Ann. Inst. Statist. Math.
Vol. 46, No. 3, 475–486 (1994)

ON ASYMPTOTIC NORMALITY OF PSEUDO LIKELIHOOD ESTIMATES FOR PAIRWISE INTERACTION PROCESSES

JENS LEDET JENSEN[1] AND HANS R. KÜNSCH[2]

[1] *Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus, DK-8000 Aarhus C, Denmark*
[2] *Seminar für Statistik, ETH, CH-8092 Zürich, Switzerland*

(Received February 22, 1993; revised August 26, 1993)

Abstract. We consider point processes defined through a pairwise interaction potential and admitting a two-dimensional sufficient statistic. It is shown that the pseudo maximum likelihood estimate can be stochastically normed so that the limiting distribution is a standard normal distribution. This result is true irrespectively of the possible existence of phase transitions. The work here is an extension of the work Guyon and Künsch (1992, *Lecture Notes in Statist.*, **74**, Springer, New York) and is based on viewing a point process interchangeably as a lattice field.

Key words and phrases: Asymptotic normality, Gibbs processes, lattice field, phase transitions, pseudo likelihood, stochastically normed.

1. Introduction

In recent years there has been considerable progress in fitting Gibbssian models to spatial point patterns. Both approximations to the maximum likelihood estimator and alternatives like the pseudo maximum likelihood estimator which are easier to compute have been developed, see e.g. Diggle *et al.* (1994) for a review and a comparison of the different estimators. For another discussion of the value of pseudo likelihood estimation we refer to the rejoinder to the discussion of Besag *et al.* (1991), from where we also have the following quote: "there are still many open problems concerning the efficiency and asymptotic normality of maximum pseudo likelihood estimators". Actually, for all the estimators there is only little known about their distribution—even asymptotically as the observation window increases. Correspondingly, inference about the unknown parameters is difficult. Asymptotic normality has been proved for the subset of parameter values satisfying the so-called Dobrushin conditions, see e.g. Guyon (1987), Jensen (1993) or Heinrich (1992). These conditions are sufficient, but not necessary for the uniqueness of the infinite volume Gibbs measure. The possible non-uniqueness of this infinite volume Gibbs measure (phase transition) lies at the heart of all difficulties. The example of the Ising model for lattice systems shows that at the transition from uniqueness to non-uniqueness long range dependence may occur. For this reason standard central limit theorems cannot be used. Recently however Guyon and Künsch (1992) have shown for lattice processes that asymptotic normality can be obtained for all easily computable estimators without mixing conditions or other detailed properties of the Gibbs measure. Instead one exploits a property that resembles that of a martingale difference sequence which holds for all values of the parameters. In this paper we extend the results of Guyon and Künsch (1992) to the pseudo maximum likelihood estimator for a large class of Gibbs point processes.

The point processes X that we will consider are given by their conditional density in a region $\Lambda \subset \mathbb{R}^d$ with respect to Poisson measure ν_Λ having unit intensity. Let x_Λ be a point configuration in Λ and y^Λ a point configuration in $\mathbb{R}^d \backslash \Lambda$. Then the measure μ is defined by

$$\frac{d\mu_\Lambda(\cdot \mid y^\Lambda)}{d\nu_\Lambda}(x_\Lambda) = Z(\zeta, \beta; y^\Lambda)\zeta^{|x_\Lambda|} \exp\{-\beta V_\Lambda(x_\Lambda \mid y^\Lambda)\}, \tag{1.1}$$

where $\beta > 0$, $\zeta > 0$, $|x_\Lambda|$ is the number of elements in x_Λ, and

$$V_\Lambda(x_\Lambda \mid y^\Lambda) = \frac{1}{2} \sum_{z_1 \neq z_2, z_i \in x_\Lambda \cup y^\Lambda, \{z_1, z_2\} \cap x_\Lambda \neq \emptyset} \phi(z_1 - z_2).$$

The potential ϕ will always be of finite range κ,

$$\phi(z) = 0 \quad \text{for} \quad |z| > \kappa,$$

and also $\phi(-z) = \phi(z)$. For this process the log pseudo likelihood function, as described in Jensen and Møller (1991), in the region Λ based on the outcome $x = x_\Lambda \cup y^\Lambda$ is

$$pl_\Lambda(\omega, \beta) = \omega|x_\Lambda| - \beta \sum_{z \in x_\Lambda} v(x \backslash z, z) - e^\omega \int_\Lambda \exp\{-\beta v(x, \xi)\} d\xi,$$

where $\omega = \log(\zeta)$ and $v(x, \xi) = \sum_{z \in x} \phi(z - \xi)$. Because of the finite range assumption we can calculate $pl_\Lambda(\omega, \beta)$ on observing $x_{\bar{\Lambda}}$ only, where $\bar{\Lambda} = \{x + y \mid x \in \Lambda, |y| \leq \kappa\}$.

