CENTRE FOR **STOCHASTIC GEOMETRY**
AND ADVANCED **BIOIMAGING**

Michaela Prokešová and Eva B. Vedel Jensen

Asymptotic Palm Likelihood Theory for Stationary Point Processes

No. 11, October 2010

Asymptotic Palm likelihood theory for stationary point processes

By Michaela Prokešová* and Eva B. Vedel Jensen†

*Charles University** *and Aarhus University*†

Abstract

In the present paper, we propose a Palm likelihood approach as a general estimating principle for stationary point processes in R^d for which the density of the second-order factorial moment measure is available in closed form. Examples of such point processes include the Neyman-Scott processes and the log Gaussian Cox processes. The computations involved in determining the Palm likelihood estimator are simple. Conditions are provided under which the Palm likelihood estimator is consistent and asymptotically normally distributed.

AMS 200 subject classifications: Primary 62M40, 62F12

Keywords and phrases: asymptotic normality, cluster processes, consistency, log Gaussian Cox processes, Neyman-Scott processes, Palm likelihood, spatial point process, strong mixing

1 Introduction

Estimation of parametric models for spatial point processes has been a very active research area in the last few years. Motivated by the need of analyzing always larger and more complicated data sets in reasonably short time, several simulation-free estimation methods based on composite likelihood and/or estimating equations have been developed as alternatives to the computationally more demanding likelihood and Bayesian methods, see [20] for a recent overview.

In the present paper we will focus on point process models for which the densities of the first and second order moment measures (and/or quantities derived from them – like the pair-correlation function g or the K-function) are available in closed form. In particular, we will consider stationary cluster processes, Cox processes and related models.

One of the estimation methods that was first suggested for such processes is the minimum contrast method based either on the K-function or the g-function, see [6], [18], [19] and references therein. In addition to stationarity, this method requires that the point process X is isotropic. A parameter θ is estimated by minimizing the discrepancy measure

$$\int_0^R [\hat{K}^c(u) - K^c(u;\theta)]^2 du \qquad \text{or} \qquad \int_0^R [\hat{g}^c(u) - g^c(u;\theta)]^2 du$$

between the estimate $\hat{K}$ or $\hat{g}$ and its theoretical value $K(\cdot;\theta)$ or $g(\cdot;\theta)$, respectively. The user specified constants c and R are used to control the sampling fluctuations in the estimates of K and g. These constants are usually chosen in some adhoc manner. Asymptotic properties of these estimates have been derived in [9] and [14].

In [1], [8] and [20], different variants of a composite likelihood approach have been suggested. They are all based on adding together individual log-likelihoods for single points or pairs of points of the point process to form a composite log-likelihood ([17]). In [1], the focus is on Gibbs point processes with exponential family likelihoods. The Papangelou conditional intensity is here used in combination with a Poisson likelihood. In [8], the density $\lambda^{(2)}$ of the second-order factorial moment measure of the point process X observed on a window W are used to form partial likelihood functions for pairs of points

$$\frac{\lambda^{(2)}(x,y;\theta)}{\iint_{W\times W}\lambda^{(2)}(u,v;\theta)dudv}, \tag{1.1}$$

yielding the following (log-)composite likelihood

$$\log CL(\theta) = \sum_{\substack{x\neq y\in X\cap W\\ |x-y|<R}} \log \frac{\lambda^{(2)}(x,y;\theta)}{\iint_{W\times W} I(|u-v|<R)\lambda^{(2)}(u,v;\theta)dudv}.$$

Here only pairs of points with distance at most R are considered and R is again a tuning constant. The estimate of θ is obtained as the maximizer of $\log CL(\theta)$. Note that since all the partial likelihoods (1.1) are normalized, $CL(\theta)$ does not depend on the intensity λ of the point process X. Thus the method effectively enables to estimate the part of the vector parameter θ which parametrizes the pair-correlation function $g(x,y) = \frac{\lambda^{(2)}(x,y)}{\lambda^2}$. Asymptotic properties of the CL-estimator have been studied in [8]. Yet, another type of composite likelihood was introduced in [20, (48)].

Very recently, an estimating procedure based on maximization of the so-called Palm likelihood has been introduced in [26] for stationary cluster processes, including Neyman-Scott processes and related models. Again the process of the differences between pairs of points of X observed in a window W is considered. In contrast to the procedure in [8] which is directly based on a normalized version of $\lambda^{(2)}$ and does not involve any point process theory, an analysis of the "difference process" is performed in [26] where it is shown that the difference process is inhomogeneous with intensity determined by the density λ_0 of the second-order reduced factorial moment measure of the process X. For further details, see Section 3 below. Since λ_0 is equal to the intensity function of the Palm distribution of the original process X, λ_0 was called the Palm intensity in the paper [26]. By approximating the process of differences by a nonstationary Poisson process with intensity proportional to λ_0, the authors arrive at the so-called Palm (log-)likelihood function defined by

$$\begin{aligned}\log L_P(\theta) = &\sum_{\substack{x\neq y\in X\cap W\\ |x-y|<R}} \log\left(|X\cap W|\lambda_0(x-y;\theta)\right)\\ &- |X\cap W|\int_{\mathbf{R}^d} I(|u|<R)\lambda_0(u;\theta)du.\end{aligned} \tag{1.2}$$

Here, $|X \cap W|$ is the number of points in $X \cap W$ and R is again a tuning constant. The estimate of θ is obtained as the maximizer of $\log L_P(\theta)$.

In the original paper [26], it was assumed that the point process X is an isotropic point process in $\mathbf{R}^2$. The defining formula (1.2) of the Palm likelihood can, however, be applied to any simple stationary point process in $\mathbf{R}^d$ with a Palm intensity $\lambda_0(\cdot;\theta)$ in closed form. Note also that the Palm likelihood estimation method belongs to the second order moment estimation methods, since

$$\lambda^{(2)}(x,y) = \lambda\lambda_0(y-x), \qquad x,y \in \mathbf{R}^d.$$

The Palm likelihood estimation method is closely related to the composite likelihood method suggested in [20]. In fact, if we in [20, (48)] replace λ by $|X \cap W|/|W|$ we arrive at the estimating equation resulting from (1.2). When we compare the Palm likelihood method to the composite likelihood method in [8] we see that the Palm likelihood method is expected to be numerically simpler because of the simpler form of the normalization term – see Section 6 for an example.

Concerning the asymptotic properties it was argued in [26] that, for the considered cluster point processes, the process of differences

$$\{x - y : x \neq y \in X \cap W\}$$

is well approximated by a nonstationary Poisson point process with intensity $|X \cap W|\,\lambda_0$ because the process of differences can be regarded as a superposition of $|X \cap W|$ (almost independent) realizations of a point process with the distribution equal to the Palm distribution of the original process X – that means with the intensity λ_0. By a superposition theorem for $|X \cap W| \to \infty$ a convergence of the suitably normalized difference process to a Poisson process can be obtained (see [21] for the argument) which would imply the consistency of the obtained Palm likelihood estimates. A formal proof of consistency was, however, not provided.

The present paper fills this gap. We provide a proof of consistency and asymptotic normality of the maximum Palm likelihood estimator. Consistency is proved under the assumption of ergodicity of the point process X. The proof of asymptotic normality is provided under the additional assumptions that the process X is strongly mixing and the strong mixing coefficient decays sufficiently fast. Moreover for cluster processes we derive some simple methods of checking sufficient conditions for the desired fast decay of the strong mixing coefficients.

The paper is organized as follows. We give the necessary notation and background information in Section 2 and introduce the Palm likelihood estimation procedure in detail in Section 3. In Sections 4 and 5, the main results of the paper are presented - the consistency and the asymptotic normality of the Palm likelihood estimator. The obtained results are further exemplified in Section 6 and discussed in Section 7. Proofs are deferred to an appendix.

2 Background

Let X denote a simple strictly stationary point process on $\mathbf{R}^d$. In the sequel, $\mathcal{B}^d$ is the Borel σ-algebra on $\mathbf{R}^d$, $|A|$ is the volume of the set $A \in \mathcal{B}^d$, ∂A its boundary

and $|\partial A|$ the $(d-1)$-dimensional surface measure of ∂A, when it exists. The origin in $\mathbf{R}^d$ will be denoted by o, $B(x,R)$ is the ball centered at $x \in \mathbf{R}^d$ with radius $R > 0$ and $\oplus$, $\ominus$ will denote Minkowski addition and substraction, respectively, with the convention that $A \oplus R = A \oplus B(o,R)$ and $A \ominus R = A \ominus B(o,R)$ for $R > 0$. The Euclidean norm of the vector x is denoted by $|x|$, for matrices we use also the Euclidean norm $|M| = (\mathrm{trace}\,(M^T M))^{\frac{1}{2}}$ and I is the indicator function.

The k-th order factorial moment measure $\alpha^{(k)}$ of the point process X is defined by the following equation

$$\int_{(\mathbf{R}^d)^k} f(u_1,\dots,u_k)\,\alpha^{(k)}(d(u_1,\dots,u_k)) = E\Big(\sum_{u_1,\dots,u_k\in X}^{\neq} f(u_1,\dots,u_k)\Big) \tag{2.1}$$

for any non–negative, Borel measurable function f on $(\mathbf{R}^d)^k$, where the sum $\sum^{\neq}$ runs over k–tuples of distinct points of X. The kth-order factorial cumulant measure $\gamma^{(k)}$ of X is a locally finite signed measure on $[(\mathbf{R}^d)^k, \mathcal{B}^{dk}]$ which is formally connected with the measures $\alpha^{(1)},\dots,\alpha^{(k)}$ by

$$\gamma^{(k)}(\times_{i=1}^k A_i) = \sum_{j=1}^k (-1)^{j-1}(j-1)! \sum_{K_1\cup\cdots\cup K_j=\{1,\dots,k\}} \prod_{i=1}^j \alpha^{(\#K_i)}\Big(\bigtimes_{k_i\in K_i} A_{k_i}\Big)$$

for bounded $A_1,\dots,A_k \in \mathcal{B}^d$, where the inner sum is taken over all partitions of the set $\{1,\dots,k\}$ in disjoint non–empty subsets $K_1,\dots,K_j$. In particular, $\alpha^{(1)}(A) = \gamma^{(1)}(A) = \lambda\,|A| = E|X\cap A|$ for $A \in \mathcal{B}^d$ where λ is called the intensity of X. For $k \geq 2$, we will assume in the sequel that the factorial moment measures have densities with respect to the Lebesgue measure on $\mathbf{R}^{dk}$ which are denoted by $\lambda^{(k)}$ and called the k-th order product densities of X (or sometimes k-th order intensity functions of X).

