© Board of the Foundation of the Scandinavian Journal of Statistics 1993. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF and 238 Main Street, Suite 501, Cambridge, MA 02142, USA. Vol. 20: 97–109 1993

Asymptotic Normality of Estimates in Spatial Point Processes

J. L. JENSEN

University of Aarhus

ABSTRACT. Gibbs models for spatial point processes defined in terms of a pair potential are studied in the Dobrushin uniqueness region. Asymptotic normality of maximum pseudo likelihood estimates and maximum likelihood estimates is established.

Key words: asymptotic normality, exponential family, Gibbs processes, maximum likelihood, pairwise interaction processes, pseudo likelihood, strong mixing

1. Introduction

Gibbs models of spatial point processes, and in particular those defined in terms of a pair potential, are a popular modelling tool in the statistical literature. This is mainly because of the suggestive interpretations of these models in terms of attraction and repulsion. Unfortunately the statistical analysis of Gibbs models is hampered by the intractability of the normalizing constant. Maximum likelihood estimates have to be found by numerical methods using simulation. Such procedures have been considered in Penttinen (1984) and recently a more simple procedure has been described in Baddeley & Moyeed (1991). Approximation methods not involving simulation have been studied by Ogata & Tanemura (1984).

To avoid the above mentioned difficulties, a pseudo likelihood function has been introduced to replace the ordinary likelihood function. The original idea is due to Besag (1975) for lattice fields, and is extended to the Strauss point process in Besag (1977). A general definition of the pseudo likelihood function for point processes was discussed in Jensen & Møller (1991) together with a proof of consistency of maximum pseudo likelihood estimates.

There seems to be almost no results available on the distributional properties of either the maximum likelihood or the maximum pseudo likelihood estimates for Gibbs point processes. It is the purpose of this paper to discuss asymptotic normality of the different estimates. The results in this paper are similar to those derived by Guyon (1987) for Gibbs lattice models. Since we are dealing with dependent observations we need estimates for the decay of the dependency. Such estimates are known in general processes only in the so-called Dobrushin uniqueness region, and therefore the results here are obtained under the same restriction on the parameters. In this region there is only one infinite volume Gibbs distribution and problems related to phase transition do not appear. The approach via bounds on the mixing coefficients seems natural although one may speculate whether the Markov structure is sufficiently exploited. Certainly, it is not exploited in this paper as effectively as in Jensen & Møller (1991) in connection with consistency of the pseudo likelihood estimates. What happens to the mixing properties when we are outside the Dobrushin uniqueness region? Typically an infinite volume Gibbs measure will be a mixture of extremal Gibbs measures (see Preston, 1976), and the mixing coefficient need not tend to zero. However, from a statistical point of view with only a part of one realization available, conditioning on this one realization coming from one of the extremal measures seems appropriate. If therefore it can be shown that the extremal measures have good mixing properties the approach of this paper can be used. Unfortunately very little is known about such questions, and there does not seem to be a general affirmative answer.

In section 2 we describe results on the decay of the dependency and central limit theorems in a general setting. Specializing to Gibbs models with a pair potential, we consider in section 3 when the results of section 2 are applicable. In section 4 we establish asymptotic normality of the maximum pseudo likelihood estimates, and in section 5 for the maximum likelihood estimates obtained from a conditional likelihood.

As mentioned above the pseudo likelihood function was introduced because of its tractability as compared to the likelihood function. But from the results in section 4 the asymptotic covariance matrix of the maximum pseudo likelihood estimate is not so easy to calculate. Perhaps then the asymptotic result in section 5 for the maximum likelihood estimate is easier to apply in practice.

2. Central limit theorem for Gibbs fields with general state space

In sections 3–5 we will transform a Gibbs point process into a lattice process by letting the lattice variable be the point configuration in a box round the lattice point. This means that the state space for the lattice variable will be more general than is the case for many of the results previously derived for lattice fields. We therefore start with some central limit theorems for lattice fields with general state space. In section 2.1 we first define the notation and derive a bound on the strong mixing coefficient based on Föllmer (1982). Combining this with Bolthausen (1982) we give a central limit theorem in section 2.2 for the stationary case. Using instead a result of Takahata (1983) we derive in section 2.3 a central limit theorem in a conditional setting.

2.1. Mixing properties in Dobrushin's uniqueness region

Let $(S, \mathscr{S})$ be a measurable space and $(E, \mathscr{E})$ the corresponding product space $E = S^I$, where I is a countable set of sites. For a point $x \in E$ we write $x = \{x_i\}_{i \in I}$ to indicate the ith coordinate x_i, and let $\check{x}_i = \{x_j\}_{j \in I \setminus \{i\}}$. Let $r(\cdot\,,\cdot)$ be a measurable metric on S and define the Vasserstein metric for probability measures on S by

$$R(v_1, v_2) = \sup_{g:\, \delta(g) < \infty} \left| \int g\, dv_1 - \int g\, dv_2 \right| \Big/ \delta(g),$$

where for functions $g: S \to \mathbb{R}$ we define $\delta(g) = \sup_{s \neq t} |g(s) - g(t)|/r(s, t)$. For a function f on E, we define for $i \in I$

$$\delta_i(f) = \sup \left\{ \frac{|f(x) - f(y)|}{r(x_i, y_i)},\ \check{x}_i = \check{y}_i \right\},$$

and then introduce the class of functions

$$L(E) = \left\{ f: E \to \mathbb{R}: \sum_i \delta_i(f) < \infty,\ |f(x) - f(y)| \leqslant \sum_i r(x_i, y_i)\delta_i(f) \text{ for all } x, y \in E \right\}. \tag{2.1}$$

Intuitively this class restricts attention to functions that depend mainly on a finite set of sites $i \in I$, i.e. $\delta_i(f)$ has to tend to zero for $i \to \infty$. Also the second inequality defining $L(E)$ is trivially satisfied for functions f with $\delta_i(f) = 0$ except for a finite set of sites $i \in I$.

Let μ be a measure on $(E, \mathscr{E})$ with regular conditional distributions $\mu_i(A|\check{x}_i) = P\{X_i \in A | X_j = x_j, j \neq i\}$ of the ith coordinate given the σ-field generated by $\{X_j\}, j \neq i$. The measure μ is called a Gibbs measure with conditional distributions $\mu_i(\cdot|\cdot)$, $i \in I$, if

$$\forall f \in L(E)\forall i \in I: \left\{x \to \int f(y)\bar{\mu}_i(dy|x)\right\} \in L(E), \tag{2.2}$$

where for $B \in \mathscr{E}$ and $x \in E$ the measure $\bar{\mu}_i(B|x) = \mu_i(\{y \in S: \tilde{x} \in B$ where $\tilde{x}_i = y$, $\tilde{x}_j = x_j$ for $j \neq i\}|\check{x}_i)$ is the product measure on E fixed at x_j for $j \neq i$ and equal to $\mu_i(\cdot|\check{x}_i)$ at site i (Föllmer, 1982). In section 3 we consider two classes of models, in examples 1 and 2 respectively, for which the condition (2.2) is satisfied. The dependency of $\mu_k(\cdot|\check{x}_k)$ on x_i is measured by

$$C_{ik} = \sup\{R(\mu_k(\cdot\ |\check{x}_k), \mu_k(\cdot\ |\check{y}_k))/r(x_i, y_i), \check{x}_i = \check{y}_i\}. \tag{2.3}$$