A basic ingredient in the approach in this paper is that a point process can be viewed interchangeably as a lattice field. Let Λ_i, $i = (i_1, \dots, i_d)$, be the cube $\{z \in \mathbb{R}^d \mid \tilde{\kappa}(i_j - \frac{1}{2}) \leq z_j \leq \tilde{\kappa}(i_j + \frac{1}{2}), j = 1, \dots, d\}$ for any chosen $\tilde{\kappa} \geq \kappa$. Setting $X_i = X_{\Lambda_i}$, $i \in \mathbb{Z}^d$, this becomes a Gibbs lattice field. We will consider estimation of (ω, β) from $pl_{\Lambda(n)}$, where $\Lambda(n) = \cup_{i \in I_n} \Lambda_i$ and where the process is observed in $\cup_{i \in \bar{I}_n} \Lambda_i$, with $\bar{I}_n = \{i + j \mid i \in I_n, |j| \leq 1\}$ and the norm is $|j| = \max\{|j_1|, \dots, |j_d|\}$. We will assume throughout that I_n increases to $\mathbb{Z}^d$ and $|\partial I_n|/|I_n| \to 0$, where $\partial I_n = \{i \in I_n \mid \exists j \notin I_n : |j - i| = 1\}$.

One of the main obstacles in the proofs below will be to prove that a certain covariance matrix is positive definite. To do this we will assume one of the following conditions on the potential ϕ:

(C1) $0 \leq \phi(z) < \infty$,

(C2) $\phi(z) = \psi(|z|)$ where $\psi(\cdot) : \mathbb{R}_+ \to \mathbb{R}$ satisfies

(i) $\psi(r) \geq -K$ for some constant K, $\psi(r) = \infty$ for $r < h$ for some constant $h > 0$, $\psi(\cdot)$ is continuously differentiable except at a finite number of points;

(ii) for all $\beta > 0$ the function $\psi'(r)e^{-\beta\psi(r)}$ is bounded;

(iii) either $\psi(r) \to a_0 \neq 0$ for $r \to \kappa$ or $\psi'(r) \to a_1 \neq 0$ for $r \to \kappa$.

(C1) says that the interaction is purely repulsive. (C2) allows an attractive interaction, but then we need a hard core. The other conditions in (C2) are technical.

The result of the paper is as follows.

THEOREM 1.1. *Let μ, defined through* (1.1), *be a stationary Gibbs point process satisfying either* (C1) *or* (C2). *Then*

$$|I_n|^{-1/2}(\hat{\omega}_n - \omega, \hat{\beta}_n - \beta) j_n V_n^{-1/2} \xrightarrow{\mathcal{D}} N_2(0, I), \tag{1.2}$$

where $(\hat{\omega}_n, \hat{\beta}_n)$ is the estimate from $pl_{\Lambda(n)}$ and j_n and V_n are given in Section 3.

The important aspect of (1.2) is that everything can be calculated fairly easily from the data. In the set up here β is one-dimensional. This is only used in an essential way in the proof of Lemma 3.3 below, and it seems likely that generalizations to the multi-dimensional case are possible. However, we do not pursue this here.

In Section 2 we describe in a general form the underlying central limit theorem used in the paper. Section 3 contains the proof of Theorem 1.1, and in Section 4 we illustrate the results by a small simulation study.

2. A central limit theorem for random fields

Let Z_i, $i \in \mathbb{Z}^d$, be a random field and let $I_n \subset \mathbb{Z}^d$. We want to consider central limit theorems for the sum $\sum_{i \in I_n} Z_i$ as I_n increases. This is a fairly recent subject, and the approach has been to impose conditions on the dependency structure through the mixing coefficients. Bolthausen (1982) seems to have obtained the minimal set of conditions needed on the strong mixing coefficient for a stationary field. Takahata (1983) has a similar result where the stationarity assumption has been relaxed. Unfortunately, it is often hard to obtain the necessary bounds on the mixing coefficients. Some results are known for Gibbs fields in the Dobrushin uniqueness region and this is exploited in Jensen (1993). Outside the Dobrushin uniqueness region very little is known. Often there is an ergodic decomposition of the measure under consideration, but there seem to be no general bounds on the mixing coefficients for the ergodic components. The example of the Ising model at the critical point shows that the mixing coefficients are in general not summable.

