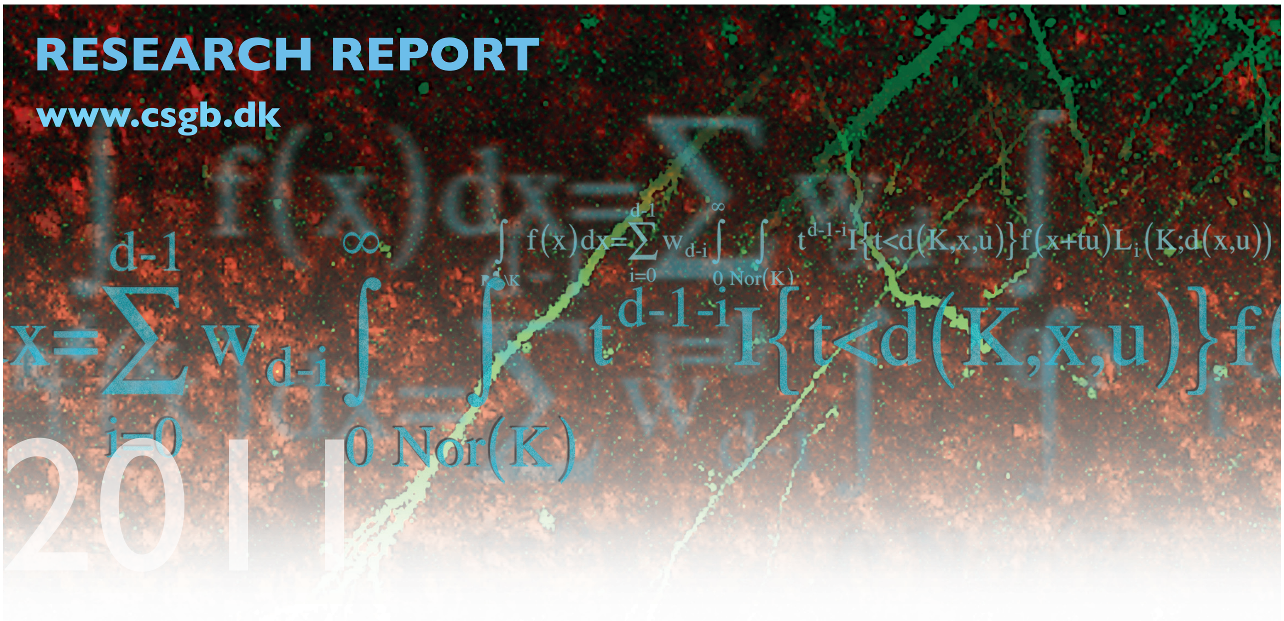
CENTRE FOR **STOCHASTIC GEOMETRY**
AND ADVANCED **BIOIMAGING**



Yongtao Guan, Abdollah Jalilian and Rasmus Waagepetersen

Optimal Estimation of the Intensity Function of a Spatial Point Process

No. 07, November 2011

Optimal Estimation of the Intensity Function of a Spatial Point Process

Yongtao Guan[1], Abdollah Jalilian[2] and Rasmus Waagepetersen[3]

[1]Department of Management Science, University of Miami,, Coral Gables, Florida 33124-6544, USA

[2]Department of Statistics, Razi University, , Bagh-e-Abrisham, Kermanshah, 67149-67346, Iran

[3]Department of Mathematical Sciences, Aalborg University,, Fredrik Bajersvej 7G, DK-9220 Aalborg, Denmark

Abstract

Although optimal from a theoretical point of view, maximum likelihood estimation for Cox and cluster point processes can be cumbersome in practice due to the complicated nature of the likelihood function and the associated score function. It is therefore of interest to consider alternative more easily computable estimating functions. We derive the optimal estimating function in a class of first-order estimating functions. The optimal estimating function depends on the solution of a certain Fredholm integral equation and reduces to the likelihood score in case of a Poisson process. We discuss the numerical solution of the Fredholm integral equation and note that a special case of the approximated solution is equivalent to a quasi-likelihood for binary spatial data. The practical performance of the optimal estimating function is evaluated in a simulation study and a data example.

Keywords: Estimating function, Fredholm integral equation, Godambe information, Intensity function, Quasi-likelihood, Spatial point process.

1 Introduction

Maximum likelihood estimation for spatial point processes such as Cox and cluster point processes is in general not easy from a computational point of view (see e.g. Møller and Waagepetersen, 2004). The intensity function on the other hand often has a simple explicit form and this enables the construction of simple estimating functions. For example, composite likelihood arguments (e.g. Møller and Waagepetersen, 2007) lead to an estimating function that is equivalent to the score of the Poisson maximum likelihood function. This provides a computationally tractable estimating

Corresponding author: Rasmus Waagepetersen, rw@math.aau.dk

function for estimation of parameters in the intensity function. Theoretical properties of the resulting estimator are well understood, see e.g. Schoenberg (2005), Waagepetersen (2007) and Guan and Loh (2007).

A drawback of the Poisson score function approach is the loss of efficiency since possible dependence between points is ignored. In the context of intensity estimation, it appears that only Mrkvička and Molchanov (2005) and Guan and Shen (2010) have tried to incorporate second-order properties (e.g. the pair correlation function which is often available in explicit forms for processes such as Cox and cluster point processes) in the estimation so as to improve efficiency. Mrkvička and Molchanov (2005) show that their proposed estimator is optimal among a class of linear, unbiased intensity estimators, where the word 'optimal' refers to minimum variance. However, their approach is restricted to a special type of spatial point processes whose intensity function is given up to an unknown scaling factor. In contrast, Guan and Shen (2010) propose a weighted estimating equation approach that is applicable to intensity functions in more general forms. However, a similar optimality result cannot be established for their approach. We show in Section 3.2 that the optimality result in Mrkvička and Molchanov (2005) is a special case of our more general result, and that the estimation method in Guan and Shen (2010) is only a crude approximation of our new approach.

For many types of correlated data other than spatial point patterns, estimating function based procedures have been widely used for model fitting when maximum likelihood estimation is computationally challenging. Examples of such data include longitudinal data (Liang and Zeger, 1986), time series data (Zeger, 1988), clustered failure time data (Gray, 2003) and spatial binary or count data (Gotway and Stroup, 1997; Lin and Clayton, 2005). For most of these methods, the inverse of a covariance matrix is used in their formulations as a way to account for the correlation in data, and optimality can be established when the so-called quasi-score estimating functions are used (Heyde, 1997). For spatial point processes, a similar covariance matrix cannot be constructed because the data are defined over a continuous spatial domain and hence are of infinite dimension. Heyde (1997) discusses generalizations of the quasi-score estimating functions to processes that are defined over continuous time and possess a special semimartingale representation. However, such a representation is generally not possible for spatial point processes.

In this paper we develop an optimal estimating function for intensity estimation that takes into account possible spatial correlation. The optimal estimating function depends on the solution of a certain Fredholm integral equation and reduces to the likelihood score in case of a Poisson process. We derive asymptotic properties of the resulting parameter estimator, and discuss the practical implementation of our proposed method based on a numerical solution of the Fredholm integral equation. We further show that a discretized version of our method is closely related to the quasi-likelihood for spatial data (Gotway and Stroup, 1997; Lin and Clayton, 2005). Our work hence not only lays the theoretical foundation for optimal intensity estimation, but also fills in a critical gap between existing literature on spatial point processes and other types of (discrete) stochastic processes. We illustrate the superior performance of our proposed approach over existing ones through a simulation study, and we apply it to some real data examples.

2 Background

2.1 Intensity and Pair Correlation Function

Let X be a point process on $\mathbb{R}^2$ and let $N(B)$ denote the number of points in $X \cap B$ for any bounded set $B \subseteq \mathbb{R}^2$. The first- and second-order moments of the counts $N(B)$ are determined by the intensity function $\rho(\cdot)$ and the pair correlation function $g(\cdot, \cdot)$, respectively, see Møller and Waagepetersen (2004). More precisely,

$$\mathbb{E}N(B) = \int_B \rho(\mathbf{u})\mathrm{d}\mathbf{u} \tag{2.1}$$

and

$$\mathbb{C}\mathrm{ov}[N(A), N(B)] = \int_{A\cap B} \rho(\mathbf{u})\mathrm{d}\mathbf{u} + \int_A \int_B \rho(\mathbf{u})\rho(\mathbf{v})[g(\mathbf{u}, \mathbf{v}) - 1]\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v} \tag{2.2}$$

for bounded sets $A, B \subseteq \mathbb{R}^2$. For convenience of exposition we assume that $g(\mathbf{u}, \mathbf{v})$ only depends on the difference $\mathbf{u}-\mathbf{v}$ since this is the common assumption in practice. In the following we thus let $g(\mathbf{r})$ denote the pair correlation function for two points $\mathbf{u}$ and $\mathbf{v}$ with $\mathbf{u} - \mathbf{v} = \mathbf{r}$. However, our proposed optimal estimating function is applicable also in the case of a non-translation invariant pair correlation function.

2.2 Composite Likelihood

Assume that the intensity function is given in terms of a parametric model $\rho(\mathbf{u}) = \rho(\mathbf{u}; \boldsymbol{\beta})$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p) \in \mathbb{R}^p$ is a vector of regression parameters. Popular choices of the parametric model include linear and log linear models, $\rho(\mathbf{u}; \boldsymbol{\beta}) = \mathbf{z}(\mathbf{u})\boldsymbol{\beta}^{\mathsf{T}}$ and $\log \rho(\mathbf{u}; \boldsymbol{\beta}) = \mathbf{z}(\mathbf{u})\boldsymbol{\beta}^{\mathsf{T}}$, where $\mathbf{z}(\mathbf{u}) = (z_1(\mathbf{u}), \ldots, z_p(\mathbf{u}))$ is a covariate vector for each $\mathbf{u} \in \mathbb{R}^2$. A first-order log composite likelihood function (Schoenberg, 2005; Waagepetersen, 2007) for estimation of $\boldsymbol{\beta}$ is given by

$$\sum_{\mathbf{u}\in X\cap W} \log \rho(\mathbf{u}; \boldsymbol{\beta}) - \int_W \rho(\mathbf{u}; \boldsymbol{\beta})\mathrm{d}\mathbf{u}, \tag{2.3}$$

where $W \subset \mathbb{R}^2$ is the observation window. This can be viewed as a limit of log composite likelihood functions for binary variables $Y_i = 1[N(B_i) > 0]$, $i = 1, \ldots, m$, where the cells B_i form a disjoint partitioning of W and $1[\cdot]$ is an indicator function (e.g. Møller and Waagepetersen, 2007). The limit is obtained when the number of cells tends to infinity and the areas of the cells tend to zero. In case of a Poisson process, the composite likelihood coincides with the likelihood function.

The composite likelihood is computationally simple and enjoys considerable popularity in particular in studies of tropical rain forest ecology where spatial point process models are fitted to huge spatial point pattern data sets of rain forest tree locations (see e.g. Shen et al., 2009; Lin et al., 2011). However, it is not statistically efficient for non-Poisson data since possible correlations between counts of points are ignored.

