, $G \in \mathbb{B}_0$, $m = 0, 1, 2, \dots,$ is defined by

$$\sigma_G^m((x)_m) = \int_{C_{G^c}} \exp(-U((x)_m, \mu))\, d\mathbb{P}(\mu),$$

and if the temperedness is taken into account. The constants γ and δ in (5) depend only on z and $\beta\Phi$ and can be used as those in (3).

LEMMA 2. *Let Φ be a function satisfying conditions* C1, C2 *and* C4, *and let $\mathbb{P}$ be a tempered Gibbsian distribution. If ψ is a positive decreasing function defined on the positive integers and the sum $\sum_{l \geq 1} l^{d-1}\psi(l)$ is finite, the following random variable is $\mathbb{P}$-integrable for every positive α*:

$$\exp\left[\alpha \sum_r n(X, r)\psi(|r|)\right]. \tag{6}$$

We can prove this result by the same argument used in the proof of Ruelle [(1970), Proposition 5.2].

LEMMA 3. *Let Φ be a function satisfying conditions* C1, C2, C3 *and* C4, *and let $\mathbb{P}$ be a tempered Gibbsian distribution. If $G \in \mathbb{B}_0$ and $\alpha > 0$, $N_G(X)^n$ as well as the following integrals are $\mathbb{P}$-integrable for $n = 0, 1, 2, \dots$ and $m = 1, 2, \dots$:*

$$\int_{G^m} U((x)_m, X)^n \exp(-\beta U((x)_m, X)) d(x)_m.$$

PROOF. Using (4) we can get

$$\mathbb{E}\{N_G(X)^n\} \leq \sum_{m \geq 1} \left[(m+1)^n - m^n\right] \exp\left[-gm^2/L^d + qm\right] < \infty.$$

Let us next prove the integrability of the integral. Let $U^+(c, \mu)$ and $U^-(c, \mu)$ be equal to $U(c, \mu)$, but with Φ being replaced by $|\Phi|$ and Φ^-, respectively. Then

$$\left|U(c, X)^n \exp(-\alpha U(c, X))\right| \leq (n/\alpha e)^n \exp(-\alpha U^-(c, X)).$$

If c is a finite configuration with $c \in \bigcup_{r \in R} F(r)$,

$$U^-(c, X) \geq U^-(c) - \sum_{r \in R} \sum_{s \in \mathbb{Z}^d} \Psi^-(|r - s|) n(c, r) n(X, s),$$

where Ψ^- is the function which appears in the definition (2) of lower regularity of Φ^- and U^- is the energy corresponding to Φ^-. Using Hölder's inequality, we obtain that the integral

$$\int \exp\left\{\alpha \sum_{r \in R} \sum_{s} \Psi^-(|r - s|) n(c, r) n(X, s)\right\} d\mathbb{P} \tag{7}$$

is bounded by the integral

$$\int \exp\left\{\alpha \sum_s \Psi_0(|s|) n(X, s)\right\} d\mathbb{P}, \tag{8}$$

where we set, with $\rho = \max_{r \in R}|r|$,

$$\Psi_0(l) = \begin{cases} \Psi^-(l-\rho), & \text{if } l \geq \rho, \\ \max_{l \leq \rho} \Psi^-(l), & \text{otherwise.} \end{cases}$$

This integral is finite by Lemma 2. □

3. Strong consistency of MPLE. The logarithm of the pseudo-likelihood for G is given by the formula

$$PL_G(z,\beta) = -zN_G(X) - \beta E_G(X) - \exp(-z)\int_G \exp(-\beta U(\{x\},X))\,dx.$$

Consider the pseudo-likelihood equations

$$\frac{\partial PL_G}{\partial z} = -N_G(X) + \exp(-z)\int_G \exp(-\beta U(\{x\},X))\,dx = 0,$$

$$\frac{\partial PL_G}{\partial \beta} = -E_G(X) + \exp(-z)\int_G U(\{x\},X)\exp(-\beta U(\{x\},X))\,dx = 0.$$

Several authors remarked that the MPLE can be interpreted as a moment-type estimator. Actually it is easy to see from the integral characterization formula of Nguyen and Zessin [(1979a), Formula (3.2)] that

$$\mathbb{E}_{z,\beta}\left\{\frac{\partial PL_G}{\partial z}(z,\beta)\right\} = \mathbb{E}_{z,\beta}\left\{\frac{\partial PL_G}{\partial \beta}(z,\beta)\right\} = 0. \tag{9}$$

Now we will define the (logarithm of) pseudo-likelihood of second order as follows:

$$\begin{aligned} PL_G^{(2)}(z,\beta) = {} & -zN_G(X)(N_G(X)-1) - \beta(N_G(X)-1)E_G(X) \\ & - \beta U(X_G) \\ & - \tfrac{1}{2}\exp(-2z)\int_{G^2}\exp(-\beta U(\{x,y\},X))\,dx\,dy. \end{aligned} \tag{10}$$

The pseudo-likelihood equation for $PL_G^{(2)}$ satisfies the same property as (9).

