The Annals of Applied Probability
1991, Vol. 1, No. 3, 445–461

PSEUDOLIKELIHOOD FOR EXPONENTIAL FAMILY MODELS OF SPATIAL POINT PROCESSES

BY JENS LEDET JENSEN AND JESPER MØLLER

University of Aarhus

The pseudolikelihood for general spatial point processes including marked point processes is derived, and some of its properties are investigated. In particular, we prove consistency of the maximum pseudolikelihood estimates for Markov processes of finite range.

1. Introduction. Many of the models used in spatial point processes have a density of the form $g(x, \theta)/Z(\theta)$, where $g(x, \theta)$ is an explicitly given function expressed in terms of interaction functions with x the data and θ a parameter, and $Z(\theta)$ is a normalizing constant that cannot be evaluated explicitly. Since $Z(\theta)$ is not known, it is difficult to perform ordinary likelihood inference. In the literature there are two ways of dealing with this problem. One considers approximations to $Z(\theta)$ either based on asymptotic arguments or based on simulations [see, in particular, Ogata and Tanemura (1984), Moyeed and Baddeley (1990) and the review given in Ripley (1988)]. In the other approach, which is the one of interest in this article, one avoids the use of the likelihood function and introduces instead a pseudolikelihood function for estimating θ. The pseudolikelihood idea originates from lattice processes [Besag (1974)], but was extended in the special case of a Strauss point process in Besag (1977). The extension is essentially based on approximating the spatial point process by a lattice process, and then using the pseudolikelihood for lattice processes. A more detailed account of this approximation is given in Besag, Milne and Zachary (1982), and based on these works Ripley (1988) states a general version of the pseudolikelihood. If in the above approximation of the spatial point process by a lattice process one does not go to the limit, it is possible to obtain a logistic likelihood instead of the pseudolikelihood [see Clyde and Strauss (1988)]. In Section 2 we derive the pseudolikelihood by a direct argument for general spatial point processes including ordinary point processes on bounded subsets of $\mathbb{R}^d$ and marked point processes such as, for example, random processes of balls. For the subclass of Markov point processes of finite range, we show in Section 3 that the maximum pseudolikelihood estimate is consistent. Three examples are briefly discussed in Section 4.

Consistency seems not to have been considered before for spatial point processes, whereas the case of a lattice process has been treated by a number of authors. Our proof resembles that of Geman and Graffigne (1986) in the

Received December 1989; revised October 1990.

AMS 1980 *subject classifications*. Primary 60G55; secondary 60J25.

Key words and phrases. Coding, consistency, exponential models, hardcore process, marked point processes, Markov point processes, pseudolikelihood, random processes of balls, spatial point processes, spatial statistics, Strauss process.

way the Markov property is used to give conditional independence. However, the other parts of the proof are different since Geman and Graffigne's proof relies on the state space of the lattice variable being finite. Guyon (1986) and Gidas (1988) also utilize strongly that the state space is compact. Quite a different and general proof in the lattice case has been given by Comets (1989). Comets uses large deviation bounds to obtain the consistency. The state space of the lattice variable need not be compact and the main restriction seems to be finiteness of a certain norm on the interaction. The underlying structure of our proof is given in the Appendix. The results in the Appendix actually give the known results for finite-range lattice processes with a finite state space in an easy way (this derivation can be found in an unpublished report by the authors).

2. Definition and first properties. In this section we give the basic notation in the paper and introduce the pseudolikelihood.

We shall be studying a finite point process X living in a space S; that is, a realization of X consists of a finite number of points in S. We let $(S, \mathscr{B}, \lambda)$ be a measure space with $\mathscr{B}$ containing all singletons and λ a diffuse measure with $\lambda(S) < \infty$. From $(S, \mathscr{B})$ we can construct the exponential space $(\Omega_S, \mathscr{F}_S)$ consisting of all finite counting measures on S. Since we shall consider only point processes X with no multiple points, we can equivalently think of Ω_S as consisting of all finite point configurations $x = \{x_1, \dots, x_n\} \subset S$ with $n = 0, 1, 2, \dots,$ and thus our point process is a random element in Ω_S. For details on the exponential space $(\Omega_S, \mathscr{F}_S)$, see Carter and Prenter (1972). For $x \in \Omega_S$ we let $n(x)$ denote the cardinality, and the empty point configuration with $n(x) = 0$ is denoted by $x = \varnothing$. Thinking of point configurations as subsets of S, we use the obvious notation such as $x \subset y$, $x \cup y$, $x \cup \xi$ and $x_A = x \cap A$ for $x, y \in \Omega_S$, $\xi \in S$ and $A \in \mathscr{B}$, and when $\xi \in x$ we also use $x \setminus \xi$.

From λ we define the Poisson measure on $(\Omega_S, \mathscr{F}_S)$ by

$$\mu_S(F) = e^{-\lambda(S)}\left[1_F(\varnothing) + \sum_{n=1}^{\infty} \frac{1}{n!} \int \cdots \int 1_F(\{x_1, \dots, x_n\}) \lambda(dx_1) \dots \lambda(dx_n)\right]$$

for $F \in \mathscr{F}_S$. Since we have assumed that λ is a diffuse measure, realizations under μ_S have almost surely no multiple points. The restriction of μ_S to Ω_A for $A \in \mathscr{B}$ is denoted μ_A.

The general setup above covers an ordinary finite point process, where S is a bounded Borel set in $\mathbb{R}^d$, $\mathscr{B}$ is the Borel σ-field and λ is the Lebesgue measure. We call this model (M1). Also covered is the model (M2) of a finite marked point process, where $S = S_p \times S_m$, $\mathscr{B} = \mathscr{B}_p \times \mathscr{B}_m$ and $\lambda = \lambda_p \times Q_m$, with $(S_p, \mathscr{B}_p, \lambda_p)$ as in model (M1), and with the mark space $(S_m, \mathscr{B}_m, Q_m)$ a probability space where $\mathscr{B}_m$ contains all singletons. Examples of marked point processes can be found in Baddeley and Møller (1989).

The statistical model is obtained by considering a class of probability measures P_θ, $\theta \in \Theta$, on Ω_S that are absolutely continuous with respect to μ_S. We call the density $f_\theta(x)$ and assume the density to be hereditary and stable for all $\theta \in \Theta$; that is,

$$(2.1) \qquad f_\theta(y) > 0 \quad \text{whenever } f_\theta(x) > 0 \text{ and } y \subset x,$$

and almost surely under μ_S,

$$(2.2) \qquad f_\theta(x) \le c_\theta K_\theta^{n(x)}$$

for some constants $c_\theta, K_\theta > 0$. The conditional density of X_A given $X_{S \setminus A} = x_{S \setminus A}$ with respect to μ_A is given by

$$(2.3) \qquad f_\theta(x_A | x_{S \setminus A}) = \frac{f_\theta(x_A \cup x_{S \setminus A})}{\int f_\theta(y \cup x_{S \setminus A}) \mu_A(dy)} \quad \text{if } f_\theta(x_{S \setminus A}) > 0,$$

and 0 otherwise. Now, let T be a fixed set in $\mathscr{B}$ and let $A_{ij} \in \mathscr{B}$, $i = 1, 2, \dots$, $j = 1, \dots, m_i$, be a nested division of T; that is, $T = \bigcup_{j=1}^{m_i} A_{ij}$, where $A_{ij} \cap A_{ij'} = \varnothing$ for $j \ne j'$ and $A_{ij} \subseteq A_{i-1, j'}$ for some j'. We assume that

$$(2.4) \qquad m_i \to \infty \quad \text{and} \quad m_i \delta_i^2 \to 0 \quad \text{as } i \to \infty,$$

with $\delta_i = \max_{1 \le j \le m_i} \lambda(A_{ij})$.