In the theorem below we replace the bounds on the mixing coefficients with an ergodicity assumption and the assumption that the conditional mean of Z_i, given the σ-field of events outside site i, is zero. The latter condition reminds one of martingale differences, although it is not clear how to bring out clearly such a comparison. The condition fits in naturally with the pseudo likelihood function, which, loosely speaking, is a product of conditional likelihood functions.

THEOREM 2.1. *Let X_i, $i \in \mathbb{Z}^d$, be a stationary and ergodic random field with $X_i \in S$ for some measurable space S, and let $f : S^{I_0} \to \mathbb{R}^k$, where $I_0 = \{j \in \mathbb{Z}^d \mid |j| \leq 1\}$. Define $Z_j = f(X_{i+j}, i \in I_0)$, $j \in \mathbb{Z}^d$, and $\Delta = E(\sum_{|j|\leq 1} Z_0^* Z_j)$, where a * denotes the transposed vector. For a region $I_n \subset \mathbb{Z}^d$ we let $S_n = \sum_{i \in I_n} Z_i$. If*

(i) $E(Z_i \mid X_j, j \neq i) = 0$, $i \in \mathbb{Z}^d$,

(ii) $E|Z_i|^3 < \infty$, *and*

(iii) I_n *increases to* $\mathbb{Z}^d$ *with* $\frac{|\partial I_n|}{|I_n|} \to 0$,

then

$$|I_n|^{-1/2} S_n \simeq N_k(0, \Delta) \qquad \text{as} \quad n \to \infty.$$

PROOF. From (i) we find the variance

$$\begin{aligned} V(|I_n|^{-1/2} S_n) &= |I_n|^{-1} \sum_{i \in I_n} E\left(Z_i^* \sum_{|j-i|\leq 1, j \in I_n} Z_j \right) \\ &= \Delta - |I_n|^{-1} \sum_{i \in \partial I_n} E\left(Z_i^* \sum_{|j-i|\leq 1, j \notin I_n} Z_j \right) \to \Delta, \end{aligned}$$

where the convergence follows from (ii) and (iii). If Δ is not positive definite some direction of $|I_n|^{-1/2} S_n$ will have a variance tending to zero, and the conclusion of the theorem will trivially be correct for those directions. We can therefore assume that Δ is positive definite.

Let $e \in \mathbb{R}^k$ with $|e| = 1$ and define

$$\begin{aligned} &Y_i = Z_i \Delta^{-1/2} e^*, \qquad \bar{S}_n = |I_n|^{-1/2} \sum_{I_n} Y_i \qquad \text{and} \\ &\bar{S}_{i,n} = |I_n|^{-1/2} \sum_{\substack{|j-i|\leq 1 \\ j \in I_n}} Y_j. \end{aligned}$$

According to Lemma 2 in Bolthausen (1982) we must prove that

$$E\{(i\lambda - \bar{S}_n) \exp(i\lambda \bar{S}_n)\} \to 0 \qquad \forall \lambda \in \mathbb{R}. \tag{2.1}$$

Following Bolthausen (1982) we write (2.1) as $E\{A_1 - A_2 - A_3\}$ with

$$\begin{aligned} A_1 &= i\lambda \exp(i\lambda \bar{S}_n) \left[1 - |I_n|^{-1/2} \sum_{I_n} Y_j \bar{S}_{j,n} \right], \\ A_2 &= |I_n|^{-1/2} \exp(i\lambda \bar{S}_n) \sum_{I_n} Y_j [1 - \exp(-i\lambda \bar{S}_{j,n}) - i\lambda \bar{S}_{j,n}] \end{aligned}$$

and

$$A_3 = |I_n|^{-1/2} \sum_{I_n} Y_j \exp[i\lambda(\bar{S}_n - \bar{S}_{j,n})].$$

From (i) we get directly that $EA_3 = 0$. For A_2 we find

$$\begin{aligned} |EA_2| &\le \frac{1}{2}|I_n|^{-3/2} \sum_{I_n} E\left\{ |Y_j| \left(\sum_{|k-j|\le 1, k\in I_n} \lambda Y_k \right)^2 \right\} \\ &\le c\lambda^2 |I_n|^{-1/2} \to 0, \end{aligned}$$

on using the moment assumption (ii).