3 An optimal first-order estimating equation

A first-order estimating function is an estimating function of the form

$$\mathbf{e}_{\mathbf{f}}(\boldsymbol{\beta}) = \sum_{\mathbf{u}\in X\cap W} \mathbf{f}(\mathbf{u}) - \int_W \mathbf{f}(\mathbf{u})\rho(\mathbf{u};\boldsymbol{\beta})\mathrm{d}\mathbf{u}, \tag{3.1}$$

where $\mathbf{f}(\mathbf{u})$ is a $1\times p$ real function that possibly depends on $\boldsymbol{\beta}$. Let $\boldsymbol{\Sigma}_{\mathbf{f}} = \mathbb{V}\mathrm{ar}\mathbf{e}_{\mathbf{f}}(\boldsymbol{\beta})$, $\mathbf{J}_{\mathbf{f}} = -\mathrm{d}\mathbf{e}_{\mathbf{f}}(\boldsymbol{\beta})/\mathrm{d}\boldsymbol{\beta}^{\mathsf{T}}$ and $\mathbf{S}_{\mathbf{f}} = \mathbb{E}\mathbf{J}_{\mathbf{f}}$. Note that $\mathbb{E}$, $\mathbb{V}\mathrm{ar}$, $\boldsymbol{\Sigma}_{\mathbf{f}}$, $\mathbf{J}_{\mathbf{f}}$ and $\mathbf{S}_{\mathbf{f}}$ all depend on $\boldsymbol{\beta}$ but we suppress this dependence in this section for ease of presentation. The matrix $\mathbf{S}_{\mathbf{f}}$ is called the sensitivity and $\mathbf{S}_{\mathbf{f}}\boldsymbol{\Sigma}_{\mathbf{f}}^{-1}\mathbf{S}_{\mathbf{f}}$ is the Godambe information. Our aim is to find a function $\boldsymbol{\phi}$ so that $\mathbf{e}_{\boldsymbol{\phi}}$ is optimal in the sense that

$$\mathbf{S}_{\boldsymbol{\phi}}\boldsymbol{\Sigma}_{\boldsymbol{\phi}}^{-1}\mathbf{S}_{\boldsymbol{\phi}} - \mathbf{S}_{\mathbf{f}}\boldsymbol{\Sigma}_{\mathbf{f}}^{-1}\mathbf{S}_{\mathbf{f}} \tag{3.2}$$

is non-negative definite for all $\mathbf{f} : W \to \mathbb{R}^p$, i.e., $\mathbf{e}_{\boldsymbol{\phi}}$ has maximal Godambe information. Let $\hat{\mathbf{e}}_{\boldsymbol{\phi}}(\boldsymbol{\beta}) = \mathbf{e}_{\mathbf{f}}(\boldsymbol{\beta})\boldsymbol{\Sigma}_{\mathbf{f}}^{-1}\boldsymbol{\Sigma}_{\mathbf{f}\boldsymbol{\phi}}$ be the optimal linear predictor of $\mathbf{e}_{\boldsymbol{\phi}}(\boldsymbol{\beta})$ given $\mathbf{e}_{\mathbf{f}}(\boldsymbol{\beta})$ where $\boldsymbol{\Sigma}_{\mathbf{f}\boldsymbol{\phi}} = \mathbb{C}\mathrm{ov}[\mathbf{e}_{\mathbf{f}}(\boldsymbol{\beta}), \mathbf{e}_{\boldsymbol{\phi}}(\boldsymbol{\beta})]$. Then

$$\mathbb{V}\mathrm{ar}[\hat{\mathbf{e}}_{\boldsymbol{\phi}}(\boldsymbol{\beta}) - \mathbf{e}_{\boldsymbol{\phi}}(\boldsymbol{\beta})] = \boldsymbol{\Sigma}_{\boldsymbol{\phi}} - \boldsymbol{\Sigma}_{\boldsymbol{\phi}\mathbf{f}}\boldsymbol{\Sigma}_{\mathbf{f}}^{-1}\boldsymbol{\Sigma}_{\mathbf{f}\boldsymbol{\phi}}$$

is non-negative definite whereby

$$\mathbf{S}_{\boldsymbol{\phi}}\boldsymbol{\Sigma}_{\boldsymbol{\phi}}^{-1}\mathbf{S}_{\boldsymbol{\phi}} - \mathbf{S}_{\boldsymbol{\phi}}\boldsymbol{\Sigma}_{\boldsymbol{\phi}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\phi}\mathbf{f}}\boldsymbol{\Sigma}_{\mathbf{f}}^{-1}\boldsymbol{\Sigma}_{\mathbf{f}\boldsymbol{\phi}}\boldsymbol{\Sigma}_{\boldsymbol{\phi}}^{-1}\mathbf{S}_{\boldsymbol{\phi}}$$

is non-negative definite too. Hence, (3.2) is non-negative definite provided

$$\mathbf{S}_{\boldsymbol{\phi}}\boldsymbol{\Sigma}_{\boldsymbol{\phi}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\phi}\mathbf{f}} = \mathbf{S}_{\mathbf{f}}$$

which holds if $\boldsymbol{\Sigma}_{\boldsymbol{\phi}\mathbf{f}} = \mathbf{S}_{\mathbf{f}}$ for all $\mathbf{f}$ (in particular, this implies $\boldsymbol{\Sigma}_{\boldsymbol{\phi}} = \boldsymbol{\Sigma}_{\boldsymbol{\phi}\boldsymbol{\phi}} = \mathbf{S}_{\boldsymbol{\phi}}$). By the Campbell formulae (e.g. Møller and Waagepetersen, 2004, Chapter 4),

$$\begin{aligned}
\boldsymbol{\Sigma}_{\boldsymbol{\phi}\mathbf{f}} &= \int_W \mathbf{f}^{\mathsf{T}}(\mathbf{u})\boldsymbol{\phi}(\mathbf{u})\rho(\mathbf{u};\boldsymbol{\beta})\mathrm{d}\mathbf{u} + \int_{W^2} \mathbf{f}^{\mathsf{T}}(\mathbf{u})\boldsymbol{\phi}(\mathbf{v})\rho(\mathbf{u};\boldsymbol{\beta})\rho(\mathbf{v};\boldsymbol{\beta})[g(\mathbf{u}-\mathbf{v})-1]\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v},\\
\mathbf{S}_{\mathbf{f}} &= \int_W \mathbf{f}^{\mathsf{T}}(\mathbf{u})\boldsymbol{\rho}'(\mathbf{u};\boldsymbol{\beta})\mathrm{d}\mathbf{u},
\end{aligned}$$

where $\boldsymbol{\rho}'(\mathbf{u};\boldsymbol{\beta}) = \mathrm{d}\rho(\mathbf{u};\boldsymbol{\beta})/\mathrm{d}\boldsymbol{\beta}$. Hence, $\mathbf{S}_{\mathbf{f}} = \boldsymbol{\Sigma}_{\boldsymbol{\phi}\mathbf{f}}$ is equivalent to

$$\int_W \mathbf{f}_{\boldsymbol{\beta}}^{\mathsf{T}}(\mathbf{u})\Big\{\boldsymbol{\rho}'(\mathbf{u};\boldsymbol{\beta}) - \boldsymbol{\phi}(\mathbf{u})\rho(\mathbf{u};\boldsymbol{\beta}) - \rho(\mathbf{u};\boldsymbol{\beta})\int_W \phi(\mathbf{v})\rho(\mathbf{v};\boldsymbol{\beta})[g(\mathbf{u}-\mathbf{v})-1]\mathrm{d}\mathbf{v}\Big\}\mathrm{d}\mathbf{u} = \mathbf{0}.$$

Assuming $\rho > 0$ we should thus choose $\boldsymbol{\phi}$ as a solution of the Fredholm integral equation (e.g. Hackbusch, 1995, Chapter 3)

$$\boldsymbol{\phi} = \frac{\boldsymbol{\rho}'}{\rho} - \mathbf{T}\boldsymbol{\phi}, \tag{3.3}$$

where $\mathbf{T}$ is the operator given by

$$(\mathbf{T}\mathbf{f})(\mathbf{u}) = \int_W t(\mathbf{u},\mathbf{v})\mathbf{f}(\mathbf{v})\mathrm{d}\mathbf{v} \quad \text{with} \quad t(\mathbf{u},\mathbf{v}) = \rho(\mathbf{v};\boldsymbol{\beta})[g(\mathbf{u}-\mathbf{v})-1]. \tag{3.4}$$

Assume that ρ and g are continuous so that $\mathbf{T}$ is compact in the space of continuous functions on W (Hackbusch, 1995, Theorem 3.2.5) and moreover that -1 is not an eigenvalue (we return to this condition in the next section). It then follows by Theorem 3.2.1 in Hackbusch (1995) that (3.3) has a unique solution

$$\boldsymbol{\phi} = (\mathbf{I}+\mathbf{T})^{-1}\frac{\boldsymbol{\rho}'}{\rho},$$

where $\mathbf{I}$ is the identity operator (or, depending on context, the identity matrix) and $(\mathbf{I}+\mathbf{T})^{-1}$ is the bounded linear inverse of $\mathbf{I}+\mathbf{T}$. We define

$$\mathbf{e}(\boldsymbol{\beta}) = \mathbf{e}_{\boldsymbol{\phi}}(\boldsymbol{\beta}) = \sum_{\mathbf{u}\in X\cap W} \boldsymbol{\phi}(\mathbf{u}) - \int_W \boldsymbol{\phi}(\mathbf{u})\rho(\mathbf{u};\boldsymbol{\beta})\mathrm{d}\mathbf{u}, \tag{3.5}$$
$$\boldsymbol{\Sigma} = \mathbb{V}\mathrm{ar}\mathbf{e}(\boldsymbol{\beta}),\ \mathbf{J} = -\mathrm{d}\mathbf{e}(\boldsymbol{\beta})/\mathrm{d}\boldsymbol{\beta}^{\mathsf{T}},\ \mathbf{S} = \mathbb{E}\mathbf{J}$$

where by the above derivations,

$$\mathbf{S} = \boldsymbol{\Sigma} = \int_W \boldsymbol{\phi}^{\mathsf{T}}(\mathbf{u})\boldsymbol{\rho}'(\mathbf{u};\boldsymbol{\beta})\mathrm{d}\mathbf{u}. \tag{3.6}$$

In the Poisson process case where $g(\cdot) = 1$, (3.5) reduces to

$$\sum_{\mathbf{u}\in X\cap W} \frac{\boldsymbol{\rho}'(\mathbf{u};\boldsymbol{\beta})}{\rho(\mathbf{u};\boldsymbol{\beta})} - \int_W \boldsymbol{\rho}'(\mathbf{u};\boldsymbol{\beta})\mathrm{d}\mathbf{u}$$

which is precisely the score of the Poisson log likelihood (2.3).

3.1 Condition for non-negative eigenvalues of T

In general it is difficult to assess the eigenvalues of $\mathbf{T}$ given by (3.4). However, suppose that $g-1$ is non-negative definite so that $\mathbf{T}^s$ is a positive operator (i.e., $\int_W \mathbf{f}^{\mathsf{T}}(\mathbf{u})(\mathbf{T}^s\mathbf{f})(\mathbf{u})\mathrm{d}\mathbf{u} \geq 0$) where $\mathbf{T}^s$ is given by the symmetric kernel

$$t^s(\mathbf{u},\mathbf{v}) = \rho(\mathbf{u};\boldsymbol{\beta})^{1/2}\rho(\mathbf{v};\boldsymbol{\beta})^{1/2}\big[g(\mathbf{u}-\mathbf{v})-1\big].$$

Then all eigenvalues of $\mathbf{T}^s$ are non-negative (Lax, 2002, Corollary 1, p. 320). In particular, -1 is not an eigenvalue. The same holds for $\mathbf{T}$ since it is easy to see that the eigenvalues of $\mathbf{T}$ coincide with those of $\mathbf{T}^s$.

The assumption of a non-negative definite $g(\cdot)-1$ is valid for the wide class of Cox point processes which in turn includes the class of cluster processes with Poisson clusters. For a Cox process driven by a random intensity function Λ, $g(u,v) = 1+\mathbb{C}\mathrm{ov}[\Lambda(\mathbf{u}),\Lambda(\mathbf{v})]/[\rho(\mathbf{u})\rho(\mathbf{v})]$ so that $g(\cdot)-1$ is non-negative definite.

