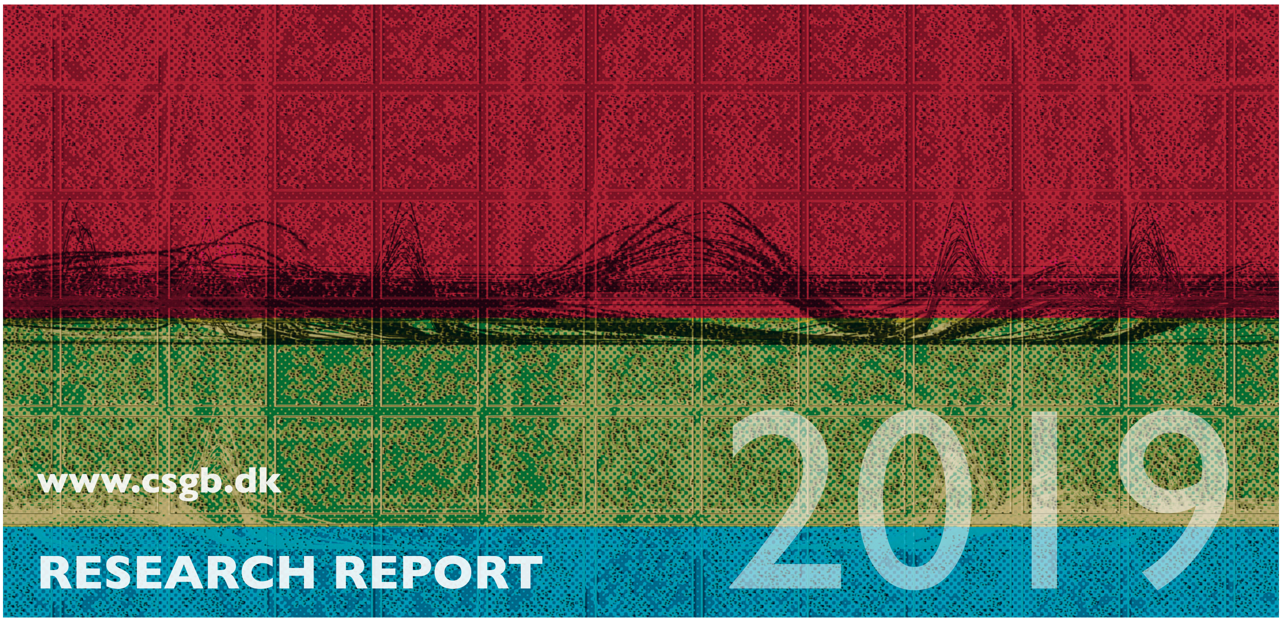
CENTRE FOR **STOCHASTIC GEOMETRY**
AND ADVANCED **BIOIMAGING**



Kristian Bjørn Hessellund, Ganggang Xu, Yongtao Guan
and Rasmus Waagepetersen

Semi-parametric multinomial logistic regression for multivariate point pattern data

No. 10, October 2019

Semi-parametric multinomial logistic regression for multivariate point pattern data

Kristian Bjørn Hessellund[1], Ganggang Xu[2], Yongtao Guan[2] and Rasmus Waagepetersen[1]

[1]Department of Mathematical Sciences, Aalborg University, Denmark
[2]Department of Management Science, University of Miami, USA

Abstract

We propose a new method for analysis of multivariate point pattern data observed in a heterogeneous environment and with complex intensity functions. We suggest semi-parametric models for the intensity functions that depend on an unspecified factor common to all types of points. This is for example well suited for analyzing spatial covariate effects on events such as street crime activities that occur in a complex urban environment. A multinomial conditional composite likelihood function is introduced for estimation of intensity function regression parameters and the asymptotic joint distribution of the resulting estimators is derived under mild conditions. Crucially, the asymptotic covariance matrix depends on the cross pair correlation functions of the multivariate point process. To make valid statistical inference without restrictive assumptions, we construct consistent non-parametric estimators for cross pair correlation function ratios. Finally, we construct standardized residual plots and predictive probability plots to validate and to visualize findings of the model. The effectiveness of the proposed methodology is demonstrated through extensive simulation studies and an application to analyzing effects of socio-economic and demographical variables on occurrences of street crimes in Washington DC.

Keywords: Conditional likelihood, Cross pair correlation functions, multinomial logistic regression, multivariate point process, semi-parametric.

1 Introduction

Multivariate point pattern data with many types of points are becoming increasingly common. Ecologists collect large data sets on locations and species of plants and animals, while police authorities gather ever increasing data sets on times, locations, and types of crimes. In epidemiology, multivariate point pattern data sets concern geo-referenced occurrences of different types of disease or bacteria. While the literature of bivariate point patterns is fairly well-developed (see e.g. the review in Waagepetersen et al., 2016), much less work has been done on statistical analysis of point patterns with more than two types of points. Diggle et al. (2005) and Baddeley et al. (2014) considered four- and six-variate multivariate Poisson processes and more recently Jalilian et al. (2015) and Waagepetersen et al. (2016) considered five- and nine-variate multivariate Cox processes. Rajala et al. (2018) and Choiruddin et al. (2019) consider penalized estimation for respectively multivariate Gibbs and log Gaussian Cox point processes for data sets containing locations of more than 80 species of rain forest trees.

This paper is concerned with statistical modeling of the first-order intensity functions of a multivariate spatial point process with an arbitrary number of types of points. For clarity of exposition we discuss our proposal in relation to the specific problem of street crime analysis where we focus on the spatial aspects of street crimes aggregated over a time span of interest, see also the data example in Section 6. To model street crime activities as a multivariate point process poses three major challenges: (1) to handle the high complexity of the first-order intensity function for each type of points; (2) to relate the street crime locations to available spatial covariates; (3) to take into account spatial correlations within and between different types of crimes. The first challenge arises because street crime activities depend in a complicated way on the layout of the city (streets, squares, malls,...) as well as the typically unknown population density at any location. Moreover, the intensity of crime activities may change abruptly from one area to neighboring areas. The second challenge arises because it is of great interest to police and criminologists to gain information on how street crime occurrences are related to demography and socio-economic variables. Finally, it is reasonable to expect spatial correlation between street crimes, which leads to the third challenge.

In this paper, inspired by the aforementioned first two challenges, we propose a semi-parametric regression model for the first-order intensity functions. Specifically, we propose a multiplicative model where the intensity function for each type of points is a product of a non-parametric component common to all types of points and a parametric component that models the influence of the covariates on the intensity function. The common non-parametric component models background factors such as population density or variation in intensity due to the layout of a city. To fit the model we propose a conditional composite likelihood function that does not depend on the non-parametric factor and is formally equivalent to multinomial logistic regression. We derive the asymptotic joint distribution of the resulting estimators and provide an estimator of the asymptotic covariance matrix. This matrix depends critically on the so-called cross pair correlation functions of the multivariate point process, which we estimate non-parametrically to avoid restrictive parametric assumptions.

Our approach is inspired by the case-control methodology introduced in Diggle and Rowlingson (1994) and further considered in Guan et al. (2008) and Xu et al. (2019). However, we do not restrict attention to the bivariate case considered in these references. Our approach also has some resemblance to Diggle et al. (2005) who considered spatially varying risks of occurrence of one type of bacteria relative to occurrence of other types. We, however, estimate relative risks using parametric models depending on covariates, where Diggle et al. (2005) applied non-parametric kernel estimation. Diggle and Rowlingson (1994), Guan et al. (2008), and Xu et al.

(2019) further assume independence between different types of points and that points of at least one type forms a Poisson process while Diggle et al. (2005) assume that all the different types of points form Poisson processes which are independent. According to the third challenge mentioned above, we do not assume that any of the point processes are Poisson and we do not assume independence between different types of points. This significantly expands the applicability of the proposed methodology to ever growing multivariate point pattern data collected in the big data era.

The rest of the paper is organized as follows. Section 2 provides an overview of multivariate point processes with focus on intensity and cross pair correlation functions. The semi-parametric model and its inference is introduced in Section 3 and theoretical investigations are given in Section 4. Simulation studies are presented in Section 5 and an application to Washington DC street crime data is given in Section 6. Concluding remarks are given in Section 7 and all technical proofs are collected in the supplementary material (Hessellund et al., 2019).

2 Background on multivariate point processes

Denote by $X = (X_1, \dots, X_p)$ a multivariate spatial point process, where X_i is a random subset of $\mathbb{R}^d$ with the property that $X_i \cap B$ is of finite cardinality for all bounded $B \subseteq \mathbb{R}^d$ and $i = 1, \dots, p$. We assume that each X_i is observed in a bounded window $W \subset \mathbb{R}^d$ and $X_i \cap X_j = \emptyset$ for any $i \neq j$. Assume that for each $m \geq 1$ and $i = 1, \dots, p$, there exists a non-negative function $\lambda_i^{(m)}(\cdot)$ such that

$$\mathrm{E} \sum_{\mathbf{u}_1,\dots,\mathbf{u}_m \in X_i}^{\neq} 1[\mathbf{u}_1 \in A_1, \dots, \mathbf{u}_m \in A_m] = \int_{\prod_{j=1}^m A_j} \lambda_i^{(m)}(\mathbf{u}_1, \dots, \mathbf{u}_m) \mathrm{d}\mathbf{u}_1 \cdots \mathrm{d}\mathbf{u}_m,$$

where $A_j \subset \mathbb{R}^d$ for $j = 1, \dots, m$, and $\sum^{\neq}$ indicates that $\mathbf{u}_1, \dots, \mathbf{u}_m$ are pairwise distinct. The function $\lambda_i^{(m)}(\cdot)$ is called the m'th order joint intensity function of X_i. For the special case with $m = 1$, the function $\lambda_i^{(1)}(\cdot)$ is referred to as the intensity and is denoted $\lambda_i(\cdot)$.

Assume further that for each $n, m \geq 1$ and $i, j = 1, \dots, p$, there exists a non-negative function $\lambda_{ij}^{(m,n)}(\cdot, \cdot)$ such that

$$\begin{aligned} &\mathrm{E} \sum_{\mathbf{u}_1,\dots\mathbf{u}_m \in X_i}^{\neq} \sum_{\mathbf{v}_1,\dots,\mathbf{v}_n \in X_j}^{\neq} 1[\mathbf{u}_1 \in A_1, \dots, \mathbf{u}_m \in A_m, \mathbf{v}_1 \in B_1, \dots, \mathbf{v}_n \in B_n] \qquad (2.1) \\ &= \int_{\prod_{j=1}^m A_j} \int_{\prod_{j=1}^m B_j} \lambda_{ij}^{(m,n)}(\mathbf{u}_1, \dots, \mathbf{u}_m, \mathbf{v}_1, \dots, \mathbf{v}_n) \mathrm{d}\mathbf{u}_1 \cdots \mathrm{d}\mathbf{u}_m \mathrm{d}\mathbf{v}_1 \cdots \mathrm{d}\mathbf{v}_n, \end{aligned}$$

where $A_k \subset \mathbb{R}^d$ and $B_l \subset \mathbb{R}^d$ for $k = 1, \dots, m$ and $l = 1, \dots, n$. The function $\lambda_{ij}^{(m,n)}(\cdot, \cdot)$ is referred to as the (m, n)'th order cross-intensity function between X_i and X_j, $i, j = 1, \dots, p$. The normalized (cross) joint intensities $g_i^{(m)}(\cdot)$ and $g_{ij}^{(n,m)}(\cdot, \cdot)$ are defined as

$$g_i^{(m)}(\mathbf{u}_1, \dots \mathbf{u}_m) = \lambda_i^{(m)}(\mathbf{u}_1, \dots, \mathbf{u}_m) \Big/ \prod_{l=1}^m \lambda_i(\mathbf{u}_l),$$

and

$$g_{ij}^{(m,n)}(\mathbf{u}_1,\dots,\mathbf{u}_m,\mathbf{v}_1,\dots,\mathbf{v}_n)=\frac{\lambda_{ij}^{(m,n)}(\mathbf{u}_1,\dots,\mathbf{u}_m,\mathbf{v}_1,\dots,\mathbf{v}_n)}{\prod_{l=1}^m\lambda_i(\mathbf{u}_l)\prod_{k=1}^n\lambda_j(\mathbf{v}_k)},\tag{2.2}$$

provided the denominators on the right hand sides are positive (otherwise we define $g_i^{(m)}(\mathbf{u}_1,\dots,\mathbf{u}_m)=0$ and $g_{ij}^{(m,n)}(\mathbf{u}_1,\dots,\mathbf{u}_m,\mathbf{v}_1,\dots,\mathbf{v}_n)=0$). For $i\neq j$, $g_{ij}^{(1,1)}(\cdot,\cdot)$ is referred to as the cross pair correlation function (cross PCF) and $g_{ii}^{(1,1)}(\cdot,\cdot)$ coincides with $g_i^{(2)}(\cdot,\cdot)$ which is known as the pair correlation function (PCF). From now on, we write $g_i(\cdot,\cdot)$ for $g_i^{(2)}(\cdot,\cdot)$ and $g_{ij}(\cdot,\cdot)$ for $g_{ij}^{(1,1)}(\cdot,\cdot)$. The notion of cross joint intensities and their normalized versions can be generalized in an obvious way to joint cross intensities $\lambda_{i_1i_2\cdots i_k}^{(n_1,\dots,n_k)}$ and normalized cross joint intensities $g_{i_1i_2\cdots i_k}^{(n_1,\dots,n_k)}$ for $X_{i_1},\dots,X_{i_k}$ for any $k\geq 1$, $\{i_1,\dots,i_k\}\subseteq\{1,2,\dots,p\}$, and integers $n_1,\dots,n_k\geq 1$.

Suppose that a point from X_i is observed at $\mathbf{u}$. Then $\lambda_j(\mathbf{v})g_{ij}(\mathbf{u},\mathbf{v})$ can be interpreted as the conditional intensity of X_j at $\mathbf{v}$ given that $\mathbf{u}\in X_i$. Thus the cross pair correlation function informs on how presence of a point in $\mathbf{u}$ affects the intensity of further points in X_j. In the special case when X_i and X_j are independent, $g_{ij}(\mathbf{u},\mathbf{v})\equiv 1$. If $X=(X_1,\dots,X_p)$ consists of independent Poisson processes, we call X a multivariate Poisson process. Then $\lambda_i^{(m)}(\mathbf{u}_1,\dots,\mathbf{u}_m)=\prod_{l=1}^m\lambda_i(\mathbf{u}_l)$ and $\lambda_{ij}^{(m,n)}(\mathbf{u}_1,\dots,\mathbf{u}_m,\mathbf{v}_1,\dots,\mathbf{v}_n)=\prod_{l=1}^m\lambda_i(\mathbf{u}_l)\prod_{k=1}^n\lambda_j(\mathbf{v}_k)$. Consequently, $g_{ij}(\mathbf{u},\mathbf{v})=1$, $i,j=1,\dots,p$, for a multivariate Poisson process which is the reference model of complete spatial independence.

Throughout the paper, we assume that the multivariate point process is second order cross-intensity reweighted isotropic meaning that $g_{ij}(\mathbf{u},\mathbf{v})$ depends only on the distance $\|\mathbf{u}-\mathbf{v}\|$. For this reason, we abuse notation and denote by $g_{ij}(r)$ the value of $g_{ij}(\mathbf{u},\mathbf{v})$ when $\|\mathbf{u}-\mathbf{v}\|=r$. We often refer to so-called Campbell's formulae. For example, by standard measure theoretical arguments, the definition of $\lambda_i^{(m)}(\cdot)$ implies

$$\mathrm{E}\sum_{\mathbf{u}_1,\dots,\mathbf{u}_m\in X_i}^{\neq} f(\mathbf{u}_1,\dots,\mathbf{u}_m)=\int_{(\mathbb{R}^d)^m} f(\mathbf{u}_1,\dots,\mathbf{u}_m)\lambda_i^{(m)}(\mathbf{u}_1,\dots,\mathbf{u}_m)\mathrm{d}\mathbf{u}_1\cdots\mathrm{d}\mathbf{u}_m$$

for any non-negative function f on $(\mathbb{R}^d)^m$. Similar Campbell formulae hold for the cross joint intensities.

3 Semi-parametric multinomial logistic regression

In this section we detail the proposed semi-parametric model and the multinomial logistic regression approach to statistical inference. Formal asymptotic considerations are deferred to Section 4.

3.1 Semi-parametric model and multinomial logistic regression

For spatial point pattern data in an environment like a city, the intensity function can be rather complex due to the city layout and variations in population density.

To overcome this difficulty, we follow Diggle and Rowlingson (1994) and assume that for each point pattern X_i, the intensity function takes the multiplicative form

$$\lambda_i(\mathbf{u}; \boldsymbol{\gamma}_i) = \lambda_0(\mathbf{u}) \exp[\boldsymbol{\gamma}_i^\mathsf{T} \mathbf{z}(\mathbf{u})], \qquad i = 1, \ldots, p, \tag{3.1}$$

where $\lambda_0(\cdot)$ is an unknown background intensity function, $\mathbf{z}(\mathbf{u})$ is a q-dimensional vector of spatial covariates at location $\mathbf{u}$, and $\boldsymbol{\gamma}_i \in \mathbb{R}^q$ is the vector of regression parameters. The background intensity $\lambda_0(\cdot)$ can be interpreted as the spatial effects of latent factors such the urban structure and population density and is assumed to be common for all point types. The model (3.1) is also closely related to the Cox regression model widely used for the conditional intensity in survival analysis (Cox, 1972).