For the last term A_1 we write, using (ii) again

$$\begin{aligned} |EA_1| &\le |\lambda| E\left| 1 - |I_n|^{-1} \sum_{I_n} Y_j R_j \right| + c|\lambda| \frac{|\partial I_n|}{|I_n|} \\ &= |\lambda| E\left| |I_n|^{-1} \sum_{I_n} (Y_j R_j - E(Y_j R_j)) \right| + c|\lambda| \frac{|\partial I_n|}{|I_n|}, \end{aligned}$$

where $R_j = \sum_{|k-j|\le 1} Y_k$ and $EY_jR_j = 1$. The condition (iii) gives that the last term here goes to zero, and the first term is seen to go to zero from the L^1-version of the ergodic theorem (Georgii (1988), Section 14.A). □

3. Proof of Theorem 1.1

We prove Theorem 1.1 through a series of lemmas. Since $pl_{\Lambda(n)} = \sum_{i\in I_n} pl_{\Lambda_i}$ we find that the two first derivatives of the log pseudo likelihood function can be written as

$$\begin{aligned} U_n(\omega,\beta) &= \frac{\partial pl_{\Lambda(n)}}{\partial(\omega,\beta)} = \sum_{i\in I_n} (g_0 \circ \theta_i, g_1 \circ \theta_i), \\ j_n(\omega,\beta) &= \frac{-\partial^2 pl_{\Lambda(n)}}{\partial(\omega,\beta)^* \partial(\omega,\beta)} = \sum_{i\in I_n} e^{\omega} \begin{pmatrix} f_0 \circ \theta_i & -f_1 \circ \theta_i \\ -f_1 \circ \theta_i & f_2 \circ \theta_i \end{pmatrix}, \end{aligned}$$

where

$$g_0(X) = |X_{\Lambda_0}| - e^{\omega} f_0(X), \qquad g_1(X) = -\sum_{z\in X_{\Lambda_0}} v(X\backslash z, z) + e^{\omega} f_1(X),$$

$$f_k(X) = \int_{\Lambda_0} v(X,\xi)^k \exp\{-\beta v(X,\xi)\} d\xi,$$

and θ_i is the translation operator.

We are going to apply Theorem 2.1 to $Z_i = (g_0 \circ \theta_i, g_1 \circ \theta_i)$. Note that g_k, $k = 0, 1$, depends on X_{Λ_i}, $|i| \le 1$, only. The first lemma will be used to check assumption (ii) in this theorem.

LEMMA 3.1. *Under the assumption* (C1) *or* (C2) *all moments of* $g_0(X)$, $g_1(X)$ *and* $f_k(X)$ *exist.*

PROOF. This is shown in Jensen (1993). □

The next lemma will give (i) of Theorem 2.1.

LEMMA 3.2. *The conditional means of* g_0 *and* g_1 *given* X^{Λ_0} *are zero,*

$$E(g_0(X) \mid X^{\Lambda_0}) = E(g_1(X) \mid X^{\Lambda_0}) = 0.$$

PROOF. This appears from Jensen and Møller (1991). □

We now turn to the variance of $U_n = U_n(\omega, \beta)$. We find as in the proof of Theorem 2.1

$$V(|I_n|^{-1/2}U_n) \to E\left\{ \begin{pmatrix} g_0 \\ g_1 \end{pmatrix} \sum_{|j|\le 1} (g_0 \circ \theta_j, g_1 \circ \theta_j) \right\},$$

where we denote the latter quantity by $\Sigma(U)$ in the following.

LEMMA 3.3. *Under either* (C1) *or* (C2) *the limiting variance* $\Sigma(U)$ *is positive definite.*