Since, for any $k \geq 2$, $\alpha^{(k)}$ is invariant under diagonal shifts there exists a corresponding reduced kth-order factorial moment measure $\alpha^{(k)}_{\mathrm{red}}$ on $[(\mathbf{R}^d)^{k-1}, \mathcal{B}^{d(k-1)}]$ which is uniquely determined by the disintegration formula

$$\begin{aligned}&\int_{(\mathbf{R}^d)^k} f(u_1,\dots,u_k)\alpha^{(k)}(d(u_1,\dots,u_k)) \\ &\quad = \lambda \int_{\mathbf{R}^d}\int_{(\mathbf{R}^d)^{k-1}} f(u_1,u_2+u_1,\dots,u_k+u_1)\,\alpha^{(k)}_{\mathrm{red}}(d(u_2,\dots,u_k))\,du_1,\end{aligned} \tag{2.2}$$

where f is as in (2.1). Similarly we may define the reduced kth-order factorial cumulant measure $\gamma^{(k)}_{\mathrm{red}}$ which turns out to be a signed measure on $[(\mathbf{R}^d)^{k-1}, \mathcal{B}^{d(k-1)}]$.

For $k = 2$, the disintegration (2.2) implies that

$$\lambda^{(2)}(x,y) = \lambda\lambda_0(y-x), \qquad x,y \in \mathbf{R}^d, \tag{2.3}$$

where the function λ_0 is the density of $\alpha^{(2)}_{\mathrm{red}}$. The function λ_0 is also called the conditional intensity or Palm intensity in the literature since λ_0 is in fact the intensity function of the Palm distribution $\mathcal{P}_0$ of the original point process X. For a detailed introduction to these notions and their properties, we refer the reader to [4].

Two popular point process characteristics can be defined from λ_0, viz. the pair correlation function

$$g(x,y) = g(y-x) = \lambda_0(y-x)/\lambda, \quad x,y \in \mathbf{R}^d, \tag{2.4}$$

and the K-function defined by

$$\lambda\, K(R) = \int_{B(o,R)} \lambda_0(u)du = E\,[|X \cap B(o,R)\backslash\{o\}|\,|\,X \cap \{o\} \neq \emptyset\,]. \tag{2.5}$$

Note that $\lambda\, K(R)$ can be interpreted as the mean number of further points from X in $B(x,R)$ centered at a typical point x of the point process X.

The Neyman-Scott process considered in [26] can be constructed as follows. Let C be a stationary Poisson point process of intensity μ. This process is called the mother process. Each mother point $c \in C$ produces a random number M of daughter points with mean ν. The daughters around c are i.i.d. with density $k(c,\cdot) = h(\cdot - c)$. The set of daughters associated with the mother c is denoted X_c. The Neyman-Scott process is then the union of the daughter clusters $X = \cup_{c\in C} X_c$. The intensity of X is $\mu\nu$. The Palm intensity of the Neyman-Scott process becomes

$$\lambda_0(z) = \mu\nu + \nu \int_{\mathbf{R}^d} h(u)h(z+u)du, \quad z \in \mathbf{R}^d. \tag{2.6}$$

If h is a Gaussian density, then (2.6) is in closed form. Other examples are given in [26].

Another class of point processes for which the Palm intensity can be obtained in closed form is the log Gaussian Cox processes ([18]). Here,

$$\lambda_0(z) = \exp(m + c(0)/2 + c(z)), \quad z \in \mathbf{R}^d,$$

where m and c are the mean and covariance function of the underlying stationary Gaussian random field, respectively.

3 Palm likelihood for stationary point processes

Let us assume that the parameter of interest of our point process model is the (vector) parameter θ and that the Palm intensity $\lambda_0(\cdot;\theta)$ is parametrized by θ. In the following, we will suppress θ in the notation if the dependence on θ is not important in the respective context.

In this section, we will review the Palm likelihood method that was introduced in [26] for stationary isotropic Neyman-Scott processes and related models. In [26], the Palm log-likelihood function was defined by

$$\begin{aligned}\log L_P(\theta) = &\sum_{\substack{x\neq y \in X\cap W \\ |x-y|<R}} \log\,(|X\cap W|\lambda_0(x-y;\theta)) \\ &- |X\cap W| \int_{\mathbf{R}^d} I(|u|<R)\lambda_0(u;\theta)du,\end{aligned} \tag{3.1}$$

where R is a chosen positive constant. The maximum Palm likelihood (MPL) estimator is obtained by maximizing $L_P(\theta)$. In the remaining part of the paper, we will consider the equivalent version of $L_P(\theta)$

$$\log L_P(\theta) = \sum_{\substack{x \neq y \in X \cap W \\ |x-y|<R}} \log\, \lambda_0(x-y;\theta) - |X \cap W| \int_{\mathbf{R}^d} I(|u|<R)\lambda_0(u;\theta)du. \tag{3.2}$$

Under the assumption that $L_P(\theta)$ is differentiable with respect to θ, the MPL-estimate is the solution to the following estimation equation

$$\frac{\mathrm{d}\log L_P(\theta)}{\mathrm{d}\theta} = 0. \tag{3.3}$$

The idea behind this estimation procedure is to use, instead of the original process X observed in the window W, the process of differences $Y = \{x-y:\ x \neq y \in X \cap W\}$ and note that the data used in the Palm likelihood are really only the difference process Y and the number of observed points $|X \cap W|$. If we for any fixed $x \in X$ let

$$Y_x = \{y-x : x \neq y \in X\}, \tag{3.4}$$

then the Palm log-likelihood in (3.2) is a sum (over $x \in X \cap W$) of Poisson log-likelihoods for the processes $Y_x \cap B(o,R)$, all assumed to have intensity function λ_0. Using the Poisson log-likelihoods implies that the higher order interactions of the processes of differences are ignored. Furthermore, since the Poisson log-likelihoods are summed, the dependence among Y_x, $x \in X \cap W$, are ignored by treating them as independent replications.

An alternative way of arriving at the Palm log-likelihood (3.1) goes as follows. Let

$$Y(R) = \{y-x : x \neq y \in X \cap W, |y-x|<R\}.$$

Evidently, $Y(R)$ is a point process contained in $B(o,R)$. The intensity function of this point process can be derived as follows. Let A be a Borel subset of $B(o,R)$. Then,

$$E(|Y(R) \cap A|) = \int_W \int_W I(y-x \in A)\lambda\lambda_0(y-x)dxdy = \int_A \gamma_W(u)\lambda\lambda_0(u)du,$$

where $\gamma_W(u) = |W \cap (W+u)|$ is the set covariance of the window W, see [25] for further details. The point process $Y(R)$ has thus an intensity function concentrated on $B(o,R)$ of the form

$$\lambda_R(u) = \gamma_W(u)\lambda\lambda_0(u), \quad u \in B(o,R).$$

The Palm log-likelihood (3.1) can now be obtained by treating $Y(R)$ as an inhomogeneous Poisson process, replacing the intensity λ of the original point process X by the observed intensity $|X \cap W|/|W|$ and approximating $\gamma_W(u)$, $u \in B(o,R)$, by $|W|$. This is a reasonable approximation for R substantially smaller than the size of the observation window W.

We can define a modified version of the Palm likelihood in which we consider only those points $x \in X$ for which $B(x,R) \subseteq W$, thus employing the inner region edge correction

$$\log L_{PU}(\theta) = \sum_{\substack{x \in X \cap (W \ominus R) \\ x \neq y \in X \\ (y-x) \in B(o,R)}} \log \lambda_0(y-x;\theta) - |X \cap (W \ominus R)| \int_{B(o,R)} \lambda_0(u;\theta) du. \tag{3.5}$$

Then the estimating (vector) equation $\frac{\mathrm{d} \log L_{PU}}{\mathrm{d}\theta} = 0$ is an unbiased estimating equation since

$$E_{\theta_0} \left(\left. \frac{\mathrm{d} \log L_{PU}(\theta)}{\mathrm{d}\theta} \right|_{\theta=\theta_0} \right) = 0, \tag{3.6}$$

where E_{θ_0} denotes the mean value with respect to the distribution with the correct parameter value θ_0. Since the difference between (3.2) and (3.5) is only in the employed edge correction (inner region or none edge correction) – i.e. only in the way the points $x \in W \backslash (W \ominus R)$ are handled, for windows W big enough with respect to R the difference in the two estimates will be negligible. Under the assumptions on the sequences of observation windows introduced in Section 4 the correctly normalized estimating equation $\frac{1}{|W|} \frac{\mathrm{d} \log L_P(\theta)}{\mathrm{d}\theta} = 0$ will be an asymptotically unbiased estimating equation.

We would like to end this paragraph by stressing that the assumption of isotropy, adopted in the original paper [26] for computational reasons, is not necessary for the formulation or validity of the MPL estimation method. Furthermore, the method is not restricted to Neyman-Scott processes and the other cluster processes discussed in [26]. The method can be used for estimation in any parametric model with a closed form of the Palm intensity $\lambda_0(\cdot;\theta)$. One very important class of such processes not considered in [26] is the log Gaussian Cox processes (see [18]).

4 Consistency of MPLE

We will assume in the sequel that the point process model is parametrized by $\theta \in \Theta$, where $\Theta \subset \mathbf{R}^q$ is compact. The true vector parameter θ_0 is assumed to be an interior point of Θ.

The asymptotics will be studied under an increasing domain setting assuming that we have a convex averaging sequence of windows $\{W_n\}_{n \in \mathbf{N}}$ – i.e. that all the windows W_n are bounded convex sets, $W_n \subseteq W_{n+1}$ for all n and the inradii

$$\rho(W_n) = \sup\{\rho : W_n \text{ contains a ball of radius } \rho\} \tag{4.1}$$

converge to ∞ as $n \to \infty$, see [4, Chapter 10] for further details. We will assume that $|W_n| = \mathcal{O}(\rho(W_n)^d)$. This implies for the convex sets $\{W_n\}$ that $|\partial W_n| = \mathcal{O}(\rho(W_n)^{d-1})$ since according to [28] we have

$$\frac{|\partial W|}{|W|} \leq \frac{d}{\rho(W)},$$

if $W \subset \mathbf{R}^d$ is a convex set.