We tackle the estimation of model (3.1) by conditional composite likelihood. Conditioned on that an event is observed at location $\mathbf{u}$, under model (3.1), the probability that it is from the point process X_i is

$$\frac{\lambda_i(\mathbf{u}; \boldsymbol{\gamma}_i)}{\sum_{k=1}^p \lambda_k(\mathbf{u}; \boldsymbol{\gamma}_k)} = \frac{\exp\left[\boldsymbol{\gamma}_i^\mathsf{T} \mathbf{z}(\mathbf{u})\right]}{\sum_{k=1}^p \exp\left[\boldsymbol{\gamma}_k^\mathsf{T} \mathbf{z}(\mathbf{u})\right]}, \qquad \boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\mathsf{T}, \ldots, \boldsymbol{\gamma}_p^\mathsf{T})^\mathsf{T}, \tag{3.2}$$

which does not depend on the background intensity $\lambda_0(\cdot)$. Clearly, the $\boldsymbol{\gamma}_k$'s are not identifiable from the probabilities (3.2). To address this issue, we pick a baseline process, say X_p, and define $\boldsymbol{\beta}_i = \boldsymbol{\gamma}_i - \boldsymbol{\gamma}_p$ for $i = 1, \ldots, p-1$. Denoting $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\mathsf{T}, \ldots, \boldsymbol{\beta}_{p-1}^\mathsf{T})^\mathsf{T}$, the conditional probabilities (3.2) become

$$\mathrm{p}_i(\mathbf{u}; \boldsymbol{\beta}) = \begin{cases} \frac{\exp[\boldsymbol{\beta}_i^\mathsf{T} \mathbf{z}(\mathbf{u})]}{1 + \sum_{k=1}^{p-1} \exp[\boldsymbol{\beta}_k^\mathsf{T} \mathbf{z}(\mathbf{u})]}, & i = 1, \ldots, p-1, \\ \frac{1}{1 + \sum_{k=1}^{p-1} \exp[\boldsymbol{\beta}_k^\mathsf{T} \mathbf{z}(\mathbf{u})]}, & i = p. \end{cases} \tag{3.3}$$

Using the new parameterization (3.3), we can evaluate the effects of the covariates $\mathbf{z}(\cdot)$ relative to the baseline process X_p similar to matched case-control studies and Cox regression in survival analysis. To estimate $\boldsymbol{\beta}$, we define the multinomial conditional composite likelihood as

$$L(\boldsymbol{\beta}) = \prod_{i=1}^p \prod_{\mathbf{u} \in X_i \cap W} \mathrm{p}_i(\mathbf{u}; \boldsymbol{\beta}).$$

This is formally equivalent to a multinomial logistic regression likelihood function. It is a composite likelihood function because it ignores possible dependencies between types of points given their locations. The log multinomial conditional composite likelihood function is of the form

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^p \sum_{\mathbf{u} \in X_i \cap W} \left[\boldsymbol{\beta}_i^\mathsf{T} \mathbf{z}(\mathbf{u}) - \log\Big(1 + \sum_{k=1}^{p-1} \exp[\boldsymbol{\beta}_k^\mathsf{T} \mathbf{z}(\mathbf{u})]\Big)\right], \tag{3.4}$$

and the conditional composite likelihood estimator is defined as $\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta})$.

3.2 Estimation of the asymptotic covariance matrix of $\widehat{\boldsymbol{\beta}}$

In this section we consider the problem of estimating the asymptotic covariance matrix of $\widehat{\boldsymbol{\beta}}$, which is challenging due to the highly complex between- and within-type correlation structure of the multivariate point process.

We denote by $\mathrm{E}(\cdot)$ and $\mathrm{Var}(\cdot)$, expectation and variance with respect to the data generating distribution of $X = (X_1, \ldots, X_p)$, where we assume the intensity function of X_i is of the form (3.1) with the parameters $\boldsymbol{\gamma}_i$ given by some specific values $\boldsymbol{\gamma}_i^* \in \mathbb{R}^q$ and we let $\boldsymbol{\beta}_i^* = \boldsymbol{\gamma}_i^* - \boldsymbol{\gamma}_p^*$ for $i = 1, \ldots, p-1$. In this section and the rest of the paper we will refer to the 'pooled' point process $X^{\mathrm{pl}} = \cup_{k=1}^{p} X_i$, whose intensity function and pair correlation function are

$$
\begin{aligned}
\lambda^{\mathrm{pl}}(\mathbf{u}; \boldsymbol{\beta}) &= \sum_{k=1}^{p} \lambda_k(\mathbf{u}; \boldsymbol{\gamma}_k) \\
\text{and} \qquad g^{\mathrm{pl}}(\mathbf{u}, \mathbf{v}; \boldsymbol{\beta}, g) &= \sum_{l=1}^{p} \sum_{l'=1}^{p} \mathrm{p}_l(\mathbf{u}; \boldsymbol{\beta}_l) \mathrm{p}_{l'}(\mathbf{v}; \boldsymbol{\beta}_l) g_{ll'}(\mathbf{u}, \mathbf{v}).
\end{aligned} \tag{3.5}
$$

The "g" inside $g^{\mathrm{pl}}(\mathbf{u}, \mathbf{v}; \boldsymbol{\beta}, g)$ signifies the dependence on the $g_{ll'}$. We use in the following the short forms $\lambda_k^*(\cdot)$, $\mathrm{p}_l^*(\cdot)$, $\lambda^{\mathrm{pl}}(\cdot)$, and $g^{\mathrm{pl}}(\cdot, \cdot)$ for $\lambda_k(\cdot; \boldsymbol{\gamma}_i^*)$, $\mathrm{p}_l(\cdot; \boldsymbol{\beta}_l^*)$, $\lambda^{\mathrm{pl}}(\cdot; \boldsymbol{\beta}^*)$, and $g^{\mathrm{pl}}(\cdot, \cdot; \boldsymbol{\beta}^*, g)$.

It is trivial to see that $\ell(\boldsymbol{\beta})$ in (3.4) is a concave function of $\boldsymbol{\beta}$ and thus maximizing $\ell(\boldsymbol{\beta})$ is equivalent to solving the estimating equation $\mathbf{e}(\boldsymbol{\beta}) = \mathbf{0}$ where

$$
\mathbf{e}(\boldsymbol{\beta}) = [\mathbf{e}_1(\boldsymbol{\beta})^\mathsf{T}, \ldots, \mathbf{e}_{p-1}(\boldsymbol{\beta})^\mathsf{T}]^\mathsf{T}, \tag{3.6}
$$

with

$$
\mathbf{e}_i(\boldsymbol{\beta}) = \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}_i} \ell(\boldsymbol{\beta}) = \sum_{\mathbf{u} \in X_i \cap W} \mathbf{z}(\mathbf{u}) - \sum_{l=1}^{p} \sum_{\mathbf{u} \in X_l \cap W} \frac{\mathbf{z}(\mathbf{u}) \exp[\boldsymbol{\beta}_i^\mathsf{T} \mathbf{z}(\mathbf{u})]}{1 + \sum_{k=1}^{p-1} \exp[\boldsymbol{\beta}_k^\mathsf{T} \mathbf{z}(\mathbf{u})]}, \tag{3.7}
$$

for $i = 1, \ldots, p-1$. According to standard estimating equation theory (see, for example, Crowder, 1986) and formally justified by Theorem 2 in Section 4.1, the asymptotic covariance matrix of $\widehat{\boldsymbol{\beta}}$ is of the form

$$
[\mathbf{S}(\boldsymbol{\beta}^*)]^{-1} \, \boldsymbol{\Sigma}(\boldsymbol{\beta}^*, g) \, [\mathbf{S}(\boldsymbol{\beta}^*)]^{-1}
$$

where $\mathbf{S}(\boldsymbol{\beta}^*) = \mathrm{E}[-\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}^\mathsf{T}} \mathbf{e}(\boldsymbol{\beta}^*)]$ is the so-called sensitivity matrix and $\boldsymbol{\Sigma}(\boldsymbol{\beta}^*, g) = \mathrm{Var}\,[\mathbf{e}(\boldsymbol{\beta}^*)]$ is the covariance matrix of $\mathbf{e}(\boldsymbol{\beta}^*)$. The "$g$" inside $\boldsymbol{\Sigma}(\boldsymbol{\beta}^*, g)$ emphasizes that $\mathrm{Var}\,[\mathbf{e}(\boldsymbol{\beta}^*)]$ depends on the underlying cross-pair correlation functions.

The explicit forms of $\mathbf{S}(\boldsymbol{\beta}^*)$ and $\boldsymbol{\Sigma}(\boldsymbol{\beta}^*, g)$ are derived in Section A of the supplementary material. The (i,j)'th block of $\mathbf{S}(\boldsymbol{\beta}^*)$ is of the form

$$
\mathbf{S}(\boldsymbol{\beta}^*)_{ij} = \begin{cases} \int_W \mathbf{Z}(\mathbf{u}, \mathbf{u}) \, [1 - \mathrm{p}_i^*(\mathbf{u})] \, \lambda_i^*(\mathbf{u}) \mathrm{d}\mathbf{u} & i = j, \\ -\int_W \mathbf{Z}(\mathbf{u}, \mathbf{u}) \mathrm{p}_j^*(\mathbf{u}) \lambda_i^*(\mathbf{u}) \mathrm{d}\mathbf{u} & i \neq j, \end{cases} \tag{3.8}
$$

for $i, j = 1, \ldots, p-1$ with $\mathbf{Z}(\mathbf{u}, \mathbf{v}) = \mathbf{z}(\mathbf{u}) \mathbf{z}(\mathbf{v})^\mathsf{T}$. The (i,j)'th block of $\boldsymbol{\Sigma}(\boldsymbol{\beta}^*, g)$ corresponding to $\mathrm{Cov}\,[\mathbf{e}_i(\boldsymbol{\beta}^*), \mathbf{e}_j(\boldsymbol{\beta}^*)]$ takes the form

$$
\begin{aligned}
\boldsymbol{\Sigma}(\boldsymbol{\beta}^*, g)_{ij} &= \mathbf{S}(\boldsymbol{\beta}^*)_{ij} \\
&\quad + \int_{W^2} \mathbf{Z}(\mathbf{u}, \mathbf{v}) \lambda_i^*(\mathbf{u}) \lambda_j^*(\mathbf{v}) g^{\mathrm{pl}}(\mathbf{u}, \mathbf{v}; \boldsymbol{\beta}^*, g) T_{ij}(\mathbf{u}, \mathbf{v}; \boldsymbol{\beta}, g) \mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v},
\end{aligned} \tag{3.9}
$$

where the function $T_{ij}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)$ is defined as

$$1+\frac{g_{ij}(\mathbf{u},\mathbf{v})}{g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)}-\sum_{l=1}^{p}\frac{[\mathrm{p}_l^*(\mathbf{v})g_{il}(\mathbf{u},\mathbf{v})+\mathrm{p}_l^*(\mathbf{u})g_{jl}(\mathbf{u},\mathbf{v})]}{g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)}. \tag{3.10}$$

By Campbell's formulae we can approximate $\mathbf{S}(\boldsymbol{\beta}^*)$ and $\boldsymbol{\Sigma}(\boldsymbol{\beta}^*,g)$ by $\widehat{\mathbf{S}}(\boldsymbol{\beta}^*)$ and $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\beta}^*,g)$, whose (i,j)'th blocks are defined as

$$\widehat{\mathbf{S}}(\boldsymbol{\beta}^*)_{ij}=\begin{cases}\sum_{\mathbf{u}\in X^{\mathrm{pl}}}\mathbf{Z}(\mathbf{u},\mathbf{u})\left[1-\mathrm{p}_i^*(\mathbf{u})\right]\mathrm{p}_i^*(\mathbf{u}) & i=j,\\ -\sum_{\mathbf{u}\in X^{\mathrm{pl}}}\mathbf{Z}(\mathbf{u},\mathbf{u})\mathrm{p}_i^*(\mathbf{u})\mathrm{p}_j^*(\mathbf{u}) & i\neq j,\end{cases} \tag{3.11}$$

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\beta}^*,g)_{ij}=\widehat{\mathbf{S}}(\boldsymbol{\beta}^*)_{ij}+\sum_{\mathbf{u},\mathbf{v}\in X^{\mathrm{pl}}:\|\mathbf{u}-\mathbf{v}\|\le R}^{\neq}\mathbf{Z}(\mathbf{u},\mathbf{v})\mathrm{p}_i^*(\mathbf{u})\mathrm{p}_j^*(\mathbf{v})T_{ij}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g), \tag{3.12}$$

for $i,j=1,\dots,p-1$. Here R denotes a 'correlation range' such that $g_{ij}(r)\approx 1$ for $r>R$. In practice we replace $\boldsymbol{\beta}^*$ by $\widehat{\boldsymbol{\beta}}$ in (3.11)–(3.12) and the notion "g" emphasizes their dependence on the underlying cross-pair correlation functions, which will be replaced by non-parametric estimators discussed in the next sections.

3.3 Naive kernel estimation of cross PCF ratios

The empirical covariance matrix (3.12) depends critically on cross PCFs which need to be estimated. The definition of a cross PCF in (2.2) suggests that its estimation requires consistent estimators of the intensity functions which are not available under the model (3.1), since $\lambda_0(\cdot)$ is unknown. However, a closer look at (3.10) reveals that for computation of (3.10) it suffices to estimate the cross pair correlation functions up to a common multiplicative factor, or, equivalently, to estimate ratios of cross PCFs, i.e.

$$g_{ij,kl}(\mathbf{u},\mathbf{v})=g_{ij}(\mathbf{u},\mathbf{v})/g_{kl}(\mathbf{u},\mathbf{v}),\qquad i,j=1,\dots,p, \tag{3.13}$$

for some arbitrary fixed pair of types of points k and l. These ratios are also of great interest in their own right as they measure the strength of correlation among two types of points relative to the strength of correlation between two other types of points. Consider the quantity

$$F_{ij}(r;b,\boldsymbol{\beta})=\sum_{\substack{\mathbf{u}\in X_i\cap W\\ \mathbf{v}\in X_j\cap W}}^{\neq}\sum\frac{k_b(\|\mathbf{u}-\mathbf{v}\|-r)}{\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta})\mathrm{p}_j(\mathbf{v};\boldsymbol{\beta})}, \tag{3.14}$$

where $k_b(\cdot)=k(\cdot/b)/b$ with $k(\cdot)$ being a kernel function defined on a bounded interval in $\mathbb{R}$ and $b>0$ is a bandwidth. Using Campbell's formula together with equation (3.3), it follows that under model (3.1),

$$\mathrm{E}[F_{ij}(r;b,\boldsymbol{\beta}^*)]=\int_{W^2}\lambda^{\mathrm{pl}}(\mathbf{u})\lambda^{\mathrm{pl}}(\mathbf{v})g_{ij}(\mathbf{u},\mathbf{v})k_b(\|\mathbf{u}-\mathbf{v}\|-r)\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v},$$

where λ^{pl} was defined in (3.5). Under suitable conditions and appropriately chosen bandwidth b, it is reasonable to expect that $F_{ij}(r; b, \widehat{\boldsymbol{\beta}}) \approx c(r) g_{ij}(r)$, where

$$c(r) = \int_{W^2} \lambda^{\mathrm{pl}}(\mathbf{u}) \lambda^{\mathrm{pl}}(\mathbf{v}) k_b(\|\mathbf{u} - \mathbf{v}\| - r) \mathrm{d}\mathbf{u} \mathrm{d}\mathbf{v},$$

is a multiplicative factor which, as desired, does not depend on ij. Consequently,

$$\hat{g}^{\mathrm{n}}_{ij,kl}(r; b, \widehat{\boldsymbol{\beta}}) = F_{ij}(r; b, \widehat{\boldsymbol{\beta}}) / F_{kl}(r; b, \widehat{\boldsymbol{\beta}}) \tag{3.15}$$

becomes an estimator of (3.13).

Note that the estimator (3.15) does not depend on the unknown background intensity $\lambda_0(\cdot)$. The superscript "n" stands for "naive" kernel estimator (a refined estimator will be introduced in the next section). Our Theorem 3 in Section 4.2 states that under mild conditions, (3.15) is consistent for $g_{ij,kl}(r)$. The naive plug-in estimator $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}}, \hat{g}^{\mathrm{n}})$ is then obtained by replacing $\boldsymbol{\beta}^*$ and the cross PCFs in (3.10) by $\widehat{\boldsymbol{\beta}}$ and the estimators (3.15) of cross PCF ratios. For the rest of the paper, we use the PCF of the baseline process X_p as the fixed denominator in (3.13), letting $k = l = p$.

3.4 Refined cross PCF ratio estimators

Even though Theorem 3 in Section 4.2 shows that the naive kernel estimator (3.15) is consistent under mild conditions, the finite sample performance of the plug-in estimators $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}}, \hat{g}^{\mathrm{n}})$ may be unsatisfactory due to high variabilities of the $\hat{g}^{\mathrm{n}}_{ij,pp}(\cdot)$'s. In particular, our numerical experiments suggest that when the number of observed points is small, some diagonal elements of the $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}}, \hat{g}^{\mathrm{n}})$ may be negative, resulting in negative estimated variances for some components of $\widehat{\boldsymbol{\beta}}$.

We notice that this phenomenon is mainly caused by the existence of a large number of negative values of $T_{ii}(\mathbf{u}, \mathbf{v}; \widehat{\boldsymbol{\beta}}, \hat{g}^{\mathrm{n}})$ when $\|\mathbf{u} - \mathbf{v}\|$ is large, leading to negative values in the diagonal of $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}}, \hat{g}^{\mathrm{n}})_{ii}$ as defined in (3.12). This issue can be resolved or alleviated by imposing constraints on the cross PCFs. In this paper, we impose the following constraints

$$g_{ij}(r) \le \sqrt{g_{ii}(r) g_{jj}(r)} \qquad \text{for } r \ge R^*,\ i, j = 1, \ldots, p, \tag{3.16}$$

for some $R^* \ge 0$. Intuitively, condition (3.16) means that for lags $r \ge R^*$, the spatial correlation between different point processes is weaker than the (geometric) average of spatial correlation within each individual point process. Condition (3.16) is not necessarily true for any multivariate point process but is indeed valid with $R^* = 0$ for a large class of multivariate log Gaussian Cox processes (Waagepetersen et al., 2016) (see also Section 5) and for a large subclass of the multivariate shot-noise Cox processes proposed in Jalilian et al. (2015).

To enforce the constraint (3.16) on the naive kernel estimators, let $\widehat{\mathbf{G}}^{\mathrm{n}}_r$ be a $p \times p$ matrix whose (i, j)'th element is $\hat{g}^{\mathrm{n}}_{ij,pp}(r; b, \widehat{\boldsymbol{\beta}})$ for some distance $r > R^*$. The refined nonparametric estimators, denoted as $\hat{g}^{\mathrm{r}}_{ij,pp}(r; b, \widehat{\boldsymbol{\beta}})$, are collected in the matrix $\widehat{\mathbf{G}}^{\mathrm{r}}_r$ obtained by

$$\widehat{\mathbf{G}}^{\mathrm{r}}_r = \arg \min_{\Theta = [\theta_{ij}]_{ij}} \|\Theta - \widehat{\mathbf{G}}^{\mathrm{n}}_r\|_F^2, \qquad \text{with } \theta_{ij} = \theta_{ji}, \theta_{pp} = 1, \theta_{ij}^2 \le \theta_{ii}\theta_{jj}, \tag{3.17}$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix.