PROOF. We must show that for any $(a, b) \neq (0, 0)$ there exists $c > 0$ and n_0 such that for $n \geq n_0$, $V((a, b) \cdot U_n) \geq c|I_n|$. We bound this variance from below by a conditional independence argument. Let $L = 3\mathbb{Z}^d$, then the X_{Λ_l}, $l \in L$, are conditionally independent given X_{Λ_l}, $l \notin L$, and each $g_k \circ \theta_j$ depends on exactly one X_{Λ_l}, $l \in L$. Hence

$$\begin{aligned} V((a,b) \cdot U_n) &\geq E\{V[(a,b) \cdot U_n \mid X_{\Lambda_l}, l \notin L]\} \\ &\geq \sum_{i \in L \cap (I_n \setminus \partial I_n)} E\left\{ V\left[\sum_{|j|\le 1} (ag_0 \circ \theta_j + bg_1 \circ \theta_j) \mid X_{\Lambda_l}, \right.\right. \\ &\qquad\qquad \left.\left. 1 \leq |l| \leq 2 \right] \right\}, \end{aligned} \tag{3.1}$$

where we use that conditionally the variance is a sum of variances and due to the stationarity the terms with $i \in L \cap (I_n \setminus \partial I_n)$ in the sum are equal. Since $|L \cap (I_n \setminus \partial I_n)| \geq c_1 |I_n|$ for n large we must show that the mean value in (3.1) is positive. This will be done by showing that the conditional variance is positive for the case $X_{\Lambda_l} = \emptyset$ for $1 \leq |l| \leq 2$. Let in this latter case $h(X_{\Lambda_0}) = \sum_{|j|\le 1}(ag_0 \circ \theta_j + bg_1 \circ \theta_j)$, so that we must show that $h(\cdot)$ is not almost surely a constant, which we will do by a counter argument. For $X_{\Lambda_0} = \emptyset$ we find $h(\emptyset) = -ae^{\omega}3^d|\Lambda_0|$. If h is a constant, we find for almost all $z_1, \ldots, z_n \in \Lambda_0$ that

$$\begin{aligned}(3.2)\qquad 0 &= h(\{z_1, \ldots, z_n\}) - h(\emptyset)\\ &= a\left\{n + e^{\omega}\int_{\tilde{\Lambda}}\left[1 - \exp\left(-\beta\sum_i \phi(z_i - \xi)\right)\right]d\xi\right\}\\ &\quad + b\left\{-\sum_{i\neq j}\phi(z_i - z_j)\right.\\ &\qquad\left. + e^{\omega}\int_{\tilde{\Lambda}}\left(\sum_i \phi(z_i - \xi)\right)\exp\left(-\beta\sum_i \phi(z_i - \xi)\right)d\xi\right\},\end{aligned}$$

where $\tilde{\Lambda} = \cup_{|i|\leq 1}\Lambda_i$.

Case 1. (C1) is satisfied: In particular for $n = 1$ we can change the variable of integration in (3.2) from ξ to $z_1 - \xi$ and use that $\phi(\xi) = 0$ if $\xi \notin \Lambda_0$ to obtain

$$b = -a\frac{1 + e^{\omega}\int_{\Lambda_0}[1 - \exp(-\beta\phi(\xi))]d\xi}{e^{\omega}\int_{\Lambda_0}\phi(\xi)\exp(-\beta\phi(\xi))d\xi} = -ac_2,$$

where $c_2 > 0$. Inserting this in (3.2) and letting $n \to \infty$ we see from the positivity of ϕ that a must be zero and so also $b = -ac_2$ is zero.

Case 2. (C2) satisfied: From (3.2) with $x = z_1$, $x + y = z_2$, $|y| > h$, we have for almost all x, y that

$$\begin{aligned}(3.3)\quad & a\left\{2 + e^{\omega}\int_{\tilde{\Lambda}}[1 - \exp(-\beta\phi(x - \xi) - \beta\phi(x + y - \xi))]d\xi\right\}\\ &\quad + be^{\omega}\int_{\tilde{\Lambda}}[\phi(x - \xi) + \phi(x + y - \xi)]\exp(-\beta\phi(x - \xi) - \beta\phi(x + y - \xi))d\xi\\ &\quad = 2b\phi(y).\end{aligned}$$

From (C2-i) the l.h.s. of (3.3) is continuous in x, y. If therefore $a_0 \neq 0$ in (C2-iii) the equation (3.3) can hold with $|y| \to \kappa$ from below and from above only if $b = 0$. If $a_0 = 0$ so that $a_1 \neq 0$ in (C2-iii) we use (C2-i, ii) to show that the l.h.s. of (3.3) is continuously differentiable, and again (3.3) can hold only if $b = 0$. When $b = 0$ we use (3.3) and (3.2) with $n = 1$,

$$(3.4)\qquad 0 = a\left\{1 + e^{\omega}\int_{\tilde{\Lambda}}[1 - \exp(-\beta\phi(z - \xi))]d\xi\right\}.$$