3.2 Relation to Existing Methods

Suppose we approximate the operator $\mathbf{T}$ by

$$(\mathbf{T}\mathbf{f})(\mathbf{u}) = \int_W \mathbf{f}(\mathbf{v})\rho(\mathbf{v};\boldsymbol{\beta})[g(\mathbf{u}-\mathbf{v})-1)]\mathrm{d}\mathbf{v} \approx \rho(\mathbf{u};\boldsymbol{\beta})\mathbf{f}(\mathbf{u})\int_W [g(\mathbf{u}-\mathbf{v})-1]\mathrm{d}\mathbf{v}. \tag{3.7}$$

This is justified if $\mathbf{f}(\mathbf{v})\rho(\mathbf{v};\boldsymbol{\beta})$ is close to $\mathbf{f}(\mathbf{u})\rho(\mathbf{u};\boldsymbol{\beta})$ for the $\mathbf{v}$ where $g(\mathbf{u}-\mathbf{v})-1$ differs substantially from zero. Then the Fredholm integral equation (3.3) can be approximated by

$$\boldsymbol{\phi} = \frac{\boldsymbol{\rho}'}{\rho} - \rho A\boldsymbol{\phi},$$

where

$$A(\mathbf{u}) = \int_W \big[g(\mathbf{u}-\mathbf{v})-1\big]\mathrm{d}\mathbf{v}.$$

We hence obtain an approximate solution $\boldsymbol{\phi} = w\boldsymbol{\rho}'/\rho$ with $w(\mathbf{u}) = [1+\rho(\mathbf{u};\boldsymbol{\beta})A(\mathbf{u})]^{-1}$. Using this approximation in (3.5) we obtain the estimating function

$$\sum_{\mathbf{u}\in X\cap W} w(\mathbf{u})\frac{\boldsymbol{\rho}'(\mathbf{u};\boldsymbol{\beta})}{\rho(\mathbf{u};\boldsymbol{\beta})} - \int_W w(\mathbf{u})\boldsymbol{\rho}'(\mathbf{u};\boldsymbol{\beta})\mathrm{d}\mathbf{u},$$

which is precisely the weighted Poisson score suggested in Guan and Shen (2010).

Mrkvička and Molchanov (2005) derived optimal intensity estimators in the situation of $\rho(\mathbf{u};\lambda) = \lambda\gamma(\mathbf{u})$ for some known function $\gamma(\mathbf{u})$ and unknown parameter $\lambda > 0$. Since λ is the only unknown parameter, a direct application of (3.3) yields

$$\lambda\boldsymbol{\phi}(\mathbf{u}) + \lambda^2\int_W \boldsymbol{\phi}(\mathbf{v})\gamma(\mathbf{v})\big[g(\mathbf{u}-\mathbf{v})-1\big]\mathrm{d}\mathbf{v} = 1,$$

which is essentially the same as the result in Theorem 2.1 of Mrkvička and Molchanov (2005).

3.3 Solution Using Neumann Series Expansion

Suppose that $\|\mathbf{T}\|_{\mathrm{op}} = \sup\{\|\mathbf{Tf}\|_\infty/\|\mathbf{f}\|_\infty : \|\mathbf{f}\|_\infty \neq 0\} < 1$ where $\|\mathbf{f}\|_\infty$ denotes the supremum norm of a continuous function $\mathbf{f}$ on W. Then we can obtain the solution $\boldsymbol{\phi}$ of (3.3) using a Neumann series expansion which may provide additional insight on the properties of $\boldsymbol{\phi}$. More specifically,

$$\boldsymbol{\phi} = \sum_{k=0}^{\infty}(-\mathbf{T})^k\frac{\boldsymbol{\rho}'}{\rho}. \tag{3.8}$$

If the infinite sum in (3.8) is truncated to the first term ($k=0$) then (3.5) becomes the Poisson score. Note that

$$\|\mathbf{T}\|_\infty \leq \sup_{\mathbf{u}\in W}\int_W |t(\mathbf{u},\mathbf{v})|\mathrm{d}\mathbf{v}.$$

Hence, a sufficient condition for the validity of the Neumann series expansion is

$$\sup_{\mathbf{u}\in W}\rho(\mathbf{u};\boldsymbol{\beta})\int_{\mathbb{R}^2}\big|g(\mathbf{r})-1\big|\mathrm{d}\mathbf{r} < 1. \tag{3.9}$$

Condition (3.9) roughly requires that $g(\mathbf{r})-1$ does not decrease too slowly to zero and/or that ρ is moderate. For example, suppose that g is the pair correlation function of a Thomas cluster process (e.g. Møller and Waagepetersen, 2004, Chapter 5),

$$g(\mathbf{r}) - 1 = \exp\big[-\|\mathbf{r}\|^2/(4\omega^2)\big]/(4\pi\omega^2\kappa), \quad \text{for some } \kappa,\omega > 0, \tag{3.10}$$

where κ is the intensity of the parent process and ω is the normal dispersal parameter. Then,

$$\int_{\mathbb{R}^2} \big|g(\mathbf{r}) - 1\big| \mathrm{d}\mathbf{r} = \frac{1}{4\pi\kappa\omega^2} \int_{\mathbb{R}^2} \exp(-\frac{\|\mathbf{r}\|^2}{4\omega^2}) \mathrm{d}\mathbf{r} = 1/\kappa$$

and (3.9) is equivalent to $\sup_{\mathbf{u}\in W} \rho(\mathbf{u};\boldsymbol{\beta}) < \kappa$. In this case, Condition (3.9) can be quite restrictive. However, the Neumann series expansion is not essential for our approach and we use it only for checking the conditions for asymptotic results; see Appendix A.

4 Asymptotic theory

Let $W_n \subset \mathbb{R}^2$ be an increasing sequence of observation windows in $\mathbb{R}^2$. We assume that the true pair correlation function is given by a parametric model $g(\mathbf{r}) = g(\mathbf{r};\boldsymbol{\psi})$ for some unknown parameter vector $\boldsymbol{\psi} \in \mathbb{R}^q$. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\psi}) \in \mathbb{R}^{p+q}$. We denote the true value of $\boldsymbol{\theta}$ by $\boldsymbol{\theta}^* = (\boldsymbol{\beta}^*, \boldsymbol{\psi}^*)$. In what follows, $\mathbb{E}$ and $\mathbb{V}\mathrm{ar}$ denote expectation and variance under the distribution corresponding to $\boldsymbol{\theta}^*$.

Introducing the dependence on n and $\boldsymbol{\theta}$ in the notation from Section 3, we have

$$\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{u},\boldsymbol{\beta}) = \Big[(\mathbf{I} + \mathbf{T}_{n,\boldsymbol{\theta}})^{-1} \frac{\boldsymbol{\rho}'(\cdot;\boldsymbol{\beta})}{\rho(\cdot;\boldsymbol{\beta})}\Big](\mathbf{u}), \quad (\mathbf{T}_{n,\boldsymbol{\theta}}\mathbf{f})(\mathbf{u}) = \int_{W_n} t_{\boldsymbol{\theta}}(\mathbf{u},\mathbf{v})\mathbf{f}(\mathbf{v})\mathrm{d}\mathbf{v}$$

and

$$t_{\boldsymbol{\theta}}(\mathbf{u},\mathbf{v}) = \rho(\mathbf{v};\boldsymbol{\beta})\big[g(\mathbf{u}-\mathbf{v};\boldsymbol{\psi}) - 1\big].$$

Following Section 5.3 we replace $\boldsymbol{\theta}$ in the kernel $t_{\boldsymbol{\theta}}$ by a preliminary estimate $\tilde{\boldsymbol{\theta}}_n = (\tilde{\boldsymbol{\beta}}_n, \tilde{\boldsymbol{\psi}}_n)$. The estimating function (3.5) then becomes $\mathbf{e}_{n,\tilde{\boldsymbol{\theta}}_n}(\boldsymbol{\beta})$ where

$$\mathbf{e}_{n,\boldsymbol{\theta}}(\boldsymbol{\beta}) = \sum_{\mathbf{u}\in X\cap W_n} \boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{u},\boldsymbol{\beta}) - \int_{W_n} \boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{u},\boldsymbol{\beta})\rho(\mathbf{u};\boldsymbol{\beta})\mathrm{d}\mathbf{u}.$$

Let $\hat{\boldsymbol{\beta}}_n$ denote the estimator obtained by solving $\mathbf{e}_{n,\tilde{\boldsymbol{\theta}}_n}(\boldsymbol{\beta}) = 0$. Further, define

$$\bar{\boldsymbol{\Sigma}}_n = |W_n|^{-1}\mathbb{V}\mathrm{ar}\mathbf{e}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*), \qquad \mathbf{J}_{n,\boldsymbol{\theta}}(\boldsymbol{\beta}) = -\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}^\mathsf{T}}\mathbf{e}_{n,\boldsymbol{\theta}}(\boldsymbol{\beta})$$
$$\text{and} \qquad \bar{\mathbf{S}}_{n,\boldsymbol{\theta}}(\boldsymbol{\beta}) = |W_n|^{-1}\mathbb{E}\mathbf{J}_{n,\boldsymbol{\theta}}(\boldsymbol{\beta}).$$

Note that $\bar{\boldsymbol{\Sigma}}_n$ and $\bar{\mathbf{S}}_{n,\boldsymbol{\theta}}(\boldsymbol{\beta})$ are 'averaged' versions of $\boldsymbol{\Sigma}_n = \mathbb{V}\mathrm{ar}\mathbf{e}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)$ and $\mathbf{S}_{n,\boldsymbol{\theta}}(\boldsymbol{\beta}) = \mathbb{E}\mathbf{J}_{n,\boldsymbol{\theta}}(\boldsymbol{\beta})$.

In Appendix B we verify the existence of a $|W_n|^{1/2}$ consistent sequence of solutions $\hat{\boldsymbol{\beta}}_n$, i.e., $|W_n|^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*)$ is bounded in probability. We further show in Appendix C that $|W_n|^{-1/2}\mathbf{e}_{n,\tilde{\boldsymbol{\theta}}_n}(\boldsymbol{\beta}^*)\bar{\boldsymbol{\Sigma}}_n^{-1/2}$ is asymptotically standard normal. The conditions needed for these results are listed in Appendix A. It then follows by a Taylor series expansion,

$$|W_n|^{-1/2}\mathbf{e}_{n,\tilde{\boldsymbol{\theta}}_n}(\boldsymbol{\beta}^*)\bar{\boldsymbol{\Sigma}}_n^{-1/2} = |W_n|^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*)\frac{\mathbf{J}_{n,\tilde{\boldsymbol{\theta}}_n}(\mathbf{b}_n)}{|W_n|}\bar{\boldsymbol{\Sigma}}_n^{-1/2}$$

for some $\mathbf{b}_n \in \mathbb{R}^p$ satisfying $\|\mathbf{b}_n - \boldsymbol{\beta}^*\| \leq \|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*\|$, and R2 and R3 in Appendix B that

$$|W_n|^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*)\bar{\mathbf{S}}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)\bar{\boldsymbol{\Sigma}}_n^{-1/2} \to N_p(0, \mathbf{I}).$$

Hence, for a fixed n and since $\bar{\boldsymbol{\Sigma}}_n = \bar{\mathbf{S}}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)$ by (3.6), $\hat{\boldsymbol{\beta}}_n$ is approximately normal with mean $\boldsymbol{\beta}^*$ and covariance matrix estimated by $|W_n|^{-1}\bar{\mathbf{S}}^{-1}_{n,(\tilde{\boldsymbol{\psi}}_n,\hat{\boldsymbol{\beta}}_n)}(\hat{\boldsymbol{\beta}}_n)$.

5 Implementation

In this section we discuss practical issues concerning the implementation of our proposed optimal estimating function.