Define the matrices $C = (C_{ik})$, $D = \Sigma_0^\infty C^n$, and the constant

$$\sigma^2 = \sup_k \int\left[\inf_{s \in S}\int r(y, s)^2\mu_k(dy|\check{x}_k)\right]\mu(dx).$$

For the set up here Föllmer has proved the following basic covariance estimate:

Theorem 2.1 (Föllmer, 1982)

If (i) $\lim_n \Sigma_i (C^n)_{ik} = 0$, $k \in I$, *and (ii)* $\sup_k \int r(x_k, y_k)^2\mu(dx) < \infty$ *for some* $y \in E$, *then*

$$|\mathrm{cov}_\mu(f, g)| = \left|\int f(x)g(x)\mu(dx) - \left(\int f(x)\mu(dx)\right)\left(\int g(x)\mu(dx)\right)\right| \leqslant \sigma^2\sum_{i,k}\delta_i(f)D_{ik}\delta_k(g),$$

for f and g in $L(E)$.

Condition (ii) implies that $\sigma^2 < \infty$, and intuitively σ^2 appears in the result of theorem 2.1 because both of f and g belong to $L(E)$, see definition (2.1).

If we assume that

$$\delta = \sup_k \sum_i C_{ik} < 1,$$

then $\Sigma_i (C^n)_{ik} \leqslant \delta^n$ and condition (i) of theorem 1 is fulfilled. In the models of examples 1 and 2 in section 3 we will see that the condition $\delta < 1$ imposes a restriction on the parameter values that can be handled. Let us now specialize to a translation invariant Gibbs measure on $\mathbb{Z}^d$ and define $\tilde{C}_k = C_{0k}$ and $\tilde{D}_k = D_{0k}$. Then we may write $C_{ik} = \tilde{C}_{i-k}$ and $D_{ik} = \tilde{D}_{i-k}$. The estimate $\Sigma_i (C^n)_{ik} \leqslant \delta^n$ then shows that $\Sigma_i \tilde{D}_i \leqslant \Sigma_0^\infty \delta^n = (1 - \delta)^{-1}$. If θ_j denotes the shift transformation on E, i.e. $(\theta_j x)_i = x_{j+1}$, we get the bound

$$\left|\sum_j \mathrm{cov}_\mu(f, f \circ \theta_j)\right| \leqslant \sigma^2\sum_j\sum_{i,k}\delta_i(f)\tilde{D}_{i-k}\delta_{j+k}(f) \leqslant \frac{\sigma^2}{1-\delta}\left(\sum_i \delta_i(f)\right)^2 < \infty \quad \text{for } f \in L(E). \tag{2.4}$$

Assume now that the metric $r(\cdot, \cdot)$ on S satisfies $r(s, t) \geqslant 1$ for $s \neq t$. Then if $f: E \to \mathbb{R}$ is an indicator function we have trivially $\delta_i(f) \leqslant 1$ and if f depends on finitely many coordinates only we also have that $f \in L(E)$. In the case of a translation invariant Gibbs measure we get from theorem 2.1 with $A_1 \in \sigma(X_i, i \in I_1)$ and $A_2 \in \sigma(X_i, i \in I_2)$

$$|\mu(A_1 \cap A_2) - \mu(A_1)\mu(A_2)| \leqslant \sigma^2 \sum_{i \in I_1, j \in I_2} \tilde{D}_{i-j}.$$

When the condition $\tilde{D}_k \leqslant c_1|k|^{-d-\gamma-\varepsilon}$ is fulfilled, for some $\varepsilon > 0$, we obtain $\Sigma_{d(0,k) \geqslant m}\tilde{D}_k \leqslant c_2 m^{-\gamma}$, where $d(i,j) = \max\{|i_l - j_l|, l = 1, \dots, d\}$. Then

$$|\mu(A_1 \cap A_2) - \mu(A_1)\mu(A_2)| \leqslant \sigma^2|I_2|c_2 m^{-\gamma} \tag{2.5}$$

whenever $d(I_1, I_2) = \inf\{d(i,j): i \in I_1, j \in I_2\} \geqslant m$, $|I_1| < \infty$ and $|I_2| < \infty$ with $|J|$ denoting the cardinality of a set J.

2.2. *Central limit theorem in the stationary case*

Let $I(n)$ be a finite subset of $\mathbb{Z}^d$ and define for a function $f: E \to \mathbb{R}$ the standardized sum

$$S_n(f) = \sum_{i \in I(n)} \left(f \circ \theta_i - \int f \, d\mu\right) \Big/ |I(m)|^{1/2}.$$

Künsch (1982a) proved that with $I(n) = [-n, n]^d$ and $f \in L(E)$ the sum $S_n(f)$ is asymptotically normally distributed under conditions on the coefficients $\tilde{C}_k$, see also Künsch (1982b). The result of Künsch (1982a) has been the inspiration for the work here. We state now instead a result for a general function f dependent on finitely many coordinates only. The boundary of the index set $I(n)$ is $\partial I(n) = \{i \in I(n): \exists j \notin I(n) \text{ with } d(i,j) = 1\}$ and we assume that

$$\frac{|\partial I(n)|}{|I(n)|} \to 0 \quad \text{for } n \to \infty. \tag{2.6}$$

In section 4 this condition will be trivially met because we only consider simple choices of $I(n)$.

Theorem 2.2
Let μ be a translation invariant Gibbs field with $\Sigma_k \tilde{C}_k < 1$, $\int r(x_0, s)^2 \mu(dx) < \infty$ for some $s \in S$, and let $f: E \to \mathbb{R}$ be a measurable function dependent on only a finite set of coordinates. If for some $\alpha > 0$, $c_1 > 0$ and $\lambda > 2d/\alpha$, we have

(*i*) $\mu(|f - \mu(f)|^{2+\alpha}) < \infty$ *and*
(*ii*) $\tilde{D}_k \leqslant c_1 |k|^{-2d-\lambda}$,

then $S_n(f)$ is asymptotically normally distributed with mean zero and variance $\Sigma_k \operatorname{cov}_\mu(f, f \circ \theta_k) < \infty$.

As usual $\mu(f)$ denotes the integral of f, i.e. $\mu(f) = \int f(x)\mu(dx)$.