We will start by showing the consistency of the unbiased version of the log Palm likelihood L_{PU}. Let us write here in detail the score function $U(\theta) = \frac{1}{|W\ominus R|}\frac{\mathrm{d}\log L_{PU}(\theta)}{\mathrm{d}\theta}$ of the Palm likelihood

$$U(\theta) = \frac{1}{|W\ominus R|}\sum_{\substack{x\in X\cap(W\ominus R)\\ x\neq y\in X\\ (y-x)\in B(o,R)}}\frac{\mathrm{d}\lambda_0(y-x;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(y-x;\theta)} - \frac{|X\cap(W\ominus R)|}{|W\ominus R|}\int_{B(o,R)}\frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}du. \tag{4.2}$$

(The score function computed from observations in the window W_n will be denoted by $U_n(\theta)$.) We can compute the mean value for the stationary point process X

$$E_{\theta_0}\Big(\sum_{\substack{x\in X\cap(W\ominus R)\\ x\neq y\in X\\ (y-x)\in B(o,R)}}\frac{\mathrm{d}\lambda_0(y-x;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(y-x;\theta)}\Big) = \int_{W\ominus R}\lambda\int_{B(o,R)}\frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(u;\theta)}\lambda_0(u;\theta_0)dudx, \tag{4.3}$$

where E_{θ_0} denotes the mean value with respect to the distribution of the point process with $\theta=\theta_0$. We see that $E_{\theta_0}U(\theta_0)=0$ for the true parameter value θ_0. Thus, the estimating equation $U_n(\theta)=0$ is indeed unbiased. The Palm likelihood estimate obtained from this equation will be denoted by $\hat{\theta}_n$.

In the theorem below, the result on consistency of $\hat{\theta}_n$ is formulated. The proof of the theorem may be found in the appendix.

Theorem 4.1. *Let X be a stationary ergodic point process observed in a convex averaging sequence $\{W_n\}_{n\in\mathbf{N}}$ of windows for which $|W_n| = \mathcal{O}(\rho(W_n)^d)$ holds. Assume that $|E_{\theta_0}U_n(\theta)| = 0$ only when $\theta=\theta_0$ and that $\frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(u;\theta)}$ and $\frac{\mathrm{d}(\int_{B(o,R)}\lambda_0(u;\theta)du)}{\mathrm{d}\theta}$ are bounded and continuous (with respect to u and θ, and θ respectively). Then $\hat{\theta}_n$ is a strongly consistent estimate of θ_0, i.e. $\hat{\theta}_n\to\theta_0$ P_{θ_0}-a.s.*

Let us now discuss the original Palm log-likelihood (3.2) without any included edge correction. The associated score function $\tilde{U}(\theta) = \frac{1}{|W|}\frac{\mathrm{d}\log L_P(\theta)}{\mathrm{d}\theta}$ takes the form

$$\tilde{U}(\theta) = \frac{1}{|W|}\sum_{\substack{x\neq y\in X\cap W\\ |x-y|<R}}\frac{\mathrm{d}\lambda_0(y-x;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(y-x;\theta)} - \frac{|X\cap W|}{|W|}\int_{B(o,R)}\frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}du. \tag{4.4}$$

Let $\tilde{\theta}_n$ denote the estimate obtained from the estimating equation $\tilde{U}_n(\theta)=0$.

For $\theta=\theta_0$ we have

$$E_{\theta_0}\frac{1}{|W|}\Big(\sum_{\substack{x\neq y\in X\cap W\\ |x-y|<R}}\frac{\mathrm{d}\lambda_0(y-x;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(y-x;\theta)}\Big) = \lambda\int_{B(o,R)}\frac{\gamma_W(u)}{|W|}\frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}\Big|_{\theta=\theta_0}du. \tag{4.5}$$

As a consequence we do not have $E_{\theta_0}\tilde{U}_n(\theta_0) = 0$. Nevertheless since

$$\sup_{z\in B(o,R)} \left| \frac{\gamma_{W_n}(z)}{|W_n|} - 1 \right| \leq \frac{R|\partial W_n|}{|W_n|} \leq \frac{dR}{\rho(W_n)}, \tag{4.6}$$

$\frac{\gamma_{W_n}(u)}{|W_n|}$ converges uniformly to 1 on the compact set $B(o,R)$ and therefore $E_{\theta_0}\tilde{U}_n(\theta_0) \to E_{\theta_0}U_n(\theta_0) = 0$ as $n \to \infty$. Therefore the proof of the consistency of $\tilde{\theta}_n$ is analogous to that of $\hat{\theta}_n$.

Theorem 4.2. *Under the same assumptions as in Theorem 4.1, $\tilde{\theta}_n$ is a strongly consistent estimate of θ_0.*

The proof of Theorem 4.2 may be found in the appendix.

5 Asymptotic normality of MPLE

We will now show the asymptotic normality of the MPLE under the assumption that the point process X is strongly mixing. Recall that for two σ-algebras $\mathcal{F}_1$, $\mathcal{F}_2$ defined on the same probability space the strong mixing coefficient is defined by

$$\alpha(\mathcal{F}_1, \mathcal{F}_2) = \sup\{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}. \tag{5.1}$$

For a stationary point process X the strong mixing coefficient $\alpha(p;k)$ quantifies the dependence between the behaviour of the point process on sets of volume at most p separated by a distance larger than or equal to k. Thus for a point process we define

$$\begin{aligned} \alpha(p;k) &= \sup\{\alpha(\mathcal{F}^X(A), \mathcal{F}^X(B)) : d(A,B) \geq k, |A| \leq p, |B| \leq p\}, \quad p > 0 \\ &= \sup\{\alpha(\mathcal{F}^X(A), \mathcal{F}^X(B)) : d(A,B) \geq k, |A| = |B| = 0\}, \qquad p = 0, \end{aligned} \tag{5.2}$$

where $\mathcal{F}^X(A)$ denotes the σ-algebra generated by $X \cap A$ and the supremum is taken over all measurable subsets A, B in $\mathcal{B}^d$.

We will assume that

$$\sup_{p\geq 0} \frac{\alpha(p;k)}{\max(p,1)} = \mathcal{O}(k^{-\epsilon}) \quad \text{for some} \quad \epsilon > d. \tag{5.3}$$

The condition (5.3) is satisfied e.g. by the log Gaussian Cox processes ([18]) if the correlation function of the underlying Gaussian field decays at a polynomial rate faster then $d + \epsilon$ and has a spectral density which is bounded below. This follows from [7, Corollary 2]. One concrete example of such correlation functions often used in practice is the class of Matérn correlation functions (including also the exponential correlation function), see e.g. [24, Section 2.7].

Concerning the Neyman-Scott processes, (5.3) is obviously satisfied for Neyman-Scott processes with a kernel density $k(c,\cdot)$ with bounded support (e.g. the Matérn cluster process). In the following lemma we show that (5.3) is also satisfied if the density has polynomially decaying tails of order $d + \epsilon$.

Lemma 5.1. *Let X be a Neyman-Scott process with mother intensity μ and mean number ν of daughter points in a cluster. Let the daughter points around a mother point in c be distributed according to the kernel density $k(c,\cdot)$. If there exists a function h such that $k(c,x) = h(x-c)$ and $h(v) = \mathcal{O}(|v|^{-\epsilon-d})$ as $|v| \to \infty$, then $\frac{\alpha(p;k)}{\max(p,1)} \leq \mathcal{O}(k^{-\epsilon})$.*

The proof of the lemma can be found in the appendix.

Remark. In the literature on point processes, see e.g. [8], [12], [22], an alternative weaker version of the strong mixing coefficient is sometimes used

$$\alpha(p;k) = \sup\{\alpha(\mathcal{F}^X(A), \mathcal{F}^X(B)) : A = B + x, d(A,B) \geq k, |A| = |B| \leq p\},$$

$p > 0$, where the supremum is taken over all compact, convex sets A and all $x \in \mathbf{R}^d$. This version of the strong mixing coefficient has been inspired by the strong mixing coefficient used in the classical paper by Rosenblat ([23]) from 1956. In the proof of the asymptotic normality below, we follow the methods of [8] (described in detail in [10]), based on the blocking technique presented in Ibragimov and Linnik [15]. In our proof, we need to use the mixing coefficient for two sets A,B where A is a union of disjoint cubes (see the proof in the appendix) and as such definitely not convex. Furthermore it is not possible to find an x such that $B \subseteq A + x$ and the distance between the sets A and $A+x$ is larger than the desired value. Thus the more general version (5.2) of the strong mixing coefficient must be used. This problem was not fully acknowledged in the papers [8] and [10].

The Neyman-Scott processes and log Gaussian Cox processes with suitably mixing driving field obey (5.3) for either definition of the mixing coefficient $\alpha(p;k)$. Thus, from a practical point of view, the definition (5.2) is not restrictive.

For the asymptotic normality of the MPL estimate we will further assume a mild moment condition on $U_n(\theta_0)$:

$$\sup_{n \in \mathbf{N}} E_{\theta_0}(|\sqrt{|W_n \ominus R|}|U_n(\theta_0)|^q) < C_q < \infty \quad \text{for some} \quad q > 2. \tag{5.4}$$

This condition is slightly stronger than the existence of the standardized asymptotic variances of $U_n(\theta)$, see also [8, p. 1505]. It is satisfied for example for the processes which have the first six reduced cumulant moment measures of finite total variation provided $\frac{\mathrm{d}\lambda_0(x;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(x;\theta)}$ is bounded for $x \in B(o,R)$ and $\theta \in \Theta$. In particular, the class of Brillinger-mixing processes (i.e. processes for which the reduced cumulant moment measures of all orders have finite total variation) obviously fulfill this condition. The Brillinger-mixing processes include among others Neyman-Scott processes for which the distribution of the size (i.e. number of points) of the cluster has finite moments of all orders. For further examples and discussions of Brillinger-mixing, see [13].