It can be shown (Section B in the supplementary material) that for $\|\mathbf{u}-\mathbf{v}\| > R^*$, the plug-in estimator with $\hat{g}^{\mathrm{r}}_{ij,pp}(\cdot)$'s satisfies

$$\min_{1\le i\le p} T_{ii}(\mathbf{u},\mathbf{v};\widehat{\boldsymbol{\beta}},\hat{g}^{\mathrm{r}}) \ge 1-\max_{1\le l\le p}\hat{g}^{\mathrm{r}}_{ll,pp}(\|\mathbf{u}-\mathbf{v}\|;b,\widehat{\boldsymbol{\beta}})/g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\widehat{\boldsymbol{\beta}},\hat{g}^{\mathrm{r}}).$$

In contrast, using the naive $\hat{g}^{\mathrm{n}}_{ij,pp}(\cdot)$'s, we can only achieve the lower bound

$$1-\left[2\max_{1\le l,l'\le p}\hat{g}^{\mathrm{n}}_{ll',pp}(\|\mathbf{u}-\mathbf{v}\|;b,\widehat{\boldsymbol{\beta}})-\min_{1\le l\le p}\hat{g}^{\mathrm{n}}_{ll,pp}(\|\mathbf{u}-\mathbf{v}\|;b,\widehat{\boldsymbol{\beta}})\right]\Big/g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\widehat{\boldsymbol{\beta}},\hat{g}^{\mathrm{n}}).$$

Note that the first lower bound above can be much larger than the second lower bound, which partly explains why the refined cross PCF ratio estimators would produce much fewer large negative $T_{ii}(\mathbf{u},\mathbf{v};\widehat{\boldsymbol{\beta}},\hat{g}^{\mathrm{r}})$ when $\|\mathbf{u}-\mathbf{v}\| > R^*$, leading to a better covariance matrix estimator. In Section 5.3, we give a more detailed demonstration through numerical examples.

Remark 1. *Our numerical investigations suggest that the refined estimator is quite robust to the choice of R^*. The simplest choice is to set $R^*=0$. Otherwise we recommend the choice $R^*=\arg\min_{r\ge 0}\{\max_i P_{ii}(r) > 0.05\}$, where $P_{ii}(r)$ is the percentage of pairs $(\mathbf{u},\mathbf{v})$ that give $T_{ii}(\mathbf{u},\mathbf{v};\widehat{\boldsymbol{\beta}},\hat{g}^{\mathrm{n}}) < 0$ within the set $\{(\mathbf{u},\mathbf{v}):\mathbf{u},\mathbf{v}\in X^{\mathrm{pl}}$ and $\|\mathbf{u}-\mathbf{v}\|\in(r-h,r+h)\}$. In other words, when the percentage of negative $T_{ii}(\mathbf{u},\mathbf{v};\widehat{\boldsymbol{\beta}},\hat{g}^{\mathrm{n}})$'s exceeds 5% around the distance R^* for any $i=1,\dots,p$, the restriction (3.16) will be enforced for $r > R^*$.*

4 Asymptotic properties

In this section we study asymptotic properties of $\widehat{\boldsymbol{\beta}}$ in the case where X is observed on a sequence of increasing windows W_n. Denote by $\mathbf{e}^{(n)}(\boldsymbol{\beta})$ the multinomial estimating function (3.6) evaluated on W_n and by $\widehat{\boldsymbol{\beta}}_n$ the sequence of estimators obtained as solutions to $\mathbf{e}^{(n)}(\boldsymbol{\beta})=\mathbf{0}$. The quantities $\boldsymbol{\gamma}^*$, $\boldsymbol{\beta}^*$, $\boldsymbol{\Sigma}_n(\boldsymbol{\beta}^*,g)$ and $\mathbf{S}_n(\boldsymbol{\beta}^*)$ are defined as in Section 3.2 with $W=W_n$ for the last two. We also define 'averaged' versions, $\bar{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}^*,g)=\boldsymbol{\Sigma}_n(\boldsymbol{\beta}^*,g)/|W_n|$ and $\bar{\mathbf{S}}_n(\boldsymbol{\beta}^*)=\mathbf{S}_n(\boldsymbol{\beta}^*)/|W_n|$. Finally, $\|\mathbf{A}\|_{\max}=\max_{ij}a_{ij}$ denotes the maximum norm of $\mathbf{A}=[a_{ij}]_{ij}$.

4.1 Consistency and asymptotic normality of $\widehat{\boldsymbol{\beta}}_n$

The following conditions are sufficient to establish the consistency of $\widehat{\boldsymbol{\beta}}_n$.

- C1 $W_1\subset W_2\subset\dots$ and $|\bigcup_{l=1}^{\infty}W_l|=\infty$.
- C2 There exists an $0<K_1<\infty$ such that $\|\mathbf{z}(\mathbf{u})\|_{\max}$, $\lambda_i^*(\mathbf{u})$ and $g_{ij}(\mathbf{u},\mathbf{v})$ are bounded above by K_1 for all $\mathbf{u},\mathbf{v}\in\bigcup_{l=1}^{\infty}W_l$ and $i,j=1,\dots,p$.
- C3 There exists an $0<K_2<\infty$ so that $\int_{\mathbb{R}^d}|g_{ij}(\mathbf{0},\mathbf{u})-1|\mathrm{d}\mathbf{u}<K_2$ for all $i,j=1,\dots,p$.
- C4 $\liminf_{n\to\infty}\lambda_{\min}\left[|W_n|^{-1}\int_{W_n}\mathbf{Z}(\mathbf{u},\mathbf{u})\lambda_i^*(\mathbf{u})\mathrm{p}_p(\mathbf{u};\boldsymbol{\beta}^*)\mathrm{d}\mathbf{u}\right]>0$ for $i=1,\dots,p-1$, where $\lambda_{\min}[A]$ denotes the minimal eigenvalue of a matrix A.

C1–C3 are mild conditions that have been widely used in the literature. C4 ensures that the averaged sensitivity matrix $\bar{\mathbf{S}}_n(\boldsymbol{\beta}^*)$ is invertible for sufficiently large n, which is commonly used in the estimating equation literature. Heuristically speaking, C4 requires that sufficient information regarding $\boldsymbol{\beta}^*$ need to be accumulated across space and it could be violated if $\mathbf{z}(\cdot)$ is close to constant.

Theorem 1. *Under conditions C1–C4, there exists a sequence of solutions $\widehat{\boldsymbol{\beta}}_n$ to the estimating equation $\mathbf{e}_n(\boldsymbol{\beta}) = \mathbf{0}$ for which*

$$\widehat{\boldsymbol{\beta}}_n \xrightarrow{p} \boldsymbol{\beta}^*, \qquad \text{as } n \to \infty.$$

The proof of Theorem 1 is given in Section C.1 of the Appendix.

Next, we proceed to establish asymptotic normality of $\widehat{\boldsymbol{\beta}}_n$. Following Biscio and Waagepetersen (2019), we define an α-mixing coefficient by regarding X as a marked point process with points in $\mathbb{R}^d$ and marks in $M = \{1, \ldots, p\}$. That is, a point $\mathbf{u}$ in X_i corresponds to a marked point $(\mathbf{u}, i)$. We then for sets $A \subseteq \mathbb{R}^d$ and $B \subseteq M$, define $X_{A,B} = X \cap A \times B$ as the set of marked points in X whose 'point parts' fall in A and whose marks fall in B.

To define the α-mixing coefficient for X we first define an α-mixing coefficient for two σ-algebras $\mathcal{F}$ and $\mathcal{G}$ on a common probability space,

$$\alpha(\mathcal{F}, \mathcal{G}) = \sup\{|\mathrm{p}(F \cap G) - \mathrm{p}(F)\mathrm{p}(G)| \colon F \in \mathcal{F}, G \in \mathcal{G}\}.$$

Define $d(\mathbf{u}, \mathbf{v}) = \max\{|u_i - v_i| \colon 1 \leq i \leq d\}$ for $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$. The marked point process α-mixing coefficient of X is then for $s, c_1, c_2 \geq 0$ given by

$$\begin{aligned}\alpha^X_{c_1,c_2}(s) = \sup\{&\alpha(\sigma(X_{E_1,M}), \sigma(X_{E_2,M})) \colon \\ &E_1 \subset \mathbb{R}^d, E_2 \subset \mathbb{R}^d, |E_1| \leq c_1, |E_2| \leq c_2, d(E_1, E_2) \geq s\},\end{aligned}$$

where $|A|$ is the Lebesgue measure of A and $d(A, B) = \inf\{d(\mathbf{u}, \mathbf{v}) \colon \mathbf{u} \in A, \mathbf{v} \in B\}$. This coefficient measures the dependence between $X \cap E_1 \times M$ and $X \cap E_2 \times M$, where E_1 and E_2 are arbitrary Borel subsets of $\mathbb{R}^d$ with volumes less than c_1 and c_2 and separated by the distance s.

The following extra conditions are needed to establish asymptotic normality.

- N1 There exists $\epsilon > 0$ such that $\alpha^X_{2,\infty}(s) = O(1/s^{d+\epsilon})$.
- N2 There exist an integer $m > 2d/\epsilon + 2$ and C_g such that $g^{(n_1,n_2,\ldots,n_k)}_{i_1 i_2 \cdots i_k}(\cdot, \ldots, \cdot) \leq C_g$ for any $\{i_1, \ldots, i_k\} \subseteq \{1, 2, \ldots, p\}$, and integers $n_1 + \cdots + n_k \leq m$.
- N3 It holds that $\liminf_{n\to\infty} \lambda_{\min}\left[\bar{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}^*, g)\right] > 0$.

N1 is a standard mixing condition that, e.g., holds for multivariate log Gaussian Cox processes with pair correlation functions of bounded range (meaning $g_{ij}(r) = 1$ when r is larger than some $0 \leq R < \infty$) or Poisson cluster point processes with sufficiently quickly decaying cluster densities. Condition N2 of bounded normalized joint cross intensities is satisfied for most multivariate point process models. N3 is a standard condition which ensures that the variance of $|W_n|^{-1}\mathbf{e}^{(n)}(\boldsymbol{\beta})$ is not degenerate for sufficiently large n.

Theorem 2. *Under conditions C1–C4 and N1–N3, as $n \to \infty$, we have that*

$$|W_n|^{1/2}\bar{\boldsymbol{\Sigma}}_n^{-1/2}(\boldsymbol{\beta}^*, g)\bar{\mathbf{S}}_n(\boldsymbol{\beta}^*)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*) \xrightarrow{d} N(0, \mathbf{I}_{(p-1)q}).$$

The proof of Theorem 2 is given in Section C.2 of the Appendix.

Theorem 2 implies that the asymptotic variance of $\widehat{\boldsymbol{\beta}}_n$ is of the form

$$|W_n|^{-1}[\bar{\mathbf{S}}_n(\boldsymbol{\beta}^*)]^{-1}\bar{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}^*, g)[\bar{\mathbf{S}}_n(\boldsymbol{\beta}^*)]^{-1} = [\mathbf{S}_n(\boldsymbol{\beta}^*)]^{-1}\boldsymbol{\Sigma}_n(\boldsymbol{\beta}^*, g)[\mathbf{S}_n(\boldsymbol{\beta}^*)]^{-1},$$

where the left hand side suggests that the variance of $\widehat{\boldsymbol{\beta}}_n$ is of order $|W_n|^{-1}$. Based on Theorem 2, one can make statistical inference regarding $\boldsymbol{\beta}^*$ and other quantities of interest. For example, as in classical multinomial regression models, one may be interested in the probability of a certain event at a given location, i.e., $\mathrm{p}_i^*(\mathbf{u})$, or the log-odds $\log \frac{\mathrm{p}_i^*(\mathbf{u})}{\mathrm{p}_p^*(\mathbf{u})} = \mathbf{z}(\mathbf{u})^\mathsf{T}\boldsymbol{\beta}_i^*$ for $i = 1, \ldots, p-1$.

Denote by $\mu(\boldsymbol{\beta}^*)$ a parameter of interest where $\mu : \mathbb{R}^{(p-1)q} \to \mathbb{R}$ is differentiable. A simple application of the Delta method gives for $0 < \alpha < 1$ the $100\alpha\%$ approximate confidence interval for $\mu(\boldsymbol{\beta}^*)$,

$$\mu(\widehat{\boldsymbol{\beta}}) \pm z_{1-\alpha/2}\sqrt{[\boldsymbol{\mu}^{(1)}(\widehat{\boldsymbol{\beta}})]^\mathsf{T}[\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\beta}})]^{-1}\widehat{\boldsymbol{\Sigma}}_n(\widehat{\boldsymbol{\beta}}, \hat{g}_n^{\mathrm{r}})[\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\beta}})]^{-1}\boldsymbol{\mu}^{(1)}(\widehat{\boldsymbol{\beta}})}, \tag{4.1}$$

where z_α is the 100α'th percentile of a standard normal distribution, $\boldsymbol{\mu}^{(1)}(\boldsymbol{\beta}) = \mathrm{d}\mu(\boldsymbol{\beta})/\mathrm{d}\boldsymbol{\beta}$, and estimators of $\boldsymbol{\beta}$ and cross PCFs have been plugged into (3.11) and (3.12), see also Sections 3.3–3.4 and Section 4.2.

4.2 Asymptotic properties of $\hat{g}^{\mathrm{n}}_{ij,kl}(r; b, \widehat{\boldsymbol{\beta}})$ and $\hat{g}^{\mathrm{r}}_{ij,kl}(r; b, \widehat{\boldsymbol{\beta}})$

Let W_n and b_n be sequences of observation windows and bandwidths, respectively. Denote by $\hat{g}^{\mathrm{n}}_{ij,kl,n}(r; b_n, \widehat{\boldsymbol{\beta}}_n)$ a sequence of estimators that is given by

$$\hat{g}^{\mathrm{n}}_{ij,kl,n}(r; b_n, \widehat{\boldsymbol{\beta}}_n) = F_{ij,n}(r; b_n, \widehat{\boldsymbol{\beta}}_n)/F_{kl,n}(r; b_n, \widehat{\boldsymbol{\beta}}_n),$$

where the $F_{ij,n}$'s are defined as in (3.14) with $W = W_n$. In this subsection, we show that $\hat{g}^{\mathrm{n}}_{ij,kl,n}(r; b_n, \widehat{\boldsymbol{\beta}}_n)$ is a consistent estimator of $g_{ij,kl}(r)$ for any $i, j = 1, \ldots, p$, under the following conditions.

K1 For $i, j = 1, \ldots, p$, the cross joint intensity $g_{ij}^{(2,2)}$ is translation invariant: $g_{ij}^{(2,2)}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_1, \mathbf{v}_2) = g_{ij}^{(2,2)}(\mathbf{0}, \mathbf{u}_2 - \mathbf{u}_1, \mathbf{v}_1 - \mathbf{u}_1, \mathbf{v}_2 - \mathbf{u}_1)$, $\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_1, \mathbf{v}_2 \in \bigcup_{l=1}^\infty W_l$, and there exists $K_3 < \infty$ so that

$$\int_{\mathbb{R}^d} |g_{ij}^{(2,2)}(\mathbf{0}, \mathbf{u}, \mathbf{v}, \mathbf{w} + \mathbf{u}) - g_{ij}(\mathbf{0}, \mathbf{v})g_{ij}(\mathbf{0}, \mathbf{w})|\mathrm{d}\mathbf{u} < K_3 \quad \text{for all } \mathbf{u}, \mathbf{v}, \mathbf{w} \in \bigcup_{l=1}^\infty W_l.$$

K2 There exists $K_4 < \infty$ so that $g_{ij}^{(m,n)}(\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{v}_1, \ldots, \mathbf{v}_n) < K_4$ for all $\mathbf{u}_m, \mathbf{v}_n \in \bigcup_{l=1}^\infty W_l$ with $m + n < 4$ and $i, j = 1, \ldots, p$.

K3 The kernel function $k(\cdot)$ has a compact support $[-1, 1]$ and the bandwidth b_n satisfies that (a) $b_n \to 0$; and (b) $|W_n|b_n \to \infty$ as $|W_n| \to \infty$.

Theorem 3. *Under conditions C2 and K1–K3, one has that*

$$\hat{g}^{\mathrm{n}}_{ij,kl,n}(r; b_n, \widehat{\boldsymbol{\beta}}) \xrightarrow{p} g_{ij,kl}(r), \qquad \text{as } n \to \infty, \text{ for } i,j,k,l = 1,\dots,p. \tag{4.2}$$

If we further assume that constraint (3.16) *holds true, then*

$$\hat{g}^{\mathrm{r}}_{ij,kl,n}(r; b_n, \widehat{\boldsymbol{\beta}}) \xrightarrow{p} g_{ij,kl}(r), \qquad \text{as } n \to \infty, \text{ for } i,j,k,l = 1,\dots,p. \tag{4.3}$$

The proof of Theorem 3 is given in Section C.3 of the Appendix.

5 Simulation studies

In this section we assess the finite sample performance of the proposed methodology through simulation studies. To evaluate our estimators we need to simulate from a model with known forms of the intensity functions and of the ratios of cross pair correlation functions. This precludes the use of multivariate Gibbs processes as considered e.g. in Rajala et al. (2018) and we consider instead a Cox process model. Specifically, the multivariate point patterns are simulated from a multivariate log-Gaussian Cox process where for $i = 1, \dots, p$, X_i has a random intensity function of the form

$$\Lambda_i(\mathbf{u}) = \lambda_0(\mathbf{u}) \exp[\gamma_{i0} + \gamma_{i1}\mathbf{z}(\mathbf{u})] \exp[\alpha_i Y(\mathbf{u}) + \sigma_i U_i(\mathbf{u}) - \alpha_i^2/2 - \sigma_i^2/2], \tag{5.1}$$

where $\lambda_0(\cdot)$ is the inhomogeneous background intensity, $\mathbf{z}(\cdot)$ is a spatial covariate, and $Y(\cdot)$ and $U_i(\cdot)$ are independent zero-mean unit variance Gaussian random fields. The spatial correlation functions of $Y(\cdot)$ and $U_i(\cdot)$ are assumed to be exponential $c_Y(\mathbf{u},\mathbf{v}) = \exp(-\|\mathbf{u}-\mathbf{v}\|/\xi)$ and $c_{U_i}(\mathbf{u},\mathbf{v}) = \exp(-\|\mathbf{u}-\mathbf{v}\|/\phi_i)$ with scale parameters ξ and ϕ_i. Conditional on the Λ_i, the X_i are independent Poisson processes. This model has a natural interpretation and can generate both positive and negative correlations between different types of points.