Definition 2.1. For $T \in \mathscr{B}$ and $x \in \Omega_S$ we define the *pseudolikelihood on* T by

$$(2.5) \qquad \mathrm{PL}_T(\theta) = \exp(-\lambda(T)) \lim_{i \to \infty} \prod_{j=1}^{m_i} f_\theta\big(x_{A_{ij}} | x_{S \setminus A_{ij}}\big).$$

We have only included the constant $\exp(-\lambda(T))$ in the right-hand side of (2.4) for convenience as seen from the following derivation of the pseudolikelihood.

Theorem 2.2. *For μ_S-a.a. $x \in \Omega_S$ the pseudolikelihood on T is well defined and given by*

$$(2.6) \qquad \mathrm{PL}_T(\theta) = \exp\{-B_\theta(x, T)\} \prod_{\xi \in x_T} b_\theta(x \setminus \xi, \xi),$$

where for $\xi \in S$,

$$(2.7) \qquad b_\theta(x, \xi) = f_\theta(x \cup \xi)/f_\theta(x) \quad \text{if} \quad f_\theta(x) > 0$$

and 0 otherwise, and for $A \in \mathscr{B}$,

$$(2.8) \qquad B_\theta(x, A) = \int_A b_\theta(x, \xi) \lambda(d\xi).$$

Proof. Since from (2.4)

$$\mu_S\Big(\big\{x | n(x_{A_{ij}}) \le 1 \text{ for all } j\big\}\Big) = \prod_{j=1}^{m_i} \big\{1 + \lambda(A_{ij})\big\} \exp\big\{-\lambda(A_{ij})\big\}$$
$$= 1 + O\big(m_i \delta_i^2\big) \to 1 \quad \text{for } i \to \infty,$$

we may assume that $n(x_{A_{ij}}) \leq 1$ for all j for i sufficiently large. Also we only need to consider the case $f_\theta(x) > 0$ because otherwise both sides of (2.6) are 0. When $f_\theta(x) > 0$ we have from (2.1) and the finiteness of x that there exists $k(x) > 0$ such that $f_\theta(x_{S \setminus A_{ij}}) \geq k(x)$ for all i and j. Using (2.2), we find

$$\int f_\theta(y \cup x_{S \setminus A}) \mu_A(dy)$$

$$= f_\theta(x_{S \setminus A}) e^{-\lambda(A)} + \int_A f_\theta(x_{S \setminus A} \cup \xi) \lambda(d\xi) e^{-\lambda(A)}$$

$$+ \omega \sum_2^\infty c_\theta \frac{(K_\theta \lambda(A))^n}{n!} e^{-\lambda(A)}$$

with $|\omega| \leq 1$, and then from (2.3) we get when $n(x_A) \leq 1$,

$$f_\theta(x_A | x_{S \setminus A}) = \frac{b_\theta(x_{S \setminus A}, x_A)^{n(x_A)} e^{\lambda(A)}}{1 + B_\theta(x_{S \setminus A}, A) + \frac{1}{k(x)} O(\lambda(A)^2)}. \tag{2.9}$$

When multiplying together terms of the form (2.8), to get the left-hand side of (2.6), we obtain the right-hand side of (2.6) by a Taylor expansion of the logarithm of the denominator of (2.8). The remainder term of the Taylor expansion is handled by the use of (2.4) and the fact that $B_\theta(x_{S \setminus A}, A) = k(x)^{-1} O(\lambda(A))$. □

When modeling spatial point processes, there is a choice between looking at what happens inside a window T not taking account of what happens outside T, or looking at the conditional density given the configuration outside T. The latter method is used to avoid edge effects [see, e.g., Ripley (1988)]. We note here that the pseudolikelihood on T is unaltered whether based on $f_\theta(x)$ or based on the conditional density $f_\theta(x_T | x_{S \setminus T})$, as appears from (2.3) and (2.7).

The main advantage of the pseudolikelihood as compared to the ordinary likelihood is that the latter usually involves a complicated normalizing constant, which cancels out in $b_\theta(x, \xi)$ and hence also in $\mathrm{PL}_T(\theta, x)$. Let us consider the case where the density $f_\theta(x)$ belongs to an exponential family. We write the density as

$$f_\theta(x) = \frac{1}{Z(\theta)} h(x) e^{\theta \cdot v(x)}, \qquad x \in \Omega_S, \theta \in \Theta, \tag{2.10}$$

where Θ is an open subset of $\mathbb{R}^k$ and $h\colon \Omega_S \to [0, \infty)$ is hereditary so that (2.1) is satisfied. Defining

$$v(x, \xi) = v(x \cup \xi) - v(x) \quad \text{and} \quad h(x, \xi) = h(x \cup \xi)/h(x)$$

for $h(x) > 0$ and 0 otherwise, we find from (2.7) that

$$b_\theta(x, \xi) = h(x, \xi) \exp\{\theta \cdot v(x, \xi)\}, \tag{2.11}$$

and the logarithm of the pseudolikelihood function (2.6) becomes

$$(2.12)\quad \mathrm{pl}_T(\theta, x) = \theta \cdot \sum_{\xi \in x_T} v(x \setminus \xi, \xi) - \int_T h(x, \xi) \exp\{\theta \cdot v(x, \xi)\} \lambda(d\xi).$$

Thus the complicated function $Z(\theta)$ from (2.10) has disappeared and $\mathrm{pl}_T(\theta, x)$ can be evaluated fairly easily using numerical integration.

For the exponential family model (2.10), we now state two properties for the pseudolikelihood function, which shows its resemblance to a true likelihood function for an exponential model.

PROPOSITION 2.3. *Let $U = S \setminus T$. Then*

(i) *$\mathrm{pl}_T(\theta, x)$ is a concave function in θ;*
(ii) *the conditional mean $E\{\mathrm{pl}_T(\theta, X_T \cup x_U) | x_U\}$ under P_{θ_0} is strictly concave in θ with a maximum at θ_0 if and only if for all $\theta \in \Theta \setminus \{\theta_0\}$:*

$$(2.13)\quad \mu_T\left\{x_T \Big| \int_T 1\left[b_\theta(x_T \cup x_U, \xi) \neq b_{\theta_0}(x_T \cup x_U, \xi)\right] \lambda(d\xi) > 0\right\} > 0.$$

PROOF. From (2.12) we find

$$(2.14)\quad -\alpha\left\{\frac{\partial^2}{\partial\theta\,\partial\theta^*}\mathrm{pl}_T(\theta, x)\right\}\alpha = \int_T \{\theta \cdot v(x, \xi)\}^2 h(x, \xi) \times \exp\{\theta \cdot v(x, \xi)\} \lambda(d\xi)$$

for any $\alpha \in \mathbb{R}^d$, which immediately gives (i). Now (2.14) is strictly positive if and only if

$$(2.15)\quad \int_T 1[\alpha \cdot v(x_T \cup x_U, \xi) \neq 0, h(x_T \cup x_U, \xi) > 0] \lambda(d\xi) > 0,$$

and $h(x, \xi) > 0$ implies $h(x) > 0$ and therefore from (2.3) and (2.10) that $f_{\theta_0}(x_T | x_U) > 0$. This gives that the conditional mean of (2.14) given x_U is strictly positive if and only if the set of x_T's satisfying (2.15) has positive μ_T measure. This must hold for all $\alpha \neq 0$ to get the strict concavity. Using (2.11) and translating "for all $\alpha \neq 0$" into "for all $\theta \neq \theta_0$" we obtain (2.13) since Θ is open. That the maximum of the conditional mean is at θ_0 is seen by showing that the derivative of the conditional mean is 0 at θ_0. □

3. Consistency of maximum pseudolikelihood estimates. In this section we prove consistency of the maximum pseudolikelihood estimate in the exponential family model (2.10), under the further assumption that the model is a Markov model of finite range.