Proof. For ease in notation we consider only the case where f is a function of x_0. The proof is the same in the general case. The result follows from Bolthausen (1982) if only we show that the mixing coefficients $\alpha_{kl}(m)$ of his paper are sufficiently small. From (2.5) and (ii) we immediately get for any $\tilde{\lambda} < \lambda$

$$\alpha_{kl}(m) \leqslant c_2 \sigma^2 k m^{-d-\tilde{\lambda}},$$

which together with (i) show that all the conditions of Bolthausen's theorem are satisfied. □

2.3. *A central limit theorem in a conditional setting*

In section 2.1 we wrote μ instead of the more proper μ_I to indicate the set of sites I. This was because we had $I = \mathbb{Z}^d$ in mind, but the result of theorem 2.1 is true for any I. In particular let $I(n)$ be a finite set of $\mathbb{Z}^d$, $J(n) = \mathbb{Z}^d \backslash I(n)$ the complementary set, and $x_{J(n)} = \{x_i, i \in J(n)\}$ a point in $S^{J(n)}$. Corresponding to the conditional distributions $\mu_i(\cdot|\check{x}_i)$, at site $i \in I_n$, we have a conditional Gibbs measure $\mu_{I(n)}(\cdot|x_{J(n)})$ on $S^{I(n)}$ given the σ-field $\sigma(X_i: i \in J(n))$. If we let $C = (C_{ik})$ be the dependency matrix defined in (2.3) for $i, k \in \mathbb{Z}^d$ and let $D = \Sigma_0^\infty C^n$, we find that D_{ik} is an upper bound to the similar coefficient defined from the sites $i \in I(n)$ only. Thus the result of theorem 2.1 reads

$$|\mathrm{cov}_{\mu_{I(n)}}(f, g)| \leqslant \sigma^2_{I(n)} \sum_{i, k \in I(n)} \delta_i(f) D_{ik} \delta_k(g)$$

for f and g in $L(S^{I(n)})$. This in turn leads to a mixing statement equivalent to (2.5) for $\mu_{I(n)}(\cdot \,|x_{J(n)})$.

Let K be a finite set of sites and f_i a function that depends on the coordinates in $i + K$ only. We will then consider the asymptotic distribution of

$$U_n = \sum_{I(n)} f_i(X)$$

when X is distributed according to $\mu_{I(n)}(\cdot) = \mu_{I(n)}(\cdot \,|x_{J(n)})$ for a fixed boundary $x_{J(n)}$.

Theorem 2.3

Let $\mu_i(\cdot \,|\check{x}_i)$ be a translation invariant specification of conditional distributions, $\Sigma\, \tilde{C}_k < 1$, and let $\tilde{\sigma}^2 = \sup_x \int r(y, s)^2 \mu_0(dy|\check{x}_0) < \infty$ for some $s \in S$. If for some c_1, c_2 and $\gamma > 3d$ we have

(i) $\mu_{I(n)}(|f_i - \mu_{I(n)}(f_i)|^3) \leqslant c_1, \quad i \in I(n), \quad n \in \mathbb{N}$

(ii) $\tilde{D}_k \leqslant c_2 |k|^{-d-\gamma}$ *then there exists c_3 such that*

$$\mathrm{var}_{\mu_{I(n)}}(U_n) \leqslant c_3 |I(m)|. \tag{2.7}$$

If furthermore there exists $c_4 > 0$ such that

(iii) $\mathrm{var}_{\mu_{I(n)}}(U_n) \geqslant c_4 |I(m)|$, *then as $I(n) \to \infty$ with $n \to \infty$*

$$(U_n - E_{\mu_{I(n)}}(U_n))/\{\mathrm{var}_{\mu_{I(n)}}(U_n)\}^{1/2} \xrightarrow{\mathcal{D}} N(0, 1). \tag{2.8}$$

Proof. Since from the assumptions $\int r(x_0, s)^2 \mu_{I(n)}(dx) \leqslant \tilde{\sigma}^2$, we can use theorem 2.1 as explained above. The result (2.5) is then true for $\mu_{I(n)}$ under assumption (ii). The assumptions in theorem 2 of Takahata (1983) are then fulfilled and (2.8) holds. In Takahata (1983) the result is formulated for a fixed sequence of random variables, but this is not used in the proof, and we can therefore apply the result to our set up. Finally, (2.7) is a standard estimate from assumption (i) and the result (2.5). □

Typically the condition (iii) in theorem 2.3 will hold if the terms in the sum U_n do not cancel with one another. Using the conditional independence approach of Jensen & Møller (1991) one should be able to prove (iii) in many cases.

3. Point processes as lattice fields

As in Klein (1982, 1984) we will now describe a point process in $\mathbb{R}^d$ as a lattice process and discuss how to check the conditions of section 2. For any bounded $\Lambda \subset \mathbb{R}^d$ we let $(S(\Lambda), \mathscr{B}(\Lambda))$ denote the measurable space of finite configurations of particles in Λ, see Preston (1976). Let $\kappa > 0$ be a fixed number, then for any $i = (i_1, \ldots, i_d) \in \mathbb{Z}^d$ we let $\Lambda_i = \{z \in \mathbb{R}^d | \kappa(i_j - \frac{1}{2}) \leqslant z_j < \kappa(i_j + \frac{1}{2})\}$. Then the space $(S(\bigcup_{i \in I} \Lambda_i), \mathscr{B}(\bigcup_{i \in I} \Lambda_i))$ is isomorphic to $\Pi_{i \in I}(S(\Lambda_i), \mathscr{B}(\Lambda_i))$, and the space Ω of locally finite point configurations in $\mathbb{R}^d$ is isomorphic to the product space obtained with $I = \mathbb{Z}^d$.

For the metric $r(\cdot, \cdot)$ in section 2.1 we choose the one used in Klein (1982), namely $r(s, t) = \rho_1(s, t) + \rho_2(s, t)$, where ρ_2 is the discrete metric on $S(\Lambda_i)$. If $|s|$ denotes the number of elements in the configuration s the metric ρ_1 is, in the case $|s| \leqslant |t|$ say, given by

$$\rho_1(s, t) = \rho_1(t, s) = \min_{\pi \in \Sigma_{n,m}} \sum_{i=1}^{n} d_e(s_i, t_{\pi(i)}) + m - n,$$

where $d_e(\cdot, \cdot)$ is euclidean distance, $n = |s|$, $m = |t|$, $n \leqslant m$, and $\Sigma_{n,m}$ is the set of all one to one mappings of $\{1, \dots, n\}$ to $\{1, \dots, m\}$. In particular we have $r(s, \varnothing) = |s| + 1$ for $s \neq \varnothing$, where $\varnothing$ is the empty point configuration.