Subtracting (3.4) twice from (3.3)—once with $z = x$ and once with $z = x + y$—we find

$$(3.5)\qquad 0 = a\int_{\tilde{\Lambda}}[1 - \exp(-\beta\phi(x - \xi))][1 - \exp(-\beta\phi(x + y - \xi))]d\xi.$$

If we take y such that $2\kappa-\epsilon < |y| < 2\kappa$ the only points contributing to the integral in (3.5) are such that $|x-\xi| > \kappa-\epsilon$ and $|x+y-\xi| > \kappa-\epsilon$. Hence if ϵ is sufficiently small we have from (C2-iii) that the integrand in (3.5) is either strictly positive or strictly negative. This implies that $a = 0$. Since the cube Λ_0 has side length $\tilde{\kappa}$ only, we actually cannot take $|y| > 2\kappa - \epsilon$ unless $\tilde{\kappa} > 2d^{-1/2}\kappa$. However, taking the sides of the cubes Λ_i twice as large changes $\Sigma(U)$ only by a factor 2^d. □

LEMMA 3.4. *Let μ be an ergodic measure. Then under* (C1) *or* (C2) *we have*

$$|I_n|^{-1/2} U_n \Sigma(U)^{-1/2} \xrightarrow{\mathcal{D}} N_2(0, I).$$

PROOF. This follows from Theorem 2.1 using Lemma 3.1 to Lemma 3.3. □

We next turn to the second derivative $j_n = j_n(\omega, \beta)$. If μ is an ergodic measure, the ergodic theorem shows that

$$|I_n|^{-1} j_n \to \Sigma(j) = e^{\omega} \begin{bmatrix} \mu(f_0) & \mu(-f_1) \\ \mu(-f_1) & \mu(f_2) \end{bmatrix}$$

almost surely.

LEMMA 3.5. *The limit $\Sigma(j)$ is positive definite.*

PROOF. We have that

$$\begin{aligned}(a,b)\Sigma(j)(a,b)^* &= e^{\omega}\mu(a^2 f_0 - 2ab f_1 + b^2 f_2) \\ &= e^{\omega}\mu\left\{\int_{\Lambda_0} [a - bv(X,\xi)]^2 \exp(-\beta v(X,\xi)) d\xi\right\},\end{aligned}$$

and since $v(X, \xi)$ is not almost surely constant this term is positive. □

We now indicate the dependence on β in f_k by writing f_k^β, and prove a uniform convergence of j_n.

LEMMA 3.6. *Let μ be ergodic and let* (C1) *or* (C2) *hold. For any $c > 0$ and any sequence $\epsilon_n \to 0$ there exists a sequence $\delta_n \to 0$ such that for any sequence (ω_n, β_n), with $|\omega_n - \omega| < c\epsilon_n$, $|\beta_n - \beta| < c\epsilon_n$, we have*

$$|I_n|^{-1} \left| \sum_{i \in I_n} (e^{\omega_n} f_k^{\beta_n} - e^{\omega} f_k^{\beta}) \circ \theta_i \right| < \delta_n$$

almost surely.

PROOF. If $|\beta_n - \beta| \le \frac{1}{2}\beta$ we can, from the conditions (C1) or (C2), find a constant a_k such that $f_k^{\beta_n}$ is bounded by a_k. Then, we also have for $|\beta_n - \beta| \le \frac{1}{2}\beta$ that $|f_k^{\beta_n} - f_k^{\beta}| \le |\beta_n - \beta| a_{k+1}$, and therefore

$$\begin{aligned}&|I_n|^{-1} \sum_{i \in I_n} |e^{\omega_n} f_k^{\beta_n} - e^{\omega} f_k^{\omega}| \circ \theta_i \\ &\quad \le |e^{\omega_n} - e^{\omega}| a_k + e^{\omega}|\beta_n - \beta| a_{k+1} \to 0\end{aligned}$$

since $\epsilon_n \to 0$. □

LEMMA 3.7. *Let μ be ergodic and let (C1) or (C2) hold. Then the maximum pseudo likelihood estimate $(\hat{\omega}_n, \hat{\beta}_n)$ converges in probability to (ω, β). Furthermore,*

$$|I_n|^{-1/2} U_n \quad \textit{and} \quad |I_n|^{-1/2}(\hat{\omega}_n - \omega, \hat{\beta}_n - \beta) j_n$$

are asymptotically equivalent.