PROPOSITION 1.

$$\mathbb{E}_{z,\beta}\left\{\frac{\partial PL_G^{(2)}}{\partial z}(z,\beta)\right\} = \mathbb{E}_{z,\beta}\left\{\frac{\partial PL_G^{(2)}}{\partial \beta}(z,\beta)\right\} = 0. \tag{11}$$

PROOF. If we apply the integral characterization formula of Nguyen and Zessin twice to the double sum $\sum_{y \in \mu, y \neq x} k(x,y,\mu)$, then

$$\begin{aligned} &\int_{\mathscr{C}} \sum_{\substack{x,y \in X \\ x \neq y}} k(x,y,X)\,d\mathbb{P}_{z,\beta} \\ &\quad = \exp(-2z)\int_{\mathbb{R}^4} dx\,dy \int_{\mathscr{C}} k(x,y,X \cup \{x,y\}) \\ &\qquad \times \exp(-\beta U(\{x,y\},X))\,d\mathbb{P}_{z,\beta}. \end{aligned} \tag{12}$$

Let $k(x, y, \mu) = 1_G(x)1_G(y)$. Then (12) is equivalent to the first equation of the proposition. Also let

$$k(x, y, \mu) = 1_G(x)1_G(y)U(\{x, y\}, \mu \setminus \{x, y\}).$$

Then (12) is equivalent to the second equation of the proposition. □

REMARK 2. Actually the pseudo-likelihood of second order is constructed so that equation (11) should hold. Note that expectations in (11) can be thought of as expectations with respect to two-point Palm measures $\mathbb{P}_{z,\beta}^{x,y}$, which are defined by

$$d\mathbb{P}_{z,\beta}^{x,y} \propto \exp(-\beta U(\{x, y\}, X))\, d\mathbb{P}_{z,\beta}.$$

The logarithm of pseudo-likelihood $PL_G^{(2)}(z, \beta)$ is a strictly concave function of (z, β) with probability 1. Therefore, the MPLE of second order $(\hat{z}, \hat{\beta})$ exists a.s.

Now we will prove the strong consistency of the MPLE. We deal with the MPLE of second order only. However, it can be seen that the same method yields the strong consistency proof of the MPLE of first order. We need the following result on convex functions.

LEMMA 4. *Let $\Delta \subset \mathbb{R}^m$ be an open convex domain and let $\{f_n\}$ be a sequence of concave functions defined on Δ. Assume that a finite limit $f(x) = \lim f_n(x)$ exists everywhere and f takes a unique maximum at x_0. Let x_n be a point which maximizes f_n. Then $x_n \to x_0$.*

PROOF. It is known that if concave functions converge on a dense subset of Δ, they converge everywhere and locally uniformly and the limit is also concave. Let $A = \{x \in \Delta; |x - x_0| = \varepsilon\}$ for each $\varepsilon > 0$. Also let $\delta = f(x_0) - \max_{x \in A} f(x) > 0$. For n large enough, $|f_n(x_0) - f(x_0)| \le \delta/2$ and

$$\max_{x \in A} |f_n(x) - f(x)| \le \delta/2.$$

Then

$$\max_{x \in A} f_n(x) \le \max_{x \in A} f(x) + \delta/2 = f(x_0) - \delta + \delta/2 \le f_n(x_0).$$

Therefore, f_n takes the maximum in the region $|x - x_0| \le \varepsilon$, that is, $|x_n - x_0| \le \varepsilon$. □

We will need the following ergodic results from the spatial ergodic theorem of Nguyen and Zessin (1979b). Integrability of relevant random variables is proved in Lemma 3.