We will study models defined in terms of a pair potential φ: $\mathbb{R}^d \to \mathbb{R}$, with $\varphi(-z) = \varphi(z)$ and $\varphi(0) = 0$, which we assume to be stable, i.e. there exists $B \geqslant 0$ such that

$$\frac{1}{2} \sum_{i,j=1}^{n} \varphi(z_i - z_j) \geqslant -nB \quad \text{for all } n, z_1, \dots, z_n. \tag{3.1}$$

When $x \in \Omega$ is a locally finite configuration we define $x_\Lambda = x \cap \Lambda \in \Omega_\Lambda$ where Ω_Λ is the space of locally finite point configurations in Λ. From the potential φ we define the interaction V_Λ by

$$V_\Lambda(x_\Lambda | y_{\tilde{\Lambda}}) = \sum_{\substack{\{z_1, z_2\} \subset x_\Lambda \cup y_{\tilde{\Lambda}} \\ \{z_1, z_2\} \cap x_\Lambda \neq \varnothing}} \varphi(z_1 - z_2),$$

for $x_\Lambda \in \Omega_\Lambda$, $y_{\tilde{\Lambda}} \in \Omega_{\tilde{\Lambda}}$ and with $\tilde{\Lambda} = \mathbb{R}^d \backslash \Lambda$ the complement of Λ. If ν_Λ is the Poisson measure on $S(\Lambda)$, with unit intensity with respect to Lebesgue measure, we define for $y \in \Omega_{\tilde{\Lambda}}$ a measure $\mu_\Lambda(\cdot | y_{\tilde{\Lambda}})$ on $S(\Lambda)$ by

$$\frac{d\mu_\Lambda(\cdot | y_{\tilde{\Lambda}})}{d\nu_\Lambda}(x_\Lambda) = Z(\zeta, \beta | y_{\tilde{\Lambda}})^{-1} \zeta^{|x_\Lambda|} \exp\{-\beta V_\Lambda(x_\Lambda | y_{\tilde{\Lambda}})\}, \tag{3.2}$$

where $Z(\zeta, \beta | y_{\tilde{\Lambda}})$ is a norming constant.

To tie this up with the notation in section 2 we let $x_i = x_{\Lambda_i}$, i.e. the configuration in Λ_i, and $\mu_i(\cdot | \{x_j\}, j \neq i) = \mu_{\Lambda_i}(\cdot | x_{\tilde{\Lambda}_i})$. We then have to study the dependency coefficients C_{ik} from (2.3) to see when the conditions of theorem 2.2 are fulfilled. We will use the approach in Klein (1982), although using slightly different assumptions and therefore obtaining slightly different results.

Lemma 3.1

(i) Assume that there exists $B_1 \geqslant 0$ *such that* $V_{\Lambda_0}(x_{\Lambda_0} | y_{\tilde{\Lambda}_0}) \geqslant -B_1 |x_{\Lambda_0}|$ *for all* x_{Λ_0} *and all* $y_{\tilde{\Lambda}_0}$. *Then*

$$\tilde{C}_i \leqslant 2e^{\omega}\{\omega + 1 - e^{-\omega}\},$$

where $\omega = \zeta |\Lambda_0| \exp(\beta B_1)$. *Furthermore* $\int r(x_0, \varnothing)^2 \mu(dx) < \infty$.

(ii) Furthermore, if there exists K_i *such that*

$$\int_{\Lambda_0} |\varphi(z - u) - \varphi(z - v)| \, dz \leqslant K_i \, d_e(u, v)$$

$$\int_{\Lambda_0} |\varphi(z - u)| \, dz \leqslant K_i$$

for $u, v \in \Lambda_i$, *then*

$$\tilde{C}_i \leqslant 2K_i \beta e^{\omega}\{\omega^2 + 2\omega\}.$$

Proof. Since $|g(x_0) - g(\varnothing)| \leqslant \delta(g) r(x_0, \varnothing)$ for a function g: $S(\Lambda_0) \to \mathbb{R}$ we find for $\delta(g) = 1$,

$$|\mu_0(g | \check{x}_0) - \mu_0(g | \check{y}_0)| \leqslant \int r(v_0, \varnothing) \left| \frac{\exp\{-\beta V_0(v_0 | \check{x}_0)\}}{Z(\beta | \check{x}_0)} - \frac{\exp\{-\beta V_0(v_0 | \check{y}_0)\}}{Z(\beta | \check{y}_0)} \right| \nu_0^{\zeta}(dv_0),$$

where $V_0(v_0|\check{x}_0) = V_{\Lambda_0}(v_0|x_{\bar{\Lambda}_0})$, v_0^ζ is Poisson measure on Λ_0 with intensity ζ times the Lebesgue measure and $Z(\beta|\check{x}_0)$ is the norming constant (3.2) for $\mu_0(\cdot\,|\check{x}_0)$. Since $Z(\beta|x) \geqslant 1 \cdot v_0^\zeta(v_0 = \varnothing) = \exp(-\zeta|\Lambda_0|)$ we get from the assumption trivially the bound

$$2\int r(v_0, \varnothing) \exp(\beta B_1|v_0| + \zeta|\Lambda_0|)v_0^\zeta(dv_0) = 2e^{\omega}\{\omega + 1 - e^{-\omega}\}$$

with $\omega = \zeta|\Lambda_0| \exp(\beta B_1)$. Finally $r(x_i, y_i) \geqslant 1$, which then gives the first result of the lemma from (2.3). That $\int r(x_0, \varnothing)^2 \mu(dx) < \infty$ follows from $r(x_0, \varnothing) = |x_0| + 1$ and $\int (|x_0| + 1)^2 \mu_0(dx_0|y) \leqslant v_0^\zeta\{(|x_0| + 1)^2 \exp(\beta B_1|x_0| - \zeta|\Lambda_0|\}$.

For the second part we shall use a trick introduced in Simon (1979). Let v_θ have density $\exp\{\theta[-\beta V_0(\cdot\,|\check{x}_0)] + (1-\theta)[-\beta V_0(\cdot\,|\check{y}_0)]\}/Z$ with respect to v_0^ζ and let $\tilde{g}(v_0) = g(v_0) - g(\varnothing)$. Then

$$\begin{aligned}|\mu_0(g|\check{x}_0) - \mu_0(g|\check{y}_0)| &= |v_1(\tilde{g}) - v_0(\tilde{g})| \\ &= \left|\int_0^1 \frac{dv_\theta(\tilde{g})}{d\theta}\, d\theta\right| \left|\int_0^1 [v_\theta(\tilde{g}q) - v_\theta(\tilde{g})v_\theta(q)]\, d\theta\right| \\ &\leqslant \int_0^1 \left|\operatorname{cov}_{v_\theta}(\tilde{g}, q)\right| d\theta \leqslant 2\int_0^1 v_\theta(|\tilde{g}\,\|q|)\, d\theta,\end{aligned} \tag{3.3}$$

where $q(\cdot) = \beta[V_0(\cdot\,|\check{x}_0) - V_0(\cdot\,|\check{y}_0)]$. Now, x and y only differ at position "i" when evaluating C_{0i}. Let $x_i = \{x_i^1, \ldots, x_i^n\}$ and $y_i = \{y_i^1, \ldots, y_i^m\}$ numbered in such a way that $\rho_1(x_i, y_i) = \Sigma_{j=1}^n d_e(x_i^j, y_i^j) + m - n$. Also let $v_0 = (v_0^1, \ldots, v_0^k)$ when $|v_0| = k$. Then

$$q(v_0) = \beta\left\{\sum_{j=1}^{n}\sum_{k=1}^{|v_0|}[\varphi(v_0^k - x_i^j) - \varphi(v_0^k - y_i^j)] + \sum_{j=n+1}^{m}\sum_{k=1}^{|v_0|}\varphi(v_0^k - y_i^j)\right\}.$$