PROOF. This follows in a standard fashion from Lemmas 3.4 and 3.6, see e.g. Sweeting (1980). Consistency has also been proved by Jensen and Møller (1991). □

From Lemmas 3.7 and 3.4 we get immediately that

$$|I_n|^{-1/2}(\hat{\omega}_n - \omega, \hat{\beta}_n - \beta) j_n \Sigma(U)^{-1/2} \xrightarrow{\sim} N_2(0, I), \tag{3.6}$$

but this is not of much practical value since $\Sigma(U)$ cannot be explicitly calculated. We therefore need to estimate $\Sigma(U)$. Define

$$V_n = |I_n|^{-1} \sum_{i \in I_n} \left\{ \begin{pmatrix} g_0 \\ g_1 \end{pmatrix} \sum_{|j| \le 1} (g_0 \circ \theta_j, g_1 \circ \theta_j) \right\} \circ \theta_i, \tag{3.7}$$

then if μ is an ergodic measure we have

$$V_n \to \Sigma(U) \tag{3.8}$$

almost surely.

PROOF OF THEOREM 1.1. If μ is an ergodic measure the result follows from (3.6) and (3.8).

If there is more than one stationary Gibbs measure, then automatically there are nonergodic Gibbs measures because the set of all Gibbs measures is convex. But any stationary Gibbs measure can be represented as a mixture of ergodic measures (Georgii (1988), Theorem 14.10). Let us write this in the way

$$\int f(x) d\mu(x) = \int \left\{ \int f(x) d\mu_e(x) \right\} dm(e),$$

for any function f, where μ_e is an ergodic measure and the outer integral is over the set of ergodic measures. Then for any bounded and continuous function h on $\mathbb{R}^2$ we find

$$\begin{aligned} \lim_n \int h\{|I_n|^{-1/2}(\hat{\omega}_n - \omega, \hat{\beta}_n - \beta) j_n V_n^{-1/2}\} d\mu \\ = \int \left\{ \int h(x,y) \frac{1}{2\pi} e^{-x^2/2 - y^2/2} dx dy \right\} dm(e) \\ = \int h(x,y) \frac{1}{2\pi} e^{-x^2/2 - y^2/2} dx dy. \end{aligned}$$

□

4. Simulations

We have simulated the process (1.1) with $\phi(x) = 1(|x| < 0.05)$ in the region $\tilde{\Lambda} = [0,2] \times [0,2]$ with $y^{\tilde{\Lambda}} = \emptyset$ and then observed the process in $\Lambda = [0.5, 1.5] \times [0.5, 1.5]$. The process is obtained as the limit of a birth and death process starting from a Poisson point configuration, see Møller (1989). We initially simulated 100.000 births and deaths and then for every new 100.000 births and deaths sampled the process 100 times. We used the parameter values $\omega = 4.8$ and $\beta = 0.2$.

To obtain a symmetric version $\tilde{V}_n$ of V_n in (3.7) we restricted the sum over j in (3.7) to $j \in I_n$. When looking at the results of the simulations it was noticed that there was a systematic difference between $(\hat{\omega}_n - \omega, \hat{\beta}_n - \beta)$ and the approximation $U_n j_n^{-1}$. This remained true when using $\hat{j}_n$ instead of j_n. For that reason we considered (1.2) with j_n replaced by $\tilde{j}_n = (j_n + \hat{j}_n)/2$. In Fig. 1 we have plotted the hundred values of the statistic

$$|I_n|^{-1/2}(\hat{\omega}_n - \omega, \hat{\beta}_n - \beta)\tilde{j}_n \tilde{V}_n^{-1/2}, \tag{4.1}$$

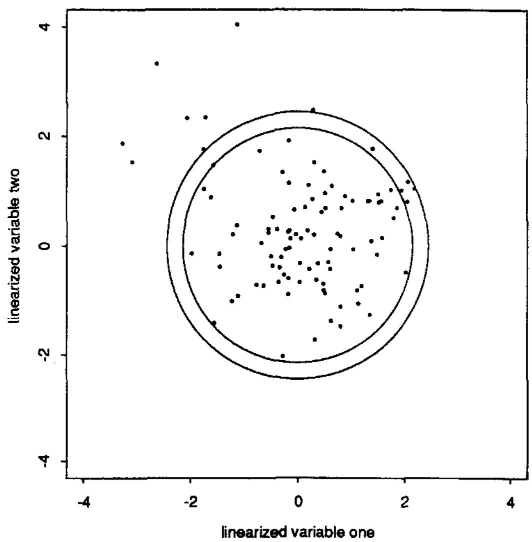

Fig. 1. The hundred simulated values of the normalized estimates (4.1). The circles give the 90% and 95% bounds from a $\chi^2(2)$-distribution.