Below, we present the theorem concerning asymptotic normality of the Palm likelihood estimator. The proof of the theorem can be found in the appendix.

Theorem 5.2. *Assume that the conditions of Theorem 4.1 are satisfied and moreover that* (5.3) *and* (5.4) *hold, the reduced factorial cumulant measures of X up to fourth order have finite total variation and that*

$$\sup_{\substack{u\in B(o,R)\\ |\theta_1-\theta_2|<\delta}} \left| \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(u;\theta)}\right)\Big|_{\theta=\theta_1} - \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(u;\theta)}\right)\Big|_{\theta=\theta_2}\right| \to 0 \quad \text{as } \delta\to 0 \tag{5.5}$$

and

$$\sup_{|\theta_1-\theta_2|<\delta} \left| \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\left(\int_{B(o,R)}\lambda_0(u;\theta)du\right)\Big|_{\theta=\theta_1} - \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\left(\int_{B(o,R)}\lambda_0(u;\theta)du\right)\Big|_{\theta=\theta_2}\right| \to 0 \quad \text{as } \delta\to 0, \tag{5.6}$$

hold.

Then $\lim_{n\to\infty}|W_n\ominus R|\operatorname{Var}_{\theta_0}(U_n(\theta)) = \Sigma(\theta)$ *exists and does not depend on the convex averaging sequence* W_n *and* $\sqrt{|W_n\ominus R|}(\hat{\theta}_n-\theta_0)$ *converges to a normal distribution with zero mean vector and covariance matrix* $M^{-1}\Sigma(\theta_0)M^{-1}$ *where*

$$\begin{aligned} M = \lambda\int_{B(o,R)}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(u;\theta)}\right)\Big|_{\theta=\theta_0}\lambda_0(u;\theta_0)du \\ - \lambda\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\left(\int_{B(o,R)}\lambda_0(u;\theta)du\right)\Big|_{\theta=\theta_0}. \end{aligned} \tag{5.7}$$

Remark. The matrix $\Sigma(\theta)$ can be expressed as a sum of mixed integrals of $\frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(u;\theta)}$ and $I_{B(o,R)}(u)$ with respect to the reduced factorial cumulant measures $\gamma^{(k)}_{\mathrm{red}}$, $k=2,3,4$ and the Lebesgue measure. The expression for these integrals seems, however, to be too complicated to be of any practical use.

Let us finish this section with a short discussion of the original log Palm likelihood (3.2) without any included edge correction. The estimate $\tilde{\theta}_n$ is derived from the estimating equation $\tilde{U}(\theta) = \frac{1}{|W|}\frac{\mathrm{d}\log L_P(\theta)}{\mathrm{d}\theta} = 0$. It follows from the proof of Theorem 5.2 that if $\sqrt{|W_n|}\,\tilde{U}_n(\theta_0)$ converges in distribution to $N(a(\theta_0),Q(\theta_0))$, then $\sqrt{|W_n|}(\tilde{\theta}_n-\theta_0)$ converges to

$$N(M^{-1}a(\theta_0), M^{-1}Q(\theta_0)M^{-1})$$

with the same M as in Theorem 5.2.

By repeating the proof of Theorem 5.2 step by step it moreover follows that $Q(\theta_0)=\Sigma(\theta_0)$, i.e. the asymptotic variance of $\sqrt{|W_n\ominus R|}\,U_n(\theta_0)$ and $\sqrt{|W_n|}\,\tilde{U}_n(\theta_0)$ is the same and that $\sqrt{|W_n|}(\tilde{U}_n(\tilde{\theta}_n) - E_{\theta_0}\tilde{U}_n(\tilde{\theta}_n))$ converges in distribution to $N(0,\Sigma(\theta_0))$.

However, it follows from the discussion at the end of Section 4 that the bias of $\tilde{U}_n(\theta_0)$ is of order $\mathcal{O}(\frac{\partial W_n}{|W_n|}) = \mathcal{O}(\rho_n^{-1})$ which is too large for convergence to 0 when multiplied by the normalization term $\sqrt{|W_n|} = \mathcal{O}(\rho_n^{d/2})$. Thus, we cannot establish a result for $\tilde{\theta}_n$ of the type presented in Theorem 5.2.

6 Example

In this section we will discuss the estimation procedure and the required assumptions for a classical point process model – the modified Thomas process [27] in $\mathbf{R}^2$.

The modified Thomas process belongs to the class of Neyman-Scott processes. The intensity of the mother process is μ. The number of points in each cluster is Poisson distributed with mean value equal to ν and each of the points is distributed according to the bivariate zero mean Gaussian density f_{σ^2} with independent components and the same variance σ^2, i.e. $f_{\sigma^2}(u) = \frac{1}{2\pi\sigma^2}\exp(-\frac{|u|^2}{2\sigma^2})$, $u \in \mathbf{R}^2$. Thus the parameter vector is $\theta = \{\mu, \nu, \sigma^2\} \in (0,\infty)^3$.

The modified Thomas process is obviously stationary and isotropic. The intensity is $\lambda = \mu\nu$, the Palm intensity is

$$\lambda_0(u;\theta) = \mu\nu + \frac{\nu}{4\pi\sigma^2}\exp\left(-\frac{|u|^2}{4\sigma^2}\right), \tag{6.1}$$

and its integral over $B(o,R)$ is equal to

$$K(R) = \pi R^2 + \frac{1}{\mu}\left(1 - \exp\left(-\frac{R^2}{4\sigma^2}\right)\right). \tag{6.2}$$

The corresponding Palm log likelihood function is thus

$$\begin{aligned}\log L_{PU}(\theta) = &\sum_{\substack{x\in X\cap(W\ominus R)\\ x\neq y\in X\\ (y-x)\in B(o,R)}} \log\left(\mu\nu + \frac{\nu}{4\pi\sigma^2}\exp\left(-\frac{|x-y|^2}{4\sigma^2}\right)\right)\\ &- |X\cap(W\ominus R)|\nu\left(\mu\pi R^2 + 1 - \exp\left(-\frac{R^2}{4\sigma^2}\right)\right).\end{aligned} \tag{6.3}$$

Its differentiation leads to the following estimating equations

$$\sum\left(\mu + \frac{1}{4\pi\sigma^2}\exp\left(-\frac{|x-y|^2}{4\sigma^2}\right)\right)^{-1} = \nu\pi R^2|X\cap(W\ominus R)|, \tag{6.4}$$

$$\sum\frac{1}{\nu} = \left(\mu\pi R^2 + 1 - \exp\left(-\frac{R^2}{4\sigma^2}\right)\right)|X\cap(W\ominus R)|, \tag{6.5}$$

$$\sum\frac{\frac{1}{\pi}\left(1-\frac{|x-y|^2}{4\sigma^2}\right)\exp\left(-\frac{|x-y|^2}{4\sigma^2}\right)}{\mu + \frac{1}{4\pi\sigma^2}\exp\left(-\frac{|x-y|^2}{4\sigma^2}\right)} = \nu R^2\exp\left(-\frac{R^2}{4\sigma^2}\right)|X\cap(W\ominus R)|, \tag{6.6}$$

where all the sums on the left-hand sides are over $x \in X\cap(W\ominus R), x \neq y \in X$, $(y-x)\in B(o,R)$.

If we define the parameter space Θ as $[a_1,b_1]\times[a_2,b_2]\times[a_3,b_3]$ for some $0 < a_i < b_i < \infty$, $i = 1,2,3$, then it is compact and the continuity and boundedness assumptions of Theorem 4.1 are fulfilled. From a practical point of view, the restriction of the parameter space to the compact set Θ is not a problem since we can always choose small and large enough a_i and b_i so that Θ covers all values of the parameters which are reasonable for a particular application.

Concerning the solution of the estimating equations we get from the second one

$$\hat{\nu} = \frac{N}{\left(\mu\pi R^2 + 1 - \exp\left(-\frac{R^2}{4\sigma^2}\right)\right)|X \cap (W \ominus R)|}, \tag{6.7}$$

where N is the number of terms in the sum of the left-hand sides of (6.4) – (6.6). Thus we only need to maximize over (μ, σ^2). The uniqueness of the joint argument of maxima $\{\hat{\mu}, \hat{\sigma}^2\}$ is, however, not clear generally.

When we turn to the assumptions of Theorem 5.2 we see that $\log \lambda_0(u;\theta)$ is continuous as a function of $\theta \in \Theta$ and is C^∞ as a function of u for any value of $\theta \in \Theta$. This implies that the uniform continuity conditions (5.5) and (5.6) are fulfilled.

The process is moreover Brillinger mixing because the Poisson distribution of the number of points in the cluster have finite moments of any order. Thus all the reduced factorial cumulant measures are finite and the assumption (5.4) is also satisfied. The mixing assumption (5.3) is satisfied with $\mathcal{O}(k^{-\epsilon})$ for any $\epsilon > d$. This follows from Lemma 5.1 since $f_{\sigma^2}(u) = \mathcal{O}(|u|^{-\epsilon-d})$ for any $\epsilon > d$. Thus all the additional assumptions of Theorem 5.2 are fulfilled.