We start with the ordinary point process (M1), where S is a bounded subset of $\mathbb{R}^d$. The Markov structure, in the sense of Ripley and Kelly (1977), is defined through a symmetric translation-invariant relation $\sim$ on $\mathbb{R}^d$. The relation is of finite range $D = \sup\{\|\xi\| \,|\, \xi \sim 0\} < \infty$, and a set $x \subset \mathbb{R}^d$ is a clique with respect to $\sim$ if $x \neq \varnothing$ and $\xi \sim \zeta$ for all distinct $\xi, \zeta \in x$. The functions h and

v in the density (2.10) are defined through translation-invariant interaction functions φ and ψ,

$$h(x) = \prod_{y \subseteq x} \varphi(y) \quad \text{and} \quad v(x) = \sum_{y \subseteq x} \psi(y), \tag{3.1}$$

with the properties

$$\varphi(x) \geq 0, \psi(x) \in \mathbb{R}^k \text{ and } \varphi(x) = 1, \psi(x) = 0 \text{ if } x \text{ is not a clique.} \tag{3.2}$$

With these definitions we find

$$h(x, \xi) = \varphi_1 \prod_{\varnothing \neq y \subseteq x} \varphi(y \cup \xi), \qquad v(x, \xi) = \psi_1 + \sum_{\varnothing \neq y \subseteq x} \psi(y \cup \xi) \tag{3.3}$$

and

$$b_\theta(x, \xi) = \varphi_1 e^{\theta \cdot \psi_1} \prod_{\varnothing \neq y \subseteq x} \varphi(y \cup \xi) \exp\{\theta \cdot \psi(y \cup \xi)\}, \tag{3.4}$$

where $\varphi_1 = \varphi(\{\xi\})$ and $\psi_1 = \psi(\{\xi\})$, which are constants because of translation invariance.

For a set $A \subseteq \mathbb{R}^d$ we will use the notation $\partial A = \{\xi \in \mathbb{R}^d \setminus A | \exists\, \zeta \in A \text{ with } \zeta \sim \xi\}$, $\bar{A} = A \cup \partial A$ and $\mathring{A} = (A^c \cup \partial A^c)^c = \{\xi \in A | \xi \nsim \zeta \;\forall\, \zeta \notin A\}$.

We will prove consistency for $n \to \infty$, where now $S = S(n)$ increases to $\mathbb{R}^d$ as $n \to \infty$. In the proof we use a "coding" principle, and for this we assume that $\mathring{S}(n)$ is a disjoint union of Borel sets A_{nij}, $i = 1, \dots, \nu$, $j = 1, \dots, m_{ni}$,

$$\mathring{S}(n) = \bigcup_{j=1}^{\nu} \bigcup_{j=1}^{m_{ni}} A_{nij} \quad \text{with } m_{ni} \to \infty \text{ as } n \to \infty, \tag{3.5}$$

such that

$$\delta = \sup_{n,i,j} \lambda(A_{nij}) < \infty \quad \text{and} \quad d(A_{nij}, A_{nik}) > D \quad \text{for } j \neq k. \tag{3.6}$$

Here $d(\cdot, \cdot)$ is the Euclidean distance between two sets. In (3.6) we have an upper bound on the sizes of A_{nij}. We also need a lower bound related to the identifiability condition (2.13). We assume that there exists a Borel set B such that with θ_0 the true parameter point,

$$\forall\, \theta \in \Theta \setminus \{\theta_0\}: \int\int_B 1\left[b_\theta(x, \xi) \neq b_{\theta_0}(x, \xi)\right] d\xi\, \mu_B(dx) > 0 \tag{3.7}$$

and

$$\forall\, n, i, j\; \exists\, \xi \in \mathbb{R}^d: B + \xi \subseteq \mathring{A}_{nij}. \tag{3.8}$$

We first study properties related to a typical set A among the A_{nij}'s. Let therefore A be a Borel set with $\lambda(A) \leq \delta$ and such that $B + \xi \subseteq \mathring{A}$ for some $\xi \in \mathbb{R}^d$. Because of the Markov property the conditional density of X_A given

$X_{S(n)\setminus A} = x_{S(n)\setminus A}$ depends on $x_{\partial A}$ only and is given by

$$(3.9)\quad f_{\theta_0}(x_A|x_{\partial A}) = \frac{1}{Z(\theta_0, x_{\partial A})} \prod_{\varnothing \neq y \subseteq x_A} \prod_{z \subseteq x_{\partial A}} \varphi(y \cup z) \exp\{\theta_0 \cdot \psi(y \cup z)\}$$

for $h(x_{\partial A}) > 0$ and 0 otherwise.

LEMMA 3.1. *Assume that for θ in an open neighborhood of θ_0 we have for all finite point configurations x with $n(x) \geq 2$,*

$$(3.10)\qquad \varphi(x) \leq 1 \quad \text{and} \quad \theta \cdot \psi(x) \leq 0.$$

Then

(i)

$$(3.11)\qquad \exists\, c_1, c_2 > 0: \|\psi(x)\| \leq -c_2\theta \cdot \psi(x)$$

for $\|\theta - \theta_0\| \leq c_1$ and $n(x) \geq 2$;

(ii)

$$(3.12)\qquad e^{-\delta} \leq Z(\theta_0, x_{\partial A}) \leq \exp\{\delta[\varphi_1 \exp(\theta_0 \cdot \psi_1)]\};$$

(iii)

$$(3.13)\qquad \exists\, c_3, c_4 > 0: -\alpha^*\left\{\frac{\partial^2}{\partial\theta\,\partial\theta^*}\mathrm{pl}_A(\theta, x_{\bar{A}})\right\}\alpha \leq c_4$$

for $\|\theta - \theta_0\| \leq c_3$ and for any α with $\|\alpha\| = 1$;

(iv)

$$(3.14)\qquad \exists\, c_5, c_6 > 0: -\alpha^*\left\{\frac{\partial^2}{\partial\theta\,\partial\theta^*}E[\mathrm{pl}_A(\theta, X_A \cup x_{\partial A})|x_{\partial A}]\right\}\alpha \geq c_6$$

for $\|\theta - \theta_0\| \leq c_5$ and for any α with $\|\alpha\| = 1$;

(v)

$$(3.15)\qquad \begin{aligned} &\exists\, c_7, c_8, c_9, c_{10} > 0: E\{|\mathrm{pl}_A(\theta, X_A \cup x_{\partial A})|^k |x_{\partial A}\} \\ &\qquad \leq c_8 c_9^k \max(m_k(c_{10}), k^k e^{-k}) \end{aligned}$$

for $\|\theta - \theta_0\| \leq c_7$ and $k \geq 1$, where m_k is the kth moment of a Poisson distribution with mean c_{10}.