Using $|\tilde{g}(v_0)| \leqslant r(v_0, \varnothing)$, $dv_\theta(v_0) \leqslant \exp(\beta B_1|v_0| + \zeta|\Lambda_0|)\, dv_0^\zeta(v_0)$, and the assumptions under (ii) we get

$$\begin{aligned}v_\theta(|\tilde{g}\|q|) &\leqslant \int r(v_0, \varnothing)\beta K_i\left(\sum_1^n d_e(x_i^j, y_i^j) + m - n\right)\exp(\beta B_1|v_0| + \zeta|\Lambda_0|)v_0^\zeta(dv_0) \\ &= \beta K_i e^{\omega}\{\omega^2 + 2\omega\}\rho_1(x_i, y_i).\end{aligned}$$

From (3.3) and (2.3) we therefore obtain

$$\tilde{C}_i = C_{0i} \leqslant 2\beta e^{\omega}\{\omega^2 + 2\omega\}K_i. \qquad \square$$

The stability condition (3.1) and the condition in lemma 3.1(i) restrict the behaviour of the potential φ. We will now consider in detail two classes of potentials that satisfy the conditions, namely positive potentials and potentials with a hard-core. If the potential is not positive the hard-core condition is natural to ensure stability. It can be relaxed by assuming that $\varphi(z)$ tends to infinity sufficiently fast for $z \to 0$, see Ruelle (1969). Note in particular that it is not possible to have $\varphi(z) < 0$ for $z \to 0$, i.e. local attraction is not possible unless there is a repulsion for very small distances.

Example 1. Finite range positive potential. Assume that $\varphi(z) = 0$ when $|z| \geqslant \kappa$, where κ is the fixed number appearing in the definition of the regions Λ_i. Then according to theorem 3.2 of Klein (1982) condition (2.2) is fulfilled. The coefficients $\tilde{C}_i = C_{0i}$ are zero unless $|i_j| \leqslant 1$ for $j = 1, \ldots, d$. Since in this case the assumption in (i) of lemma 3.1 is fulfilled with $B_1 = 0$, we get from the lemma with $K = \max\{K_i, |i_j| \leqslant 1, j = 1, \ldots, d\}$

$$\sum_i \tilde{C}_i \leqslant 2(3^d - 1)\min\{e^{\zeta\kappa^d}[\zeta\kappa^d + 1 - e^{-\zeta\kappa^d}],\, \beta K e^{\zeta\kappa^d}[(\zeta\kappa^d)^2 + 2\zeta\kappa^d]\} \tag{3.4}$$

and

$$\sum_k \tilde{C}_k |k|^\gamma < \infty \quad \forall \gamma > 0.$$

In the particular case of a Strauss process on $\mathbb{R}^2$ with $\varphi(z) = 1(|z| < \kappa)$ it is trivial to see that the second part of condition (ii) in lemma 3.1 is fulfilled with $K_i = |\Lambda_0| = \kappa^2$. The first part is more complicated, but a calculation shows that we can take $K_i = 2\kappa$. Thus the upper bound K in (3.4) can be taken as $\max(2\kappa, \kappa^2)$. For a fixed β (3.4) provides an upper bound for ζ, in order that theorem 2.2 is applicable, and in particular this upper bound tends to infinity as $\beta \to 0$, i.e. in the limit of a Poisson process.

Example 2. Hard-core potential. We assume that $\varphi(x) = \infty$ for $|x| < h$, which means that no two points are allowed to be closer than h. Furthermore we assume that there exists a decreasing function $\psi: [\kappa, \infty) \to \mathbb{R}$ such that

$$|\varphi(z)| \leqslant \psi(|z|) \quad \text{for} \quad |z| \geqslant \kappa \quad \text{and} \quad \lambda = \sum_{i: d_e(\Lambda_0, \Lambda_i) \geqslant \kappa} \psi(d_e(\Lambda_0, \Lambda_i)) < \infty.$$

According to theorem 3.3 of Klein (1982), condition (2.2) is now fulfilled.

In this case we can restrict attention to the spaces $S_N(\Lambda_i)$ of point configuration with at most N points, where N is the maximal number of points in Λ_i with all interpoint distances greater than h. Using the stability condition (3.1), we find

$$V_{\Lambda_0}(v_0|y) \geqslant -B|v_0| - |v_0| N(\lambda + 3^d - 1) = -B_1|v_0|,$$

and condition (i) of lemma 3.1 is fulfilled.

For conditions (ii) of lemma 3.1 we assume that K_i can be taken as $K\psi(d_e(\Lambda_0, \Lambda_i))$ for some constant K, where we define $\psi(0) = \psi_0$. In this case we end up with

$$\sum_i \tilde{C}_i \leqslant 2K\{\lambda + (3^d - 1)\psi_0\} \min\{e^\omega(\omega + 1 - e^{-\omega}), \beta e^\omega(\omega^2 + 2\omega)\},$$

where $\omega = \zeta\kappa^d \exp(\beta B_1)$.

In the particular case of the potential in $\mathbb{R}^2$,

$$\varphi(z) = \begin{cases} \infty & |z| < h \\ -1 & h \leqslant |z| < \kappa \\ 0 & |z| \geqslant \kappa \end{cases}$$

we can take $\psi(r) \equiv 0$, $\psi_0 = 1$ and $K = \max(2\kappa, \kappa^2)$, as in example 1. The constant B_1 can be estimated by $9N$.

As example 1 and example 2 show, we can obtain asymptotic normality under μ for a large class of functions using theorem 2.2. Asymptotic normality of the minimal sufficient statistic $(|X_\Lambda|, V_\Lambda(X_\Lambda, \varnothing))$, under the model (3.2) with $y_{\bar{\Lambda}} = \varnothing$ and with Λ increasing to $\mathbb{R}^d$, was considered in Jensen (1991a). Also the conditional distribution of $V_\Lambda(X_\Lambda|\varnothing)$ given $|X_\Lambda|$ was considered. The results were based on proving uniform convergence of the cumulant transform and its derivatives, and can easily be translated into results about asymptotic normality of the maximum likelihood estimate from the model (3.2) with $y_{\bar{\Lambda}} = \varnothing$. See also section 5 below.