The process $Y(\cdot)$ can be viewed as an unobserved factor that affects all types of points and hence induces spatial correlations both within and between different types of points. The latent Gaussian process $U_i(\cdot)$ is a type-specific factor that only affects the i'th type of points. Conditional on $\lambda_0(\cdot)$ and $\mathbf{z}(\cdot)$, $\mathrm{E}[\Lambda_i(\mathbf{u})] = \lambda_0(\mathbf{u}) \exp[\gamma_{i0} + \gamma_{i1}\mathbf{z}(\mathbf{u})]$ and the cross PCF between X_i and X_j is of the form

$$g_{ij}(r; \boldsymbol{\theta}) = \exp\left[\alpha_i\alpha_j \exp(-r/\xi) + 1[i=j]\sigma_i^2 \exp(-r/\phi_i)\right], \tag{5.2}$$

where $\boldsymbol{\theta} = (\alpha_1, \dots, \alpha_p, \xi, \sigma_1^2, \dots, \sigma_p^2, \phi_1, \dots, \phi_p) \in \mathbb{R}^{3p+1}$. For $i \neq j$, $\alpha_i\alpha_j > 0$ (< 0) implies positive (negative) correlation between points from X_i and X_j whereas $\alpha_i\alpha_j = 0$ implies that X_i and X_j are independent given $\lambda_0(\cdot)$ and $\mathbf{z}(\cdot)$.

5.1 Simulation settings

More specifically, we consider the multivariate log-Gaussian Cox process with $p = 4$ and observed within a sequence of increasing square windows $W_l = [0, l] \times [0, l]$, $1 \le l \le 2$. The baseline intensity function in (5.1) is $\lambda_0(\mathbf{u}) = \exp[0.5V(\mathbf{u}) - 0.5^2/2]$, where $V(\mathbf{u})$ is a realization of zero-mean unit variance Gaussian random field with the exponential correlation function and a scale parameter 0.05. The spatial covariate $\mathbf{z}(\mathbf{u})$ is chosen as an independent copy of $V(\mathbf{u})$, see Figure 1(a)–(b).

The parameters for the multivariate log-Gaussian Cox process are listed in Table 1, where the intercept parameters γ_{i0}^*, $i = 1, \dots, p$, are chosen so that there are on average N_i points in the point pattern X_i in W_1 with the N_i's specified in Table 1. We use X_p as the baseline point process and consider three parameters of interest: the intercepts $\beta_{0i}^* = \gamma_{0i}^* - \gamma_{0p}^*$, the slopes $\beta_{1i}^* = \gamma_{1i}^* - \gamma_{1p}^*$, and the log-odds $\theta_i^*(\mathbf{u}) = \log \frac{\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)}{\mathrm{p}_p(\mathbf{u};\boldsymbol{\beta}^*)} = \beta_{0i}^* + \beta_{1i}^* \mathbf{z}(\mathbf{u})$, for $i = 1, \dots, p-1$. The log-odds $\theta_i^*(\mathbf{u})$ represent the elevated (or reduced) likelihood of a point in X_i at location $\mathbf{u}$ with an observed covariate $\mathbf{z}(\mathbf{u})$ relative to the probability of a point in X_p at $\mathbf{u}$. For the log odds we consider $\mathbf{z}(\mathbf{u}) = 0.5$. The α_i's are chosen such that there are positive and negative spatial correlations among the X_i's. The resulting PCFs and cross PCFs show (Figure 1(c)) strong spatial between and within dependence.

Table 1: The true parameters for the multivariate LGCP.

X	α_i	σ_i^2	ξ	ϕ_i	γ_{i0}^*	γ_{i1}^*	N_i
X_1	0.5	0.5	0.1	0.05	5.17	0	150
X_2	-0.4	0.5	0.1	0.05	5.44	0.3	200
X_3	0.6	0.5	0.1	0.05	5.88	-0.6	300
X_4	-0.3	0.5	0.1	0.05	6.13	0.6	400

In the following Section 5.2 we evaluate estimation accuracies for the parameters of interest and the coverage probabilities of their associated confidence intervals. The performances of the non-parametric cross PCF estimators proposed in Section 3.3–3.4 are considered in Section 5.3.

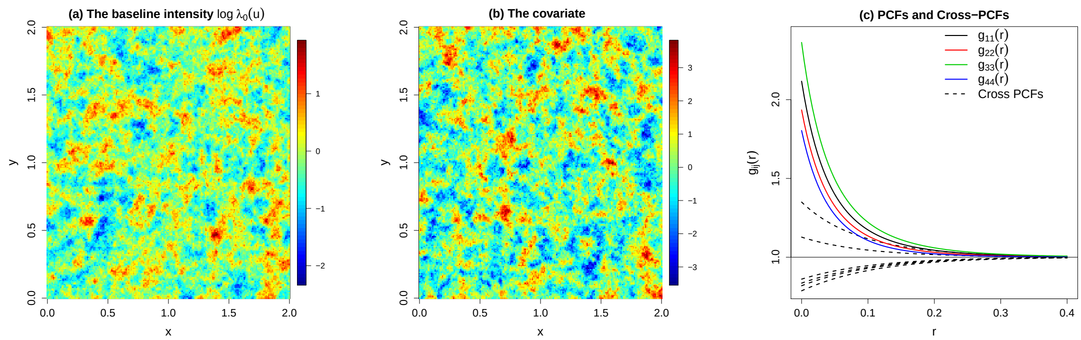


Figure 1: The log-background intensity (left panel); The spatial covariate (middle panel); The true PCFs and cross PCFs (right panel).

5.2 Estimation accuracies and coverage probabilities

In this simulation study, we evaluate estimation accuracies for the The log odds $\theta_i^*(\mathbf{u})$ are estimated by replacing the $\boldsymbol{\beta}_i$'s in the definition of the $\theta_i^*(\mathbf{u})$'s by their estimates $\widehat{\boldsymbol{\beta}}_i$. Four types of confidence intervals are investigated, denoted $\mathrm{CI}_{\hat{g}^{\mathrm{n}}}$, $\mathrm{CI}_{\hat{g}^{\mathrm{r}}}$, $\mathrm{CI}_{g^{\mathrm{Poisson}}}$, and $\mathrm{CI}_{g^{\mathrm{true}}}$. All confidence intervals are constructed using (4.1) with the

Table 2: Estimation accuracies and coverage probabilities of confidence intervals.

		Bias	SE	$\mathrm{CI}_{\hat{g}^{\mathrm{n}}}$ 90%	$\mathrm{CI}_{\hat{g}^{\mathrm{n}}}$ 95%	$\mathrm{CI}_{\hat{g}^{\mathrm{r}}}$ 90%	$\mathrm{CI}_{\hat{g}^{\mathrm{r}}}$ 95%	$\mathrm{CI}_{g^{\mathrm{Poisson}}}$ 90%	$\mathrm{CI}_{g^{\mathrm{Poisson}}}$ 95%	$\mathrm{CI}_{g^{\mathrm{true}}}$ 90%	$\mathrm{CI}_{g^{\mathrm{true}}}$ 95%
W_1	$\widehat{\beta}_{01}$	−0.002	0.246	66.1	71.8	87.0	92.6	47.6	54.6	89.3	93.8
	$\widehat{\beta}_{02}$	0.002	0.155	66.5	72.2	93.6	97.1	62.6	70.5	90.5	94.9
	$\widehat{\beta}_{03}$	0.002	0.254	67.0	74.4	84.7	90.8	39.3	45.5	89.7	94.4
	$\widehat{\beta}_{11}$	−0.001	0.135	88.6	94.4	88.2	94.4	68.5	77.6	90.4	95.9
	$\widehat{\beta}_{12}$	0.002	0.105	89.5	94.4	89.1	94.6	75.7	83.3	90.5	95.4
	$\widehat{\beta}_{13}$	−0.001	0.127	87.4	93.5	86.9	92.6	63.9	73.3	89.6	94.5
	$\hat{\theta}_1$	−0.003	0.246	68.4	75.4	87.0	93.0	44.3	52.2	89.8	94.5
	$\hat{\theta}_2$	−0.008	0.157	70.5	77.6	92.2	96.0	61.4	70.6	89.7	95.3
	$\hat{\theta}_3$	−0.002	0.261	72.6	80.7	86.2	91.3	44.1	51.3	90.8	94.5
W_2	$\widehat{\beta}_{01}$	−0.001	0.131	82.1	89.2	86.7	92.3	46.1	52.7	88.0	93.8
	$\widehat{\beta}_{02}$	−0.006	0.080	83.3	90.4	92.3	95.7	62.0	68.5	89.6	94.7
	$\widehat{\beta}_{03}$	0.005	0.137	81.5	88.3	86.2	92.3	34.8	42.6	87.9	94.0
	$\widehat{\beta}_{11}$	−0.002	0.067	91.2	96.0	91.6	96.0	71.1	80.4	91.6	96.4
	$\widehat{\beta}_{12}$	−0.001	0.054	89.7	95.5	89.7	95.5	78.1	85.8	90.5	95.6
	$\widehat{\beta}_{13}$	−0.001	0.067	88.7	95.4	88.8	95.4	63.6	72.6	89.2	95.4
	$\hat{\theta}_1$	−0.002	0.130	83.9	88.7	88.0	92.4	45.4	52.4	88.8	94.2
	$\hat{\theta}_2$	−0.006	0.083	84.0	89.7	91.2	96.0	59.4	69.0	89.2	95.2
	$\hat{\theta}_3$	0.005	0.143	83.7	88.2	86.0	91.8	40.2	47.9	88.7	93.9

sensitivity and the covariance matrices estimated using equations (3.11) and (3.12) with $R = 0.4$ but with different choices of cross PCF estimators. The $\mathrm{CI}_{\hat{g}^{\mathrm{n}}}$ and $\mathrm{CI}_{\hat{g}^{\mathrm{r}}}$ use respectively the "naive" and "refined" kernel cross PCF ratio estimators (3.15) and (3.17). The R^* used for the "refined" kernel estimators is obtained with the data-driven procedure in Remark 1. The $\mathrm{CI}_{g^{\mathrm{Poisson}}}$ is obtained by assuming $g_{ij}(\cdot) \equiv 1$ for $i, j = 1, \ldots, p$, and $\mathrm{CI}_{g^{\mathrm{true}}}$ is constructed using the true $g_{ij}(\cdot)$'s. The coverage probabilities of $\mathrm{CI}_{g^{\mathrm{true}}}$ serve as bench marks while $\mathrm{CI}_{g^{\mathrm{Poisson}}}$ may reveal potential problems of using multivariate Poisson point process models in presence of spatial correlations. Summary statistics based on 1000 simulations are given in Table 2 and also illustrated in Figure 2.

The first column in Table 2 shows that the parameter estimates are close to unbiased. Further, as predicted by Theorem 2, the standard errors are approximate halved when the observation window is increased from W_1 to the four times larger W_2. The coverage probabilities of $\mathrm{CI}_{g^{\mathrm{true}}}$ are all close to the nominal levels, suggesting that statistical inferences based on Theorem 2 are valid provided all cross PCF functions are correctly specified. On the contrary, in almost all cases, $\mathrm{CI}_{g^{\mathrm{Poisson}}}$ suffers from severe undercoverage that may lead to wrong conclusions in practical applications.Confidence intervals based on the "naive" kernel estimator of cross PCF ratios, i.e. $\mathrm{CI}_{\hat{g}^{\mathrm{n}}}$, achieve nominal levels for all slope parameters but suffer from serious undercoverage for intercepts and the log-odds when the observation window is small ($W_1 = [0,1] \times [0,1]$). The undercoverage of $\mathrm{CI}_{\hat{g}^{\mathrm{n}}}$ becomes much less severe when the window expand to $W_2 = [0,2] \times [0,2]$. Finally, confidence intervals based on the "refined" cross PCF ratio estimators, i.e. $\mathrm{CI}_{\hat{g}^{\mathrm{r}}}$, can effectively correct the undercoverage of $\mathrm{CI}_{\hat{g}^{\mathrm{n}}}$ and achieve nominal levels for all parameters of interest. This suggests that it is important to apply the modification proposed in Section 3.4 for practical applications with only limited sample sizes.

Figure 2 paints a more complete picture of how estimation accuracies and coverage probabilities change as W_l expands. The root mean squared error (RMSE) of all estimators decrease as the window size increases, supporting our theoretical findings in Section 4.1. Figure 2 also reveals that while the coverage probabilities of $\mathrm{CI}_{\hat{g}^{\mathrm{n}}}$ for intercepts and log-odds are getting closer to the nominal level as W_l expands, the undercoverage of $\mathrm{CI}_{g^{\mathrm{Poisson}}}$ does not improve at all. This emphasizes the importance of taking into account spatial correlations to make valid statistical inferences. Lastly, the coverage probabilities of $\mathrm{CI}_{\hat{g}^{\mathrm{r}}}$ are close to the nominal level for all parameters and window sizes and only slightly worse than those of $\mathrm{CI}_{g^{\mathrm{true}}}$. Therefore, we recommend $\mathrm{CI}_{\hat{g}^{\mathrm{r}}}$ for practical use.

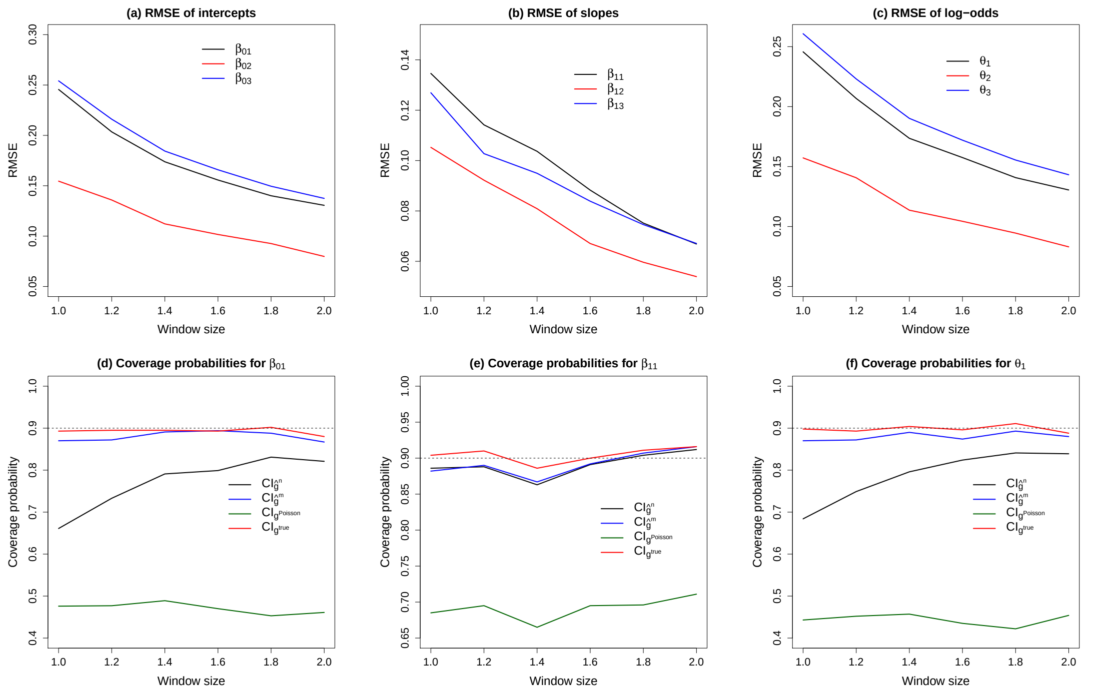


Figure 2: Top panels: the root mean squared errors (RMSE) of multinomial composite likelihood estimators; Bottom panels: coverage probabilities of various confidence intervals. Observation windows range from W_1 to W_2.

5.3 Performances of kernel estimators of cross PCF ratios

Following the last paragraph in Section 3.3, Figure 3 illustrates the performances of the "naive" kernel estimator and "refined" kernel estimator for some ratios $g_{ij}(r)/g_{44}(r)$ under various window sizes. Both estimators are unbiased and the "refined" kernel estimator has a slightly narrower 95% probability band than the "naive" kernel estimator for the window W_1. As the observation window is increased from W_1 to W_2, all probability bands become much tighter, supporting the theoretical findings in Theorem 3. However, we do not observe appreciable differences between these two non-parametric estimators in terms of estimation accuracies.

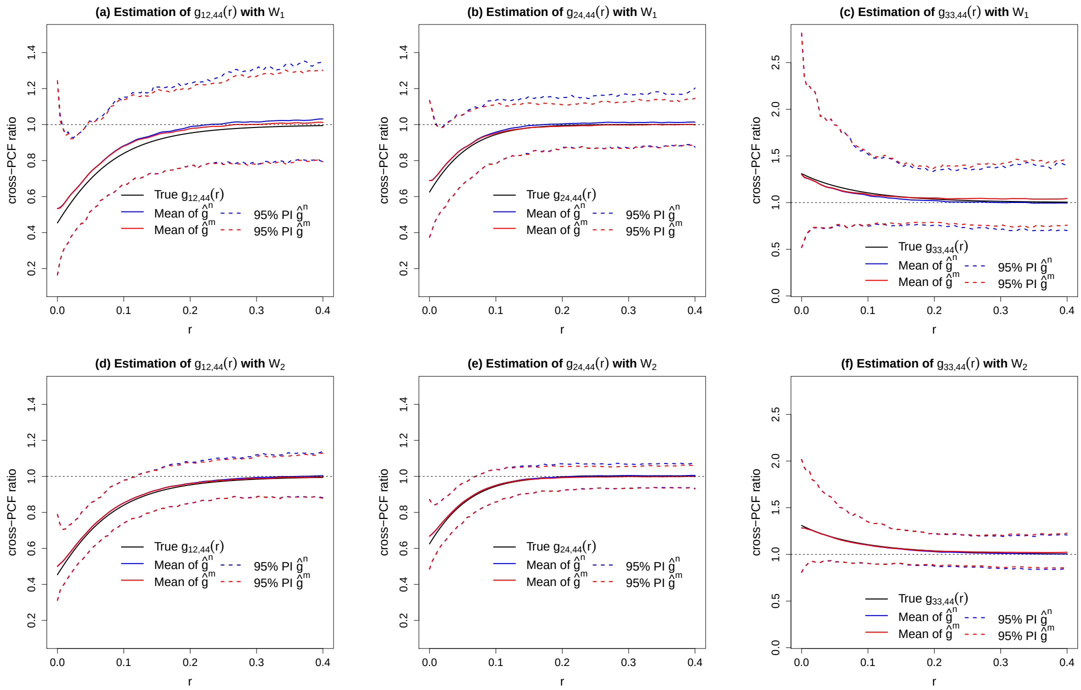


Figure 3: Means of estimated cross PCF ratios and point-wise 95% probability intervals for cross PCF ratios. Upper row: W_1, lower row: W_2.