PROOF. (i) Let (3.10) hold for $\|\theta - \theta_0\| < \varepsilon$. If $\theta_0 = 0$, then (3.10) implies that $\psi(x) = 0$ and (3.11) holds trivially. If $\theta_0 \neq 0$ we take $k < \min(\varepsilon, \|\theta_0\|)$ and define $\cos\alpha = \min_{\|\theta - \theta_0\| \leq k} (\theta/\|\theta\|) \cdot (\theta_0/\|\theta_0\|)$. Then (3.10) implies that $(\theta_0/\|\theta_0\|) \cdot (-\psi(x)/\|\psi(x)\|) \geq \sin\alpha$ and therefore

$$(\theta/\|\theta\|) \cdot (-\psi(x)/\|\psi(x)\|) \geq \sin(\alpha/2) \quad \text{if } (\theta_0/\|\theta_0\|) \cdot (\theta/\|\theta\|) > \cos(\alpha/2).$$

We thus obtain (3.11) by taking $c_1 < k$ such that $(\theta/\|\theta\|) \cdot (\theta_0/\|\theta_0\|) > \cos(\alpha/2)$ for $\|\theta - \theta_0\| < c_1$ and $c_2 = \{(\|\theta_0\| - c_1)\sin(\alpha/2)\}^{-1}$.

(ii) The density (3.9) is with respect to μ_A and so

$$Z(\theta_0, x_{\partial A}) \geq \mu_A\{n(X_A) = 0\} = \exp(-\lambda(A)) \geq e^{-\delta}.$$

Similarly from (3.10) we get

$$Z(\theta_0, x_{\partial A}) \leq \int \{\varphi_1 \exp(\theta_0 \cdot \psi_1)\}^{n(x)} \mu_A(dx)$$
$$= \exp\{\lambda(A)[\varphi_1 \exp(\theta_0 \cdot \psi_1) - 1]\},$$

which gives (3.12).

(iii) Putting together (2.12), (3.3) and (3.4), we have

$$\begin{aligned} \mathrm{pl}_A(\theta, x_{\bar{A}}) &= \theta \cdot \left\{ n(x_A)\psi_1 + \sum_{\xi \in x_A} \sum_{\varnothing \neq y \subseteq x_{\bar{A}} \setminus \xi} \psi(y \cup \xi) \right\} \\ &\quad - \varphi_1 \exp(\theta \cdot \psi_1) \int_A \prod_{\varnothing \neq y \subseteq x_{\bar{A}}} \varphi(y \cup \xi) \exp\{\theta \cdot \psi(y \cup \xi)\}\, d\xi. \end{aligned} \tag{3.16}$$

The second derivative is then easily obtained, and for any α with $\|\alpha\| = 1$ we get

$$-\alpha^* \left\{ \frac{\partial^2}{\partial\theta\, \partial\theta^*} \mathrm{pl}_A(\theta, x_{\bar{A}}) \right\} \alpha \leq \varphi_1 \exp(\theta \cdot \psi_1) \int_A \left\| \psi_1 + \sum_{\varnothing \neq y \subseteq x_{\bar{A}}} \psi(y \cup \xi) \right\|^2$$
$$\times \prod_{\varnothing \neq y \subseteq x_{\bar{A}}} \varphi(y \cup \xi) \exp\{\theta \cdot \psi(y \cup \xi)\}\, d\xi.$$

Then using the φ-part of (3.10) and (3.11), we find the bound

$$\varphi_1 \exp(\theta \cdot \psi_1) \int_A \left(\|\psi_1\| - c_2 \sum_{\varnothing \neq y \subseteq x_{\bar{A}}} \psi(y \cup \xi) \cdot \theta \right)^2 \exp\left\{ \sum_{\varnothing \neq y \subseteq x_{\bar{A}}} \psi(y \cup \xi) \cdot \theta \right\} d\xi$$
$$\leq 4c_2^2 \exp\{\theta \cdot \psi_1 + \|\psi_1\|/c_2 - 2\}\delta \leq c_4$$

for $\|\theta - \theta_0\| \leq c_1$ from (3.11).

(iv) We now again consider the second derivative of (3.16) and integrate with respect to the conditional density (3.9). Replacing $Z(\theta_0, x_{\partial A})$ by the upper bound in (3.12), to be denoted by k, and changing the integral over $\Omega_A \times A$ to an integral over $\Omega_{\mathring{A}} \times \mathring{A}$, we get, using (3.2),

$$\begin{aligned} &- \alpha^* \left\{ \frac{\partial^2}{\partial\theta\, \partial\theta^*} E\left[\mathrm{pl}_A(\theta, x_{\bar{A}}) | x_{\partial_A}\right] \right\} \\ &\quad \geq \frac{1}{k} \psi_1 \exp(\theta \cdot \psi_1) \int_{\Omega_{\mathring{A}}} \int_{\mathring{A}} \{\alpha \cdot v(x, \xi)\}^2 \prod_{\varnothing \neq y \subseteq x} \varphi(y)\varphi(y \cup \xi) \\ &\quad \times \exp\{\theta_0 \cdot \psi(y) + \theta \cdot \psi(y \cup \xi)\}\, d\xi\, \mu_A(dx). \end{aligned} \tag{3.17}$$

Now choose $\zeta \in \mathbb{R}^d$ such that $B + \zeta \subseteq \mathring{A}$. From the translation invariance we can change the integral in (3.17) over $\Omega_{\mathring{A}} \times \mathring{A}$ to an integral over $\Omega_{\mathring{A} - \zeta} \times (\mathring{A} - \zeta)$, and then to an integral over $\Omega_B \times B$ giving a lower bound. Using also

$\mu_{A-\zeta}(dx) = \mu_{A-\zeta}(\Omega_B)\mu_B(dx) \geq e^{-\delta}\mu_B(dx)$ for $x \in \Omega_B$, we get instead of (3.17)

$$\frac{e^{-\delta}}{k}\varphi_1 \exp(\theta \cdot \psi_1)\int_{\Omega_B}\int_B \{\alpha \cdot v(x,\xi)\}^2 \prod_{\varnothing \neq y \subseteq x} \varphi(y)\varphi(y \cup \xi) \times \exp\{\theta_0 \cdot \psi(y) + \theta \cdot \psi(y \cup \xi)\}\, d\xi\, \mu_B(dx). \tag{3.18}$$

This is a continuous function of θ and α, and as in the proof of Proposition 2.2 it is seen that (3.18) is strictly positive under condition (3.7). Thus for any c_5 there exists $c_6 > 0$, which is a lower bound to (3.18) for all $\|\theta - \theta_0\| \leq c_5$ and $\|\alpha\| = 1$. The bound c_6 is independent of $x_{\partial A}$ except for the fact that we have used the conditional density (3.9) corresponding to $h(x_{\partial A}) = 0$. However, $h(x_{\partial A}) = 0$ implies that $h(x) = 0$ and so this event has probability 0.

(v) From (3.10) and (3.12) we have the bound

$$f_{\theta_0}(x_A | x_{\partial A}) \leq e^{\delta}\varphi_1^{n(x_A)} \exp\left\{ \sum_{\varnothing \neq y \subseteq x_A} \sum_{z \subseteq x_{\partial A}} \psi(y \cup z) \cdot \theta_0 \right\}.$$

Using this and (3.10) again, it is easy to establish an upper bound to the conditional kth-order moment of the term with $n(x_A)$ and the integral term in (3.16) giving rise to the $m_k(c_{10})$ term in (3.15). For the remaining term in (3.16) we find, using (3.11), that for $\|\theta - \theta_0\| \leq c_1$,

$$E\left\{ \left| \sum_{\xi \in X_A} \sum_{\varnothing \neq y \subseteq X_{\bar{A}} \setminus \xi} \psi(y \cup \xi) \cdot \theta \right|^k \Big| x_{\partial A} \right\}$$
$$\leq \left(\|\theta_0\| + c_1\right)^k c_2^k e^{\delta} \int \left(\varphi_1 e^{\theta_0 \cdot \psi_1}\right)^{n(x)} \{-\eta(x, x_{\partial A})\} \exp\{\eta(x, x_{\partial A})\} \mu_A(dx),$$

with

$$\eta(x, x_{\partial A}) = \sum_{\varnothing \neq y \subseteq x} \sum_{z \subseteq x_{\partial A}:\, n(y \cup z) \geq 2} \psi(y \cup z) \cdot \theta_0.$$