4. Asymptotic normality of maximum pseudo likelihood estimates

In this section we again restrict attention to models of the form (3.2) defined through a pair potential φ. If the process is of finite range κ and observed in a window Λ it is customary to consider either the likelihood function obtained from (3.2) using the conditional distribution of $X_{\mathring{\Lambda}}$ given $X_{\Lambda \setminus \mathring{\Lambda}}$, where $\mathring{\Lambda} = \{\xi \in \Lambda: d_e(\xi, \eta) > \kappa \text{ for all } \eta \notin \Lambda\}$ is the interior of Λ, or the likelihood function obtained from (3.2) on putting $y_{\bar{\Lambda}} = \emptyset$. In both cases the analysis is made very difficult by not knowing the normalizing constant Z explicitly. For this reason the pseudo likelihood function has been introduced. The pseudo likelihood was originally introduced by Besag (1975) for lattice processes as a product of the one site conditional densities. A general formulation together with more references can be found in Jensen & Møller (1991). In the set up here, the log pseudo likelihood function becomes

$$pl_\Lambda(\alpha, \beta) = \alpha|x_\Lambda| - \beta \sum_{z \in x_\Lambda} v(x\backslash z, z) - e^{\omega} \int_\Lambda \exp\{-\beta v(x, \xi)\}\, d\xi \tag{4.1}$$

where $\omega = \log(\zeta)$, $v(x, \xi) = \Sigma_{z \in x}\, \varphi(\xi - z)$, and $x\backslash z$ signifies that the point z has been deleted from the configuration x. If the interaction is of finite range we can evaluate $pl_{\bar{\Lambda}}$ on observing x_Λ. In the non-finite range case we will have to specify $x_{\bar{\Lambda}}$ in order to evaluate (4.1). In statistical applications we only have a finite part of the configuration available and for that reason the finite range models will be of main interest. Most likely the results given in this paper for a finite range potential can be generalized to the infinite range case using ideas similar to those of Gidas (1988), but we do not pursue this here.

In the following we consider the finite range case in detail and obtain distributional results under the infinite volume Gibbs measure μ. Thus $\varphi(z) = 0$ for $|z| > \kappa$ and let Λ_i be defined as in section 3. In order not to overburden the notation we take $I(n) = [-(n+1), n+1]^d \subset \mathbb{Z}^d$ and $\mathring{I}(n) = [-n, n]^d \subset \mathbb{Z}^d$, although more general sets can be treated in the same way. We observe the process in $\Lambda(n) = \bigcup_{i \in I(n)} \Lambda_i$ and consider estimation from $pl_{\mathring{\Lambda}(n)}$, where $\mathring{\Lambda}(n) = \bigcup_{i \in \mathring{I}(n)} \Lambda_i$. It is clear from (4.1) that $pl_{\mathring{\Lambda}(n)} = \Sigma_{i \in \mathring{I}(n)}\, pl_{\Lambda_i}$ and so the first and second derivatives of $pl_{\mathring{\Lambda}(n)}$ can be written as a sum over $\mathring{I}(n)$,

$$U_n(\omega, \beta) = \frac{\partial pl_{\mathring{\Lambda}(n)}}{\partial(\omega, \beta)} = \sum_{i \in \mathring{I}(n)} \left(|x_{\Lambda_i}| - e^{\omega} f_0 \circ \theta_i,\ - \sum_{z \in x_{\Lambda_i}} v(x\backslash z, z) + e^{\omega} f_1 \circ \theta_i \right)$$

$$J_n(\omega, \beta) = \frac{-\partial^2 pl_{\mathring{\Lambda}(n)}}{\partial(\omega, \beta)^* \partial(\omega, \beta)} = \sum_{i \in \mathring{I}(n)} e^{\omega} \begin{pmatrix} f_0 \circ \theta_i & -f_1 \circ \theta_i \\ -f_1 \circ \theta_i & f_2 \circ \theta_i \end{pmatrix},$$

where

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

Let $(\hat{\omega}, \hat{\beta})$ be a maximum pseudo likelihood estimate, i.e. $U_n(\hat{\omega}, \hat{\beta}) = 0$, when a solution exists, otherwise $(\hat{\omega}, \hat{\beta})$ is arbitrarily defined. If we write $U_n = \Sigma_{\mathring{I}(n)}\, (g_0 \circ \theta_i, g_1 \circ \theta_i)$ we define

$$G = \sum_{i \in \mathbb{Z}^d} \begin{pmatrix} \operatorname{cov}_\mu(g_0, g_0 \circ \theta_i) & \operatorname{cov}_\mu(g_0, g_1 \circ \theta_i) \\ \operatorname{cov}_\mu(g_0, g_1 \circ \theta_i) & \operatorname{cov}_\mu(g_1, g_1 \circ \theta_i) \end{pmatrix} \quad \text{and} \quad H = \begin{pmatrix} \mu(f_0) & -\mu(f_1) \\ -\mu(f_1) & \mu(f_2) \end{pmatrix}$$

Theorem 4.1

Let φ be a positive potential of finite range κ. If $\zeta = \exp(\omega) > 0$ and $\beta > 0$ satisfies $\Sigma \tilde{C}_k < 1$, then as $n \to \infty$

$$(2n+1)^{-d/2}(\hat{\omega} - \omega, \hat{\beta} - \beta) \xrightarrow{\mathcal{D}} N_2(0, H^{-1}GH^{-1}/\zeta^2).$$

Proof. We must prove that (i) $(2n+1)^{-d/2} U_n(\omega, \beta) \xrightarrow{\mathcal{D}} N_2(0, G)$ and (ii) that $(2n+1)^{-d} J_n(\tilde{\omega}, \tilde{\beta}) \to e^{\omega} H$ for $(\tilde{\omega}, \tilde{\beta})$ converging to (ω, β).

For part (i) we use theorem 2.3. Since $\tilde{C}_k = 0$ unless $|k_i| \leqslant 1$, $i = 1, \ldots, d$, we have $\tilde{D}_k \leqslant c_1 |k|^{-2d-\lambda}$ for any λ. We therefore only have to prove that $\mu(|g_0|^{2+\alpha}) < \infty$ and $\mu(|g_1|^{2+\alpha}) < \infty$ for some $\alpha > 0$. Since $v(x, \xi) \geqslant 0$ we have $0 \leqslant f_k \leqslant a_k$ for some constant a_k.

From the estimate $d\mu_0(v_0|x) \leqslant \exp(\zeta|\Lambda_0|)\, dv_0^\zeta(v_0)$, we now immediately find that $\mu(|g_0|^l) < \infty$ for any l. To study g_1 we note $0 \leqslant \Sigma_{z \in x_{\Lambda_0}} v(x\backslash z, z) \leqslant 2V_0(x_0|\check{x}_0)$ and

$$\mu(V_0(x_0|\check{x}_0)^l) \leqslant v_0^\zeta\left(e^{\zeta|\Lambda_0|}\left(\frac{l}{\beta}\right)^l e^{-l}\right) < \infty$$

for any l.