5.1 Numerical Approximation

To estimate $\boldsymbol{\phi}$, consider the numerical approximation

$$(\mathbf{T}\boldsymbol{\phi})(\mathbf{u}) = \int_W t(\mathbf{u}, \mathbf{v})\boldsymbol{\phi}(\mathbf{v})\mathrm{d}\mathbf{v} \approx \sum_{i=1}^{m} t(\mathbf{u}, \mathbf{u}_i)\boldsymbol{\phi}(\mathbf{u}_i)w_i, \tag{5.1}$$

where $\mathbf{u}_i, i = 1, \ldots, m$, are quadrature points with associated weights w_i. An estimate $\hat{\boldsymbol{\phi}}(\mathbf{u}_i)$ of $\boldsymbol{\phi}(\mathbf{u}_i)$ is obtained by solving the linear equations,

$$\boldsymbol{\phi}(\mathbf{u}_i) + \sum_{j=1}^{m} t(\mathbf{u}_i, \mathbf{u}_j)\boldsymbol{\phi}(\mathbf{u}_j)w_j = \frac{\boldsymbol{\rho}'(\mathbf{u}_i; \boldsymbol{\beta})}{\rho(\mathbf{u}_i; \boldsymbol{\beta})}, \quad i = 1, \ldots, m.$$

The Nyström approximate solution of (3.3) is simply

$$\hat{\boldsymbol{\phi}}(\mathbf{u}) = \frac{\boldsymbol{\rho}'(\mathbf{u}; \boldsymbol{\beta})}{\rho(\mathbf{u}; \boldsymbol{\beta})} - \sum_{i=1}^{m} t(\mathbf{u}, \mathbf{u}_i)\hat{\boldsymbol{\phi}}(\mathbf{u}_i)w_i. \tag{5.2}$$

Provided the quadrature scheme is convergent, it follows by Lemma 4.7.4, Lemma 4.7.6 and Theorem 4.7.7 in Hackbusch (1995) that $\|\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}\|_\infty$ converges to zero as $m \to \infty$. This justifies the use of the Nyström method to obtain an approximate solution of the Fredholm integral equation.

Replacing $\boldsymbol{\phi}$ in (3.5) by $\hat{\boldsymbol{\phi}}$, we obtain the estimating function

$$\sum_{\mathbf{u} \in X \cap W} \hat{\boldsymbol{\phi}}(\mathbf{u}) - \int_W \hat{\boldsymbol{\phi}}(\mathbf{u})\rho(\mathbf{u}; \boldsymbol{\beta})\mathrm{d}\mathbf{u}. \tag{5.3}$$

In many cases, the integral in (5.3) has to be numerically approximated. Although a more general quadrature rule could be used, we for simplicity adopt the same rule used to approximate $(\mathbf{T}\boldsymbol{\phi})(\mathbf{u})$. Then, (3.5) is approximated by

$$\hat{\mathbf{e}}(\boldsymbol{\beta}) = \sum_{\mathbf{u} \in X \cap W} \hat{\boldsymbol{\phi}}(\mathbf{u}) - \sum_{i=1}^{m} \hat{\boldsymbol{\phi}}(\mathbf{u}_i)\rho(\mathbf{u}_i; \boldsymbol{\beta})w_i. \tag{5.4}$$

To estimate $\boldsymbol{\beta}$, we solve $\hat{\mathbf{e}}(\boldsymbol{\beta}) = 0$ iteratively using Fisher scoring. Suppose that the current estimate is $\boldsymbol{\beta}^{(l)}$. Then $\boldsymbol{\beta}^{(l+1)}$ is obtained by the Fisher scoring update

$$\boldsymbol{\beta}^{(l+1)} = \boldsymbol{\beta}^{(l)} + \hat{\mathbf{e}}(\boldsymbol{\beta}^{(l)})\hat{\mathbf{S}}^{-1}, \tag{5.5}$$

where

$$\hat{\mathbf{S}} = \sum_{i=1}^{m} \hat{\boldsymbol{\phi}}(\mathbf{u}_i)^{\mathsf{T}} \boldsymbol{\rho}'(\mathbf{u}_i; \boldsymbol{\beta}) w_i.$$

is the numerical approximation of the sensitivity matrix $\mathbf{S} = \int_W \boldsymbol{\phi}^{\mathsf{T}}(\mathbf{u})\boldsymbol{\rho}'(\mathbf{u}; \boldsymbol{\beta})\mathrm{d}\mathbf{u}$.

The simplest quadrature scheme is Riemann quadrature in which case (5.4) and (5.5) takes the form of quasi-likelihood and iterative generalized least squares, respectively, see Section 5.2.

5.2 Implementation as quasi-likelihood

Suppose that we are using the simple Riemann quadrature in (5.1). Then the w_i's correspond to areas of some sets B_i that partition W and for each i, $\mathbf{u}_i \in B_i$. Let Y_i denote the number of events in B_i and define $\mu_i = \rho(\mathbf{u}_i; \boldsymbol{\beta})w_i$. If the B_i's are sufficiently small so that the Y_i's are binary then (5.4) is approximately equal to

$$\sum_{i=1}^{m} \hat{\boldsymbol{\phi}}(\mathbf{u}_i)(Y_i - \mu_i). \tag{5.6}$$

Further, by (2.1) and (2.2), $\mathbb{E}Y_i \approx \mu_i$ and

$$\begin{aligned}\mathbb{C}\mathrm{ov}(Y_i, Y_j) &= 1(i = j)\int_{B_i} \rho(\mathbf{u}; \boldsymbol{\beta})\mathrm{d}\mathbf{u} + \int_{B_i \times B_j} \rho(\mathbf{u}; \boldsymbol{\beta})\rho(\mathbf{v}; \boldsymbol{\beta})\big[g(\mathbf{u} - \mathbf{v}) - 1\big]\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v} \\ &\approx V_{ij} = \mu_i 1(i = j) + \mu_i\mu_j\big[g(\mathbf{u}_i, \mathbf{u}_j) - 1\big].\end{aligned}$$

Define $\mathbf{Y} = (Y_i)_i$, $\boldsymbol{\mu} = (\mu_i)_i$ and $\mathbf{V} = [V_{ij}]_{ij}$. Then $\mathbb{E}\mathbf{Y} \approx \boldsymbol{\mu}$ and $\mathbb{C}\mathrm{ov}\mathbf{Y} \approx \mathbf{V}$. Moreover, from (5.2), $[\hat{\boldsymbol{\phi}}(\mathbf{u}_i)]_i = \mathbf{V}^{-1}\mathbf{D}$ where $\mathbf{D} = \mathrm{d}\boldsymbol{\mu}^{\mathsf{T}}/\mathrm{d}\boldsymbol{\beta}$ is the $m \times p$ matrix of partial derivatives $\mathrm{d}\mu_i/\mathrm{d}\beta_j$. Hence, (5.6) becomes

$$(\mathbf{Y} - \boldsymbol{\mu})\mathbf{V}^{-1}\mathbf{D}, \tag{5.7}$$

which is formally a quasi-likelihood score for spatial data $\mathbf{Y}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{V}$ (Gotway and Stroup, 1997).

Similarly, $\hat{\mathbf{S}} = \mathbf{D}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{D}$ and substituting $\hat{\mathbf{e}}$ in (5.5) by (5.7), we obtain the iterative generalized least squares equation

$$(\boldsymbol{\beta}^{(l+1)} - \boldsymbol{\beta}^{(l)})\mathbf{D}(\boldsymbol{\beta}^{(l)})^{\mathsf{T}}\mathbf{V}(\boldsymbol{\beta}^{(l)})^{-1}\mathbf{D}(\boldsymbol{\beta}^{(l)}) = [\mathbf{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}^{(l)})]\mathbf{V}(\boldsymbol{\beta}^{(l)})^{-1}\mathbf{D}(\boldsymbol{\beta}^{(l)}), \tag{5.8}$$

where we have used the notation $\mathbf{D}(\boldsymbol{\beta})$, $\mathbf{V}(\boldsymbol{\beta})$ and $\boldsymbol{\mu}(\boldsymbol{\beta})$ to emphasize the dependence of $\mathbf{D}$, $\mathbf{V}$, and $\boldsymbol{\mu}$ on $\boldsymbol{\beta}$.

5.3 Preliminary Estimation of Intensity and Pair Correlation

Using the notation from Section 5.2, $\mathbf{V} = \mathbf{V}_{\boldsymbol{\mu}}^{1/2}(\mathbf{I}+\mathbf{G})\mathbf{V}_{\boldsymbol{\mu}}^{1/2}$ where $\mathbf{V}_{\boldsymbol{\mu}} = \mathrm{Diag}(\mu_i)$ and

$$G_{ij} = \sqrt{\mu_i \mu_j}\big[g(\mathbf{u}_i, \mathbf{u}_j) - 1\big]$$

so that $\mathbf{G} = [G_{ij}]_{ij}$ is the matrix analogue of the symmetric operator $\mathbf{T}^s$ from Section 3.1. In general g is unknown and must be replaced by an estimate. Moreover it is advantageous if $\mathbf{G}$ is fixed in order to avoid the computational burden of repeated matrix inversion in the generalized least squares iterations (5.8).

As in Section 4, we assume that $g(\mathbf{r}) = g(\mathbf{r}; \boldsymbol{\psi})$ where $g(\cdot; \boldsymbol{\psi})$ is a translation invariant parametric pair correlation function model. We replace $\boldsymbol{\psi}$ and $\boldsymbol{\beta}$ inside $\mathbf{G}$ by preliminary estimates $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\psi}}$ which are fixed during the iterations (5.8). The estimates $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\psi}}$ can be obtained using the two-step approach in Waagepetersen and Guan (2009) where $\tilde{\boldsymbol{\beta}}$ is obtained from the composite likelihood function and $\tilde{\boldsymbol{\psi}}$ is a minimum contrast estimate based on the K-function. If translation invariance can not be assumed, $\boldsymbol{\psi}$ may instead be estimated by using a second-order composite likelihood as in Jalilian et al. (2011).

5.4 Tapering

The matrix $\mathbf{V}$ can be of very high dimension. However, many entries in $\mathbf{V}$ are very close to zero and we can therefore approximate $\mathbf{V}$ by a sparse matrix $\mathbf{V}_{\text{taper}}$ obtained by tapering (e.g. Furrer et al., 2006). More precisely, we replace $\mathbf{G}$ in $\mathbf{V}$ by a matrix $\mathbf{G}_{\text{taper}}$ obtained by assigning zero to entries G_{ij} below a suitable threshold. We then compute a sparse matrix Cholesky decomposition, $\mathbf{I} + \mathbf{G}_{\text{taper}} = \mathbf{L}\mathbf{L}^{\mathsf{T}}$. Then $(\mathbf{Y}-\boldsymbol{\mu})\mathbf{V}_{\boldsymbol{\mu}}^{-1/2}(\mathbf{I}+\mathbf{G}_{\text{taper}})^{-1}$ can be easily computed by solving the equation $\mathbf{x}\mathbf{L}\mathbf{L}^{\mathsf{T}} = (\mathbf{Y}-\boldsymbol{\mu})\mathbf{V}_{\boldsymbol{\mu}}^{-1/2}$ in terms of $\mathbf{x}$ using forward and back substitution for the sparse Cholesky factors $\mathbf{L}$ and $\mathbf{L}^{\mathsf{T}}$, respectively.

In practice, it is often assumed that $g(\mathbf{r}) = g_0(\|\mathbf{r}\|)$ for some function g_0. If g_0 is a decreasing function of $\|\mathbf{r}\|$ then we may define the entries in $\mathbf{G}_{\text{taper}}$ as $G_{ij}1[\|\mathbf{u}_i - \mathbf{u}_j\| \le d_{\text{taper}}]$, where d_{taper} solves $[g_0(d)-1]/[g_0(0)-1] = \epsilon$ for some small ϵ. That is, we replace entries G_{ij} by zero if $g_0(\|\mathbf{u}_i - \mathbf{u}_j\|) - 1$ is below some small percentage of the maximal value $g_0(0) - 1$.