Using (3.10), we obtain the bound

$$\left(\|\theta_0\| + c_1\right)^k c_2^k e^{\delta} \int \left(\varphi_1 e^{\theta_0 \cdot \psi_1}\right)^{n(x)} k^k e^{-k} \mu_A(dx),$$

which is finite and independent of θ and $x_{\partial A}$. Also, since $\lambda(A) \leq \delta$ we get an upper bound independent of A and of the form given in (3.15). □

THEOREM 3.2. *Assume the setup for the ordinary point process* (M1) *to be as in* (3.1), (3.2) *and* (3.5)–(3.8). *Assume either the boundedness condition* (3.10) *or that there exist constants $N, K < \infty$ such that when $n(x) \geq 2$,*

$$\varphi(x) \leq K, \|\psi(x)\| \leq K \text{ and } n\left(X_{\bar{A}_{nij}}\right) \leq N, \qquad \forall\, n, i, j, \tag{3.19}$$

the latter holding almost surely under μ_S. Then the consistency result (A.10) *holds for the maximum pseudolikelihood estimate obtained with $Y_n(\theta) = \mathrm{pl}_{\mathring{S}(n)}(\theta, X)$ and the estimate is unique with a probability tending to* 1.

PROOF. With the notation from Theorem A.2, we let $Y_{nij}(\theta) = \mathrm{pl}_{A_{nij}}(\theta, X)$ and then $Y_n(\theta) = \mathrm{pl}_{\mathring{S}(n)}(\theta, X)$. With $\mathscr{F}_{ni}$ generated by $X_{S(n) \setminus A_{ni}}$, where $A_{ni} = \bigcup_{j=1}^{m_{ni}} A_{nij}$, we get (A.1) from the Markov property, the finite-range assumption and (3.6). We thus have to prove (A.12) and (A.9).

If (3.10) holds the upper bound in (A.12) follows from (3.13), the lower bound follows from (3.14) and condition (A.9) follows from (3.15).

The uniqueness can be proved in the following way. From (2.14) it appears that $Y_n(\theta)$ is either strictly concave at the point θ for all θ-values or for no θ-values. Therefore if $Y_n(\theta)$ is not strictly concave in the direction α at the maximum point $\hat{\theta}_n$, the function $Y_n(\hat{\theta}_n + t\alpha)$ will be linear in t, which contradicts statement (A.8). Since (A.8) was proved to hold with a probability tending to 1 in Theorem A.2, we get the uniqueness.

We then have to consider the case where (3.19) holds instead of (3.10). From (3.16) we find, using (3.6),

$$|Y_{nij}(\theta)| \le N\theta \cdot \psi_1 + N2^{N-1}K\|\theta\| + \varphi_1 e^{\theta \cdot \psi_1} K^{2^N} e^{\|\theta\| K 2^N} \delta, \tag{3.20}$$

with K in (3.19) chosen greater than 1. This immediately shows that (A.9) holds. Similarly, we find

$$-\alpha \left\{ \frac{\partial^2}{\partial\theta\,\partial\theta^*} Y_{nij}(\theta) \right\} \alpha^* \le \varphi_1 e^{\theta \cdot \psi_1} \left(\|\psi_1\| + 2^N K \right)^2 K^{2^N} e^{\|\theta\| K 2^N} \delta$$

for $\|\alpha\| = 1$, whereby the upper bound in (A.12) follows. Finally, we note that for the conditional density (3.9), we have

$$Z(\theta_0, x_{\partial A}) \le \varphi_1^N K^{2^N} \exp\{ N\psi_1 \cdot \theta_0 + K^{2^N} \|\theta_0\| \} = \tilde{k},$$

and the proof for the lower bound in (A.12) parallels that of (3.14) with k replaced by $\tilde{k}$. The uniqueness is proved as above. □

REMARK. Condition (3.19) is of relevance in connection with hardcore models. The proof above actually shows that we can replace (3.19) by the weaker condition that in an open neighborhood of θ_0 we have for $n(x) \ge 2$,

$$\|\psi(x)\| \le K \max\{1, -\theta \cdot \psi(x)\}. \tag{3.21}$$

An example where (3.20) holds but not (3.19) is a one-dimensional model with $\theta_0 > 0$, $\varphi(\{\xi, \zeta\}) = 1(\|\xi - \zeta\| > \delta)$ for some $\delta > 0$, $\psi(\{\xi, \zeta\}) = g(\|\xi - \zeta\|)$ with $g(\cdot) \le K$ and $g(u) \downarrow -\infty$ for $u \downarrow \delta$, and finally $\varphi(x) = 1$ and $\psi(x) = 0$ if $n(x) > 2$.

We now turn to the marked point process (M2) with the points in $S_p(n)$, a bounded Borel set increasing with n, and the marks in the space S_m, which does not depend on n. The relation $\sim$ is now a symmetric relation on $\mathbb{R}^d \times S_m$, the interaction functions φ and ψ satisfy (3.2), and $\sim$, φ and ψ are invariant under translations in $\mathbb{R}^d$. The relation $\sim$ is of finite range in the sense $D = \sup\{\|\xi\| \mid \exists\ r, s \in S_m\colon (\xi, r) \sim (0, s)\} < \infty$. For $A \subseteq \mathbb{R}^d$ we define

$\partial A = \{\xi \in \mathbb{R}^d \setminus A | \exists\ \zeta \in A,\ r, s \in S_m\colon (\xi, r) \sim (\zeta, s)\}$, and $\bar{A}$ and $\mathring{A}$ is then defined as before. We let $\mathring{S}(n) = \mathring{S}_p(n) \times S_m$ and assume

$$\mathring{S}_p(n) = \bigcup_{i=1}^{\nu} \bigcup_{j=1}^{m_{ni}} A_{nij}, \tag{3.22}$$

where the A_{nij}'s are disjoint sets satisfying (3.5) and (3.6). The identifiability condition (3.7) becomes

$$\int \int_B \left\{ \int_{S_m} 1\left[b_\theta(x, (\xi, r)) \neq b_{\theta_0}(x, (\xi, r)) \right] Q_m(dr) \right\} d\xi\, \mu_{B \times S_m}(dx) > 0 \tag{3.23}$$

for all $\theta \in \Theta \setminus \{\theta_0\}$.

THEOREM 3.3. *Let the marked point process satisfy* (3.2) *and assume that* (3.22) *together with* (3.5) *and* (3.6) *and* (3.23) *hold. Assume that either the boundedness condition* (3.10) *or* (3.19) *holds, then the consistency result* (*A*.10) *holds for the maximum pseudolikelihood estimate obtained with* $Y_n(\theta) = \mathrm{pl}_{\mathring{S}(n)}(\theta, X)$.

PROOF. The proof is analogous to the proof of Theorem 3.2. □

REMARK (Strong consistency). Strong consistency can be proved by using Theorem A.3 instead of Theorem A.2, in the case where the point process considered is the restriction to $S(n)$ of an infinite-volume point process. Condition (A.11) can be established from (3.15) under condition (3.10) and from (3.20) under condition (3.19).