For comparison let us also have a look on the estimating equations derived by differentiation of the log-composite likelihood ([8])

$$\begin{aligned}\log CL(\theta) = &\sum_{\substack{x\in X\cap(W\ominus R)\\ x\neq y\in X\\ (y-x)\in B(o,R)}} \log\left(\mu\nu + \frac{\nu}{4\pi\sigma^2}\exp\left(-\frac{|x-y|^2}{4\sigma^2}\right)\right)\\ &- N\log\left[\mu\nu|W\ominus R|\left(\mu\nu\pi R^2 + \nu\left(1-\exp\left(-\frac{R^2}{4\sigma^2}\right)\right)\right)\right].\end{aligned} \tag{6.8}$$

Here we get

$$\sum \frac{2\mu + \frac{1}{4\pi\sigma^2}\exp\left(-\frac{|x-y|^2}{4\sigma^2}\right)}{\mu + \frac{1}{4\pi\sigma^2}\exp\left(-\frac{|x-y|^2}{4\sigma^2}\right)} = \frac{N\left(2\mu + 1 - \exp\left(-\frac{R^2}{4\sigma^2}\right)\right)}{\pi R^2\mu + 1 - \exp\left(-\frac{R^2}{4\sigma^2}\right)}, \tag{6.9}$$

$$\frac{2}{\nu}N = \frac{2}{\nu}N, \tag{6.10}$$

$$\sum \frac{\frac{1}{\pi}\left(1 - \frac{|x-y|^2}{4\sigma^2}\right)\exp\left(-\frac{|x-y|^2}{4\sigma^2}\right)}{\mu + \frac{1}{4\pi\sigma^2}\exp\left(-\frac{|x-y|^2}{4\sigma^2}\right)} = \frac{NR^2\exp\left(-\frac{R^2}{4\sigma^2}\right)}{\pi\mu R^2 + 1 - \exp\left(-\frac{R^2}{4\sigma^2}\right)}. \tag{6.11}$$

Thus we see that the parameter ν cannot be identified by the composite likelihood estimation and we have to estimate it from the intensity λ by

$$\hat{\nu} = \frac{\hat{\lambda}}{\mu} = \frac{|X \cap W|}{\mu|W|}.$$

The third estimating equation is the same for the composite likelihood and Palm likelihood estimation (when we plug $\hat{\nu}$ into (6.6) for the Palm likelihood case) and the first estimating equation (6.9) is a more complicated reweighted version of (6.4).

Finaly let us remark here that there is a similarity between the estimate (6.7) of $\hat{\nu}$ for the Palm likelihood case and the estimate for the composite likelihood case. Namely (6.7) also uses the theoretical and the estimated value of an intensity (i.e. mean number of) but now it is not the intensity of the points of X in W but the intensity of the pairs of points of X in W.

7 Discussion

In the present paper we have considered parameter estimation in stationary point process models by means of Palm likelihood. We have proved consistency and asymptotic normality of the maximum Palm likelihood estimator.

For the proof of asymptotic normality we have imposed strong mixing conditions. These may be difficult to verify in practice. However, for the important class of Neyman-Scott processes we have derived a sufficient condition which is easy to verify, see Lemma 5.1. For the class of log Gaussian Cox processes, the mixing conditions for the point process are implied by the mixing conditions for the driving Gaussian field, which have been treated in the literature (see [7]).

An alternative proof of asymptotic normality could be based on the central limit theorem of Bolthausen [2] for stationary α-mixing random fields. The required mixing assumptions would be more restrictive in this case.

For evaluating the efficiency of the estimator $\hat{\theta}_n$, an estimate of the variance matrix $\Sigma(\theta_0)$ is needed. Since $\Sigma(\theta_0)$ is equal to a complicated expression involving integrals with respect to the measures $\gamma_{\mathrm{red}}^{(3)}$ and $\gamma_{\mathrm{red}}^{(4)}$, which are typically not available in a feasible form, an approximation of $\Sigma(\theta_0)$ needs to be determined. This can be done by means of simulation. We can produce independent realizations X_l, $l \in \{1, \ldots, N\}$ from the fitted model and approximate $\Sigma(\theta_0)$ by the sample variance matrix of the score functions $U_n(\hat{\theta}_n)$ computed for each of the replications X_l. Or if the original data are large enough we can use subsampling methods for the estimation of $\Sigma(\theta_0)$ (see e.g. [12], [22] for further information).

Appendix

For $\epsilon > 0$, let $R_\epsilon = \{|\theta - \theta_0| < \epsilon\}$ and let

$$\omega_n(\delta) = \sup_{|\theta_1 - \theta_2| < \delta} \{|U_n(\theta_1) - E_{\theta_0} U_n(\theta_1) - U_n(\theta_2) + E_{\theta_0} U_n(\theta_2)|\} \tag{A.1}$$

be the modulus of continuity of $U_n(\theta) - E_{\theta_0} U_n(\theta)$.

For the proof of Theorem 4.1 we need a stronger version of [3, Theorem 3.1]. Let us state it here as a lemma.

Lemma A.1. *Let Θ be compact. Suppose that the following conditions are satisfied for any $\epsilon > 0$*

$$\inf_{\Theta \setminus R_\epsilon} |E_{\theta_0} U_n(\theta)| \geq C_\epsilon \text{ for some } C_\epsilon > 0 \text{ and all } n > N \text{ for some fixed } N > 0, \quad \text{(A.2)}$$

$$\sup_{\Theta} |U_n(\theta) - E_{\theta_0} U_n(\theta)| \to 0 \quad P_{\theta_0} - a.s., \quad \text{(A.3)}$$

then $\hat{\theta}_n \to \theta_0$ P_{θ_0}-a.s.

Proof. It suffices to show that for all $\epsilon > 0$

$$P_{\theta_0} \left(\exists n_0 \, \forall n \geq n_0 : \{\theta \in \Theta : U_n(\theta) = 0\} \subseteq R_\epsilon \right) = 1.$$

Let

$$S_{mn} = \{\theta \in \Theta : |E_{\theta_0} U_n(\theta)| \leq m\},$$

where $m > 0$. From (A.2) we find for $m < C_\epsilon$ and $n > N$ that $S_{mn} \subseteq R_\epsilon$. Let us choose such an $m < C_\epsilon$. It therefore suffices to show that for all $\epsilon > 0$

$$P_{\theta_0} \left(\exists n_0 \, \forall n \geq n_0 : \{\theta \in \Theta : U_n(\theta) = 0\} \subseteq S_{mn} \right) = 1. \quad \text{(A.4)}$$

For θ with $U_n(\theta) = 0$, we have

$$|E_{\theta_0} U_n(\theta)| \leq |E_{\theta_0} U_n(\theta) - U_n(\theta)| + U_n(\theta) \leq \sup_{\Theta} |U_n(\theta) - E_{\theta_0} U_n(\theta)|.$$

Using (A.3), (A.4) follows. □

Moreover for checking the assumption (A.3) we derive a stronger version of [3, Lemma 3.2].

Lemma A.2. *Let Θ be compact and assume that*

$$|U_n(\theta) - E_{\theta_0} U_n(\theta)| \to 0 \; P_{\theta_0} - a.s. \text{ for any } \theta \in \Theta, \quad \text{(A.5)}$$

$$\begin{aligned} &\text{there exists a sequence } \{\epsilon_k\}_{k \in \mathbf{N}},\ \epsilon_k \xrightarrow[k \to 0]{} 0, \text{ such that} \\ &P_{\theta_0}\big(\limsup_{n \to \infty} \omega_n(\tfrac{1}{k}) \geq \epsilon_k \big) = 0 \text{ for each } k \in \mathbf{N}. \end{aligned} \quad \text{(A.6)}$$

Then $\sup_\Theta |U_n(\theta) - E_{\theta_0} U_n(\theta)| \to 0$ P_{θ_0}-a.s.

Proof. We want to show that for all $\delta > 0$

$$P_{\theta_0}\Big(\exists n_0 \, \forall n \geq n_0 : \sup_{\Theta} |U_n(\theta) - E_{\theta_0} U_n(\theta)| < \delta \Big) = 1.$$

The proof is by a standard covering argument (see e.g. [11]). For a fixed $\delta > 0$ we choose k such that $\delta > 2\epsilon_k$ and cover the compact space Θ with N balls $B(\theta_i, \frac{1}{k})$, $i = 1, \ldots, N$. Let $D = \{\limsup_{n \to \infty} \omega_n(\frac{1}{k}) < \epsilon_k\}$. Then, by (A.6), $P_{\theta_0}(D) = 1$. In other words, P_{θ_0}-a.s. there exists n_0 such that for $n \geq n_{01}$,

$$\omega_n\left(\tfrac{1}{k}\right) < \epsilon_k < \tfrac{\delta}{2}.$$

Furthermore for $\theta \in B(\theta_i, \frac{1}{k})$ we have

$$\begin{aligned}|U_n(\theta) - E_{\theta_0}U_n(\theta)| &\le |U_n(\theta) - E_{\theta_0}U_n(\theta) - U_n(\theta_i) + \ E_{\theta_0}U_n(\theta_i)| \\ &\quad + |U_n(\theta_i) - E_{\theta_0}U_n(\theta_i)| \\ &\le \omega_n\left(\tfrac{1}{k}\right) + \max_{i=1,\dots,N}\{|U_n(\theta_i) - E_{\theta_0}U_n(\theta_i)|\}.\end{aligned}$$

It follows that

$$\sup_{\Theta}|U_n(\theta) - E_{\theta_0}U_n(\theta)| \le \omega_n\left(\tfrac{1}{k}\right) + \max_{i=1,\dots,N}\{|U_n(\theta_i) - E_{\theta_0}U_n(\theta_i)|\}.$$

Using (A.5), P_{θ_0}-a.s. there exists n_{02} such that for $n \ge n_{02}$

$$\max_{i=1,\dots,N}\{|U_n(\theta_i) - E_{\theta_0}U_n(\theta_i)|\} < \tfrac{\delta}{2}.$$

The result of the lemma now follows by choosing $n_0 = \max\{n_{01}, n_{02}\}$. □

Proof of Theorem 4.1. It suffices to show that the conditions of Lemma A.1 are satisfied. First notice that

$$E_{\theta_0}U_n(\theta) = \lambda\Big(\int_{B(o,R)} \frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(u;\theta)}\lambda_0(u;\theta_0)du - \frac{\mathrm{d}\int_{B(o,R)}\lambda_0(u;\theta)du}{\mathrm{d}\theta}\Big)$$

is bounded and continuous (even uniformly continuous on the compact set Θ) with respect to θ from the assumptions and it does not depend on the observation window W_n. Thus from the assumptions that Θ is compact and $|E_{\theta_0}U_n(\theta)| = 0$ only for θ_0 we get that (A.2) holds, in fact for all n.

To show (A.3) we will use Lemma A.2. From the ergodicity of the point process X and the form of $U_n(\theta)$ (being just a sum over pairs of points closer than R of some continuous, bounded function over the convex averaging sequence of $\{W_n\}$), it follows that $U_n(\theta) \to E_{\theta_0}U_n(\theta)$ almost surely for any fixed θ (see e.g. [4, pp. 335–338]). Thus (A.5) holds true.