LEMMA 5. *Let $\{G_n\}$ be a regular and increasing sequence of convex sets which expands to $\mathbb{R}^d$. Suppose the potential function Φ is superstable and* *lower regular. Assume that the corresponding Gibbsian distribution* $\mathbb{P}$ *is ergodic. Then, with probability* 1,

$$|G_n|^{-1} N_{G_n}(X) \to \mathbb{E}\{N_F(X)\},$$

$$|G_n|^{-1} E_{G_n}(X),\ 2|G_n|^{-1} U(X_{G_n}) \to \mathbb{E}\{E_F(X)\}, \tag{13}$$

$$|G_n|^{-1} \int_{G_n} f(x, X)\, dx \to \mathbb{E}\left\{\int_F f(x, X)\, dx\right\}, \tag{14}$$

$$\frac{1}{|G_n|^2} \int_{G_n^2} \exp(-U(\{x, y\}, X))\, dx\, dy \to \left[\mathbb{E}\left\{\int_F \exp(-U(\{x\}, X))\, dx\right\}\right]^2, \tag{15}$$

where $f(x, \mu)$ *is any nonnegative function such that the expectation on the right-hand side of* (14) *exists.*

PROOF. The convergence (13) is proved in Nguyen and Zessin (1979b). Let us prove (15). By condition C4 there exists $r > 0$ for each $\varepsilon > 0$ such that $|1 - e^{-\beta\Phi(x)}| \le \varepsilon$ if $|x| \ge r$. There is also a constant k such that $|1 - e^{-\beta\Phi(x)}| \le k$ for every x. Then

$$\begin{aligned} &\left| \int_{G^2} \exp(-\beta U(\{x, y\}, \mu))\, dx\, dy \right. \\ &\left. - \int_{G^2} \exp(-\beta U(\{x\}, \mu) - \beta U(\{y\}, \mu))\, dx\, dy \right| \\ &\le \varepsilon \left(\int_G \exp(-\beta U(\{x\}, \mu))\, dx \right)^2 \\ &\quad + k \int_G dx \int_{b(r, x)} \exp(-\beta U(\{x\}, \mu) - \beta U(\{y\}, \mu))\, dy, \end{aligned}$$

where $b(r, x)$ is the closed ball with center at x and radius r. On the other hand, from Schwartz's inequality, the second integral on the right-hand side of the last inequality is bounded by

$$|b(r, 0)| \int_{G \oplus b(r, 0)} \exp(-\beta U(\{x\}, \mu))\, dx,$$

where the symbol $\oplus$ stands for the vectorial sum of two sets. Therefore, the proof can be completed if we use Lemma 5. □

Now we are ready to prove our main result.

PROPOSITION 2. *Let a potential function* Φ *satisfy the four conditions* C1, C2, C3 *and* C4. *Let* $\{\mathbb{P}_\theta\}$, $\theta = (z, \beta) \in \Theta$, *be a family of stationary and tempered Gibbsian distributions corresponding to* $(z, \beta\Phi)$, *where* Θ *is an open* *subset of* $(-\infty,\infty)\times(0,\infty)$. *Assume* $\{G_n\}\subset\mathbb{B}_0$ *is a sequence of regular convex sets expanding to* $\mathbb{R}^d$. *If* $\hat{\theta}_n=(\hat{z}_n,\hat{\beta}_n)$ *denotes the MPLE of second order calculated from the observation* X_{G_n}, *it is strongly consistent.*

PROOF. The set Ξ of tempered stationary Gibbsian distributions for a superstable and lower regular potential is nonempty, convex and compact (with respect to the uniform convergence topology of density functions) and is a Choquet simplex; see Ruelle (1970). Hence, each $\mathbb{P}\in\Xi$ can be uniquely represented as a mean of extremal (ergodic) elements Ξ^* of Ξ as

$$\mathbb{P}=\int_{\Xi^*}\mathbb{Q}\,dJ(\mathbb{Q}).$$

Therefore, if random variables X_n converge to 0 $\mathbb{Q}$-a.s. for each ergodic $\mathbb{Q}$, they also converge to 0 $\mathbb{P}$-a.s.