For part (ii) we first note that since $0 \leqslant f_k \leqslant a_k$ we have from theorem 2.3 that $\operatorname{var}\{(2n+1)^{-d/2}\Sigma_{I(n)} f_k \circ \theta_i\}$ converges to a finite value. By Markov's inequality we obtain then that $(2n+1)^{-d}\Sigma_{I(n)} f_k \circ \theta_i$ converges in probability to $\mu(f_k)$. We now indicate the dependency of β by f_k^β. We want to show that the convergence to $\mu(f_k)$ is uniform for $|\tilde{\beta} - \beta| < c(2n+1)^{-d/2}$ for any $c > 0$. We use the estimate

$$|f_k^{\tilde{\beta}} - f_k^\beta| \leqslant |\tilde{\beta} - \beta| f_{k+1}^{1/2\beta}, \tag{4.2}$$

which holds when $|\tilde{\beta} - \beta| < \frac{1}{2}\beta$ because $v(x, \xi) \geqslant 0$. Thus

$$\left|(2n+1)^{-d}\sum_{I(n)}(f_k^{\tilde{\beta}} - f_k^\beta)\circ\theta_i\right| < c(2n+1)^{-d/2}\left\{(2n+1)^{-d}\sum_{I(n)} f_{k+1}^{1/2\beta}\circ\theta_i\right\},$$

where the term in the parenthesis, as above, converges in probability. Part (ii) has then been proved and the result of the theorem follows in a standard fashion, see e.g. Sweeting (1980). □

Theorem 4.2

Let φ be a hard core potential, i.e. $\varphi(z) = \infty$ for $|z| \leqslant h$, of finite range κ and with $M = -\inf_z \varphi(z) < \infty$. If $\zeta = \exp(\omega) > 0$ and $\beta > 0$ satisfies $\Sigma\tilde{C}_k < 1$, then as $n \to \infty$

$$(2n+1)^{-d/2}(\hat{\omega} - \omega, \hat{\beta} - \beta) \xrightarrow{\mathcal{D}} N_2(0, H^{-1}GH^{-1}/\zeta^2).$$

Proof. As in part (i) of the proof of theorem 4.1, we only have to show that $\mu(|g_0|^{2+\alpha}) < \infty$ and $\mu(|g_1|^{2+\alpha}) < \infty$ for some $\alpha > 0$. As explained in example 3.2 we can restrict attention to the spaces $S_N(\Lambda_i)$ with N sufficiently large. Since the finite range assumption gives $v(x, \xi) \geqslant -3^d NM$, we find that $|f_k| \leqslant a_k$ for some constant a_k. From the calculations in example 3.2, we have that $d\mu_0(v_0|y) \leqslant \exp\{\zeta|\Lambda_0| + B_1|v_0|\}\, dv_0^\zeta(v_0)$, and therefore $\mu(|g_0|^l) < \infty$ for any l. To study the moments of g_1 we write $\varphi = \varphi^+ - \varphi^-$, where φ^+ and φ^- are the positive and negative parts, respectively. Similarly we then have $v(x, \xi) = v^+(x, \xi) - v^-(x, \xi)$ and $V_0(v_0|\check{x}_0) = V_0^+(v_0|\check{x}_0) - V_0^-(v_0|\check{x}_0)$. We then use

$$\sum_{z \in x_{\Lambda_0}} v(x\backslash z, z) = \sum_{z \in x_{\Lambda_0}} v^+(x\backslash z, z) - \sum_{z \in x_{\Lambda_0}} v^-(x\backslash z, z).$$

Here the last term is bounded by $3^d N^2 M$ and the first term is less than $2V_0^+(x_0|\check{x}_0)$. Since also $V_0^-(x_0|\check{x}_0) \leqslant 3^d N^2 M$, we find

$$\mu\{V_0^+(x_0|\check{x}_0)^l\} \leqslant v_0^\zeta\left\{e^{\zeta|\Lambda_0| + 3^d N^2 M}\left(\frac{l}{\beta}\right)^l e^{-l}\right\} < \infty$$

for any l.

For part (ii) we proceed exactly as in the proof of theorem 4.1. Instead of (4.2) we write

$$\begin{aligned}|f_k^{\tilde{\beta}} - f_k^\beta| &\leqslant |\tilde{\beta} - \beta| \int_{\Lambda_0} |v(x, \xi)|^{k+1} \exp\left\{-\frac{1}{2}\beta v^+(x, \xi) + \frac{3}{2}\beta v^-(x, \xi)\right\} d\xi \\ &\leqslant |\tilde{\beta} - \beta|\tilde{a}_k\end{aligned}$$

for some constant $\tilde{a}_k$. The proof is then as before. □

5. Conditional models and asymptotic normality of maximum likelihood estimates

In this section we consider the asymptotic properties of the maximum likelihood estimates obtained from the conditional model (3.2). If $y_{\bar{\Lambda}} = \varnothing$ the asymptotic distribution under $\mu_\Lambda(\cdot\,|\varnothing)$ can be obtained from theorem 2.3, but as mentioned at the end of section 3 this kind of result can also be obtained from Jensen (1991a). Here we consider a general value of $y_{\bar{\Lambda}}$ and obtain asymptotic normality of the conditional maximum likelihood estimates under the infinite volume Gibbs measure μ via theorem 2.3.

We use the set up of section 4 for the finite range case and observe the process in $\bigcup_{I(n)} \Lambda_i$, where now $I(n) = [-(n+1), n+1]^d \subset \mathbb{Z}^d$. We then use the conditional likelihood function obtained from (3.2) with $\Lambda = \bigcup_{I(n)} \Lambda_i$. The first and second derivatives of the log likelihood function p_n^c become with $\kappa(\omega, \beta | x_{\bar{\Lambda}}) = \log Z(e^\omega, \beta | x_{\bar{\Lambda}})$,

$$U_n^c(\omega, \beta) = \frac{\partial p_n^c}{\partial(\omega, \beta)} = \left\{ |x_\Lambda| - \frac{\partial \kappa}{\partial \omega}, -V_\Lambda(x_\Lambda | x_{\bar{\Lambda}}) - \frac{\partial \kappa}{\partial \beta} \right\},$$

and

$$J_n^c(\omega, \beta) = -\frac{\partial^2 p_n^c}{\partial(\omega, \beta)^* \partial(\omega, \beta)} = \begin{pmatrix} \dfrac{\partial^2 \kappa}{\partial \omega^2} & \dfrac{\partial^2 \kappa}{\partial \omega \partial \beta} \\ \dfrac{\partial^2 \kappa}{\partial \omega \partial \beta} & \dfrac{\partial^2 \kappa}{\partial \beta^2} \end{pmatrix} = \mathrm{var}_\mu\,(U_n^c | x_{\bar{\Lambda}}).$$

Let $(\hat{\omega}^c, \hat{\beta}^c)$ be the conditional maximum likelihood estimates, $U_n^c(\hat{\omega}^c, \hat{\beta}^c) = 0$.

To study the convergence properties of $J_n^c(\tilde{\omega}, \tilde{\beta})$ we need to look at the derivative of J_n^c, which becomes the third order cumulants in the conditional distribution $\mu_\Lambda(\cdot\,|x_{\bar{\Lambda}})$. We therefore interject the following lemma. Let X_i, $i \in \mathbb{Z}^d$, be a random field with the following mixing property

$$|P(A_1 \cap A_2) - P(A_1)P(A_2)| \leqslant c_1 |I_1| |I_2| \{d(I_1, I_2)\}^{-\lambda}$$

for some $\lambda > 8d + 4$, where $A_j = \sigma(X_i, i \in I_j)$. Let Y_i be a vector function of X_i such that

$$E \| Y_i \|^4 \leqslant \beta_4 \quad \text{for all } i,$$

where we take $\beta_4 \geqslant 1$.