When $\mathbf{V}$ in (5.8) is replaced by $\mathbf{V}_{\text{taper}}$ we obtain the following estimate of the covariance matrix of $\hat{\boldsymbol{\beta}}$:

$$\mathbf{S}_{\text{taper}}^{-1}\mathbf{D}^{\mathsf{T}}\mathbf{V}_{\text{taper}}^{-1}\mathbf{V}\ \mathbf{V}_{\text{taper}}^{-1}\mathbf{D}\mathbf{S}_{\text{taper}}^{-1} \tag{5.9}$$

where $\mathbf{S}_{\text{taper}} = \mathbf{D}^{\mathsf{T}}\mathbf{V}_{\text{taper}}^{-1}\mathbf{D}$. Note that it is not required to invert the non-sparse covariance matrix $\mathbf{V}$ in order to compute (5.9).

6 Simulation study and data example

To examine the performance of our optimal intensity estimator, we carry out a simulation study under the Guan and Shen (2010) setting. We moreover apply our estimator to three tropical rain forest data sets. We use the quasi-likelihood implementation of our estimator as described in Sections 5.2-5.4 and hence for convenience we use in the following the term quasi-likelihood for our approach.

6.1 Simulation Study

In the simulation study, following Guan and Shen (2010), realizations of Cox processes are generated on a square window W. Each simulation involves first the generation of a zero-mean Gaussian random field $\mathbf{Z} = \{Z(\mathbf{u})\}_{\mathbf{u}\in W}$ with exponential covariance function $c(\mathbf{u}) = \exp(-\|\mathbf{u}\|/0.1)$ and then the generation of an inhomogeneous Thomas process given $\mathbf{Z}$ with intensity function $\rho(\mathbf{u};\boldsymbol{\beta}) = \exp\left[\beta_0 + \beta_1 Z(\mathbf{u})\right]$ and clustering parameter $\boldsymbol{\psi} = (\kappa, \omega)$, cf. (3.10). For each simulation $\boldsymbol{\beta} = (\beta_0, \beta_1)$ is estimated using composite likelihood (CL), weighted composite likelihood (WCL), and quasi-likelihood (QL). The clustering parameter $\boldsymbol{\psi}$ is estimated using minimum contrast estimation based on the K-function (e.g. Section 10.1 in Møller and Waagepetersen, 2004).

The simulation window is either $W = [0,1]^2$ or $W = [0,2]^2$. The mean square error (MSE) of the CL, WCL and QL estimates is computed using 1000 simulations for each combination of different clustering levels (i.e., different expected numbers of clusters $\kappa^* = 100$ or 200 and different cluster radii $\omega^* = 0.02$ or 0.04), inhomogeneity levels ($\beta_1^* = 0.5$ or 1), and expected number of points (400 in the case of $W = [0,1]^2$ and 1600 in the case of $W = [0,2]^2$). The integral terms in the CL, WCL and QL estimating equations are approximated using a 50×50 grid for $W = [0,1]^2$ and a 100×100 grid for $W = [0,2]^2$. Tapering for QL is carried out as described in Section 5.4 using d_{taper} obtained with $\epsilon = 0.01$ for each estimated pair correlation function $g(\cdot;\hat{\boldsymbol{\psi}})$. For WCL we use $A(\mathbf{u}) \approx K(d_{\text{taper}};\hat{\boldsymbol{\psi}}) - \pi d_{\text{taper}}^2$ where

$$K(t;\psi) = \int_{\|\mathbf{r}\|\leq t} g(\mathbf{r};\boldsymbol{\psi})\mathrm{d}\mathbf{r}.$$

Table 1 shows the reduction in MSE for the WCL and QL estimators relative to the CL estimator. The reductions show that one can obtain more efficient estimates of the intensity function by taking into account the correlation structure of the process. As expected from the theoretical results, the QL estimator has superior performance compared with both the CL and the WCL estimators in all cases. The improvement over the CL estimator is especially substantial in the more clustered (corresponding to small κ^* and ω^*) and more inhomogeneous (corresponding to $\beta_1^* = 1$) cases where the largest reduction is 68.5 %. As we alluded in Section 3.2, the performance of the WCL estimator may rely on the validity of the approximation (3.7). In case of a longer dependence range, the approximation is expected to be less accurate and this explains the large drop in the efficiency of the WCL estimator relative to the CL estimator when ω^* increases from 0.02 to 0.04. In particular, the WCL estimator does not appear to perform any better than the CL estimator when $\boldsymbol{\psi}^* = (200, 0.04)$. In contrast, the QL estimator still gives significant reductions in MSE of size 10 % to 26 % depending on the value of β_1^* and W.

Table 1: Reduction (%) in MSE (summed for β_0 and β_1) for WCL and QL relative to CL.

	$W = [0,1]^2$				$W = [0,2]^2$			
	$\beta_1^* = 0.5$		$\beta_1^* = 1.0$		$\beta_1^* = 0.5$		$\beta_1^* = 1.0$	
$\boldsymbol{\psi}^* = (\kappa^*, \omega^*)$	WCL	QL	WCL	QL	WCL	QL	WCL	QL
(100, 0.02)	15.6	35.9	41.4	59.3	17.2	39.7	52.2	68.5
(100, 0.04)	1.5	34.4	14.2	42.2	11.9	38.9	13.6	55.1
(200, 0.02)	4.9	15.4	20.2	34.0	8.6	19.9	26.3	40.0
(200, 0.04)	−3.5	16.5	3.0	26.2	2.0	10.3	−7.5	18.0

6.2 Data Example

Figure 1 shows the spatial locations of three tree species, *Acalypha diversifolia* (528 trees), *Lonchocarpus heptaphyllus* (836 trees) and *Capparis frondosa* (3299 trees), in a 1000 m × 500 m observation window on Barro Colorado Island (Condit et al., 1996; Condit, 1998; Hubbell and Foster, 1983). We moreover consider ten covariates: pH, elevation (dem), slope gradient (grad), multiresolution index of valley bottom flatness (mrvbf), incoming mean solar radiation (solar), topographic wetness index (twi) as well as soil contents of copper (Cu), potassium (K), mineralized nitrogen (Nmin) and phosphorus (P).

We fit a Cox process model with a log-linear intensity function including all ten covariates to each of the three tree species using CL, WCL and QL. For each species we fit the following pair correlation functions of normal variance mixture type (Jalilian et al., 2011):

$$g(\mathbf{r};\boldsymbol{\psi}) = 1 + c(\mathbf{r};\boldsymbol{\psi}), \quad \mathbf{r} \in \mathbb{R}^2,$$

where the covariance function $c(\mathbf{r};\boldsymbol{\psi})$ is either Gaussian

$$c(\mathbf{r};(\sigma^2,\alpha)) = \sigma^2 \exp\big[-(\|\mathbf{r}\|/\alpha)^2\big],$$

Matérn (K_ν is the modified Bessel function of the second kind)

$$c(\mathbf{r};(\sigma^2,\alpha,\nu)) = \sigma^2 \frac{(\|\mathbf{r}\|/\alpha)^\nu K_\nu(\|\mathbf{r}\|/\alpha)}{2^{\nu-1}\Gamma(\nu)},$$

or Cauchy

$$c(\mathbf{r};(\sigma^2,\alpha)) = \sigma^2\big[1 + (\|\mathbf{r}\|/\alpha)^2\big]^{-3/2}.$$

These covariance function represent very different tail-behaviour ranging from light (Gaussian), exponential (Matérn), to heavy tails (Cauchy). The pair correlation function obtained with the Gaussian covariance function is just a reparametrization of the Thomas process pair correlation function (3.10). For the Matérn covariance we consider three different values of the shape parameter $\nu = 0.25$, 0.5 and 1. With $\nu = 0.5$ the exponential model $c[\mathbf{r};(\sigma^2,\alpha,0.5)] = \sigma^2\exp(-\|\mathbf{r}\|/\alpha)$ is obtained while $\nu = 0.25$ and 1 yields respectively a log convex and a log concave covariance function. The WCL and QL estimations were implemented as in the simulation study but using a 200 × 100 grid for the numerical quadrature.

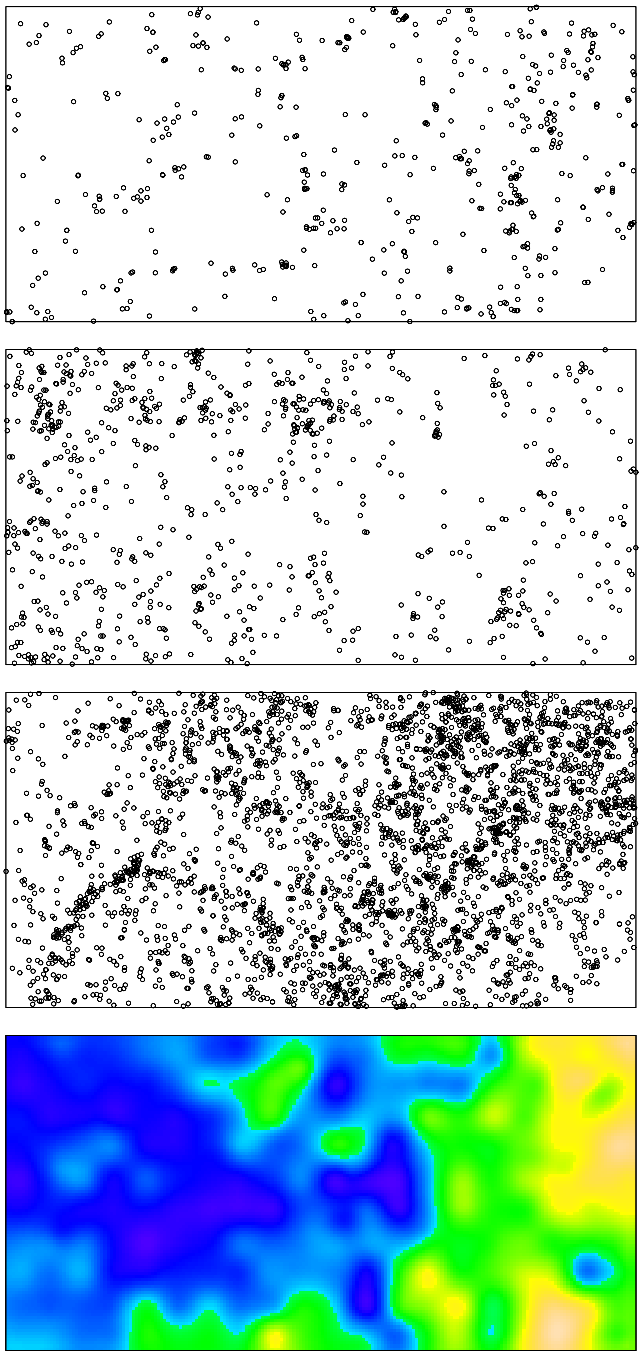


Figure 1: Locations of Acalypha, Lonchocarpus, and Capparis trees and image of interpolated potassium content in the surface soil (from top to bottom).