For (A.6) we observe, that for any $\epsilon > 0$ there exists $\delta > 0$ such that

$$\sup_{\substack{u\in B(o,R)\\ |\theta_1-\theta_2|<\delta}}\left|\frac{\mathrm{d}\lambda_0(u;\theta_1)}{\mathrm{d}\theta}\frac{1}{\lambda_0(u;\theta_1)} - \frac{\mathrm{d}\lambda_0(u;\theta_2)}{\mathrm{d}\theta}\frac{1}{\lambda_0(u;\theta_2)}\right| < \frac{\epsilon}{6\lambda^2 K(R;\theta_0)}, \tag{A.7}$$

where $K(R;\theta_0) = \frac{1}{\lambda}\int_{B(o,R)}\lambda_0(u;\theta_0)du$ is the K-function, and

$$\sup_{|\theta_1-\theta_2|<\delta}\left|\frac{\mathrm{d}\int_{B(o,R)}\lambda_0(u;\theta_1)du}{\mathrm{d}\theta} - \frac{\mathrm{d}\int_{B(o,R)}\lambda_0(u;\theta_2)du}{\mathrm{d}\theta}\right| < \frac{\epsilon}{6\lambda}, \tag{A.8}$$

due to the continuity assumptions on $\frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(u;\theta)}$ and $\frac{\mathrm{d}(\int_{B(o,R)}\lambda_0(u;\theta)du)}{\mathrm{d}\theta}$. It follows that we can construct a sequence $\{\epsilon_k\}$ such that (A.7) and (A.8) hold for every $k \in \mathbf{N}$ when we take $\delta = \frac{1}{k}$ and $\epsilon = \epsilon_k$. Moreover since $E_{\theta_0}U_n(\theta)$ is uniformly continuous on Θ we can modify the sequence $\{\epsilon_k\}$ in such a way that

$$\sup_{|\theta_1-\theta_2|<\frac{1}{k}}\left|E_{\theta_0}U_n(\theta_1) - E_{\theta_0}U_n(\theta_2)\right| < \frac{\epsilon_k}{3}, \tag{A.9}$$

holds true for all k (and any n).

Let

$$A_n = \frac{1}{|W_n \ominus R|} \sum_{\substack{x \in X \cap (W_n \ominus R) \\ x \neq y \in X \\ (y-x) \in B(o,R)}} 1.$$

Then,

$$\begin{aligned}
\sup_{|\theta_1-\theta_2|<\frac{1}{k}} & \left|U_n(\theta_1) - E_{\theta_0}U_n(\theta_1) - U_n(\theta_2) + E_{\theta_0}U_n(\theta_2)\right| \\
&< \frac{\epsilon_k}{6\lambda^2 K(R;\theta_0)} A_n + \frac{\epsilon_k}{6\lambda} \frac{|X \cap (W_n \ominus R)|}{|W_n \ominus R|} \\
&\quad + \sup_{|\theta_1-\theta_2|<\frac{1}{k}} \left|E_{\theta_0}U_n(\theta_1) - E_{\theta_0}U_n(\theta_2)\right|
\end{aligned}$$

holds for every k and n. Thus, for any k, we have

$$\begin{aligned}
& P_{\theta_0} \limsup_{n\to\infty} \left(\omega_n(\tfrac{1}{k}) \geq \epsilon_k\right) \\
&\quad \leq P_{\theta_0}\Big(\limsup_{n\to\infty} A_n > 2\lambda^2 K(R,\theta_0)\Big) \\
&\qquad + P_{\theta_0}\Big(\limsup_{n\to\infty} \frac{|X \cap (W_n \ominus R)|}{|W_n \ominus R|} > 2\lambda\Big) \\
&\qquad + P_{\theta_0}\Big(\limsup_{n\to\infty} \sup_{|\theta_1-\theta_2|<\frac{1}{k}} |E_{\theta_0}U_n(\theta_1) - E_{\theta_0}U_n(\theta_2)| > \frac{\epsilon_k}{3}\Big) = 0.
\end{aligned}$$

The first two terms are equal to 0 because $\frac{|X\cap(W\ominus R)|}{|W\ominus R|}$ converges to λ and A_n converges to $\lambda^2 K(R;\theta_0)$ almost surely from the ergodicity of the process X. The third term is equal to 0 from (A.9). Thus (A.6) holds true and from Lemma A.2 we obtain (A.3) which completes the proof. □

Proof of Theorem 4.2. The proof is analogous to the proof of Theorem 4.1. We need to show that the conditions (A.2) and (A.3) are fulfilled for $\tilde{U}_n(\theta)$ instead of $U_n(\theta)$.

First observe that

$$\begin{aligned}
& \left|E_{\theta_0}U_n(\theta) - E_{\theta_0}\tilde{U}_n(\theta)\right| \qquad (A.10) \\
&\quad \leq \lambda \int_{B(o,R)} \left|\frac{\gamma_{W_n}(u)}{|W_n|} - 1\right| \left|\frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta} \frac{1}{\lambda_0(u;\theta)}\right| \lambda_0(u;\theta_0)du \\
&\quad \leq \lambda K(R;\theta_0) \max_{u\in B(o,R),\theta\in\Theta} \left|\frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta} \frac{1}{\lambda_0(u;\theta)}\right| \sup_{z\in B(o,R)} \left|\frac{\gamma_{W_n}(z)}{|W_n|} - 1\right| \\
&\quad < C_\epsilon/2,
\end{aligned}$$

for all n larger than some $N > 0$ from the continuity assumptions and from (4.6), where C_ϵ is the constant from (A.2) in proof of Theorem 4.1. Now combining (A.2) in the proof of Theorem 4.1 with (A.10) we get that (A.2) holds also for $\tilde{U}_n(\theta)$ for $\tilde{C}_\epsilon = C_\epsilon/2$.

The validity of (A.3) follows in exactly the same way as the validity of (A.3) in the proof in Theorem 4.1. □

Proof of Lemma 5.1. Let us consider two sets A and B with $d(A,B) \geq k$ and $|A|$, $|B| \leq p$. Let us rewrite the cluster process X as $X = \bigcup_{c \in C} X_c$, where X_c is the cluster centered around a mother point located at c and C is the stationary Poisson process of mothers.

We denote $X_1 = \bigcup_{c \in A \oplus \frac{k}{2}} X_c$ and $X_2 = X \backslash X_1$. Then X_1 and X_2 are independent processes and $X = X_1 \cup X_2$. Let G_1, G_2 be sets of point configurations and $C_1 = \{X \cap A \in G_1\}$, $C_2 = \{X \cap B \in G_2\}$ be arbitrary fixed events from $\mathcal{F}^X(A), \mathcal{F}^X(B)$, respectively. Moreover we let $D_1 = \{X_1 \cap B = \emptyset\}$, $D_2 = \{X_2 \cap A = \emptyset\}$ and $D = D_1 \cap D_2$.

Then $P(C_1 \cap C_2) = P(C_1 \cap C_2 \cap D) + P(C_1 \cap C_2 \cap D^C)$ and $P(C_1 \cap C_2 \cap D) = P(X_1 \cap A \in G_1, X_1 \cap B = \emptyset) P(X_2 \cap B \in G_2, X_2 \cap A = \emptyset)$. Similarly

$$\begin{aligned} &P(C_1)P(C_2) \\ &\quad = P(X_1 \cap A \in G_1, X_1 \cap B = \emptyset) P(X_2 \cap B \in G_2, X_2 \cap A = \emptyset) P(D) \\ &\qquad + P(C_1 \cap D) P(C_2 \cap D^C) + P(C_1 \cap D^C) P(C_2 \cap D) \\ &\qquad + P(C_1 \cap D^C) P(C_2 \cap D^C). \end{aligned}$$

Thus

$$|P(C_1 \cap C_2) - P(C_1)P(C_2)| \leq 4P(D^C) \leq 4P(D_1^C) + 4P(D_2^C),$$

and

$$P(D_1^C) \leq E|X_1 \cap B| = \mu\nu \int_{A \oplus \frac{k}{2}} \int_B k(c,u) du dc \leq \mu\nu |B| \int_{\mathbf{R}^d \backslash B(o, \frac{k}{2})} h(v) dv,$$

since the distance of $c \in A \oplus \frac{k}{2}$ and $x \in B$ is always larger or equal to $\frac{k}{2}$. Similarly

$$P(D_2^C) \leq \mu\nu |A| \int_{\mathbf{R}^d \backslash B(o, \frac{k}{2})} h(v) dv.$$

Thus

$$\frac{\alpha(p;k)}{\max(p,1)} \leq \mu\nu \left(\frac{p}{p} + \frac{p}{p} \right) \int_{\mathbf{R}^d \backslash B(o, \frac{k}{2})} h(v) dv = \mathcal{O}\Big(\int_{\frac{k}{2}}^{\infty} v^{d-1-d-\epsilon} dv \Big) = \mathcal{O}(k^{-\epsilon}),$$

where we at the second equality have used change into polar coordinates. This concludes the proof. □

Proof of Theorem 5.2. To show the existence of

$$\lim_{n \to \infty} |W_n \ominus R| \operatorname{Var}_{\theta_0} U_n(\theta),$$

let us write in detail that

$$
\begin{aligned}
&|W_n \ominus R| \operatorname{Var}_{\theta_0} U_n(\theta) \\
&= \frac{1}{|W_n \ominus R|} \operatorname{Var}_{\theta_0} \Bigg(\sum_{x,y \in X}^{\neq} I_{W_n \ominus R}(x) I_{B(o,R)}(x-y) \frac{\mathrm{d}\lambda_0(y-x;\theta)}{\mathrm{d}\theta} \frac{1}{\lambda_0(y-x;\theta)} \Bigg) \\
&\quad - \frac{(2\int_{B(o,R)} \frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta} du)^T}{|W_n \ominus R|} \\
&\quad \times \operatorname{Cov}_{\theta_0} \Bigg(\sum_{x,y \in X}^{\neq} I_{W_n \ominus R}(x) I_{B(o,R)}(x-y) \frac{\mathrm{d}\lambda_0(y-x;\theta)}{\mathrm{d}\theta} \frac{1}{\lambda_0(y-x;\theta)}, \sum_{z \in X} I_{W_n \ominus R}(z) \Bigg) \\
&\quad + \frac{\operatorname{Var}_{\theta_0}(X \cap |W_n \ominus R|)}{|W_n \ominus R|} \Big(\int_{B(o,R)} \frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta} du \Big)^T \Big(\int_{B(o,R)} \frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta} du \Big).
\end{aligned}
$$

The last term converges to

$$
\lambda(1 + \gamma^{(2)}_{\mathrm{red}}(\mathbf{R}^d)) \left(\int_{B(o,R)} \frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta} du \right)^T \left(\int_{B(o,R)} \frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta} du \right).
$$

The variance and covariance in the first and second term can be expressed as integrals with respect to the factorial moment measures. Using the relations between the moment and cumulant measures and disintegration of $\gamma^{(k)}$ for stationary processes, the terms can be further reexpressed as a combination of mixed integrals with respect to the reduced factorial cumulant measures up to fourth order. Examining them one by one, it can be shown that under the assumptions of boundedness and continuity of $\frac{\mathrm{d}\lambda_0(x;\theta)}{\mathrm{d}\theta} \frac{1}{\lambda_0(x;\theta)}$ and the finiteness of the total variation of $\gamma^{(k)}_{\mathrm{red}}, k = 2, 3, 4$, the normalization by $1/|W_n \ominus R|$ is the correct one to make them all converge.