To shed more light on why $\mathrm{CI}_{\hat{g}^{\mathrm{r}}}$ outperforms $\mathrm{CI}_{\hat{g}^{\mathrm{n}}}$, we study the diagonal blocks of the covariance matrix estimator (3.12) and focus on the diagonal elements in each $\widehat{\mathbf{\Sigma}}(\boldsymbol{\beta}^*, g)_{ii}$ corresponding to the intercept, which can be rewritten as

$$\widehat{\tau}_i(\boldsymbol{\beta}^*, g) = \sum_{\mathbf{u}\in X^{\mathrm{pl}}} [1 - \mathrm{p}_i^*(\mathbf{u})]\mathrm{p}_i^*(\mathbf{u}) + \sum_{l=1}^{L-1} \phi_{i,l}(\boldsymbol{\beta}^*, g), \tag{5.3}$$

for $i = 1, \dots, p-1$, where for $L \geq 1$ and $l = 1, \dots, L-1$,

$$\phi_{i,l}(\boldsymbol{\beta}^*, g) = \sum_{\mathbf{u},\mathbf{v}\in X^{\mathrm{pl}}}^{\neq} \mathrm{p}_i^*(\mathbf{u})\mathrm{p}_i^*(\mathbf{v}) T_{ii}(\mathbf{u}, \mathbf{v}; \boldsymbol{\beta}^*, g) 1(r_l < \|\mathbf{u} - \mathbf{v}\| \leq r_{l+1}), \tag{5.4}$$

for an equally-spaced partition $0 = r_1 < r_2 < \cdots < r_L = R$ of the interval $[0, R]$ and with $T_{ii}(\mathbf{u}, \mathbf{v}; \boldsymbol{\beta}^*, g)$ defined in (3.10).

The $\widehat{\tau}_i(\boldsymbol{\beta}^*, g)$'s are estimators of variances. However, after plugging in $\widehat{\boldsymbol{\beta}}$ and estimated cross PCFs, the resulting $\widehat{\tau}_i(\widehat{\boldsymbol{\beta}}, \hat{g})$'s are not guaranteed to be positive. This issue is especially severe for the "naive" kernel cross PCF ratio estimators. Figure 4 compares the means of $\phi_i(r; \boldsymbol{\beta}^*, g)$, $\phi_i(r; \widehat{\boldsymbol{\beta}}, \hat{g}^{\mathrm{n}})$ and $\phi_i(r; \widehat{\boldsymbol{\beta}}, \hat{g}^{\mathrm{r}})$ based on

1000 simulations together with point-wise 95% probability bands for $i = 1, 2, 3$. While there exist little differences between the means of $\phi_i(r; \widehat{\boldsymbol{\beta}}, \hat{g}^{\mathrm{n}})$ and $\phi_i(r; \widehat{\boldsymbol{\beta}}, \hat{g}^{\mathrm{r}})$, the low quantiles of $\phi_i(r; \widehat{\boldsymbol{\beta}}, \hat{g}^{\mathrm{n}})$ can take very large negative values, which may lead to a small and even negative value of $\widehat{\tau}_i(\widehat{\boldsymbol{\beta}}, \hat{g}^{\mathrm{n}})$. In contrast, the lower quantiles of $\phi_i(r; \widehat{\boldsymbol{\beta}}, \hat{g}^{\mathrm{r}})$ are always close to 0, and thus the associated $\widehat{\tau}_i(\widehat{\boldsymbol{\beta}}, \hat{g}^{\mathrm{r}})$'s are bounded away from negative values. Since the estimated covariance matrix of $\widehat{\boldsymbol{\beta}}$ takes the form $[\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\beta}})]^{-1}\widehat{\boldsymbol{\Sigma}}_n(\widehat{\boldsymbol{\beta}}, \hat{g})[\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\beta}})]^{-1}$, it is generally the case that larger diagonal elements in $\widehat{\boldsymbol{\Sigma}}_n(\widehat{\boldsymbol{\beta}}, \hat{g})$ leads to larger estimated variances for $\widehat{\boldsymbol{\beta}}$. Therefore, $\mathrm{CI}_{\hat{g}^{\mathrm{r}}}$ tends to achieve higher coverage probability than that of $\mathrm{CI}_{\hat{g}^{\mathrm{n}}}$.

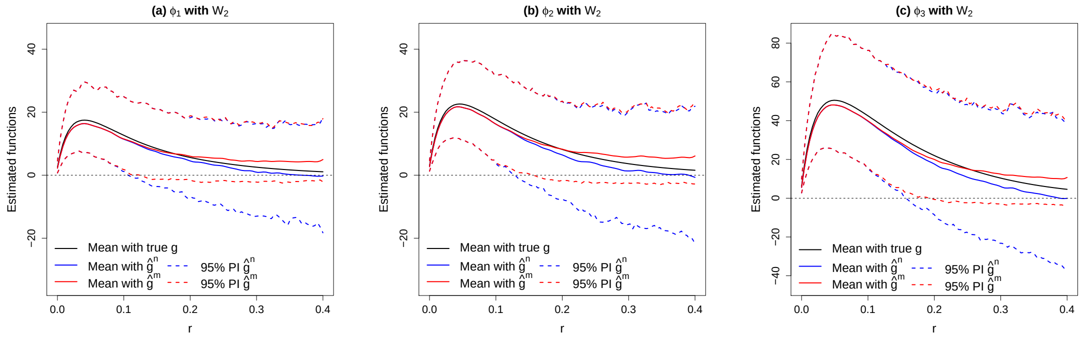


Figure 4: Means of $\phi_{i,l}(\boldsymbol{\beta}^*, g)$, $\phi_{i,l}(\boldsymbol{\beta}, \hat{g}^{\mathrm{n}})$ and $\phi_{i,l}(\widehat{\boldsymbol{\beta}}, \hat{g}^{\mathrm{r}})$ against r_l, $l = 1, \dots, L$, $i = 1, 2, 3$, and point-wise 95% probability bands.

6 Washington DC street crime data

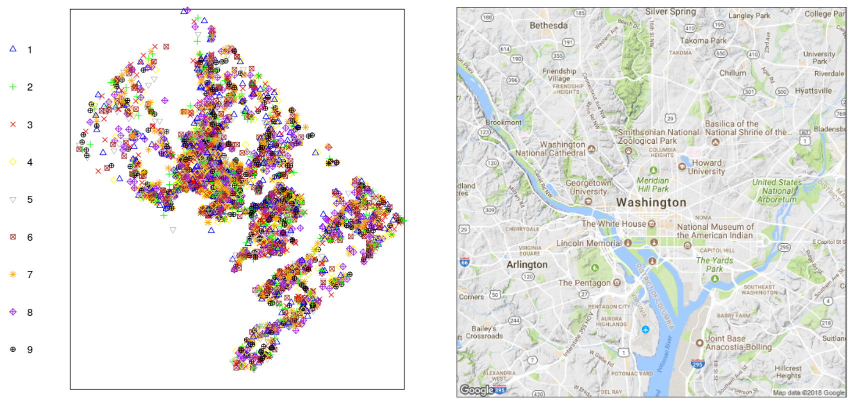


Figure 5: Left: street crimes locations ($n = 5378$); Right: a map of Washington DC.

Figure 5 shows spatial locations of nine types of street crimes committed in Washington DC in January and February 2017, which can be downloaded from `http://opendata.dc.gov/datasets/crime-incidents-in-2017`. Nine types of street crime are included: (1) Other theft, (2) Robbery, (3) Theft from automobile, (4) Motor vehicle theft, (5) Assault with weapon, (6) Sex abuse, (7) Arson, (8) Burglary and (9) Homicide. The numbers of each crime type are $n_1 = 2254$, $n_2 = 366$,

Table 3: List of spatial covariates.

Name	Definition
1. % African	Square root of percentage of African American residents
2. % Hispanic	Square root of percentage of Hispanic residents
3. % Male	Square root of percentage of male residents with age 18-24
4. % HouseRent	Percentage of housing units occupied by renters
5. % Bachelor	Percentage of residents over age 25 with a bachelor's degree
6. MedIncome	Logarithm of median annual per capita income (in $1000)
7. Pdist	Logarithm of the distance to the nearest police station

$n_3 = 1832$, $n_4 = 335$, $n_5 = 332$, $n_6 = 44$, $n_7 = 1$, $n_8 = 259$ and $n_9 = 14$. We omit the rare street crimes "Sex abuse", "Arson" and "Homicide". Using spatial covariates similar to those suggested in Reinhart and Greenhouse (2018), the first 6 spatial covariates listed in Table 3 are obtained from US census data and are constant within each of 179 census tracts partitioning Washington DC, see also Section 6.2. We calculated ourselves the last covariate (distance to nearest police station) which varies smoothly across the city. Square root and log transformations have been applied to some covariates to achieve approximate normal distributions.

6.1 Inference regarding regression coefficients and cross PCFs

Using model (3.1), we assume that the intensity of each street crime is given by

$$\lambda_i(\mathbf{u}; \boldsymbol{\gamma}_i) = \lambda_0(\mathbf{u}) \exp[\gamma_{i0} + \gamma_{i1} z_1(\mathbf{u}) + \cdots + \gamma_{i7} z_7(\mathbf{u})], \qquad i = 1, \ldots, 5, 8.$$

where the $z_k(\cdot)$'s are listed in Table 3. The common first street crime "Other theft" is used as the baseline. The regression parameters are estimated by maximizing the composite likelihood (3.4). The asymptotic standard errors and p-values are computed with $R = 3$ km and either of two types of cross PCFs: using the "refined" kernel estimator $\hat{g}^{\mathrm{r}}$ proposed in Section (3.4) with $b = 0.2$km, or assuming all $g_{ij}(\cdot) \equiv 1$ ("Poisson") for any $i, j = 1, \ldots, 5, 8$. The R^* used for the "refined" kernel estimators is obtained through the data-driven procedure outlined in Remark 1. Estimated regression coefficients, standard deviations, and p-values are summarized in Table 4, and estimated PCF ratios and cross PCF ratios are illustrated in Figure 6.

Figure 6(a) indicates that within and between clustering for crimes types other than "Other theft" is less strong than for "Other theft" up to around 250 meters. After that some crime types appear to be more clustered than "Other theft" but the difference in clustering strength vanishes around 3km distance. In particular, Figure 6 suggests that a multivariate Poisson model is not appropriate for the street crime data.

In Table 4, the Poisson model as expected always gives smaller standard errors for all coefficients. As a result, more regression coefficients appear to be statistically significant at the $\alpha = 0.05$ level (highlighted in blue) compared to those for the proposed method where cross PCFs are estimated from the data. In some cases, the two

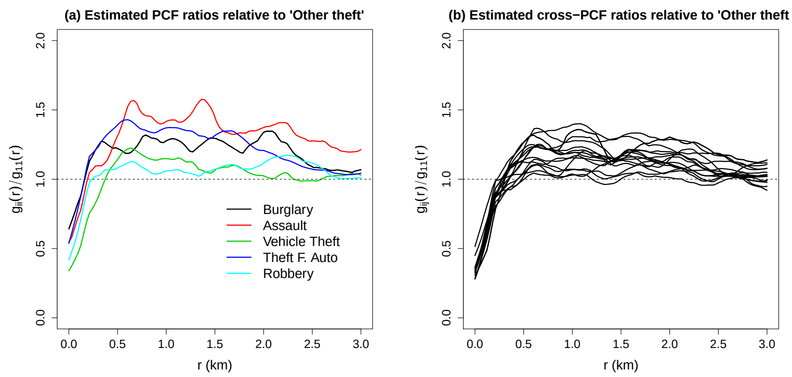


Figure 6: (a) Estimated PCF ratios $g_{ii}(r)/g_{11}(r)$ for $i = 2, \ldots, 5, 8$; (b) estimated cross PCF ratios $g_{ij}(r)/g_{11}(r)$ for $i, j = 2, \ldots, 5, 8$ and $i \neq j$.

methods reach contradictory conclusions. For example, the covariate "% HouseRent" is significant under the Poisson model (p-value 0.028) when comparing "Theft from auto" to the baseline process "Other theft", while the proposed model asserts otherwise with a p-value of 0.352. In such cases, considering the strong spatial correlations displayed in Figure 6, we argue that the proposed method is more reliable.

Based on the proposed method, all estimated coefficients for "% HouseRent" are negative and many of them are significant, suggesting that when "% HouseRent" is large, "Other theft" becomes relatively more frequent compared to all other crime types. Second, no covariate elevates or reduces the relative risk of "Robbery" compared to "Other theft" and no covariate other than "% HouseRent" is significant for the relative risk between "Motor vehicle theft" and "Other theft". Third, "Theft from automobile" tend to occur more often in a neighborhood with more African American/Hispanic population, less young male percentage and residents with relatively low education level, as compared to "Other theft". Fourth, "Assault with weapon" is more likely to occur in a neighborhood with low young male population and low income levels compared to "Other theft". Finally, compared to "Other theft", "Burglary" tends to occur more in areas with low African American population, low education level and larger distance to a police station.

6.2 Residual analysis

In this subsection, we perform a residual analysis to assess goodness of model fit. We divide the data according to the 179 census tracts in Washington DC. Denoting the census tracts $A_1, A_2, \ldots, A_K$, $K = 179$, we define the raw residual for the i'th type of street crime in A_k as

$$\hat{\varepsilon}_{i,k}(\widehat{\boldsymbol{\beta}}) = \sum_{\mathbf{u} \in X_i} I(\mathbf{u} \in A_k) - \sum_{\mathbf{u} \in X^{\mathrm{pl}} \cap A_k} \mathrm{p}_i(\mathbf{u}; \widehat{\boldsymbol{\beta}}), \tag{6.1}$$

for $i = 1, \ldots, p$ and $k = 1, \ldots, K$. Equation (6.1) is essentially a restricted version (within A_k) of the intercept component of $\mathbf{e}_i(\widehat{\boldsymbol{\beta}})$ defined in (3.7). By definition of $\widehat{\boldsymbol{\beta}}$, $\mathbf{e}_i(\widehat{\boldsymbol{\beta}}) = \mathbf{0}$, implying $\sum_{k=1}^{K} \hat{\varepsilon}_{i,k} = 0$ for $i = 1, \ldots, p$. If the model fits the data reasonably well, one should expect most $\hat{\varepsilon}_{i,k}$ to be relatively close to 0.

Table 4: Estimated regression coefficients, standard errors, and p-values for street crime data.

			Std. err.		P-values	
Street crime	Covariate	Coef.	$\hat{g}^{\mathrm{r}}$	Poisson	$\hat{g}^{\mathrm{r}}$	Poisson
Robbery ($n_2 = 366$)	% African	0.894	0.867	0.697	0.302	0.199
	% Hispanic	0.669	0.685	0.499	0.329	0.180
	% Male	0.141	1.183	0.962	0.905	0.884
	% HouseRent	−0.783	0.442	0.352	0.077	**0.026**
	% Bachelor	−1.130	0.970	0.760	0.244	0.137
	MedIncome	−0.071	0.371	0.304	0.847	0.814
	Pdist	0.176	0.108	0.086	0.102	**0.040**
Theft from automobile ($n_3 = 1832$)	% African	2.318	0.813	0.346	**0.004**	**<0.0001**
	% Hispanic	2.369	0.760	0.286	**0.002**	**<0.0001**
	% Male	−2.332	1.049	0.500	**0.026**	**<0.0001**
	% HouseRent	−0.412	0.444	0.188	0.352	**0.028**
	% Bachelor	2.936	0.891	0.417	**0.001**	**<0.0001**
	MedIncome	−0.461	0.339	0.164	0.174	**0.004**
	Pdist	0.071	0.107	0.047	0.508	0.131
Motor vehicle theft ($n_4 = 335$)	% African	−0.451	0.872	0.702	0.605	0.520
	% Hispanic	−0.556	0.724	0.533	0.443	0.297
	% Male	−0.139	1.174	0.962	0.906	0.885
	% HouseRent	−1.295	0.443	0.355	**0.003**	**0.0003**
	% Bachelor	−1.767	0.993	0.785	0.075	**0.024**
	MedIncome	−0.174	0.361	0.300	0.630	0.563
	Pdist	0.205	0.113	0.089	0.070	**0.022**
Assult with weapon ($n_5 = 332$)	% African	1.346	1.004	0.806	0.180	0.095
	% Hispanic	−0.101	0.794	0.541	0.898	0.851
	% Male	−2.76	1.358	1.132	**0.042**	**0.0145**
	% HouseRent	−1.229	0.494	0.377	**0.013**	**0.001**
	% Bachelor	−0.619	1.124	0.839	0.582	0.461
	MedIncome	−0.798	0.391	0.314	**0.041**	**0.011**
	Pdist	0.145	0.122	0.088	0.235	0.100
Burglary ($n_8 = 259$)	% African	−2.332	1.187	0.801	**0.050**	**0.003**
	% Hispanic	−0.029	0.983	0.583	0.977	0.961
	% Male	0.776	1.555	1.039	0.618	0.455
	% HouseRent	−1.930	0.670	0.376	**0.001**	**<0.0001**
	% Bachelor	−3.374	1.327	0.875	**0.011**	**0.001**
	MedIncome	−0.352	0.432	0.300	0.415	0.240
	Pdist	0.359	0.168	0.105	**0.033**	**0.0006**

Following the arguments in Section (3.2) leading to (3.12), we can approximate the variance of $\hat{\varepsilon}_{i,k}(\boldsymbol{\beta}^*)$ by

$$\hat{\sigma}^2_{i,k}(\boldsymbol{\beta}^*, g) = \sum_{\mathbf{u}\in X^{\mathrm{pl}}\cap A_k} [1 - \mathrm{p}_i^*(\mathbf{u})]\mathrm{p}_i^*(\mathbf{u}) + \sum_{\mathbf{u},\mathbf{v}\in X^{\mathrm{pl}}\cap A_k}\sum^{\mathbf{u}\neq\mathbf{v}} \mathrm{p}_i^*(\mathbf{u})\mathrm{p}_i^*(\mathbf{v})T_{ii}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*, g),$$

where $T_{ii}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*, g)$ is defined in (3.10). Consequently, by replacing $\boldsymbol{\beta}$ and cross PCFs by their estimates, the standardized residual can be defined as

$$\hat{\epsilon}_{i,k}(\widehat{\boldsymbol{\beta}}) = \hat{\varepsilon}_{i,k}(\widehat{\boldsymbol{\beta}})/\hat{\sigma}_{i,k}(\widehat{\boldsymbol{\beta}}, \hat{g}^{\mathrm{r}}), \tag{6.2}$$

for $i = 1, \ldots, p$ and $k = 1, \ldots, K$.