Figure 2 shows $c(\cdot;\hat{\boldsymbol{\psi}}) = g(\cdot;\hat{\boldsymbol{\psi}}) - 1$ for the best fitting (in terms of the minimum contrast criterion for the corresponding K-function) pair correlation functions: Cauchy for Acalypha, Matérn ($\nu = 0.5$) for Loncocharpus and Matérn ($\nu = 0.25$) for Capparis. The so-called integral ranges for these fitted functions, which are obtained by integrating $c(\cdot;\hat{\boldsymbol{\psi}})/c(\mathbf{0};\hat{\boldsymbol{\psi}})$ on $\mathbb{R}^2$ (e.g. Chilès and Delfiner, 1999), are 2037, 2759, and 3320 for Acalypha, Loncocharpus and Capparis, respectively. Moreover, the tapering distances are 21, 52 and 110 for the three species, respectively. The in-

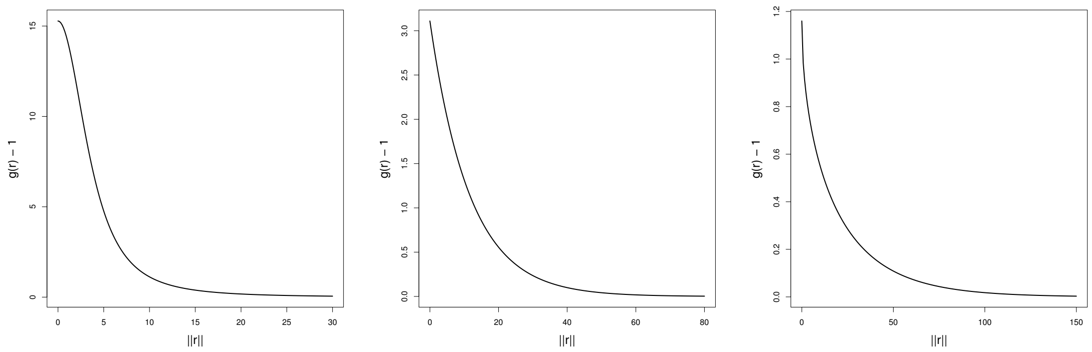


Figure 2: Best fitting covariance functions $c(\cdot;\hat{\boldsymbol{\psi}}) = g(\cdot;\hat{\boldsymbol{\psi}}) - 1$ for Acalypha (left), Loncocharpus (middle), and Capparis (right).

tegral ranges and tapering distances show that the dependence range is the largest for Capparis and smallest for Acalypha. The difference in the dependence ranges is likely caused by the distinct seed dispersal modes of the three species (Wright et al., 2007). Specifically, the seeds are dispersed by exploding capsules for Acalypha, by the wind for Lonchocarpus and by birds and mammals for Capparis. Seidler and Plotkin (2006) hypothesized that the modes of seed dispersal are reflected in the spatial patterns of tree locations with tight clusters for exploding capsules, loose clusters for bird and mammal dispersal and tightness of clustering somewhere in between for species with wind dispersal.

Table 2 shows the CL, WCL and QL estimates, where the latter two estimates were obtained by using the best fitting pair correlation models. Backward model selection was carried out for all methods and the models shown in Table 2 only contain covariates that were retained for at least one of the backward model selections. In terms of WCL and QL, the resulting regression parameter estimates are nearly identical for Acalypha, are in slightly less agreement for Lonchocarpus, but are very different for Capparis. In particular, the QL estimate in case of slope gradient is more than twice larger than the WCL estimate. The distinct levels of agreement between the estimates from these two methods might be due to the difference in the dependence ranges of these three species, as is also suggested by our simulation study. Our main findings in terms of significance of the covariates also vary among the three different methods. For Acalypha, elevation is found to be significant by CL at the 5 % level but not so by either WCL or QL. For Lonchocarpus, the QL approach suggests that phosphorus is significant but topographic wetness index is not, whereas CL and WCL suggest the opposite. For Capparis, slope gradient is found to be significant by QL but not so by either CL or WCL. In all cases, the smallest estimated standard errors are obtained with QL which is consistent with our developed theory of optimality.

Table 2: Estimates of regression parameters and their estimated standard errors (in parantheses) using CL, WCL and QL. * indicates significance at the 5 % level.

Species/Covariance		$\hat{\boldsymbol{\beta}}$
Acalypha	CL	$-6.91 + 0.021\texttt{dem} + 0.0047\texttt{K}$ $(77.34^*, 9.77^*, 1.153^*) \times 10^{-3}$
	WCL	$-6.90 + 0.017\texttt{dem} + 0.0046\texttt{K}$ $(77.23^*, 9.57, 1.137^*) \times 10^{-3}$
Cauchy $\hat{\boldsymbol{\psi}} = (15.28, 4.61)$	QL	$-6.90 + 0.016\texttt{dem} + 0.0047\texttt{K}$ $(77.09^*, 9.54, 1.133^*) \times 10^{-3}$
Lonchocarpus	CL	$-6.49 - 0.021\texttt{Nmin} - 0.11\texttt{P} - 0.59\texttt{pH} - 0.11\texttt{twi}$ $(81.06^*, 7.45^*, 58.78, 282.89^*, 53.19^*) \times 10^{-3}$
	WCL	$-6.49 - 0.023\texttt{Nmin} - 0.098\texttt{P} - 0.58\texttt{pH} - 0.12\texttt{twi}$ $(80.75^*, 7.04^*, 56.67, 272.24^*, 51.49^*) \times 10^{-3}$
Matérn $\hat{\boldsymbol{\psi}} = (3.11, 11.62, 0.5)$	QL	$-6.49 - 0.023\texttt{Nmin} - 0.12\texttt{P} - 0.55\texttt{pH} - 0.084\texttt{twi}$ $(80.15^*, 6.95^*, 55.23^*, 266.10^*, 45.47) \times 10^{-3}$
Capparis	CL	$-5.07 + 0.028\texttt{dem} - 1.10\texttt{grad} + 0.0043\texttt{K}$ $(79.54^*, 9.98^*, 1200.36, 1.16^*) \times 10^{-3}$
	WCL	$-5.07 + 0.028\texttt{dem} - 0.91\texttt{grad} + 0.0042\texttt{K}$ $(79.43^*, 9.61^*, 1141.97, 1.14^*) \times 10^{-3}$
Matérn $\hat{\boldsymbol{\psi}} = (1.16, 21.37, 0.25)$	QL	$-5.10 + 0.019\texttt{dem} - 2.50\texttt{grad} + 0.0039\texttt{K}$ $(77.77^*, 8.86^*, 935.02^*, 1.02^*) \times 10^{-3}$

7 Discussion

We develop theory and methods for optimal estimation of the intensity function of a spatial point process. Our proposed optimal intensity estimation method only requires the specification of the intensity function and a pair correlation function. Moreover, the estimation of the regression parameters can be expected to be quite robust towards misspefication of the pair correlation function since the resulting estimating equation is unbiased for any choice of pair correlation function. In the data example we considered pair correlation functions obtained from covariance functions of normal variance mixture type. Alternatively one might consider pair correlation functions of the log Gaussian Cox process type (Møller et al., 1998), i.e., $g(\mathbf{r}) = \exp\left[c(\mathbf{r})\right]$, where $c(\cdot)$ is an arbitrary covariance function.

If a log Gaussian Cox process is deemed appropriate, a computationally feasible alternative to our approach is to use the method of integrated nested Laplace approximation (INLA Rue et al., 2009) to implement Bayesian inference. However, in order to apply INLA it is required that the Gaussian field can be approximated well by a Gaussian Markov random field and this can limit the choice of covariance function. For example, the accurate Gaussian Markov random field approximations in Lindgren et al. (2011) of Gaussian fields with Matérn covariance functions are restricted to integer ν in the planar case. In contrast, our approach is not subject to such limitations and can also be applied to non-log Gaussian Cox processes.

We finally note that for the Nyström approximate solution of the Fredholm equation we used the simplest possible quadrature scheme using a Riemann sum for a fine grid. This entails a minimum of assumptions regarding the integrand but at the expense of a typically high-dimensional covariance matrix $\mathbf{V}$. There may hence be scope for further development considering more sophisticated numerical quadrature schemes.

Acknowledgements

Abdollah Jalilian and Rasmus Waagepetersen's research was supported by the Danish Natural Science Research Council, grant 09-072331 'Point process modeling and statistical inference', and by Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from the Villum Foundation. Yongtao Guan's research was supported by NSF grant DMS-0845368 and by NIH grant 1R01DA029081-01A1.

The BCI forest dynamics research project was made possible by National Science Foundation grants to Stephen P. Hubbell: DEB-0640386, DEB-0425651, DEB-0346488, DEB-0129874, DEB-00753102, DEB-9909347, DEB-9615226, DEB-9615226, DEB-9405933, DEB-9221033, DEB-9100058, DEB-8906869, DEB-8605042, DEB-8206992, DEB-7922197, support from the Center for Tropical Forest Science, the Smithsonian Tropical Research Institute, the John D. and Catherine T. MacArthur Foundation, the Mellon Foundation, the Celera Foundation, and numerous private individuals, and through the hard work of over 100 people from 10 countries over the past two decades. The plot project is part of the Center for Tropical Forest Science, a global network of large-scale demographic tree plots.

The BCI soils data set were collected and analyzed by J. Dalling, R. John, K. Harms, R. Stallard and J. Yavitt with support from NSF DEB021104, 021115, 0212284, 0212818 and OISE 0314581, STRI and CTFS. Paolo Segre and Juan Di Trani provided assistance in the field. The covariates dem, grad, mrvbf, solar and twi were computed in SAGA GIS by Tomislav Hengl (`http://spatial-analyst.net/`).

References

Jean-Paul Chilès and Pierre Delfiner. *Geostatistics - modeling spatial uncertainty.* Probability and Statistics. Wiley, 1999.

R. Condit. *Tropical Forest Census Plots.* Springer-Verlag and R. G. Landes Company, Berlin, Germany and Georgetown, Texas, 1998.

R. Condit, S. P. Hubbell, and R. B. Foster. Changes in tree species abundance in a neotropical forest: impact of climate change. *Journal of Tropical Ecology*, 12:231–256, 1996.

Reinhard Furrer, Marc G. Genton, and Douglas Nychka. Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15: 502–523, 2006.

C. A. Gotway and W. W. Stroup. A generalized linear model approach to spatial data analysis and prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, 2: 157–178, 1997.

J. Gray, Robert. Weighted estimating equations for linear regression analysis of clustered failure time data. *Lifetime Data Analysis*, 9(2):123–138, 2003.

Y. Guan and J. M. Loh. A thinned block bootstrap procedure for modeling inhomogeneous spatial point patterns. *Journal of the American Statistical Association*, 102:1377–1386, 2007.

Y. Guan and Y. Shen. A weighted estimating estimation approach for inhomogeneous spatial point processes. *Biometrika*, 2010. to appear.

Y. Guan, M. Sherman, and J. A. Calvin. A nonparametric test for spatial isotropy using subsampling. *Journal of the American Statistical Society*, 99:810–821, 2004.

Wolfgang Hackbusch. *Integral equations - theory and numerical treatment*. Birkhäuser, 1995.

C. C. Heyde. *Quasi-likelihood and its application - a general approach to optimal parameter estimation*. Springer Series in Statistics. Springer, 1997.

S. P. Hubbell and R. B. Foster. Diversity of canopy trees in a neotropical forest and implications for conservation. In S. L. Sutton, T. C. Whitmore, and A. C. Chadwick, editors, *Tropical Rain Forest: Ecology and Management*, pages 25–41. Blackwell Scientific Publications, Oxford, 1983.

I. A. Ibramigov and Y. V. Linnik. *Independent and stationary sequences of random variables*. Groningen: Wolters-Noordhoff, 1971.

A. Jalilian, Y. Guan, and R. Waagepetersen. Decomposition of variance for spatial Cox processes. submitted, 2011.

Peter D. Lax. *Functional analysis*. Wiley, 2002.

K. Liang and S. L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73:13–22, 1986.

Pei-Sheng Lin and Murray K. Clayton. Analysis of binary spatial data by quasi-likelihood estimating equations. *Annals of Statistics*, 33:542–555, 2005.

Yi-Ching Lin, Li-Wan Chang, Kuoh-Cheng Yang, Hsiang-Hua Wang, and I-Fang Sun. Point patterns of tree distribution determined by habitat heterogeneity and dispersal limitation. *Oecologia*, 165:175–184, 2011.

Finn Lindgren, HÃĕvard Rue, and Johan Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society B*, 73:423–498, 2011.

J. Møller and R. P. Waagepetersen. *Statistical inference and simulation for spatial point processes*. Chapman and Hall/CRC, Boca Raton, 2004.