To prove the asymptotic normality of the MPL estimate let us denote by $U_n'(\theta)$ the derivative $\frac{\mathrm{d}U_n(\theta)}{\mathrm{d}\theta}$. From the mean value theorem we have that

$$
U_n(\hat{\theta}_n) = U_n(\theta_0) + U_n'(\theta_n^*)(\hat{\theta}_n - \theta_0) = 0, \tag{A.11}
$$

for some $\theta_n^* = \theta_0 + Q(\hat{\theta}_n - \theta_0)$, where Q is a diagonal matrix with diagonal elements between 0 and 1. Thus for proving the asymptotic normality of $\hat{\theta}_n$ it is by Slutsky's lemma enough to prove:

$$
U_n'(\theta_n^*) \to M \text{ in probability}, \tag{A.12}
$$

$$
\sqrt{|W_n \ominus R|}\, U_n(\theta_0) \text{ converges in distribution to } N(0, \Sigma(\theta_0)). \tag{A.13}
$$

To show (A.12) we observe that

$$
|U_n'(\theta_n^*) - M| \le |U_n'(\theta_n^*) - U_n'(\theta_0)| + |U_n'(\theta_0) - M|. \tag{A.14}
$$

Now $|U_n'(\theta_0) - M|$ converges to 0 almost surely from the ergodicity of X since $M = E_{\theta_0} U_n'(\theta_0)$.

To show that $|U_n'(\theta_n^*) - U_n'(\theta_0)|$ converges to 0 in probability, let us denote by $m^f(\delta)$ the supremum from the formula (5.5) taken over $|\theta_1 - \theta_2| < \delta$ and by $m^h(\delta)$

the supremum from the formula (5.6) taken again over $|\theta_1 - \theta_2| < \delta$. Since both of the $m^f(\delta)$ and $m^h(\delta)$ converge to 0 as $\delta \to 0$ we can find for a given $\epsilon > 0$ a $\Delta > 0$ so that

$$m^f(\delta) < \frac{\epsilon}{4}\frac{1}{\lambda^2 K(R;\theta_0)} \qquad \text{and} \qquad m^h(\delta) < \frac{\epsilon}{4}\frac{1}{\lambda} \tag{A.15}$$

holds for all $\delta < \Delta$. Then similarly like in the end of the proof of Theorem 4.1 we have that

$$\sup_{|\theta_1-\theta_0|<\Delta} |U_n'(\theta_1) - U_n'(\theta_0)| < \frac{\epsilon}{4}\frac{1}{\lambda^2 K(R;\theta_0)} A_n + \frac{\epsilon}{4}\frac{1}{\lambda}\frac{|X \cap (W_n \ominus R)|}{|W_n \ominus R|}, \tag{A.16}$$

and thus from the ergodicity of X for a given $\eta > 0$ we can find an N_1 large enough so that

$$P_{\theta_0}\left(|U_n'(\theta_1) - U_n'(\theta_0)| > \epsilon\right) < \tfrac{\eta}{2}, \tag{A.17}$$

for all $n > N_1$ and θ_1 closer than Δ to θ_0.

Now we have to remember that since $\hat{\theta}_n$ converges to θ_0 almost surely and thus also in probability from Theorem 4.1, so does θ_n^*. And thus for a given $\eta > 0$ we can find an N_2 such that

$$P_{\theta_0}(|\theta_n^* - \theta_0| > \Delta) < \tfrac{\eta}{2} \qquad \text{for all} \;\; n > N_2, \tag{A.18}$$

which together with (A.17) gives that for a given $\epsilon > 0$ and $\eta > 0$

$$P_{\theta_0}\left(|U_n'(\theta_n^*) - U_n'(\theta_0)| > \epsilon\right) < \eta \qquad \text{for all} \;\; n > \max(N_1, N_2). \tag{A.19}$$

Thus $|U_n'(\theta_n^*) - U_n'(\theta_0)|$ converges in probability to 0 which completes the proof of (A.12).

To prove (A.13) we will use a blocking method similar to the one used in [10].

Let α and η be positive constants such that $2d/(d+\epsilon) < \eta < \alpha < 1$, let $\rho_n = \rho(W_n)$, $l_n = \rho_n^\alpha$ and $m_n = \rho_n^\alpha - \rho_n^\eta$.

For a fixed n let us cover $\mathbf{R}^d$ by the union of disjoint d-dimensional cubes $\{K_n^j\}$ of sidelength l_n. Let $C_n^j \subset K_n^j$ be the closed cube with the same center as K_n^j but with sidelength m_n. In the sequel we will consider the collection $\{C_n^j, j \in J_n\}$ of all cubes contained in W_n. Note that the distance between any two distinct cubes C_n^j, $C_n^{j'}$ is at least ρ_n^η which goes to infinity as n increases. Thus by the strong mixing, $X \cap C_n^j$ and $X \cap C_n^{j'}$ become asymptotically independent and furthermore the volume (and thus the observed information available in these sets) of W_n and $\bigcup_{j\in J_n} C_n^j$ are of the same order. This is the main idea of the blocking method. For the formal development of the argument we need some more notation.

Let $U_n(\theta_0)_i$ denote the i-th component of the score function $U_n(\theta_0)$ and let

$$\begin{aligned}
S_n &= \sqrt{|W_n \ominus R|}\, U_n(\theta_0)_i, \\
s_n^j &= \sqrt{|C_n^j \ominus R|}\, U(\theta_0; C_n^j)_i, \\
s_n &= \Big(\sum_{j\in J_n} s_n^j\Big) \Big/ \sqrt{k_n}, \\
s_n' &= \Big(\sum_{j\in J_n} s_n'^j\Big) \Big/ \sqrt{k_n}
\end{aligned}$$

where $U(\theta_0; C_n^j)_i$ is the i-th component of the score function $U_n(\theta_0)$ with W_n replaced by C_n^j, k_n is the number of elements in J_n and $\{s_n'^j, j \in J_n\}$ is a collection of independent identically distributed random variables with the same distribution as s_n^j.

To prove that $\sqrt{|W_n \ominus R|}\, U_n(\theta_0)_i$ converges in distribution to $N(0, \sigma_i(\theta_0))$, where $\sigma_i(\theta_0) = (\Sigma(\theta_0))_{i,i}$, it is enough to show the following three facts:

$$(S_n - s_n) \to 0 \text{ in probability}, \tag{A.20}$$

$$(\phi_n(t) - \phi_n'(t)) \to 0 \text{ for all } t \in \mathbf{R}, \tag{A.21}$$

$$s_n' \to N(0, \sigma_i(\theta_0)) \text{ in distribution}, \tag{A.22}$$

where ϕ_n denotes the characteristic function of s_n and ϕ_n' of s_n', respectively.

Since $E_{\theta_0} S_n = E_{\theta_0} s_n = 0$, it is enough to show that $\mathrm{Var}_{\theta_0}(S_n - s_n) \to 0$ to prove (A.20). Obviously

$$|W_n \ominus R| \geq k_n |C_n^j \ominus R|,$$

and $W_n \ominus R \subseteq W_n \subseteq (\partial W_n \oplus \sqrt{d} l_n) \cup (\cup_{j \in J_n} K_n^j)$, thus

$$|W_n \ominus R| \leq |\partial W_n \oplus (\sqrt{d}\, l_n)| + k_n |K_n^j|.$$

Since $|\partial W_n \oplus (\sqrt{d}\, l_n))| = \mathcal{O}((\rho_n)^{(d-1)+\alpha})$ and $k_n \leq |W_n \oplus (\sqrt{d}\, l_n))|/l_n^d = \mathcal{O}(\rho_n^{d(1-\alpha)})$ we have

$$\lim_{n\to\infty} \frac{|\partial W_n \oplus (\sqrt{d}\, l_n))|}{k_n |C_n^j \ominus R|} \leq \lim_{n\to\infty} \frac{\mathcal{O}(\rho_n^{d-1+\alpha})}{\mathcal{O}(\rho_n^{d(1-\alpha)} m_n^d)} = \lim_{n\to\infty} \mathcal{O}(\rho_n^{\alpha-1}) = 0,$$

and

$$\lim_{n\to\infty} \frac{k_n |K_n^j|}{k_n |C_n^j \ominus R|} = \lim_{n\to\infty} \frac{l_n^d}{m_n^d} = 1.$$

Thus

$$\lim_{n\to\infty} \frac{|W_n \ominus R|}{k_n |C_n^j \ominus R|} = 1. \tag{A.23}$$