Lemma 5.1
There exists a constant c_2, dependent on c_1 and λ only, such that all third order cumulants of $\Sigma_{I(n)} Y_i$ are bounded by

$$c_2(2n+1)^d \beta_4^2.$$

Proof. This kind of bound goes back to Bulinski & Zhurbenko (1976). The slightly different dependence on the moments is due to a truncation technique in Götze & Hipp (1983). The details can be found in Jensen (1986). □

Theorem 5.2
Assume either the conditions of theorem 4.1 *or theorem* 4.2. *Furthermore, if the eigenvalues of $J_n^c(\omega, \beta)/(2n+1)^d$ are bounded away from zero we have under the infinite volume Gibbs measure μ*

$$(\hat{\omega}^c - \omega, \hat{\beta}^c - \beta)\ \mathrm{var}_\mu\,(U_n^c | x_{\bar{\Lambda}})^{1/2} \xrightarrow{\mathcal{D}} N_2(0, I).$$

Proof. First we express U_n^c as a sum. Obviously $|x_\Lambda| = \Sigma_{\tilde{I}(n)} |x_{\Lambda_i}|$. Define

$$\tilde{V}_i = \sum_{\{z_1, z_2\} \subset x_{\Lambda_i}} \varphi(z_1 - z_2) + \frac{1}{2} \sum_{\substack{z_1 \in x_{\Lambda_i} \\ z_2 \in x_{\Lambda \setminus \Lambda_i}}} \varphi(z_1 - z_2) + \sum_{\substack{z_1 \in x_{\Lambda_i} \\ z_2 \in x_{\bar{\Lambda}}}} \varphi(z_1 - z_2),$$

then $V_\Lambda(x_\Lambda | x_{\bar{\Lambda}}) = \Sigma_{\tilde{I}(n)} \tilde{V}_1$. Using theorem 2.3 and proceeding as in the proofs of theorems 4.1 and 4.2, we get asymptotic normality of U_n in the conditional model $\mu_\Lambda(\,\cdot\, | x_{\bar{\Lambda}})$.

We must then show that $J_n^c(\tilde{\omega}, \tilde{\beta})/(2n+1)^d$ is close to $J_n^c(\omega, \beta)/(2n+1)^d$ for $(\tilde{\omega}, \tilde{\beta})$ close to (α, β). Since the dependency coefficients C_{ik} are continuous functions of (ω, β), we can use the same mixing estimate (2.5) for all the conditional models $\mu_\Lambda(\,\cdot\, | x_{\bar{\Lambda}}; \tilde{\omega}, \tilde{\beta})$ with $|(\tilde{\omega} - \omega, \tilde{\beta} - \beta)| < \varepsilon$, say. Then lemma 5.1 shows that the difference between $J_n^c(\tilde{\omega}, \tilde{\beta})/(2n+1)^d$ and $J_n^c(\omega, \beta)/(2n+1)^d$ is of order $|(\tilde{\omega} - \omega, \tilde{\beta} - \beta)|$, as the derivatives of $J_n^c(\omega, \beta)$ with respect to ω and β are third order cumulant in the conditional model. Standard arguments now give the result of the theorem under the set of conditional models $\mu_\Lambda(\,\cdot\, | x_{\bar{\Lambda}})$. This in turn implies the same result under the measure μ. □

When the parameter β is of primary interest with ζ a nuisance parameter it is customary to consider the conditional distribution given the number of points $|X_{I(n)}|$. However, it is not clear how to use the methods of this paper for this conditional distribution, i.e. it is not clear how to derive an equivalent statement to the mixing statement (2.5). Instead perhaps one can improve the results here to obtain local limit results, and then approximate the conditional distribution by approximating both the joint density and the marginal density. For an example along these lines see Jensen (1991b).

Acknowledgements

It is a pleasure to thank Hans Künsch for allowing me to see his note (Künsch, 1982a) and for pointing out the work of Klein. The careful reading of two referees is also acknowledged.

References

Baddeley, A. J. & Moyeed, R. A. (1991). Stochastic approximation of the MLE for a spatial point pattern. *Scand J. Statist.* **18,** 39–50.

Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician* **24,** 179–195.

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

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

Bulinski, A. V. & Zhurbenko, I. G. (1976). The central limit theorem for random fields. *Soviet Math. Dokl.* **17,** 14–17.

Föllmer, H. (1982). A covariance estimate for Gibbs measures. *J. Funct. Anal.* **46,** 387–395.

Gidas, B. (1988). Consistency of maximum likelihood and pseudo-likelihood estimators for Gibbs distributions. In *Stochastic differential systems, stochastic control theory and applications* (eds W. Fleming & P. L. Liars), 129–145. Springer, New York.

Götze, F. & Hipp, C. (1983). Asymptotic expansions for sums of weakly dependent random vectors. *Z. Wahrscheinlichkeits theor. Verw. Geb.* **64,** 211–240.

Guyon, X. (1987). Estimation d'un champ par pseudo-vraisemblance conditionelle: étude asymptotique et application au cas Markovieu. In *Spatial processes and spatial time series analysis* (ed. F. Droesbeke). *Proceedings of the 6th Franco–Belgian meeting of statisticians*, November 1985. Universitaires Saint-Louis, Bruxelles.

Jensen, J. L. (1986). A note on the work of Götze and Hipp concerning asymptotic expansions for sums of weakly dependent random vectors. Memoirs No. 10, Department of Theoretical Statistics, University of Aarhus.

Jensen, J. L. (1991a). A note on asymptotic normality in the thermodynamic limit at low densities. *Adv. Appl. Math.* **12,** 387–399.

Jensen, J. L. (1991b). A note on asymptotic expansions for sums over a weakly dependent random field.

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

Klein, D. (1982). Dobrushin uniqueness techniques and the decay of correlations in continuum statistical mechanics. *Commun. Math. Phys.* **86,** 227–246.

Klein, D. (1984). Convergence of grand canonical Gibbs measures. *Commun. Math. Phys.* **92,** 295–308.

Künsch, H. (1982a). The central limit theorem for Gibbs fields with weak interaction. Unpublished note.

Künsch, H. (1982b). Decay of correlation under Dobrushin's uniqueness condition and its applications. *Commun. Math. Phys.* **84,** 207–222.

Ogata, Y. & Tanemura, M. (1984). Likelihood analysis of spatial point patterns. *J. Roy. Statist. Soc. Ser.* B **46,** 496–518.

Penttinen, A. (1984). Modelling interactions in spatial point patterns: parameter estimation by the maximum likelihood method. *Jyväskylä Studies in Computer Science, Economics and Statistics 7.*

Preston, C. (1976). *Random fields.* Lecture Notes in Mathematics 534. Springer, Berlin.

Ruelle, D. (1969). *Statistical Mechanics. Rigorous Results.* W. A. Benjamin, Amsterdam.

Simon, B. (1979). A remark on Dobrushin's uniqueness theorem. *Commun. Math. Phys.* **68,** 183–185.

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 Mathematical J.* **31,** 67–77.

Received June 1991, in final form February 1992

Jens Ledet Jensen, Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus, DK-8000 Aarhus C, Denmark.