4. Examples.

EXAMPLE 1 (Approximation by lattice processes). In applications it may be useful to approximate the pseudolikelihood as follows. Consider the model (M1) and suppose $f_\theta(\cdot)$ is μ_S-a.s. continuous in the sense that $f_\theta(\{x_1, \dots, x_n\})$ is continuous at $(x_1, \dots, x_n) \in \mathbb{R}^{dn}$ for μ_S-a.a. $\{x_1, \dots, x_n\} \in \Omega_S$ with $n > 0$. Let $\xi_{ij} \in A_{ij}$ be a fixed point where A_{ij}, $j = 1, \dots, m_i$, $i = 1, 2, \dots,$ are sets as in Definition 2.1 such that $\max_j \operatorname{diam}(A_{ij}) \to 0$ as $i \to \infty$, where "diam" denotes the diameter of a set. Define

$$\mathrm{PL}_T^{(i)}(\theta, x) = \prod_{j\colon \xi_{ij} \in T} \frac{b_\theta\left(\{\xi_{ij}\colon k \neq j, n_{ik} = 1\}, \xi_{ij}\right)}{1 + b_\theta\left(\{\xi_{ik}\colon k \neq j, n_{ik} = 1\}, \xi_{ij}\right)\lambda(A_{ij})} \tag{4.1}$$

for $i = 1, 2, \dots,$ where $n_{ij} = n(x_{A_{ij}})$. Taylor-expanding the logarithm of the denominator in (4.2) and using the continuity of $f_\theta(\cdot)$, we get

$$\mathrm{PL}_T(\theta, x) = \lim_{i \to \infty} \mathrm{PL}_T^{(i)}(\theta, x). \tag{4.2}$$

This can be interpreted as a limit of pseudolikelihoods for processes $X^{(i)}$, $i = 1, 2, \dots,$ which essentially are lattice processes that converge weakly to X [cf., e.g., Ripley (1988) and Särkka (1989)]. The process $X^{(i)}$ is constructed such that $n(X^{(i)}_{A_{ij}}) \le 1$ for all j and the density $f^{(i)}_\theta$ with respect to μ_S is given by

$$(4.3) \qquad f^{(i)}_\theta(x)/f^{(i)}_\theta(\varnothing) = f_\theta(\{\xi_{ij}: n_{ij} = 1\})/f_\theta(\varnothing).$$

Hence

$$\mathrm{PL}^{(i)}_T(\theta, x) = \prod_{j:\,\xi_{ij}\in T} P_\theta\Big(n\big(X^{(i)}_{A_{ij}}\big) = n_{ij} | n\big(X^{(i)}_{A_{ik}}\big) = n_{ik},\, k \neq j\Big)/\lambda(A_{ij})^{n_{ij}},$$

where the product of the numerator corresponds to Besag's pseudolikelihood for the binary lattice process $\{n(X^{(i)}_{A_{ij}}): j = 1, \dots, m_i\}$. It follows immediately from (4.3) that

$$(4.4) \qquad f_\theta(x) = \lim_{i\to\infty} f^{(i)}_\theta(x) \quad \text{for } \mu_S\text{-a.a. } x \in \Omega_S.$$

Finally, let us briefly consider the model (M2) under the following conditions. Suppose $S = S_p \times S_m$ is a metric space such that $\max_j \operatorname{diam}(A_{ij}) \to 0$ as $i \to \infty$, where $A_{ij} = B_{ij} \times C_{ij}$ are sets as in Definition 2.1 with $B_{ij} \in \mathscr{B}_p$ and $C_{ij} \in \mathscr{B}_m$. Suppose also that Q_m is absolutely continuous with respect to some finite measure λ_m on $\mathscr{B}_m$ and let q_m denote the density. Now, define

$$\mathrm{PL}^{(i)}_\theta(\theta, x) = \prod_{j:\,\xi_{ij}\in T} \frac{\big[b_\theta(\{\xi_{ik}: k \neq j,\, n_{ik} = 1\}, \xi_{ij}) q_m(r_{ij})\big]^{n_{ij}}}{1 + b_\theta(\{\xi_{ik}: k \neq j,\, n_{ik} = 1\}, \xi_{ij})\lambda(A_{ij})},$$

where r_{ij} is the mark of $x_{A_{ij}}$. Then, if the process $X^{(i)}$ is constructed as before and $f_\theta(\cdot)$ and $q_m(\cdot)$ are μ_S-a.s. continuous, (4.2) and (4.4) remain true.

EXAMPLE 2 (Pairwise interaction model). The special case of a Markov model (3.1) with $\varphi(x) = 1$ and $\psi(x) = 0$ for $n(x) \ge 3$ is called a pairwise interaction process. If we let $\varphi_1 = 1$, $\psi_1 = (1, 0)$ and $\psi(\{\xi, \eta\}) = (0, \tilde{\psi}(\{\xi, \eta\}))$, we get the exponential family of order 2

$$(4.5) \qquad \begin{aligned} f_\theta(x) = {} & \frac{1}{Z(\theta)}\Big\{\prod_{\{x_i, x_j\}\subset x} \varphi(\{x_i, x_j\})\Big\} \\ & \times \exp\Big\{\theta_1 n(x) + \theta_2 \sum_{\{x_i, x_j\}\subset x} \tilde{\psi}(\{x_i, x_j\})\Big\}. \end{aligned}$$

The log pseudolikelihood from (3.4) and (2.12) becomes

$$\begin{aligned} \mathrm{pl}_T(\theta, x) = {} & \theta_1 n(x_T) + \sum_{\xi\in x_T} \{\ln t(x_T \setminus \xi, \xi) + \theta_2 s(x_T \setminus \xi, \xi)\} \\ & - e^{\theta_1}\int_T t(x_T, \xi) \exp\{\theta_2 s(x_T, \xi)\}\, d\xi, \end{aligned}$$

where $t(x, \xi) = \prod_{\eta \in x} \varphi(\{\eta, \xi\})$ and $s(x, \xi) = \sum_{\eta \in x} \psi(\{\eta, \xi\})$. Setting the derivative of pl_T equal to 0, the two equations for determining $\hat{\theta}_1$ and $\hat{\theta}_2$ can be separated as

$$(4.6)\quad \frac{\int_T s(x_T, \xi) t(x_T, \xi) \exp\{\theta_2 s(x_T, \xi)\}\, d\xi}{\int_T t(x_T, \xi) \exp\{\theta_2 s(x_T, \xi)\}\, d\xi} = \frac{1}{n(x_T)} \sum_{\xi \in x_T} s(x_T \setminus \xi, \xi)$$

and

$$\theta_1 = \ln\left\{ n(x_T) \Big/ \int_T t(x_T, \xi) \exp\{\theta_2 s(x_T, \xi)\}\, d\xi \right\}.$$

We note that (4.6) can be interpreted as the likelihood equation for an exponential family on T with canonical statistic $s(x_T, \cdot\,)$.

Alternatively, under (M1) the pseudolikelihood can be approximated by (4.2) which yields

$$\mathrm{PL}_T(\theta, x) \cong \prod_{j:\, \xi_{ij} \in T} \frac{\exp\left[n_{ij}\left(\theta_1 + \theta_2 s\left(\{\xi_{ik}\colon k \neq j, n_{ik} = 1\}, \xi_{ij}\right)\right)\right]}{1 + \lambda(A_{ij}) \exp\left[\theta_1 + \theta_2 s\left(\{\xi_{ik}\colon k \neq j, n_{ik} = 1\}, \xi_{ij}\right)\right]}.$$

This can be interpreted as the likelihood function for a logistic dose response model if the A_{ij}'s have the same size $\lambda(A_{ij}) = a_i$ for all j.