J. Møller and R. P. Waagepetersen. Modern statistics for spatial point processes. *Scandinavian Journal of Statistics*, 34:643–684, 2007.

J. Møller, A. R. Syversveen, and R. P. Waagepetersen. Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25:451–482, 1998.

T. Mrkvička and I. Molchanov. Optimisation of linear unbiased intensity estimators for point processes. *Annals of the Institute of Statistical Mathematics*, 57:71–81, 2005.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society B*, 71:319–392, 2009.

F. P. Schoenberg. Consistent parametric estimation of the intensity of a spatial-temporal point process. *Journal of Statistical Planning and Inference*, 128:79–93, 2005.

T. G. Seidler and J. B. Plotkin. Seed dispersal and spatial pattern in tropical trees. *PLoS Biology*, 4:2132–2137, 2006.

Guochun Shen, Mingjian Yu, Xin-Sheng Hu, Xiangcheng Mi, Haibao Ren, I-Fang Sun, and Keping Ma. Species-area relationships explained by the joint effects of dispersal limitation and habitat heterogeneity. *Ecology*, 90:3033–3041, 2009.

R. Waagepetersen. An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics*, 63:252–258, 2007.

R. Waagepetersen and Y. Guan. Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society, Series B*, 71:685–702, 2009.

S. J. Wright, A. Hernandéz, and R. Condit. The bushmeat harvest alters seedling banks by favoring lianas, large seeds, and seeds dispersed by bats, birds, and wind. *Biotropica*, 39:363–371, 2007.

Scott L. Zeger. A regression model for time series of counts. *Biometrika*, 75:621–629, 1988.

Appendix A. Conditions and lemmas

To verify the existence of a $|W_n|^{1/2}$ consistent sequence of solutions $\hat{\boldsymbol{\beta}}_n$, we assume that the following conditions are satisfied:

C1: $\rho(\mathbf{u};\boldsymbol{\beta}) = \rho(\mathbf{z}(\mathbf{u})\boldsymbol{\beta}^\mathsf{T})$ where $\rho(\cdot) > 0$ is twice continuously differentiable and $\sup_{\mathbf{u}\in\mathbb{R}^2} \|\mathbf{z}(\mathbf{u})\| < K_1$ for some $K_1 < \infty$.

C2: for some $0 < K_2 < \infty$, $\int_{\mathbb{R}^2} \left|g(\mathbf{r};\boldsymbol{\psi}^*) - 1\right| \mathrm{d}\mathbf{r} \leq K_2$.

C3: $\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{u},\boldsymbol{\beta})$ is differentiable with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, and for $|\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{u},\boldsymbol{\beta})|$, $|\mathrm{d}\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{u},\boldsymbol{\beta})/\mathrm{d}\boldsymbol{\beta}|$ and $|\mathrm{d}\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{u},\boldsymbol{\beta})/\mathrm{d}\boldsymbol{\theta}|$, the supremum over $\mathbf{u} \in \mathbb{R}^2$, $\boldsymbol{\beta} \in b(\boldsymbol{\beta}^*, K_3)$, $\boldsymbol{\theta} \in b(\boldsymbol{\theta}^*, K_3)$ is bounded for some $K_3 > 0$, where $b(\mathbf{x}, r)$ denotes the ball centered at $\mathbf{x}$ with radius $r > 0$.

C4: $|W_n|^{1/2}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*)$ is bounded in probability.

C5: $\liminf_n l_n > 0$, where for each n, l_n denotes the minimal eigenvalue of

$$\bar{\mathbf{S}}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*) = |W_n|^{-1}\mathbb{E}\mathbf{J}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*) = |W_n|^{-1} \int_{W_n} \boldsymbol{\phi}_{n,\boldsymbol{\theta}^*}(\mathbf{u})^\mathsf{T} \boldsymbol{\rho}'(\mathbf{u};\boldsymbol{\beta}^*)\mathrm{d}\mathbf{u}.$$

Condition C1 and C2 imply L1 and L2 below.

L1: for $\rho(\mathbf{u};\boldsymbol{\beta})$, $\boldsymbol{\rho}'(\mathbf{u};\boldsymbol{\beta})$ and $\boldsymbol{\rho}''(\mathbf{u};\boldsymbol{\beta})$, the supremum over $\mathbf{u} \in \mathbb{R}^2$, $\boldsymbol{\beta} \in b(\boldsymbol{\beta}^*, K_3)$, $\boldsymbol{\theta} \in b(\boldsymbol{\theta}^*, K_3)$ is bounded.

L2: for a function $h : \mathbb{R}^2 \to \mathbb{R}$,

$$\mathbb{V}\mathrm{ar} \sum_{\mathbf{u} \in X \cap W_n} h(\mathbf{u}) \leq |W_n| \big[1 + \sup_{\mathbf{u} \in W_n} \rho(\mathbf{u}; \boldsymbol{\beta}^*) K_2\big] \sup_{\mathbf{u} \in W_n} h(\mathbf{u})^2 \sup_{\mathbf{u} \in W_n} \rho(\mathbf{u}; \boldsymbol{\beta}^*).$$

In particular, $|W_n|^{-1} \mathbb{V}\mathrm{ar} \sum_{\mathbf{u} \in X \cap W_n} h(\mathbf{u})$ is bounded when h is bounded.

The condition C3 is not so easy to verify in general due to the abstract nature of the function $\boldsymbol{\phi}_{n,\boldsymbol{\theta}}$. However, it can be verified e.g. assuming that $\boldsymbol{\phi}_{n,\boldsymbol{\theta}}$ can be expressed using the Neumann series. Condition C4 holds under conditions specified in Waagepetersen and Guan (2009) (including e.g. C1 and C2). Condition C5 is not unreasonable since

$$\bar{\mathbf{S}}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*) = |W_n|^{-1} \int_{W_n} \left[\frac{\boldsymbol{\rho}'(\mathbf{u}; \boldsymbol{\beta}^*)}{\rho(\mathbf{u}; \boldsymbol{\beta}^*)^{1/2}} \right]^{\mathsf{T}} \left[(\mathbf{I} + \mathbf{T}^s_{n,\boldsymbol{\theta}^*})^{-1} \frac{\boldsymbol{\rho}'(\cdot; \boldsymbol{\beta}^*)}{\rho(\cdot; \boldsymbol{\beta}^*)^{1/2}} \right] (\mathbf{u}) \mathrm{d}\mathbf{u}$$

and $(\mathbf{I} + \mathbf{T}^s_{n,\boldsymbol{\theta}^*})^{-1}$ is a positive operator (see Section 3.1). Since $\bar{\boldsymbol{\Sigma}}_n = \bar{\mathbf{S}}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)$, C5 also implies

L3: $\liminf_n l_n > 0$ where for each n, l_n denotes the minimal eigenvalue of $\bar{\boldsymbol{\Sigma}}_n$.

To prove the asymptotic normality of $|W_n|^{-1/2} \mathbf{e}_{n,\tilde{\boldsymbol{\theta}}_n}(\boldsymbol{\beta}^*) \bar{\boldsymbol{\Sigma}}_n^{-1/2}$, we assume that the following additional conditions are satisfied:

N1: $W_n = nA$ where $A \subset (0,1] \times (0,1]$ is the interior of a simple closed curve with nonempty interior.

N2: $\sup_p \frac{\alpha(p;k)}{p} = \mathrm{O}(k^{-\epsilon})$ for some $\epsilon > 2$, where $\alpha(p;k)$ is the strong mixing coefficient (Rosenblatt, 1956). For each p and k, the mixing condition measures the dependence between $X \cap E_1$ and $X \cap E_2$ where E_1 and E_2 are arbitrary Borel subsets of $\mathbb{R}^2$ each of volume less than p and at distance k apart.

N3: for some $K_4 < \infty$ and $k = 3, 4$,

$$\sup_{\mathbf{u}_1 \in \mathbb{R}^2} \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} \big| Q_k(\mathbf{u}_1, \cdots, \mathbf{u}_k) \big| \mathrm{d}\mathbf{u}_2 \cdots \mathrm{d}\mathbf{u}_k < K_4,$$

where Q_k is the k-th order cumulant density function of X (e.g. Guan and Loh, 2007).

Conditions N1–N3 correspond to conditions (2), (3) and (6), respectively, in Guan and Loh (2007). See this paper for a discussion of the conditions.

Appendix B. Existence of a $|W_n|^{1/2}$ consistent $\hat{\boldsymbol{\beta}}_n$

We use Theorem 2 and Remark 1 in Waagepetersen and Guan (2009) to show the existence of a $|W_n|^{1/2}$ consistent sequence of solutions $\hat{\boldsymbol{\beta}}_n$. Let $\|\mathbf{A}\|_M = \sup_{ij} |a_{ij}|$ for a matrix $\mathbf{A} = [a_{ij}]_{ij}$. With $\mathbf{V}_n = |W_n|^{1/2} \bar{\boldsymbol{\Sigma}}_n^{1/2}$ we need to verify the following results:

R1: $\|\mathbf{V}_n^{-1}\|_M \to 0$.

R2: For any $d > 0$,

$$\sup_{\boldsymbol{\beta}:\|(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\mathbf{V}_n\|\leq d} \|\mathbf{V}_n^{-1}\big[\mathbf{J}_{n,\tilde{\boldsymbol{\theta}}_n}(\boldsymbol{\beta}) - \mathbf{J}_{n,\tilde{\boldsymbol{\theta}}_n}(\boldsymbol{\beta}^*)\big]\mathbf{V}_n^{-1}\|_M$$

converges to zero in probability.

R3: $\|\mathbf{J}_{n,\tilde{\boldsymbol{\theta}}_n}(\boldsymbol{\beta}^*)/|W_n| - \bar{\mathbf{S}}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)\|_M$ converges to zero in probability.

R4: $\mathbf{e}_{n,\tilde{\boldsymbol{\theta}}_n}(\boldsymbol{\beta}^*)\mathbf{V}_n^{-1}$ is bounded in probability.

R5: $\liminf_n l_n > 0$ where

$$l_n = \inf_{\|\mathbf{x}\|=1} \mathbf{x}\bar{\boldsymbol{\Sigma}}_n^{-1/2}\bar{\mathbf{S}}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)\bar{\boldsymbol{\Sigma}}_n^{-1/2}\mathbf{x}^\mathsf{T}.$$

We now demonstrate that R1–R5 hold under the conditions C1–C5 listed in Appendix A. For each of the results below the required conditions or previous results are indicated in square brackets.

R1 [C3, L1–L3]: By C3, L1 and L2 the entries in $\bar{\boldsymbol{\Sigma}}_n$ are bounded from below and above. Moreover, by L3 the determinant of $\bar{\boldsymbol{\Sigma}}_n$ is bounded below by $l^p > 0$.