Standardized residuals for all census tracts in Washington DC are illustrated in Figure 7. One census tract that does not have any reported street crime activities in January and February 2017 is indicated by black color. Most standardized residuals are inside the range of $[-3, 3]$ for all six types of street crimes, indicating an adequate model fit. Finally, the apparent strong spatial correlations among the residuals further support the use of the proposed method.

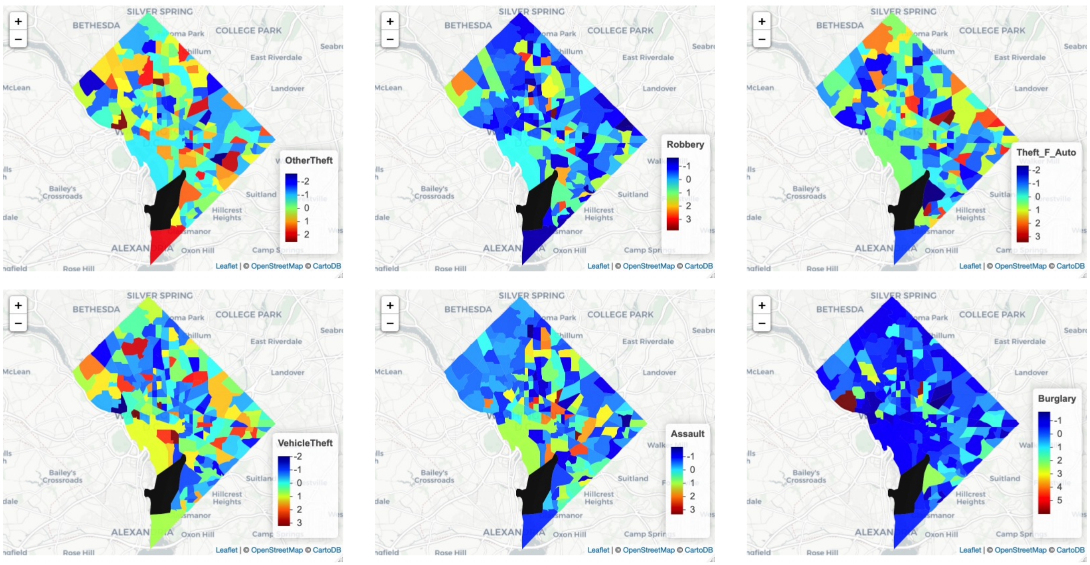


Figure 7: Standardized residuals for 179 census tracts for six types of street crimes.

6.3 Conditional probability maps

We conclude the data analysis by creating a series of conditional probability maps. For any location $\mathbf{u}$, using the fitted $\widehat{\boldsymbol{\beta}}$, we can compute $\mathrm{p}_i(\mathbf{u}, \widehat{\boldsymbol{\beta}})$ for $i = 1, \ldots, p$, using (3.3). Figure 8 shows the $\mathrm{p}_i(\mathbf{u}, \widehat{\boldsymbol{\beta}})$, $i = 1, \ldots, 5, 8$ computed at the 5378 observed crime locations. Recall that given a street crime occurs at location $\mathbf{u}$, $\mathrm{p}_i(\mathbf{u}, \widehat{\boldsymbol{\beta}})$ is the fitted probability that the crime is of the i'th type. The strong spatial patterns in these conditional probabilities are remarkable. For instance, in the southeast part of the city (southeast to the Anacostia river), given a crime occurs, it is much more likely to be of type "Robbery" or "Assault" than in other parts of the city. In contrast, "Theft from automobile" is more likely to be reported in the middle and northern part of the city while the hot spot for "Other theft" is located in the middle-west part of the city.

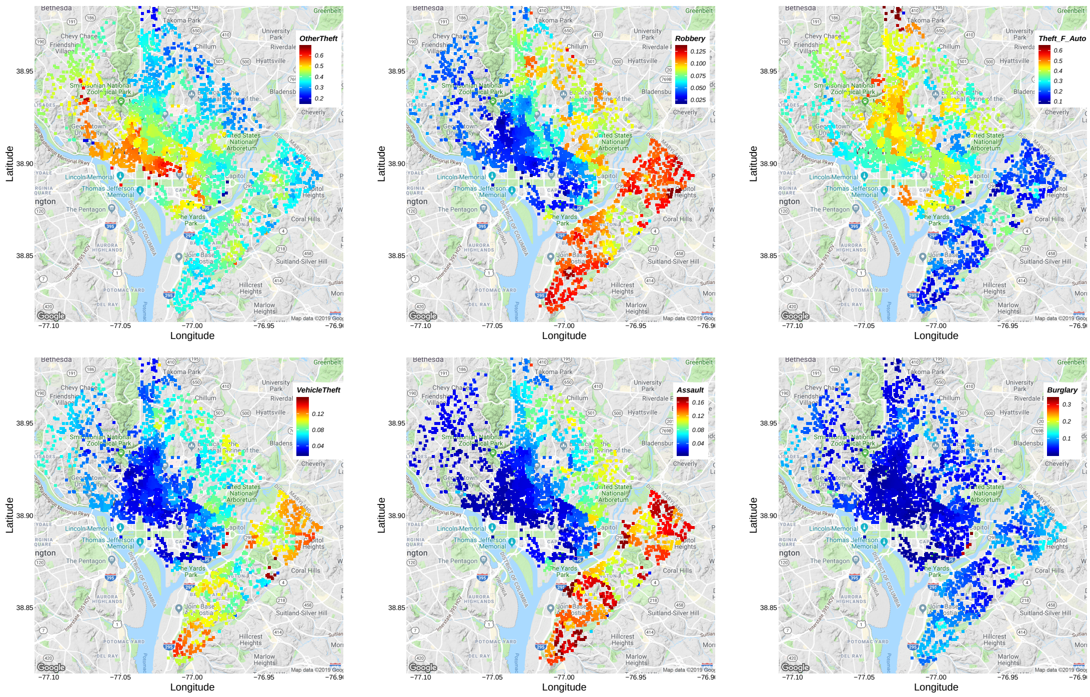


Figure 8: Estimated conditional probability maps for Washington DC.

7 Concluding remarks

We propose a flexible semi-parametric model for multivariate point pattern data. The non-parametric component of the model takes into account features of the multivariate intensity function that are difficult to model or specify while the parametric part facilitates a study of effects of covariates on relative risks of occurrence of different types of points. Moreover, from the parametric part of the model it is possible to construct interesting conditional probability maps.

Our multinomial logistic composite likelihood estimation approach does not require knowledge of the non-parametric model component. It is moreover well founded theoretically since we established the asymptotic properties of the estimation approach in a very general setting that does not require any independence assumptions, neither within or between the different types of points.

Our non-parametric estimation approach allows to estimate cross PCFs up to a common multiplicative factor. This is sufficient for estimating the covariance matrix of regression parameter estimates and for inferring ratios of cross PCFs. However, to infer individual cross PCFs, it seems necessary to introduce parametric models for the cross PCFs. We plan to pursue this in future work.

Our methodology is applicable in very diverse fields. Our example application is within criminology where the estimated conditional probability maps disclose a remarkable structure in the occurrence of various types of street crimes in Washington DC. Other obvious areas of applications are disease mapping in epidemiology and studies of spatial distributions of plant and animal species in ecology. Our approach can further be extended to space-time multivariate point pattern data, which have attracted much interest in various research areas including criminology, see e.g. the thorough review in the the recent paper Reinhart and Greenhouse (2018).

Acknowledgment

Kristian Bjørn Hessellund and Rasmus Waagepetersen were supported by The Danish Council for Independent Research | Natural Sciences, grant DFF – 7014-00074 'Statistics for point processes in space and beyond', and by the 'Centre for Stochastic Geometry and Advanced Bioimaging', funded by grant 8721 from the Villum Foundation.

Appendix: Supplement

A Sensitivity and covariance matrices for $\mathbf{e}(\boldsymbol{\beta})$

Theorem 1. *The sensitivity matrix of the estimating function* $\mathbf{e}(\boldsymbol{\beta})$ *is a symmetric* $(p-1)q \times (p-1)q$ *matrix* $\mathbf{S}(\boldsymbol{\beta})$*, where the diagonal blocks are given by:*

$$\mathbf{S}(\boldsymbol{\beta})_{ii} = \int_W \mathbf{Z}(\mathbf{u},\mathbf{u}) \left[1 - \mathrm{p}_i(\mathbf{u};\boldsymbol{\beta})\right] \mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}) \lambda^{\mathrm{pl}}(\mathbf{u}) \mathrm{d}\mathbf{u}$$

for $i = 1, \ldots, p-1$ *and the off-diagonal blocks are given by:*

$$\mathbf{S}(\boldsymbol{\beta})_{ij} = -\int_W \mathbf{Z}(\mathbf{u},\mathbf{u}) \mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}) \mathrm{p}_j(\mathbf{u};\boldsymbol{\beta}) \lambda^{\mathrm{pl}}(\mathbf{u}) \mathrm{d}\mathbf{u}$$

for distinct $i, j = 1, \ldots, p-1$*. When* $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ *these results simplify to*

$$\mathbf{S}(\boldsymbol{\beta}^*)_{ii} = \int_W \mathbf{Z}(\mathbf{u},\mathbf{u}) \left[1 - \mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)\right] \lambda_i^*(\mathbf{u}) \mathrm{d}\mathbf{u}$$

and

$$\mathbf{S}(\boldsymbol{\beta}^*)_{ij} = -\int_W \mathbf{Z}(\mathbf{u},\mathbf{u}) \mathrm{p}_j(\mathbf{u};\boldsymbol{\beta}^*) \lambda_i^*(\mathbf{u}) \mathrm{d}\mathbf{u}.$$

Proof. Some straightforward algebra yields that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}_i^{\mathsf{T}}} \mathbf{e}_i(\boldsymbol{\beta}) &= \nabla_{\boldsymbol{\beta}_i^{\mathsf{T}}} \sum_{\mathbf{u} \in X_i \cap W} \mathbf{z}(\mathbf{u}) - \sum_{l=1}^{p} \sum_{\mathbf{u} \in X_l \cap W} \mathbf{z}(\mathbf{u}) \nabla_{\boldsymbol{\beta}_i^{\mathsf{T}}} \mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}) \\
&= -\sum_{l=1}^{p} \sum_{\mathbf{u} \in X_l \cap W} \mathbf{Z}(\mathbf{u},\mathbf{u}) \frac{\exp[\boldsymbol{\beta}_i^{\mathsf{T}} \mathbf{z}(\mathbf{u})] \left(1 + \sum_{k=1}^{p-1} \exp[\boldsymbol{\beta}_k^{\mathsf{T}} \mathbf{z}(\mathbf{u})] - \exp[\boldsymbol{\beta}_i^{\mathsf{T}} \mathbf{z}(\mathbf{u})]\right)}{\left(1 + \sum_{k=1}^{p-1} \exp[\boldsymbol{\beta}_k^{\mathsf{T}} \mathbf{z}(\mathbf{u})]\right)^2} \\
&= -\sum_{l=1}^{p} \sum_{\mathbf{u} \in X_l \cap W} \mathbf{Z}(\mathbf{u},\mathbf{u}) \mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}) \left[1 - \mathrm{p}_i(\mathbf{u};\boldsymbol{\beta})\right]. \qquad \text{(A.1)}
\end{aligned}$$

The expectation of (A.1) negated is by Campbell's formula:

$$\mathbf{S}(\boldsymbol{\beta})_{ii} = \mathrm{E}\Big[-\frac{\partial}{\partial \boldsymbol{\beta}_i^\mathsf{T}}\mathbf{e}_i(\boldsymbol{\beta})\Big] = \int_W \mathbf{Z}(\mathbf{u},\mathbf{u})[1-\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta})]\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta})\lambda^{\mathrm{pl}}(\mathbf{u})\mathrm{d}\mathbf{u}.$$

Similarly, for the off-diagonal blocks of the Hessian matrix of $\ell(\boldsymbol{\beta})$, we have that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}_j^\mathsf{T}}\mathbf{e}_i(\boldsymbol{\beta}) &= \nabla_{\boldsymbol{\beta}_j^\mathsf{T}} \sum_{\mathbf{u}\in X_i\cap W}\mathbf{z}(\mathbf{u}) - \sum_{l=1}^{p}\sum_{\mathbf{u}\in X_l\cap W}\mathbf{z}(\mathbf{u})\nabla_{\boldsymbol{\beta}_j^\mathsf{T}}\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta})\\ &= \sum_{l=1}^{p}\sum_{\mathbf{u}\in X_l\cap W}\mathbf{Z}(\mathbf{u},\mathbf{u})\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta})\mathrm{p}_j(\mathbf{u};\boldsymbol{\beta}). \qquad \text{(A.2)}\end{aligned}$$

The expectation of (A.2) negated is

$$\mathbf{S}(\boldsymbol{\beta})_{ij} = \mathrm{E}\Big[-\frac{\partial}{\partial \boldsymbol{\beta}_j^\mathsf{T}}\mathbf{e}_i(\boldsymbol{\beta})\Big] = -\int_W \mathbf{Z}(\mathbf{u},\mathbf{u})\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta})\mathrm{p}_j(\mathbf{u};\boldsymbol{\beta})\lambda^{\mathrm{pl}}(\mathbf{u})\mathrm{d}\mathbf{u}.$$

Finally, when $\boldsymbol{\beta} = \boldsymbol{\beta}^*$, we have that $\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)\lambda^{\mathrm{pl}}(\mathbf{u}) = \lambda_i^*(\mathbf{u})$ due to (3.2), which completes the proof of Theorem 1. □

Theorem 2. *The variance-covariance matrix of* $\mathbf{e}(\boldsymbol{\beta}^*)$ *is a* $(p-1)q\times(p-1)q$ *matrix with blocks given by* (3.9) *for* $i,j=1,\dots,p-1$.

Proof. It is straightforward to prove that $\mathrm{E}[\mathbf{e}(\boldsymbol{\beta}^*)] = \mathbf{0}$ by showing that

$$\begin{aligned}\mathrm{E}[\mathbf{e}_i(\boldsymbol{\beta}^*)] &= \int_W \mathbf{z}(\mathbf{u})\lambda_i^*(\mathbf{u})\mathrm{d}\mathbf{u} - \sum_{l=1}^{p}\int_W \mathbf{z}(\mathbf{u})\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)\lambda_l^*(\mathbf{u})\mathrm{d}\mathbf{u}\\ &= \int_W z(\mathbf{u})\lambda_i^*(\mathbf{u})\mathrm{d}\mathbf{u} - \int_W z(\mathbf{u})\lambda_i^*(\mathbf{u})\mathrm{d}\mathbf{u} = \mathbf{0},\end{aligned}$$

where the second last equality follows from the fact that $\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)\lambda^{\mathrm{pl}}(\mathbf{u}) = \lambda_i^*(\mathbf{u})$ due to (3.2).