Thus, in order to show that $\mathrm{Var}_{\theta_0}(S_n - s_n) \to 0$, it is enough to show that

$$\mathrm{Var}_{\theta_0} \Bigg(\sum_{\substack{x \in X \cap (W_n \ominus R) \\ x \neq y \in X \\ (y-x) \in B(o,R)}} \frac{\left(\frac{\mathrm{d}\lambda_0(y-x;\theta)}{\mathrm{d}\theta} \frac{1}{\lambda_0(y-x;\theta)} \Big|_{\theta=\theta_0} \right)_i}{\sqrt{|W_n \ominus R|}} \tag{A.24}$$

$$- \sum_{j=1}^{k_n} \sum_{\substack{x \in X \cap (C_n^j \ominus R) \\ x \neq y \in X \\ (y-x) \in B(o,R)}} \frac{\left(\frac{\mathrm{d}\lambda_0(y-x;\theta)}{\mathrm{d}\theta} \frac{1}{\lambda_0(y-x;\theta)} \Big|_{\theta=\theta_0} \right)_i}{\sqrt{|W_n \ominus R|}} \Bigg) \to 0,$$

and

$$\mathrm{Var}_{\theta_0} \left(\frac{|X \cap (W_n \ominus R)| - \sum_{j=1}^{k_n} |X \cap (C_n^j \ominus R)|}{\sqrt{|W_n \ominus R|}} \left(\frac{\mathrm{d} \int_{B(o,R)} \lambda_0(r;\theta) dr}{\mathrm{d}\theta} \Big|_{\theta=\theta_0} \right)_i \right) \to 0. \tag{A.25}$$

Let us denote by V_n the set $(W_n \ominus R)\backslash(\bigcup_{j=1}^{k_n}(C_n^j \ominus R))$. Then from the boundedness assumptions of Theorem 4.1 the variance from (A.25) is bounded from above by

$$\text{const}\frac{\lambda|V_n|}{|W_n \ominus R|}(1+\gamma_{\text{red}}^{(2)}(V_n)) \le \text{const}'\frac{\lambda|V_n|}{|W_n \ominus R|}, \tag{A.26}$$

when $\gamma_{\text{red}}^{(2)}$ has finite total variation. Similarly but with a substantialy larger amount of algebra it is possible to derive an upper bound for the variance from (A.24) of the same form $\text{const}\frac{\lambda|A_n|}{|W_n \ominus R|}$ where the constant is a combination of the total variations of $\gamma_{\text{red}}^{(2)}, \gamma_{\text{red}}^{(3)}$ and $\gamma_{\text{red}}^{(4)}$.

The proof of (A.20) is complete by observing that

$$\frac{|V_n|}{|W_n \ominus R|} = \frac{|W_n \ominus R| - k_n|C_n^j \ominus R|}{|W_n \ominus R|} \;\to\; 0, \quad \text{as } n \to \infty, \tag{A.27}$$

according to (A.23).

To show (A.21) we will use the mixing assumptions. Let us define

$$V_j = \exp\left(\iota t \frac{s_n^j}{\sqrt{k_n}}\right), \tag{A.28}$$

where ι denotes the imaginary unit. Then,

$$\phi_n(t) = E\Big(\prod_{j\in J_n} V_j\Big), \qquad \phi_n'(t) = \prod_{j\in J_n} EV_j, \tag{A.29}$$

and

$$|\phi_n(t) - \phi_n'(t)| \le \sum_{j=1}^{k_n-1}\left|E\left(\prod_{s=1}^{j+1} V_s\right) - E\left(\prod_{s=1}^{j} V_s\right)EV_{j+1}\right|. \tag{A.30}$$

If we denote $Z_j = \prod_{s=1}^{j} V_s$ and $Y_j = V_j$, then obviously $Z_j \in \mathcal{F}^X(\cup_{s=1}^{j} C_n^s)$, $Y_j \in \mathcal{F}^X(C_n^{j+1})$, $|\cup_{s=1}^{j} C_n^s| = j(m_n)^d$, $|C_n^{j+1}| = (m_n)^d$. Recall that $d(\cup_{s=1}^{j} C_n^s, C_n^{j+1}) \ge (\rho_n)^\eta$. Since both random variables Z_j and Y_j are bounded in absolute value by 1, we obtain the following bound on their covariance by means of the strong mixing coefficient (see e.g. Lemma 1.2.1 in [29])

$$\text{Cov}(Z_j, Y_j) \le 4\alpha(j(m_n)^d, (\rho_n)^\eta) \le \mathcal{O}(j(\rho_n^\alpha - \rho_n^\eta)^d(\rho_n)^{-\eta\epsilon}) = \mathcal{O}(j(\rho_n)^{\alpha d - \eta\epsilon}). \tag{A.31}$$

Finally from the obvious observation $k_n \le |W_n|/(\rho_n)^{\alpha d} = \mathcal{O}((\rho_n)^{d-\alpha d})$ and from (A.30) we find

$$|\phi_n(t) - \phi_n'(t)| \le k_n\mathcal{O}(k_n(\rho_n)^{\alpha d - \eta\epsilon}) \le \mathcal{O}((\rho_n)^{2d - \alpha d - \eta\epsilon}), \tag{A.32}$$

which under the assumptions we made about α and η goes to 0 as $\rho_n \to \infty$ and (A.21) is proved.

(A.22) is just an application of the Lyapunov central limit theorem.

Finally since the convergence of $\sqrt{|W_n \ominus R|}\, c \cdot U_n(\theta_0) \to N(0, c\,\Sigma(\theta_0)c^T)$ in distribution for any $c \in \mathbf{R}^q$ follows the same type of derivations as above, the proof of the asymptotic normality of the vector $\sqrt{|W_n \ominus R|}\, U_n(\theta_0)$ follows directly from the Cramér-Wold device. □

Acknowledgements

This work was supported by projects GAČR 201/08/P100 and 201/10/0472 from the Czech Science Foundation and MSM0021620839 financed by the Ministry of Education of the Czech Republic, and by Centre for Stochastic Geometry and Advanced Bioimaging, funded by the Villum Foundation.

References

[1] Baddeley, A. J. and Turner, R. (2000). Practical maximum pseudolikelihood for spatial point processes. *Aust. N. Z. J. Stat.* **42**, 283–322.

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

[3] Crowder, M. J. (1986). On Consistency and Inconsistency of Estimating Equations. *Econometric Theory* **2**, 305–330.

[4] Daley, D. J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes*. 1st edition. New York: Springer Verlag.

[5] Davidson, J. (1994). *Stochastic limit theory*. New York: Oxford University Press.

[6] Diggle, P. J. (2003). *Statistical Analysis of Spatial Point Patterns*. New York: Oxford University Press.

[7] Doukhan, P. (1994). *Mixing: Properties and Examples*. New York: Springer Verlag.

[8] Guan, Y. (2006). A composite likelihood approach in fitting spatial point process models. *J. Am. Stat. Assoc.* **101**, 1502–1512.

[9] Guan, Y., Sherman, M. (2007). On least squares fitting for stationary spatial point processes. *J. Roy. Statist. Soc. Ser. B* **69**, 31ï£¡-49.

[10] Guan, Y., Sherman, M., Calvin J. A. (2007). On asymptotic properties of the mark variogram estimator of a marked point process. *Journal of Statistical Planning and Inference* **137**, 148–161.

[11] Guyon, X. (1995). *Random fields on a network*. New York: Springer Verlag.

[12] Heagerty, P. J. and Lumley, T. (2000). Window subsampling of estimating functions with application to regression models. *J. Am. Stat. Assoc.* **95**, 197–211.

[13] Heinrich, L. (1988). Asymptotic Gaussianity of some estimators for reduced factorial moment measures and product densities of stationary poisson cluster processes. *Statistics* **19**, 87–106.

[14] Heinrich, L. (1992). Minimum contrast estimates for parameters of spatial ergodic point processes. In *Transactions of the 11th Prague Conference on Random Processes, Information Theory and Statistical Decision Functions*, Prague: Academic Publishing House.

[15] Ibragimov, I. A. and Linnik, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff.

[16] Jolivet, E. (1981). Central limit theorem and convergence of empirical processes of stationary point processes. In *Point Processes and Queueing Problems* (P.Bartfai and J.Tomko eds.), Amsterdam: North–Holland.

[17] Lindsay, B.G. (1988). Composite Likelihood Methods. *Contemporary Mathematics* **80**, 221–239.

[18] Møller, J., Syversveen, A.R. and Waagepetersen, R.P. (1998). Log Gaussian Cox processes. *Scand. J. Statist.* **25**, 451–482.

[19] Møller, J. and Waagepetersen, R.P. (2003). *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton: Chapman & Hall/CRC.

[20] Møller, J. and Waagepetersen, R. P. (2007). Modern statistics for spatial point processes. *Scand. J. Statist.* **34**, 643–684.

[21] Ogata, Y. and Katsura, K. (1991). Maximum likelihood estimates of fractal dimension for random spatial patterns. *Biometrika* **78**, 463–474.

[22] Politis, D. N. and Sherman, M. (2001). Moment estimation for statistics from marked point processes. *J. Roy. Statist. Soc. Ser. B* **63**, 261–275.

[23] Rosenblatt, M. (1956). A Central Limit Theorem and a Strong Mixing Condition. *Proceedings of the National Academy of Sciences* **42**, 43–47.

[24] Stein, M. L. (1999). *Interpolation of Spatial Data*. New York: Springer Verlag.

[25] Stoyan, D., Kendall, W. S. and Mecke, J. (1995). *Stochastic Geometry and its Applications*. Second edition. Chichester: Wiley.

[26] Tanaka, U., Ogata, Y., Stoyan, D. (2007). Parameter Estimation and Model Selection for Neyman-Scott Point Processes. *Biometrical Journal* **49**, 1–15.

[27] Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology, *Biometrika* **36**, 18–25.

[28] Wills, J.M. (1970). Zum Verhältnis von Volumen zu Oberfläche bei Convexen Körpern. *Archiv der Mathematik* **21**, 557–560.

[29] Zhengyan, L. and Chuanrong, L. (1996). *Limit Theory for Mixing Dependent Random Variables*. Dordrecht: Kluwer Academic Publishers.

Department of Probability and Statistics
Faculty of Mathematics and Physics
Charles University
Sokolovská 83
186 75 Prague
Czech Republic
Email: prokesov@karlin.mff.cuni.cz

Centre for Stochastic Geometry
and Advanced Bioimaging
Department of Mathematical Sciences
Aarhus University
Ny Munkegade 118, Building 1530
DK-8000 Aarhus C
Denmark
Email: eva@imf.au.dk