where $\tilde{V}_n^{-1/2}$ was taken to be an upper triangular matrix. Four points in the upper left hand corner are somewhat extreme. Of these three are associated with the lowest values of the numbers of points in the observation window. Included in the figure are also the 90% and 95% confidence circles.

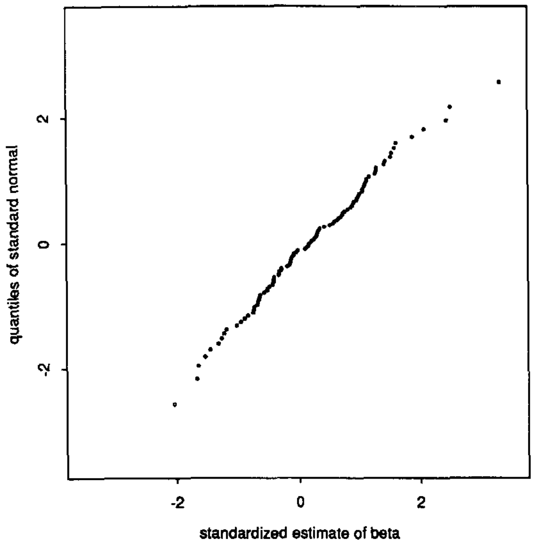

Fig. 2. The orderd values of the statistic (4.2) plotted against the quantiles of a standard normal distribution.

In Theorem 1.1 we have given the asymptotic normality for $(\hat{\omega}_n, \hat{\beta}_n)$. We can also formulate an equivalent statement for $\hat{\beta}_n$ only, i.e.

$$|I_n|^{-1/2}(\hat{\beta}_n - \beta)\{(\tilde{\jmath}_n^{-1})_2^* \tilde{V}_n(\tilde{\jmath}_n^{-1})_2\}^{-1/2} \xrightarrow{\mathcal{D}} N(0,1), \tag{4.2}$$

where A_2 indicate the second column of a two by two matrix A. In Fig. 2 we have compared the hundred values of the statistic in (4.2) with the quantiles of a standard normal distribution. As can be seen the normal approximation is a satisfactory approximation.

References

Besag, J., York, J. and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion), *Ann. Inst. Statist. Math.*, **43**, 1–59.

Bolthausen, E. (1982). On the central limit theorem for stationary mixing random fields, *Ann. Probab.*, **10**, 1047–1050.

Diggle, P., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D. and Tanemura, M. (1994). On parameter estimation for pairwise interaction point processes, *Internat. Statist. Rev.*, **62**, 99–117.

Georgii, H. O. (1988). *Gibbs Measures and Phase Transitions*, Walter de Gruyter, Berlin.

Guyon, X. (1987). Estimation d'un champ par pseudo-vraisemblance conditionelle: étude asymptotique et application au cas Markovien, *Spatial Processes and Spatial Time Series Analysés* (ed. F. Droesbeke), Proceedings of the 6th Franco-belgian meeting of statisticians, November 1985, Universitaires Saint-Louis, Bruxelles.

Guyon, X. and Künsch, H. R. (1992). Asymptotic comparison of estimators in the Ising model, In Stochastic Models, Statistical Methods and Algorithms in Image Analysis (eds. P. Barone, A. Frigessi and M. Piccioni), *Lecture Notes in Statist.*, **74**, 177–198, Springer, New York.

Heinrich, L. (1992). Mixing properties of Gibbsian point processes and asymptotic normality of Takacs-Fiksel estimates, Preprint 92-051, Universität Bielefeld, Germany.

Jensen, J. L. (1993). Asymptotic normality of estimates in spatial point processes, *Scand. J. Statist.*, **20**, 97–109.

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

Møller, J. (1989). On the rate of convergence of spatial birth-and-death processes, *Ann. Inst. Statist. Math.*, **41**, 565–581.

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

Takahata, H. (1983). On the central limit theorem for weakly dependent random fields, *Yokohama Math. J.*, **31**, 67–77.