The Strauss process [Strauss (1975) and Kelly and Ripley (1976)] is obtained when $\varphi(\{\xi, \zeta\}) = 1$ and $\psi(\{\xi, \zeta\}) = 1(\|\xi - \zeta\| \le D)$ for a fixed number D. This process is defined for $\theta_1 \in \mathbb{R}$ and $\theta_2 \le 0$, where $\theta_2 = 0$ corresponds to the Poisson process, and is a Markov model of finite range D. If we take B to be a ball of radius greater than 0, it is trivial to see that the identifiability condition (3.7) is satisfied by looking at those x with $n(x) = 1$. Since $\theta_2 \le 0$ we have that (3.10) holds and we obtain the consistency from Theorem 3.2.

For the modified Strauss process with hardcore $0 \le \delta < D$, we have $\psi(\cdot)$ as above and furthermore $\varphi(\{\xi, \zeta\}) = 1(\|\xi - \zeta\| \ge \delta)$. This is again a Markov model of finite range D, and the model is defined for $\theta_1 \in \mathbb{R}$ and $\theta_2 \in \mathbb{R}$. The identifiability condition holds as above when B is a ball of radius greater than δ, and the consistency is obtained from Theorem 3.2 since condition (3.19) holds.

EXAMPLE 3 (Strauss-like marked point process). We consider a marked point process where the mark space is $S_m = [0, L]$ for some fixed number L. We think of $(\xi, r) \in S_p \times S_m$ as a ball with center at ξ and radius r. Restricting ourselves to pairwise interacting models (4.1), we take $\varphi(\{(\xi, r), (\zeta, s)\}) = 1$ and $\psi(\{(\xi, r), (\zeta, s)\}) = 1(\|\xi - \zeta\| \le r + s)$. This is a Markov model with respect to the relation $(\xi, r) \sim (\zeta, s) \Leftrightarrow \|\xi - \zeta\| \le r + s$, which is of finite range $D = 2L$. Note that the canonical statistic that goes with θ_2 in the density (4.5) is the number of pairs of balls that intersect one another.

The density is defined for $\theta_1 \in \mathbb{R}$ and $\theta_2 \le 0$, and so (3.10) holds trivially. The identifiability (3.23) holds as in Example 1 when B is a ball of radius greater than 0. The consistency is then established from Theorem 3.3.

APPENDIX

When proving consistency a standard method is to impose conditions that ensure uniform convergence of the second derivative of the likelihood function. The idea of the proof here is to use concavity to obtain uniform convergence of the pseudolikelihood function itself. Also the coding idea is used; that is, we impose a certain conditional independence structure.

Let $(E_n, \mathscr{E}_n, P_n)$ be a measure space and let $Y_{nij}(\theta) = Y_{nij}(\theta, e)$ be random functions concave in $\theta \in \Theta$, where $\Theta \subseteq \mathbb{R}^k$ is open and $i = 1, \dots, \nu$, $j = 1, \dots, m_{ni}$, with $m_{ni} \to \infty$ as $n \to \infty$. Let $\mathscr{F}_{ni}$ be a sub-σ-field of $\mathscr{E}_n$ and define $Z_{nij}(\theta) = Z_{nij}(\theta, e) = E(Y_{nij}(\theta) | \mathscr{F}_{ni})$. We will assume that

(A.1) $\quad Y_{ni1}(\theta), \dots, Y_{nim_{ni}}(\theta)$ are conditionally independent given $\mathscr{F}_{ni}$,

and for some fixed $\theta_0 \in \Theta$ there exist constants c_1, c_2 and c_3 such that almost surely

$$-c_1\|\theta - \theta_0\|^2 \le Z_{nij}(\theta) - Z_{nij}(\theta_0) \le -c_2\|\theta - \theta_0\|^2 \quad \text{for } \|\theta - \theta_0\| < c_3. \tag{A.2}$$

LEMMA A.1. *Assume that* (A.2) *holds and define*

$$G_{ni}(\theta) = \frac{1}{m_{ni}} \sum_{j=1}^{m_{ni}} \big(Y_{nij}(\theta) - Z_{nij}(\theta_0)\big)$$

and $g_{ni}(\theta)$ similarly with $Y_{nij}(\theta)$ replaced by $Z_{nij}(\theta)$. Then for any $\varepsilon > 0$ and $0 < \delta < c_3$, there exist $s < \infty$ and $\theta_1, \dots, \theta_s$ with $\|\theta_i - \theta_0\| < c_3$ such that

$$\sup_{\|\theta - \theta_0\| \le \delta} |G_{ni}(\theta) - g_{ni}(\theta)| \le \varepsilon \tag{A.3}$$

holds on $\bigcap_{r=0}^{s} A_{ni}(\theta_r)$, where

$$A_{ni}(\theta) = \{e \in E_n | |G_{ni}(\theta) - g_{ni}(\theta)| \le \varepsilon/2\}. \tag{A.4}$$

PROOF. First we take $\theta_1, \dots, \theta_{s_1}$ with $\delta < \|\theta_i - \theta_0| < c_3$ such that $\{\theta | \|\theta - \theta_0\| \le \delta\}$ is contained in the convex hull of $\{\theta_1, \dots, \theta_{s_1}\}$. Since (A.2) implies

$$-c_1\|\theta - \theta_0\|^2 \le g_{ni}(\theta) \le -c_2\|\theta - \theta_0\|^2 \quad \text{for } \|\theta - \theta_0\| < c_3, \tag{A.5}$$

it is possible, from the concavity, to obtain that

$$-c_1 c_3^2 - \frac{\varepsilon}{2} \le G_{ni}(\theta) \le c_1 c_3^2 + \varepsilon \quad \text{for } \|\theta - \theta_0\| \le \delta \tag{A.6}$$

on the set $\bigcap_{r=0}^{s_1} A_{ni}(\theta_r)$. The upper bound is established with the help of the lower bound and using that $e \in A_{ni}(\theta_0)$. According to the proof of Theorem 10.4 in Rockafeller (1970), there exists α, which depends on the bounds in

(A.6) only, such that

$$|G_{ni}(\tilde{\theta}_1) - G_{ni}(\tilde{\theta}_2)| \le \alpha\|\tilde{\theta}_1 - \tilde{\theta}_2\| \tag{A.7}$$

for all $\tilde{\theta}_1, \tilde{\theta}_2$ with $\|\tilde{\theta}_i - \theta_0\| \le \delta$. Using the same argument, this inequality also holds with $G_{ni}(\cdot)$ replaced by $g_{ni}(\cdot)$.

Now take $\theta_{s_1+1}, \dots, \theta_s$ with $\|\theta_i - \theta_0\| \le \delta$ such that for all $\|\theta - \theta_0\| \le \delta$ there exists a θ_{s_1+r} with $\|\theta - \theta_{s_1+r}\| \le \varepsilon/(4\alpha)$. Then on $\bigcap_{r=0}^{s} A_{ni}(\theta_r)$ we have, according to (A.4) and (A.7), that

$$\begin{aligned}|G_{ni}(\theta) - g_{ni}(\theta)| &\le |G_{ni}(\theta) - G_{ni}(\theta_{s_1+r})| + |G_{ni}(\theta_{s_1+r}) - g_{ni}(\theta_{s_1+r})| \\ &\quad + |g_{ni}(\theta_{s_1+r}) - g_{ni}(\theta)| \\ &\le 2\alpha\frac{\varepsilon}{4\alpha} + \frac{\varepsilon}{2} = \varepsilon\end{aligned}$$

for all $\|\theta - \theta_0\| \le \delta$. □

If (A.3) holds with $\varepsilon < \frac{1}{2}c_2\delta^2$ and $\delta < c_3$, we get from (A.5) and the concavity that

$$m_{ni}G_{ni}(\theta) < m_{ni}G_{ni}(\theta_0) \quad \text{for } \|\theta - \theta_0\| > \delta.$$

If this holds for $i = 1, \dots, \nu$, we get with $Y_n(\theta) = \sum_{i,j} Y_{nij}(\theta)$,

$$Y_n(\theta) < Y_n(\theta_0) \quad \text{for } \|\theta - \theta_0\| > \delta \tag{A.8}$$

and therefore $M_n \subseteq \{\theta \mid \|\theta - \theta_0\| \le \delta\}$, where

$$M_n = \left\{\theta \in \Theta \mid Y_n(\theta) = \sup_{\tilde{\theta} \in \Theta} Y_n(\tilde{\theta})\right\}.$$

It is now easy to establish the following two theorems.