The diagonal blocks of $\mathrm{Var}[\mathbf{e}(\boldsymbol{\beta}^*)]$ are then given by $\mathrm{Var}[\mathbf{e}_i(\boldsymbol{\beta}^*)] = \mathrm{E}[\mathbf{e}_i(\boldsymbol{\beta}^*)\mathbf{e}_i(\boldsymbol{\beta}^*)^\mathsf{T}]$ for $i=1,\dots,p-1$, where $\mathbf{e}_i(\boldsymbol{\beta}^*)\mathbf{e}_i(\boldsymbol{\beta}^*)^\mathsf{T}$ is

$$\begin{aligned}&\mathbf{e}_i(\boldsymbol{\beta}^*)\mathbf{e}_i(\boldsymbol{\beta}^*)^\mathsf{T}\\ &= \sum_{\mathbf{u}\in X_i\cap W}\mathbf{z}(\mathbf{u})\sum_{\mathbf{v}\in X_i\cap W}\mathbf{z}(\mathbf{v})^\mathsf{T} - \sum_{\mathbf{u}\in X_i\cap W}\mathbf{z}(\mathbf{u})\sum_{l=1}^{p}\sum_{\mathbf{v}\in X_l\cap W}\mathbf{z}(\mathbf{v})^\mathsf{T}\mathrm{p}_i(\mathbf{v};\boldsymbol{\beta}^*)\\ &\quad -\sum_{l=1}^{p}\sum_{\mathbf{u}\in X_l\cap W}\mathbf{z}(\mathbf{u})\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)\sum_{\mathbf{v}\in X_i\cap W}\mathbf{z}(\mathbf{v})^\mathsf{T}\\ &\quad +\sum_{l=1}^{p}\sum_{\mathbf{u}\in X_l\cap W}\mathbf{z}(\mathbf{u})\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)\sum_{l=1}^{p}\sum_{\mathbf{v}\in X_l\cap W}\mathbf{z}(\mathbf{v})^\mathsf{T}\mathrm{p}_i(\mathbf{v};\boldsymbol{\beta}^*), \qquad \text{(A.3)}\end{aligned}$$

where the first term in (A.3) is

$$\sum_{\mathbf{u}\in X_i\cap W}\mathbf{z}(\mathbf{u})\sum_{\mathbf{v}\in X_i\cap W}\mathbf{z}(\mathbf{v})^\mathsf{T} = \sum_{\mathbf{u},\mathbf{v}\in X_i\cap W}^{\neq}\mathbf{Z}(\mathbf{u},\mathbf{v}) + \sum_{\mathbf{u}\in X_i\cap W}\mathbf{Z}(\mathbf{u},\mathbf{u}),$$

and the second and third terms are of the form

$$
\begin{aligned}
&\sum_{\mathbf{u}\in X_i\cap W}\mathbf{z}(\mathbf{u})\sum_{l=1}^{p}\sum_{\mathbf{v}\in X_l\cap W}\mathbf{z}(\mathbf{v})^{\mathsf{T}}\mathrm{p}_i(\mathbf{v};\boldsymbol{\beta}^*)\\
&\quad=\sum_{\substack{l=1\\ l\neq i}}^{p}\sum_{\substack{\mathbf{u}\in X_i\cap W\\ \mathbf{v}\in X_l\cap W}}\mathbf{Z}(\mathbf{u},\mathbf{v})\mathrm{p}_i(\mathbf{v};\boldsymbol{\beta}^*)\\
&\qquad+\sum_{\mathbf{u},\mathbf{v}\in X_i\cap W}^{\neq}\mathbf{Z}(\mathbf{u},\mathbf{v})\mathrm{p}_i(\mathbf{v};\boldsymbol{\beta}^*)+\sum_{\mathbf{u}\in X_i\cap W}\mathbf{Z}(\mathbf{u},\mathbf{u})\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)
\end{aligned}
\tag{A.4}
$$

and the fourth term is

$$
\begin{aligned}
&\sum_{l=1}^{p}\sum_{\mathbf{u}\in X_l\cap W}\mathbf{z}(\mathbf{u})\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)\sum_{l'=1}^{p}\sum_{\mathbf{v}\in X_{l'}\cap W}\mathbf{z}(\mathbf{v})^{\mathsf{T}}\mathrm{p}_i(\mathbf{v};\boldsymbol{\beta}^*)\\
&\quad=\sum_{\substack{l,l'=1\\ l\neq l'}}^{p}\sum_{\substack{\mathbf{u}\in X_l\cap W\\ \mathbf{v}\in X_{l'}\cap W}}\mathbf{Z}(\mathbf{u},\mathbf{v})\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)\mathrm{p}_i(\mathbf{v};\boldsymbol{\beta}^*)\\
&\qquad+\sum_{l=1}^{p}\sum_{\mathbf{u},\mathbf{v}\in X_l\cap W}^{\neq}\mathbf{Z}(\mathbf{u},\mathbf{v})\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)\mathrm{p}_i(\mathbf{v};\boldsymbol{\beta}^*)\\
&\qquad+\sum_{l=1}^{p}\sum_{\mathbf{u}\in X_l\cap W}\mathbf{Z}(\mathbf{u},\mathbf{u})[\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)]^2.
\end{aligned}
\tag{A.5}
$$

The variance of $\mathbf{e}_i(\boldsymbol{\beta}^*)$ is by Campbell's formula

$$
\begin{aligned}
&\mathrm{Var}[\mathbf{e}_i(\boldsymbol{\beta}^*)]\\
&\quad=\int_{W^2}\mathbf{Z}(\mathbf{u},\mathbf{v})\lambda_i^*(\mathbf{u})\lambda_i^*(\mathbf{v})g_{ii}(\mathbf{u},\mathbf{v})\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}\\
&\qquad-\int_{W^2}\mathbf{Z}(\mathbf{u},\mathbf{v})\mathrm{p}_i(\mathbf{v};\boldsymbol{\beta}^*)\lambda_i^*(\mathbf{u})\sum_{l=1}^{p}\lambda_l^*(\mathbf{v})g_{il}(\mathbf{u},\mathbf{v})\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}\\
&\qquad-\int_{W^2}\mathbf{Z}(\mathbf{u},\mathbf{v})\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)\lambda_i^*(\mathbf{v})\sum_{l=1}^{p}\lambda_l^*(\mathbf{u})g_{il}(\mathbf{u},\mathbf{v})\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}\\
&\qquad+\int_{W^2}\mathbf{Z}(\mathbf{u},\mathbf{v})\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)\mathrm{p}_i(\mathbf{v};\boldsymbol{\beta}^*)\sum_{l,l'=1}^{p}\sum\lambda_l^*(\mathbf{u})\lambda_{l'}^*(\mathbf{v})g_{ll'}(\mathbf{u},\mathbf{v})\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}+\mathbf{S}_{ii}(\boldsymbol{\beta}^*)\mathrm{d}\mathbf{u}\\
&\quad=\int_{W^2}\mathbf{Z}(\mathbf{u},\mathbf{v})\lambda_i^*(\mathbf{u})\lambda_i^*(\mathbf{v})g_{ii}(\mathbf{u},\mathbf{v})\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}\\
&\qquad-\int_{W^2}\mathbf{Z}(\mathbf{u},\mathbf{v})\lambda_i^*(\mathbf{u})\lambda_i^*(\mathbf{v})\Big[\sum_{l=1}^{p}[\mathrm{p}_l(\mathbf{v};\boldsymbol{\beta}^*)g_{il}(\mathbf{u},\mathbf{v})+\mathrm{p}_l(\mathbf{u};\boldsymbol{\beta}^*)g_{il}(\mathbf{u},\mathbf{v})]\Big]\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}\\
&\qquad+\int_{W^2}\mathbf{Z}(\mathbf{u},\mathbf{v})\lambda_i^*(\mathbf{u})\lambda_i^*(\mathbf{v})g^{\mathrm{pl}}(\mathbf{u},\mathbf{v})\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}+\mathbf{S}_{ii}(\boldsymbol{\beta}^*)
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbf{S}_{ii}(\boldsymbol{\beta}^*) + \int_{W^2} \mathbf{Z}(\mathbf{u},\mathbf{v})\lambda_i^*(\mathbf{u})\lambda_i^*(\mathbf{v})[g^{\mathrm{pl}}(\mathbf{u},\mathbf{v}) + g_{ii}(\mathbf{u},\mathbf{v})]\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v} \\
&\quad - \int_{W^2} \mathbf{Z}(\mathbf{u},\mathbf{v})\lambda_i^*(\mathbf{u})\lambda_i^*(\mathbf{v})\Big[\sum_{l=1}^{p}[\mathrm{p}_l(\mathbf{v};\boldsymbol{\beta}^*)g_{il}(\mathbf{u},\mathbf{v}) + \mathrm{p}_l(\mathbf{u};\boldsymbol{\beta}^*)g_{il}(\mathbf{u},\mathbf{v})]\Big]\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}
\end{aligned}
$$

The calculation of $\mathrm{Cov}[\mathbf{e}_i(\boldsymbol{\beta}^*), \mathbf{e}_j(\boldsymbol{\beta}^*)]$ is similar to (A.3) with the first term replaced by $\sum_{\mathbf{u}\in X_i\cap W}\mathbf{z}(\mathbf{u})\sum_{\mathbf{v}\in X_j\cap W}\mathbf{z}(\mathbf{v})^\mathsf{T} = \sum_{\mathbf{u}\in X_i\cap W, \mathbf{v}\in X_j\cap W}\mathbf{z}(\mathbf{u})\mathbf{z}(\mathbf{v})^\mathsf{T}$. Following the same calculations as for the diagonal blocks of $\mathrm{Var}[\mathbf{e}(\boldsymbol{\beta}^*)]$, the off-diagonal blocks are given by

$$
\begin{aligned}
&\mathrm{Cov}[\mathbf{e}_i(\boldsymbol{\beta}^*), \mathbf{e}_j(\boldsymbol{\beta}^*)] \\
&\quad= \mathbf{S}_{ij}(\boldsymbol{\beta}^*) + \int_{W^2} \mathbf{Z}(\mathbf{u},\mathbf{v})\lambda_i^*(\mathbf{u})\lambda_j^*(\mathbf{v})[g^{\mathrm{pl}}(\mathbf{u},\mathbf{v}) + g_{ij}(\mathbf{u},\mathbf{v})]\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v} \\
&\qquad - \int_{W^2} \mathbf{Z}(\mathbf{u},\mathbf{v})\lambda_i^*(\mathbf{u})\lambda_j^*(\mathbf{v})\Big[\sum_{l=1}^{p}\big[\mathrm{p}_l(\mathbf{v};\boldsymbol{\beta}^*)g_{il}(\mathbf{u},\mathbf{v}) + \mathrm{p}_l(\mathbf{u};\boldsymbol{\beta}^*)g_{jl}(\mathbf{u},\mathbf{v})\big]\Big]\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}.
\end{aligned}
$$

The result now follows by the definition of T_{ij} in (3.10). □

B Lower bound for $T_{ii}(\mathbf{u}, \mathbf{v}; \boldsymbol{\beta}^*, g)$

The proof of lower bounds for $T_{ii}(\mathbf{u}, \mathbf{v}; \widehat{\boldsymbol{\beta}}, g^{\mathrm{n}})$ and $T_{ii}(\mathbf{u}, \mathbf{v}; \widehat{\boldsymbol{\beta}}, g^{\mathrm{r}})$ are identical to the proof of $T_{ii}(\mathbf{u}, \mathbf{v}; \boldsymbol{\beta}^*, g)$, therefore, we only provide proofs of $T_{ii}(\mathbf{u}, \mathbf{v}; \boldsymbol{\beta}^*, g)$.

B.1 Lower bounds under constraint (3.16)

$$
\begin{aligned}
&T_{ii}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g) \\
&\quad= 1 + \frac{g_{ii}(\mathbf{u},\mathbf{v})}{g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)} - \sum_{l=1}^{p}\frac{[\mathrm{p}_l^*(\mathbf{v})g_{il}(\mathbf{u},\mathbf{v}) + \mathrm{p}_l^*(\mathbf{u})g_{il}(\mathbf{u},\mathbf{v})]}{g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)} \\
&\quad= 1 + \frac{g_{ii}(\mathbf{u},\mathbf{v})\sum_{l=1}^{p}[\mathrm{p}_l^*(\mathbf{v}) + \mathrm{p}_l^*(\mathbf{u})]/2}{g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)} \\
&\qquad - \sum_{l=1}^{p}\frac{[\mathrm{p}_l^*(\mathbf{v}) + \mathrm{p}_l^*(\mathbf{u})]\,g_{il}(\mathbf{u},\mathbf{v})}{g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)} \\
&\quad\geq 1 + \frac{g_{ii}(\mathbf{u},\mathbf{v})\sum_{l=1}^{p}[\mathrm{p}_l^*(\mathbf{v}) + \mathrm{p}_l^*(\mathbf{u})]/2}{g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)} \\
&\qquad - \sum_{l=1}^{p}\frac{[\mathrm{p}_l^*(\mathbf{v}) + \mathrm{p}_l^*(\mathbf{u})]\sqrt{g_{ii}(\mathbf{u},\mathbf{v})g_{ll}(\mathbf{u},\mathbf{v})}}{g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)} \\
&\quad= 1 + \sum_{l=1}^{p}\frac{[\mathrm{p}_l^*(\mathbf{u}) + \mathrm{p}_l^*(\mathbf{v})](g_{ii}(\mathbf{u},\mathbf{v}) - 2\sqrt{g_{ii}(\mathbf{u},\mathbf{v})g_{ll}(\mathbf{u},\mathbf{v})})}{2g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)} \\
&\quad= 1 + \sum_{l=1}^{p}\frac{[\mathrm{p}_l^*(\mathbf{u}) + \mathrm{p}_l^*(\mathbf{v})]\big[\big(\sqrt{g_{ii}(\mathbf{u},\mathbf{v})} - \sqrt{g_{ll}(\mathbf{u},\mathbf{v})}\big)^2 - g_{ll}(\mathbf{u},\mathbf{v})\big]}{2g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)}
\end{aligned}
$$

$$\geq 1-\sum_{l=1}^{p}\frac{[\mathrm{p}_l^*(\mathbf{u})+\mathrm{p}_l^*(\mathbf{v})]g_{ll}(\mathbf{u},\mathbf{v})}{2g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)}$$
$$\geq 1-\frac{\max_{1\leq l\leq p} g_{ll}(\mathbf{u},\mathbf{v})}{g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)}$$

Since the right-hand side of the last inequality does not depend on i, we have that

$$\min_{1\leq i\leq p} T_{ii}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)\geq 1-\frac{\max_{1\leq l\leq p} g_{ll}(\mathbf{u},\mathbf{v})}{g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)}.$$

B.2 Lower bounds without constraint (3.16)

$$T_{ii}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)=1+\frac{g_{ii}(\mathbf{u},\mathbf{v})}{g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)}-\sum_{l=1}^{p}\frac{[\mathrm{p}_l^*(\mathbf{v})g_{il}(\mathbf{u},\mathbf{v})+\mathrm{p}_l^*(\mathbf{u})g_{il}(\mathbf{u},\mathbf{v})]}{g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)}$$
$$\geq 1+\frac{g_{ii}(\mathbf{u},\mathbf{v})}{g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)}-\frac{2\max_{1\leq l\leq p} g_{il}(\mathbf{u},\mathbf{v})}{g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)}.$$

Therefore,

$$\min_{1\leq i\leq p} T_{ii}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)\geq 1+\frac{\min_{1\leq i\leq p} g_{ii}(\mathbf{u},\mathbf{v})}{g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)}-\frac{2\max_{1\leq i\leq p}\max_{1\leq l\leq p} g_{il}(\mathbf{u},\mathbf{v})}{g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)}$$
$$\geq 1-\frac{2\max_{1\leq l,l'\leq p} g_{ll'}(\mathbf{u},\mathbf{v})-\min_{1\leq i\leq p} g_{ii}(\mathbf{u},\mathbf{v})}{g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)}.$$

C Proofs regarding consistency and asymptotic normality

In the following proofs we several times refer to auxiliary lemmas stated in Section D.

C.1 Proof of Theorem 1

For ease of notation we use the abbreviations $\mathrm{p}_i(\mathbf{u})$ and $\mathrm{p}_i^*(\mathbf{u})$ for $\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta})$ and $\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)$, respectively. To prove Theorem 1 we invoke Theorem 2 in Waagepetersen and Guan (2009) with $V_n=|W_n|^{1/2}\mathbf{I}_q$ where $\mathbf{I}_q$ is the $q\times q$ identity matrix and we define $\bar{\mathbf{J}}(\boldsymbol{\beta})=-\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}^{\mathsf{T}}}\mathbf{e}_n(\boldsymbol{\beta})$ to be the 'average' observed information over W_n. It suffices to verify the following conditions

- W1 There exists a $t'>0$ such that $t'_n\geq t'$ for all sufficiently large n where

$$t'_n=\inf_{\|\mathbf{x}\|=1}\mathbf{x}^{\mathsf{T}}\bar{\mathbf{S}}_n(\boldsymbol{\beta}^*)\mathbf{x}.$$

- W2 As $n\to\infty$, $\|\bar{\mathbf{J}}_n(\boldsymbol{\beta})-\bar{\mathbf{S}}_n(\boldsymbol{\beta})\|_{\max}$ converges to zero in probability for any $\boldsymbol{\beta}\in\mathbb{R}^q$.
- W3 For any $\delta>0$, $\sup_{\||W_n|^{-1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\|\leq\delta}\|\bar{\mathbf{J}}_n(\boldsymbol{\beta})-\bar{\mathbf{J}}_n(\boldsymbol{\beta}^*)\|_{\max}\to 0$ in probability as $n\to\infty$.
- W4 As $n\to\infty$, the sequence $|W_n|^{-1/2}\mathrm{e}^{(n)}(\boldsymbol{\beta}^*)$ is bounded in probability.

Proof of W1. Consider a unit length $(p-1)q \times 1$ vector $\mathbf{x} = [\mathbf{x}_1^\mathsf{T}, \ldots, \mathbf{x}_{p-1}^\mathsf{T}]^\mathsf{T}$ with sub-vectors $\mathbf{x}_i = (x_{i1}, \ldots, x_{iq})^\mathsf{T}$ for $i = 1, \ldots, p-1$. By (3.8),

$$\bar{\mathbf{S}}_n(\boldsymbol{\beta}^*) = |W_n|^{-1} \int_{W_n} \mathbf{A}(\mathbf{u}) \otimes \mathbf{Z}(\mathbf{u}, \mathbf{u}) \lambda^{\mathrm{pl}}(\mathbf{u}) \mathrm{d}\mathbf{u},$$

where $\otimes$ denotes the Kronecker product and $\mathbf{A}(\mathbf{u}) = [a_{ij}(\mathbf{u})]_{ij}$ is a $(p-1) \times (p-1)$ symmetric matrix of the form

$$\mathbf{A}(\mathbf{u}) = \begin{bmatrix} [1 - \mathrm{p}_1^*(\mathbf{u})]\mathrm{p}_1^*(\mathbf{u}) & -\mathrm{p}_1^*(\mathbf{u})\mathrm{p}_2^*(\mathbf{u}) & \cdots & -\mathrm{p}_1^*(\mathbf{u})\mathrm{p}_{p-1}^*(\mathbf{u}) \\ -\mathrm{p}_1^*(\mathbf{u})\mathrm{p}_2^*(\mathbf{u}) & [1 - \mathrm{p}_2^*(\mathbf{u})]\mathrm{p}_2^*(\mathbf{u}) & & -\mathrm{p}_2^*(\mathbf{u})\mathrm{p}_{p-1}^*(\mathbf{u}) \\ \vdots & & \ddots & \vdots \\ -\mathrm{p}_1^*(\mathbf{u})\mathrm{p}_{p-1}^*(\mathbf{u}) & \cdots & & [1 - \mathrm{p}_{p-1}^*(\mathbf{u})]\mathrm{p}_{p-1}^*(\mathbf{u}) \end{bmatrix}. \tag{C.1}$$