Let $\theta_0=(z_0,\beta_0)$ be the true parameter. Define two functions

$$A(\beta)=\mathbb{E}_{\theta_0}\left\{\int_F\exp(-\beta U(\{x\},X))\,dx\right\},$$

$$B(\beta)=-A'(\beta)=\mathbb{E}_{\theta_0}\left\{\int_F U(\{x\},X)\exp(-\beta U(\{x\},X))\,dx\right\}.$$

Then from Lemma 5 and the remark above,

$$\frac{1}{|G_n|^2}PL^{(2)}_{G_n}(\theta)\to -z\exp(-2z_0)A^2(\beta_0)$$

$$-\beta\exp(-2z_0)A(\beta_0)B(\beta_0)-\tfrac{1}{2}\exp(-2z)A^2(\beta_0)$$

$\mathbb{P}_{\theta_0}$-a.s. for each θ. We can assume that the last convergence holds for all θ with probability 1. Let $f(\theta)$ be the right-hand side of the last relation. Then it can be shown that

$$\frac{\partial f}{\partial z}(\theta_0)=\frac{\partial f}{\partial\beta}(\theta_0)=0.$$

Furthermore, the matrix of second derivatives of f is seen to be negative definite. Therefore, f is a strict concave function and has the unique maximum at θ_0. Then the assertion follows from Lemma 4. □

We can prove the strong consistency of the MPLE of Besag using a similar argument.

PROPOSITION 3. *Under the same conditions as in Proposition* 2, *the MPLE of first order is strongly consistent.*

The assumption of temperedness in the last two propositions is never too restrictive. If the potential is hard core, the distance of any two points of sample configurations cannot be smaller than the hard-core distance with probability 1. Therefore, it is trivially tempered. We can also show the following result.

PROPOSITION 4. *If Φ is superstable, lower regular and nonnegative, the corresponding Gibbsian distributions are tempered.*

PROOF. Let $G \in \mathbb{B}_0$ and $h(\mu)$ be an $\mathscr{F}_G$-measurable nonnegative function. Then from the DLR equation for Gibbsian distributions,

$$\int h(X)\, d\mathbb{P} \leq Z_G \int_{\mathscr{C}_G} h(X)\, d\mathbb{P}_G,$$

where Z_G is the grand partition function for $\mathbb{P}_G$. It is known that if $G = G_l = \sum_{|r| \leq l} F(l)$, there is a constant p such that

$$\lim_{l \to +\infty} |G_l|^{-1} \log Z_{G_l} \to p;$$

see Ruelle [(1970), Theorem 3.3]. If we combine this result with (3), we can show that there is a constant ρ such that

$$\mathbb{P}\left\{ \sum_{|r| \leq l} n(X, r)^2 \geq N^2 (2L + 1)^d \right\} \leq \exp\left[-(\gamma N^2 - \rho)(2l + 1)^d \right].$$

Hence

$$\mathbb{P}(S_N) \geq 1 - \sum_{l=0}^{\infty} \exp\left[-(\gamma N^2 - \rho)(2l + 1)^d \right] \to 1 \quad \text{as } N \to \infty. \qquad \square$$

REFERENCES

BESAG, J., YORK, J. C. and MOLLIÉ, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Ann. Inst. Statist. Math.* **43** 1–59.

COMETS, F. (1992). On consistency of a class of estimators for exponential families of Markov random fields on the lattice. *Ann. Statist.* **20** 455–468.

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

GEORGII, H.-O. (1994). Large deviations and the equivalence of ensembles for Gibbsian particle systems with superstable interaction. *Probab. Theory Related Fields* **99** 171–195.

GUYON, X. and KÜNSCH, H. R. (1992). Asymptotic comparison of estimators in the Ising model. *Stochastic Models, Statistical Methods, and Algorithms in Image Analysis. Lecture Notes in Statist.* **74** 177–198. Springer, Berlin.

JENSEN, J. L. and KÜNSCH, H. R. (1994). On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. *Ann. Inst. Statist. Math.* **46** 475–486.

JENSEN, J. L. and MØLLER, J. (1991). Pseudolikelihood for exponential family models of spatial point processes. *Ann. Appl. Probab.* **1** 445–461.

NGUYEN, X. X. and ZESSIN, H. (1979a). Integral and differential characterizations of the Gibbs process. *Math. Nachr.* **88** 105–115.

NGUYEN, X. X. and ZESSIN, H. (1979b). Ergodic theorems for spatial processes. *Z. Wahrsch. Verw. Gebiete* **48** 133–158.

PRESTON, C. (1976). *Random Fields. Lecture Notes in Math.* **534**. Springer, Berlin.

RUELLE, D. (1970). Superstable interactions in classical statistical mechanics. *Comm. Math. Phys.* **18** 127–159.

RUELLE, D. (1983). *Statistical Mechanics: Rigorous Results.* Addison-Wesley, Reading, MA.

FACULTY OF INTEGRATED ARTS AND SCIENCES
HIROSHIMA UNIVERSITY
HIGASHI-HIROSHIMA CITY
HIROSHIMA, 724
JAPAN