THEOREM A.2. *Assume* (A.1) *and* (A.2) *and that there exists a constant* c_4 *such that for all* $n = i = 1, \dots, \nu$,

$$\frac{1}{m_{ni}}\sum_{j=1}^{m_{ni}} \operatorname{Var}(Y_{nij}(\theta) \mid \mathscr{F}_{ni}) < c_4 \quad \textit{for } \|\theta - \theta_0\| < c_3. \tag{A.9}$$

Then for any $\varepsilon > 0$,

$$P_n\Big(\sup_{\theta \in M_n} \|\theta - \theta_0\| > \varepsilon\Big) \to 0 \quad \textit{as } n \to \infty. \tag{A.10}$$

PROOF. We must prove that the probability that (A.3) holds tends to 1. From (A.1), (A.9) and Markov's inequality, we get

$$P_n(|G_{ni}(\theta) - g_{ni}(\theta)| > \varepsilon \mid \mathscr{F}_{ni}) \le \frac{1}{m_{ni}}c_4/\varepsilon^2 \quad \text{for } \|\theta - \theta_0\| < c_3.$$

Since this tends to 0 we have $P_n(A_{ni}(\theta_r)) \to 1$ and thus $P(\bigcap_{r=0}^{s} A_{ni}(\theta_r)) \to 1$, and the result follows from Lemma A.1. □

We now turn to almost sure convergence and assume $E_n \equiv E$ and P_n is the restriction of a fixed probability measure P to $\mathscr{E}_n$.

THEOREM A.3. *Assume* (A.1) *and* (A.2) *and that there exist positive constants* c_4 *and* c_5 *such that for all* n *and* $i = 1, \dots, \nu$,

$$\text{(A.11)} \qquad \prod_{j=1}^{m_{ni}} E\Big\{\exp\big[t\big(Y_{nij}(\theta) - Z_{nij}(\theta)\big)\big] | \mathscr{F}_{ni}\Big\} \le \exp\big(c_4 m_{ni} t^2\big)$$

for $|t| < c_5$ *and* $\|\theta - \theta_0\| < c_3$. *Let* $\hat{\theta}_n \in M_n$. *Then*

$$\hat{\theta}_n \to \theta_0 \quad \textit{almost surely}.$$

PROOF. We must now prove that (A.3) holds as $n \to \infty$ almost surely. From Lemma A.1 we only have to show that $G_{ni}(\theta) - g_{ni}(\theta) \to 0$ almost surely for $i = 1, \dots, \nu$ and a fixed number of θ-values. From (A.11) we get

$$E \exp\big[t m_{ni}\big(G_{ni}(\theta) - g_{ni}(\theta)\big)\big] \le \exp\big(c_4 m_{ni} t^2\big)$$

and the usual exponential estimate [Révész (1967)] then gives that for any $\varepsilon > 0$ there exist $c_3 > 0$ and $\rho_\varepsilon < 1$ such that

$$P\big(|G_{ni}(\theta) - g_{ni}(\theta)| > \varepsilon\big) \le c_\varepsilon \rho_\varepsilon^{m_{ni}}.$$

Finally, we conclude from the Borel–Cantelli lemma that

$$G_{ni}(\theta) - g_{ni}(\theta) \to 0 \quad \text{almost surely}. \qquad \square$$

If $Z_{nij}(\theta)$ is twice differentiable in θ and $(\partial/\partial\theta) Z_{nij}(\theta_0) = 0$, condition (A.2) will of course be satisfied if

$$\text{(A.12)} \quad \sup -\alpha^* \frac{\partial^2}{\partial\theta\,\partial\theta^*} Z_{nij}(\theta)\alpha < \infty \quad \text{and} \quad \inf -\alpha^* \frac{\partial^2}{\partial\theta\,\partial\theta^*} Z_{nij}(\theta)\alpha < 0,$$

where the supremum and the infimum are over all (n, i, j), α and θ with $\|\alpha\| = 1$ and $\|\theta - \theta_0\| < c_3$.

REFERENCES

BADDELEY, A. and MØLLER, J. (1989). Nearest-neighbour Markov point processes and random sets. *Internat. Statist. Rev.* **57** 89–121.

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

BESAG, J. (1977). Some methods of statistical analysis for spatial data. *Bull. Inst. Internat. Statist.* **47** 77–92.

BESAG, J., MILNE, R. and ZACHARY, S. (1982). Point process limits of lattice processes. *J. Appl. Probab.* **19** 210–216.

CARTER, D. S. and PRENTER, P. M. (1972). Exponential spaces and counting processes. *Z. Wahrsch. Verw. Gebiete* **21** 1–19.

CLYDE, M. and STRAUSS, D. (1988). Logistic regression for spatial pair-potential model. Preprint, Dept. Statistics, Univ. California.

COMETS, F. (1989). On consistency of a class of estimators for exponential families of Markov random fields on the lattice. Preprint, Université Paris-X et Laboratorie de Statistique Appliquée, Univ. Paris-Sud.

GEMAN, S. and GRAFFIGNE, C. (1986). Markov random fields image models and their applications to computer vision. In *Proc. Internat. Congress of Mathematics 1986* (M. Gleason, ed.). Amer. Math. Soc., Providence, R.I.

GIDAS, B. (1988). Consistency of maximum likelihood and pseudolikelihood estimators for Gibbs distributions. In *Stochastic Differential Systems, Stochastic Control Theory and Applications*. IMA **10**. Springer, New York.

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

KELLY, F. P. and RIPLEY, B. D. (1976). A note on Strauss' model for clustering. *Biometrika* **63** 357–360.

MOYEED, R. A. and BADDELEY, A. (1989). Stochastic approximation of the MLE for a spatial point pattern. Technical Report BS-R8926, Math. Centrum, Amsterdam.

OGATA, Y. and TANEMURA, M. (1984). Likelihood analysis of spatial point patterns. *J. Roy. Statistics Soc. Ser. B* **46** 496–518.

RÉVÉSZ, P. (1967). *The Laws of Large Numbers*. Akadémiai Kiadó, Budapest.

RIPLEY, B. D. (1988). *Statistical Inference for Spatial Processes*. Cambridge Univ. Press.

RIPLEY, B. D. and KELLY, F. P. (1977). Markov point processes. *J. London Math. Soc.* **15** 188–192.

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

SÄRKKA, A. (1989). On parameter estimation of Gibbs point processes through the pseudo-likelihood method. Technical Report, Dept. Statistics, Univ. Jyväskylä.

STRAUSS, D. J. (1975). A model for clustering. *Biometrika* **62** 467–475.

DEPARTMENT OF THEORETICAL STATISTICS
INSTITUTE OF MATHEMATICS
UNIVERSITY OF AARHUS
DK-8000 AARHUS
DENMARK