R2 [R1, C3, L1, L2, C4]: We show that

$$\sup_{(\boldsymbol{\theta},\boldsymbol{\beta}):\|(\boldsymbol{\theta}-\boldsymbol{\theta}^*,\boldsymbol{\beta}-\boldsymbol{\beta}^*)|W_n|^{1/2}\|\leq d} \||W_n|^{-1}\big[\mathbf{J}_{n,\boldsymbol{\theta}}(\boldsymbol{\beta}) - \mathbf{J}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)\big]\|_M$$

converges to zero in probability. Note

$$|W_n|^{-1}\mathbf{J}_{n,\boldsymbol{\theta}}(\boldsymbol{\beta}) = \mathbf{L}_{n,\boldsymbol{\theta}}(\boldsymbol{\beta}) + \mathbf{M}_{n,\boldsymbol{\theta}}(\boldsymbol{\beta})$$

where

$$\mathbf{L}_{n,\boldsymbol{\theta}}(\boldsymbol{\beta}) = -\sum_{\mathbf{u}\in X} \mathbf{f}_{1,n,\boldsymbol{\theta}}(\mathbf{u},\boldsymbol{\beta}) \text{ and } \mathbf{M}_{n,\boldsymbol{\theta}}(\boldsymbol{\beta}) = \int_{\mathbb{R}^2} \mathbf{f}_{2,n,\boldsymbol{\theta}}(\mathbf{u},\boldsymbol{\beta})$$

with

$$\mathbf{f}_{1,n,\boldsymbol{\theta}}(\mathbf{u},\boldsymbol{\beta}) = \frac{1[\mathbf{u}\in W_n]}{|W_n|}\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}^\mathsf{T}}\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{u},\boldsymbol{\beta})$$

and

$$\mathbf{f}_{2,n,\boldsymbol{\theta}}(\mathbf{u},\boldsymbol{\beta}) = \frac{1[\mathbf{u}\in W_n]}{|W_n|}\big[\rho(\mathbf{u};\boldsymbol{\beta})\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}^\mathsf{T}}\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{u},\boldsymbol{\beta}) + \boldsymbol{\rho}'(\mathbf{u};\boldsymbol{\beta})^\mathsf{T}\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{u},\boldsymbol{\beta})\big].$$

Define

$$h_{i,n}(\mathbf{u}) = \sup_{(\boldsymbol{\theta},\boldsymbol{\beta}):\|(\boldsymbol{\theta}-\boldsymbol{\theta}^*,\boldsymbol{\beta}-\boldsymbol{\beta}^*)|W_n|^{1/2}\|\leq d} |\mathbf{f}_{i,n,\boldsymbol{\theta}}(\mathbf{u},\boldsymbol{\beta}) - \mathbf{f}_{i,n,\boldsymbol{\theta}^*}(\mathbf{u},\boldsymbol{\beta}^*)|,\ i=1,2$$

and note that $h_{i,n}(\mathbf{u})$ converge to zero as $n\to\infty$. Then

$$\sup_{(\boldsymbol{\theta},\boldsymbol{\beta}):\|(\boldsymbol{\theta}-\boldsymbol{\theta}^*,\boldsymbol{\beta}-\boldsymbol{\beta}^*)|W_n|^{1/2}\|\leq d} |\mathbf{M}_{n,\boldsymbol{\theta}}(\boldsymbol{\beta}) - \mathbf{M}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)| \leq \int_{\mathbb{R}^2} h_{1,n}(\mathbf{u})\mathrm{d}\mathbf{u}$$

where the right hand side converges to zero by dominated convergence. Moreover,

$$\sup_{(\boldsymbol{\theta},\boldsymbol{\beta}):\|(\boldsymbol{\theta}-\boldsymbol{\theta}^*,\boldsymbol{\beta}-\boldsymbol{\beta}^*)|W_n|^{1/2}\|\le d} \big|\mathbf{L}_{n,\boldsymbol{\theta}}(\boldsymbol{\beta}) - \mathbf{L}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)\big| \le \sum_{\mathbf{u}\in X} h_{2,n}(\mathbf{u}) \le \\ \Big|\sum_{\mathbf{u}\in X} h_{2,n}(\mathbf{u}) - \mathbb{E}\sum_{\mathbf{u}\in X} h_{2,n}(\mathbf{u})\Big| + \Big|\mathbb{E}\sum_{\mathbf{u}\in X} h_{2,n}(\mathbf{u})\Big|.$$

The first term on the right hand side converges to zero in probability by Chebyshev's inequality and the second term converges to zero by dominated convergence.

R3 [R1, L1, L2, C4]:

$$|W_n|^{-1}\mathbf{J}_{n,\tilde{\boldsymbol{\theta}}_n}(\boldsymbol{\beta}^*) - \bar{\mathbf{S}}_n(\boldsymbol{\beta}^*) = \\ |W_n|^{-1}\big[\mathbf{J}_{n,\tilde{\boldsymbol{\theta}}_n}(\boldsymbol{\beta}^*) - \mathbf{J}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)\big] + \big[|W_n|^{-1}\mathbf{J}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*) - \bar{\mathbf{S}}_n(\boldsymbol{\beta}^*)\big]$$

It follows from the proof of R2 that the first term on the right hand side converges to zero in probability. The last term converges to zero in probability by Chebyshev's inequality.

R4 [C3, L1, L2, C4]: Since $\mathbb{V}\mathrm{ar}\mathbf{e}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)\mathbf{V}_n^{-1}$ is the identity matrix, $\mathbf{e}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)\mathbf{V}_n^{-1}$ is bounded in probability by Chebyshev's inequality. The result then follows by showing that $|W_n|^{-1/2}\big[\mathbf{e}_{n,\tilde{\boldsymbol{\theta}}_n}(\boldsymbol{\beta}^*) - \mathbf{e}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)\big]$ converges to zero in probability. Let

$$\mathbf{f}_n(\boldsymbol{\theta}) = |W_n|^{-1}\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}^\mathsf{T}}\mathbf{e}_{n,\boldsymbol{\theta}}(\boldsymbol{\beta}^*) = \\ |W_n|^{-1}\Big[\sum_{\mathbf{u}\in X\cap W_n}\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}^\mathsf{T}}\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{u},\boldsymbol{\beta}^*) - \int_{W_n}\rho(\mathbf{u};\boldsymbol{\beta}^*)\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}^\mathsf{T}}\boldsymbol{\phi}_{n,\boldsymbol{\theta}}(\mathbf{u},\boldsymbol{\beta}^*)\mathrm{d}\mathbf{u}\Big].$$

Then

$$|W_n|^{-1/2}\big[\mathbf{e}_{n,\tilde{\boldsymbol{\theta}}_n}(\boldsymbol{\beta}^*) - \mathbf{e}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)\big] = |W_n|^{1/2}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*)\mathbf{f}_n(\mathbf{t}_n)$$

where $\|\mathbf{t}_n-\boldsymbol{\theta}^*\| \le \|\tilde{\boldsymbol{\theta}}_n-\boldsymbol{\theta}^*\|$ and the factor $|W_n|^{1/2}(\tilde{\boldsymbol{\theta}}_n-\boldsymbol{\theta}^*)$ is bounded in probability. Further,

$$\mathbf{f}_n(\mathbf{t}_n) = \mathbf{f}_n(\mathbf{t}_n) - \mathbf{f}_n(\boldsymbol{\theta}^*) + \mathbf{f}_n(\boldsymbol{\theta}^*)$$

where $\mathbf{f}_n(\boldsymbol{\theta}^*)$ converges to zero in probability by Chebyshev's inequality and $\mathbf{f}_n(\mathbf{t}_n) - \mathbf{f}_n(\boldsymbol{\theta}^*)$ converges to zero in probability along the lines of the proof of R2.

R5 [C5, L3]: Follows directly from C5 and L3.

Appendix C. Asymptotic normality of $|W_n|^{-1/2}\mathbf{e}_{n,\tilde{\boldsymbol{\theta}}_n}(\boldsymbol{\beta}^*)\boldsymbol{\Sigma}_n^{-1/2}$

By the proof of R4 it suffices to show that $|W_n|^{-1/2}\mathbf{e}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)\bar{\boldsymbol{\Sigma}}_n^{-1/2}$ is asymptotically normal. To do so we use the blocking technique used in Guan and Loh (2007). Specifically, Condition N1 implies that there is a sequence of windows $W_n^B = \cup_{i=1}^{k_n} W_n^i$ given for each n by a union of $m_n \times m_n$ subsquares W_n^i, $i = 1, \cdots, k_n$, such that

$|W_n^B|/|W_n| \to 1$, $m_n = \mathrm{O}(n^\alpha)$ and the inter-distance between any two neighbouring subsquares is of order n^η for some $4/(2+\epsilon) < \eta < \alpha < 1$. Let

$$\mathbf{e}^B_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}) = \sum_{\mathbf{u} \in X \cap W_n^B} \boldsymbol{\phi}_{n,\boldsymbol{\theta}^*}(\mathbf{u};\boldsymbol{\beta}) - \int_{W_n^B} \boldsymbol{\phi}_{n,\boldsymbol{\theta}^*}(\mathbf{u};\boldsymbol{\beta})\rho(\mathbf{u};\boldsymbol{\beta})\mathrm{d}\mathbf{u} \equiv \sum_{i=1}^{k_n} \mathbf{e}^{B,i}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}),$$

where

$$\mathbf{e}^{B,i}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}) = \sum_{\mathbf{u} \in X \cap W_n^i} \boldsymbol{\phi}_{n,\theta^*}(\mathbf{u};\boldsymbol{\beta}) - \int_{W_n^i} \boldsymbol{\phi}_{n,\boldsymbol{\theta}^*}(\mathbf{u};\boldsymbol{\beta})\rho(\mathbf{u};\boldsymbol{\beta})\mathrm{d}\mathbf{u}.$$

Define

$$\tilde{\mathbf{e}}^B_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}) = \sum_{i=1}^{k_n} \tilde{\mathbf{e}}^{B,i}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}),$$

where the $\tilde{\mathbf{e}}^{B,i}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta})$'s are independent and for each i and n, $\tilde{\mathbf{e}}^{B,i}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta})$ is distributed as $\mathbf{e}^{B,i}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta})$. Let $\bar{\boldsymbol{\Sigma}}^B_n = |W_n^B|^{-1}\mathbb{V}\mathrm{ar}\mathbf{e}^B_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)$ and $\tilde{\boldsymbol{\Sigma}}^B_n = |W_n^B|^{-1}\mathbb{V}\mathrm{ar}\tilde{\mathbf{e}}^B_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)$. We need to verify the following results:

S1: $||\tilde{\boldsymbol{\Sigma}}^B_n - \bar{\boldsymbol{\Sigma}}^B_n||_M \to 0$ and $||\bar{\boldsymbol{\Sigma}}^B_n - \bar{\boldsymbol{\Sigma}}_n||_M \to 0$ as $n \to \infty$,

S2: $|W_n^B|^{-1/2}\tilde{\mathbf{e}}^B_{n,\theta^*}(\boldsymbol{\beta}^*)\big(\tilde{\boldsymbol{\Sigma}}^B_n\big)^{-1/2}$ is asymptotically standard normal,

S3: $|W_n^B|^{-1/2}\mathbf{e}^B_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)\big(\tilde{\boldsymbol{\Sigma}}^B_n\big)^{-1/2}$ has the same asymptotic distribution as $|W_n^B|^{-1/2}\tilde{\mathbf{e}}^B_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)\big(\bar{\boldsymbol{\Sigma}}^B_n\big)^{-1/2}$,

S4: $\||W_n^B|^{-1/2}\mathbf{e}^B_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*) - |W_n|^{-1/2}\mathbf{e}_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)\|$ converges to zero in probability.

S1 [C2, C3, N1]: This follows from the proof of Theorem 2 in Guan and Loh (2007).

S2 [C2, C3, N3]: Conditions C2, C3 and N3 imply $\mathbb{E}[\tilde{\mathbf{e}}^i_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta})^4]$ is bounded (see the proof of Lemma 1 in Guan and Loh, 2007). Thus, S2 follows from an application of Lyapunov's central limit theorem.

S3 [N2]: this follows by bounding the difference between the characteristic functions of $|W_n^B|^{-1/2}\mathbf{e}^B_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)$ and $|W_n^B|^{-1/2}\tilde{\mathbf{e}}^B_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)$ using techniques in Ibramigov and Linnik (1971) and secondly applying the mixing condition N2, see also Guan et al. (2004).

S4 [C1–C3, C5, N1]: Recall that $|W_n^B|/|W_n| \to 1$ due to N1. By C5 we only need to show $\mathbb{V}\mathrm{ar}\big[\mathbf{e}_{n,\theta^*}(\boldsymbol{\beta}^*) - \mathbf{e}^B_{n,\boldsymbol{\theta}^*}(\boldsymbol{\beta}^*)\big]/|W_n| \to 0$. This is implied by conditions C1–C3 and $|W_n^B|/|W_n| \to 1$.