Defining now $b_i(\mathbf{u}) = \mathbf{z}(\mathbf{u})^\mathsf{T}\mathbf{x}_i$,

$$\begin{aligned} \mathbf{x}^\mathsf{T}\bar{\mathbf{S}}_n(\boldsymbol{\beta}^*)\mathbf{x} &= |W_n|^{-1} \int_{W_n} \sum_{i,j=1}^{p-1} a_{ij}(\mathbf{u}) b_i(\mathbf{u}) b_j(\mathbf{u}) \lambda^{\mathrm{pl}}(\mathbf{u}) \mathrm{d}\mathbf{u} \\ &= |W_n|^{-1} \int_{W_n} \mathbf{b}^\mathsf{T}(\mathbf{u}) \mathbf{A}(\mathbf{u}) \mathbf{b}(\mathbf{u}) \lambda^{\mathrm{pl}}(\mathbf{u}) \mathrm{d}\mathbf{u}, \end{aligned}$$

where $\mathbf{b}(\mathbf{u}) = [b_1(\mathbf{u}), \ldots, b_{p-1}(\mathbf{u})]^\mathsf{T}$. Using Lemma 1 in the supplementary material, it immediately follows that

$$\begin{aligned} \mathbf{b}^T(\mathbf{u})\mathbf{A}(\mathbf{u})\mathbf{b}(\mathbf{u}) \geq \mathrm{p}_p^*(\mathbf{u}) \sum_{i=1}^{p-1} \mathbf{b}_i^2(\mathbf{u}) \mathrm{p}_i^*(\mathbf{u}) &= \mathrm{p}_p^*(\mathbf{u}) \sum_{i=1}^{p-1} \mathbf{x}_i^\mathsf{T}[\mathbf{z}(\mathbf{u})\mathbf{z}(\mathbf{u})^\mathsf{T} \mathrm{p}_i^*(\mathbf{u})]\mathbf{x}_i \\ &= \mathbf{x}^\mathsf{T}\mathbf{Z}_D(\mathbf{u})\mathbf{x}, \end{aligned}$$

where $\mathbf{Z}_D(\mathbf{u})$ is the $(p-1)q \times (p-1)q$ block-diagonal matrix with diagonal blocks $\mathbf{z}(\mathbf{u})\mathbf{z}(\mathbf{u})^\mathsf{T}\mathrm{p}_i^*(\mathbf{u})\mathrm{p}_p^*(\mathbf{u})$. Therefore, it follows that

$$\begin{aligned} \mathbf{x}^\mathsf{T}\bar{\mathbf{S}}_n(\boldsymbol{\beta}^*)\mathbf{x} &\geq \mathbf{x}^\mathsf{T}\big[|W_n|^{-1} \int_{W_n} \mathbf{Z}_D(\mathbf{u}) \lambda^{\mathrm{pl}}(\mathbf{u}) \mathrm{d}\mathbf{u}\big]\mathbf{x} \\ &\geq \lambda_{\min}\big[|W_n|^{-1} \int_{W_n} \mathbf{Z}_D(\mathbf{u}) \lambda^{\mathrm{pl}}(\mathbf{u}) \mathrm{d}\mathbf{u}\big] \\ &= \min_{1 \leq i \leq p-1} \big\{\lambda_{\min}\big[|W_n|^{-1} \int_{W_n} \mathbf{z}(\mathbf{u})\mathbf{z}(\mathbf{u})^\mathsf{T} \mathrm{p}_i^*(\mathbf{u}) \mathrm{p}_p^*(\mathbf{u}) \lambda^{\mathrm{pl}}(\mathbf{u}) \mathrm{d}\mathbf{u}\big]\big\} \\ &= \min_{1 \leq i \leq p-1} \big\{\lambda_{\min}\big[|W_n|^{-1} \int_{W_n} \mathbf{z}(\mathbf{u})\mathbf{z}(\mathbf{u})^\mathsf{T} \lambda_i^*(\mathbf{u}) \mathrm{p}_p^*(\mathbf{u}) \mathrm{d}\mathbf{u}\big]\big\}, \end{aligned}$$

where the second-last equality follows from the block-diagonal structure of the matrix $\mathbf{Z}_D(\mathbf{u})$. The result now follows from C4. □

Proof of W2. The proof of W2 follows directly from Lemma 2 in the supplementary material and a straightforward application of Chebyshev's inequality. □

Proof of W3. Consider the following inequality

$$\begin{aligned}
&\|\bar{\mathbf{J}}_n(\boldsymbol{\beta}^*) - \bar{\mathbf{J}}_n(\boldsymbol{\beta})\|_{\max} \\
&\qquad \le \|\bar{\mathbf{J}}_n(\boldsymbol{\beta}^*) - \bar{\mathbf{S}}_n(\boldsymbol{\beta}^*)\|_{\max} + \|\bar{\mathbf{S}}_n(\boldsymbol{\beta}^*) - \bar{\mathbf{S}}_n(\boldsymbol{\beta})\|_{\max} + \|\bar{\mathbf{S}}_n(\boldsymbol{\beta}) - \bar{\mathbf{J}}_n(\boldsymbol{\beta})\|_{\max}.
\end{aligned}$$

The first and third terms on the right hand side tends to zero as $n \to \infty$ by W2. To show that the second term tends to zero for $|W_n|^{-1/2}\|(\boldsymbol{\beta} - \boldsymbol{\beta}^*)\| \le \delta$, we use Theorem 1 in the supplementary material and consider first the diagonal blocks of $\bar{\mathbf{S}}_n(\boldsymbol{\beta}^*)$ and $\bar{\mathbf{S}}_n(\boldsymbol{\beta})$:

$$\begin{aligned}
&\|\bar{\mathbf{S}}_n(\boldsymbol{\beta}^*)_{ii} - \bar{\mathbf{S}}_n(\boldsymbol{\beta})_{ii}\|_{\max} \\
&\quad \le \frac{1}{|W_n|}\int_{W_n} \|\mathbf{Z}(\mathbf{u},\mathbf{u})\|_{\max}\lambda^{\mathrm{pl}}(\mathbf{u})\big|[1-\mathrm{p}_i^*(\mathbf{u})]\mathrm{p}_i^*(\mathbf{u}) - [1-\mathrm{p}_i(\mathbf{u})]\mathrm{p}_i(\mathbf{u})\big|\mathrm{d}\mathbf{u} \\
&\quad \le \frac{1}{|W_n|}\int_{W_n} K_1^3 p\big|[1-\mathrm{p}_i^*(\mathbf{u})]\mathrm{p}_i^*(\mathbf{u}) - [1-\mathrm{p}_i(\mathbf{u})]\mathrm{p}_i(\mathbf{u})\big|\mathrm{d}\mathbf{u},
\end{aligned}$$

where the last inequality follows from condition C2. Let $t_{i,\boldsymbol{\beta}}(\mathbf{u}) = [1-\mathrm{p}_i(\mathbf{u})]\mathrm{p}_i(\mathbf{u})$. By straightforward calculations it can be shown that $\|\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}}t_{i,\tilde{\boldsymbol{\beta}}}(\mathbf{u})\|_{\max} \le C$ for some constant $C > 0$ under condition C2 for any $\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\| \le |W_n|^{1/2}\delta$, which in turn gives that

$$\frac{|t_{i,\boldsymbol{\beta}^*}(\mathbf{u}) - t_{i,\boldsymbol{\beta}}(\mathbf{u})|}{|W_n|} \le \frac{\|\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}}t_{i,\tilde{\boldsymbol{\beta}}}(\mathbf{u})\|_{\max}\|\boldsymbol{\beta}^* - \boldsymbol{\beta}\|}{|W_n|} \le \frac{C}{|W_n|^{1/2}}\frac{\|\boldsymbol{\beta}^* - \boldsymbol{\beta}\|}{|W_n|^{1/2}},$$

where the second fraction is bounded by δ while the first fraction converge to 0 as $n \to \infty$. Therefore, $\|\bar{\mathbf{S}}_n(\boldsymbol{\beta}^*)_{ii} - \bar{\mathbf{S}}_n(\boldsymbol{\beta})_{ii}\|_{\max} \to 0$ when $n \to \infty$, $i = 1,\dots,p$. Similarly, $\|\bar{\mathbf{S}}_n(\boldsymbol{\beta}^*)_{ij} - \bar{\mathbf{S}}_n(\boldsymbol{\beta})_{ij}\|_{\max} \to 0$ when $n \to \infty$ for any $i \ne j = 1,\dots,p-1$. □

Proof of W4. By Theorem 2,

$$\begin{aligned}
\boldsymbol{\Sigma}_n(\boldsymbol{\beta}^*, g)_{ij} &= \mathrm{Var}[\mathrm{e}^{(n)}(\boldsymbol{\beta}^*)] \\
&= \int_{W_n}\int_{W_n} \mathbf{Z}(\mathbf{u},\mathbf{v})\lambda_i^*(\mathbf{u})\lambda_j^*(\mathbf{u})h_{ij}(\mathbf{u},\mathbf{v})\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v} + \mathbf{S}_{ij}(\boldsymbol{\beta}^*),
\end{aligned}$$

where the functions

$$\begin{aligned}
h_{ij}(\mathbf{u},\mathbf{v}) = g_{ij}(\mathbf{u},\mathbf{v}) &+ \sum_{l=1}^{p}\sum_{l'=1}^{p}\mathrm{p}_l^*(\mathbf{u})\mathrm{p}_{l'}^*(\mathbf{v})g_{ll'}(\mathbf{u},\mathbf{v}) \\
&- \sum_{l=1}^{p}\mathrm{p}_l^*(\mathbf{v})g_{il}(\mathbf{u},\mathbf{v}) - \sum_{l=1}^{p}\mathrm{p}_l^*(\mathbf{u})g_{jl}(\mathbf{u},\mathbf{v}),
\end{aligned} \tag{C.2}$$

$i, j = 1,\dots,p-1$, can be bounded as follows:

$$\begin{aligned}
|h_{ij}(\mathbf{u},\mathbf{v})| = \Big| g_{ij}(\mathbf{u},\mathbf{v}) - 1 &+ \sum_{l=1}^{p}\sum_{l'=1}^{p}\mathrm{p}_l^*(\mathbf{u})\mathrm{p}_{l'}^*(\mathbf{v})\,[g_{ll'}(\mathbf{u},\mathbf{v}) - 1] \\
&- \sum_{l=1}^{p}\mathrm{p}_l^*(\mathbf{v})[g_{il}(\mathbf{u},\mathbf{v}) - 1] - \sum_{l=1}^{p}\mathrm{p}_l^*(\mathbf{u})[g_{jl}(\mathbf{u},\mathbf{v}) - 1]\Big|
\end{aligned}$$

$$\le |g_{ij}(\mathbf{u},\mathbf{v})-1| + \sum_{l=1}^{p}\sum_{l'=1}^{p}|g_{ll'}(\mathbf{u},\mathbf{v})-1|$$
$$+ \sum_{l=1}^{p}\big[|g_{il}(\mathbf{u},\mathbf{v})-1| + |g_{jl}(\mathbf{u},\mathbf{v})-1|\big]$$
$$\le 4\sum_{l=1}^{p}\sum_{l'=1}^{p}|g_{ll'}(\mathbf{u},\mathbf{v})-1|.$$

Therefore, under conditions C2 and C3 and recalling that g_{ij} is isotropic, it is straightforward to show that

$$\begin{aligned}
&\|\boldsymbol{\Sigma}_n(\boldsymbol{\beta}^*, g)_{ij}\|_{\max} \\
&\quad \le K_1^4 \int_{W_n}\int_{W_n} |h_{ij}(\mathbf{u},\mathbf{v})| \mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v} + K_1^3|W_n| \\
&\quad \le 4K_1^4 \int_{W_n}\int_{W_n} \sum_{l=1}^{p}\sum_{l'=1}^{p}|g_{ll'}(\mathbf{u},\mathbf{v})-1| \mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v} + K_1^3|W_n| \\
&\quad \le 4K_1^4 \int_{W_n} \sum_{l=1}^{p}\sum_{l'=1}^{p} K_2 \mathrm{d}\mathbf{u} + K_1^3|W_n| = (4K_1^4K_2p^2 + K_1^3)|W_n|,
\end{aligned}$$

which implies that $\|W_n\|^{-1}\|\boldsymbol{\Sigma}_n(\boldsymbol{\beta}^*, g)_{ij}\|_{\max}$ is asymptotically bounded for any $i, j = 1, \ldots, p-1$. It then follows from Chebychev's inequality that $|W_n|^{-1/2}\mathrm{e}^{(n)}(\boldsymbol{\beta}^*)$ is (element-wise) bounded in probability as $n \to \infty$. □

C.2 Proof of Theorem 2

By a first-order Taylor-expansion of $\mathbf{e}^{(n)}(\widehat{\boldsymbol{\beta}}) = \mathbf{0}$ around $\boldsymbol{\beta}^*$,

$$\begin{aligned}
\mathbf{0} &= \mathbf{e}^{(n)}(\boldsymbol{\beta}^*) - J_n(\tilde{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \\
&= \mathbf{e}^{(n)}(\boldsymbol{\beta}^*) - [\mathbf{J}_n(\tilde{\boldsymbol{\beta}}) - \mathbf{S}_n(\boldsymbol{\beta}^*)](\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) - \mathbf{S}_n(\boldsymbol{\beta}^*)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)
\end{aligned}$$

where $\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\| \le \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|$. Under condition C4, $\bar{\mathbf{S}}_n^{-1}(\boldsymbol{\beta}^*)$ is well defined for sufficiently large n and thus

$$\begin{aligned}
\sqrt{|W_n|}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) &= \bar{\mathbf{S}}_n^{-1}(\boldsymbol{\beta}^*)[|W_n|^{-1/2}\mathbf{e}^{(n)}(\boldsymbol{\beta}^*)] \\
&\quad - \bar{\mathbf{S}}_n^{-1}(\boldsymbol{\beta}^*)[\bar{\mathbf{J}}_n(\tilde{\boldsymbol{\beta}}) - \bar{\mathbf{S}}_n(\boldsymbol{\beta}^*)][|W_n|^{-1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)] \\
&= \bar{\mathbf{S}}_n^{-1}(\boldsymbol{\beta}^*)[|W_n|^{-1/2}\mathbf{e}^{(n)}(\boldsymbol{\beta}^*)] + o_p(1),
\end{aligned} \tag{C.3}$$

where the last equality follows from W1–W3 in the proof of Theorem 1 and the conclusion of Theorem 1. Therefore, to prove Theorem 2, it suffices to prove the asymptotic normality of $|W_n|^{-1/2}\mathbf{e}^{(n)}(\boldsymbol{\beta}^*)$.

Let for $\mathbf{l} \in \mathbb{Z}^{\mathrm{d}}$, $C(\mathbf{l})$ denote the unit volume hypercube centered around $\mathbf{l}$ and let $\mathcal{D}_n = \{\mathbf{l} \in \mathbb{Z}^d : \; C_n(\mathbf{l}) \cap W_n \neq \emptyset\}$. Furthermore, define

$$\mathbf{Z}_n(\mathbf{l}) = [(\mathbf{Z}_n^{(1)}(\mathbf{l}))^\mathsf{T}, \ldots, (\mathbf{Z}_n^{(p-1)}(\mathbf{l}))^\mathsf{T}]^\mathsf{T}$$

where

$$\mathbf{Z}_n^{(i)}(\mathbf{l}) = \sum_{\mathbf{u}\in X_i\cap C(\mathbf{l})\cap W_n} \mathbf{z}(\mathbf{u}) - \sum_{l=1}^{p} \sum_{\mathbf{u}\in X_l\cap C(\mathbf{l})\cap W_n} \mathbf{z}(\mathbf{u}) \frac{\exp[\boldsymbol{\beta}_i^{*\mathsf{T}}\mathbf{z}(\mathbf{u})]}{1+\sum_{k=1}^{p-1}\exp[\boldsymbol{\beta}_k^{*\mathsf{T}}\mathbf{z}(\mathbf{u})]}$$

is the restriction of $\mathbf{e}_i(\boldsymbol{\beta}^*)$ to $C(\mathbf{l})\cap W_n$.

The asymptotic normality of $|W_n|^{-1/2}\mathbf{e}^{(n)}(\boldsymbol{\beta}^*)$ then follows from Theorem 3.1 in Biscio and Waagepetersen (2019) provided the following holds

$\mathcal{H}1$ $W_1 \subset W_2 \subset \dots$ and $|\bigcup_{l=1}^{\infty} W_l| = \infty$.

$\mathcal{H}2$ There exists $\epsilon > 0$ such that $\alpha_{2,\infty}^X(s) = O(1/s^{d+\epsilon})$.

$\mathcal{H}3$ There exists $\tau > 2d/\epsilon$ such that $\sup_{n\in\mathbb{N}}\sup_{\mathbf{l}\in\mathcal{D}_n} \mathrm{E}\|\mathbf{Z}_n(\mathbf{l}) - \mathrm{E}\mathbf{Z}_n(\mathbf{l})\|^{2+\tau} < \infty$.

$\mathcal{H}4$ We have $0 < \liminf_{n\to\infty} \lambda_{\min}[\bar{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}^*, g)]$, where $\lambda_{\min}(M)$ denotes the smallest eigenvalue of a symmetric matrix M.

Conditions $\mathcal{H}1$ and $\mathcal{H}2$ are assumed in conditions C1 and N1, and condition $\mathcal{H}4$ is ensured by condition N3. Therefore, it suffices to show that $\mathcal{H}3$ holds under the conditions of Theorem 2.

Proof of $\mathcal{H}3$. Since $\mathrm{E}[\mathbf{Z}_n(\mathbf{l})] = \mathbf{0}$ and $m-2 > 2d/\epsilon$ it suffices to show that $\sup_{n\in\mathbb{N}}\sup_{\mathbf{l}\in\mathcal{D}_n} \mathrm{E}\|\mathbf{Z}_n(\mathbf{l})\|^m < \infty$ (i.e. we take $\tau = m-2$). Moreover, letting $Z_{n,j}^{(i)}(\mathbf{l})$ denote the jth component of $\mathbf{Z}_n^{(i)}(\mathbf{l})$,

$$\mathrm{E}\|\mathbf{Z}_n(\mathbf{l})\|^m \le (q(p-1))^{m/2} \sum_{i=1}^{p-1}\sum_{j=1}^{q} \mathrm{E}|Z_{n,j}^{(i)}(\mathbf{l})|^m$$

so we just need to show the boundedness of $\mathrm{E}|Z_{n,j}^{(i)}(\mathbf{l})|^m$.

The binomial formula $(x+y)^m = \sum_{k=0}^m \binom{m}{k} x^k y^{m-k}$ gives that

$$\begin{aligned}
&\mathrm{E}[|Z_{n,j}^{(i)}(\mathbf{l})|^m] \\
&= \sum_{k=0}^{m}\binom{m}{k}\mathrm{E}\Big[\Big(\sum_{\mathbf{u}\in X_i\cap C(\mathbf{l})\cap W_n} z_j(\mathbf{u})\Big)^k \Big(-\sum_{l=1}^{p}\sum_{\mathbf{u}\in X_l\cap C(\mathbf{l})\cap W_n} z_j(\mathbf{u})\mathrm{p}_i^*(\mathbf{u})\Big)^{m-k}\Big] \\
&\le \sum_{k=0}^{m}\binom{m}{k}\mathrm{E}\Big[\Big(\sum_{\mathbf{u}\in X_i\cap C(\mathbf{l})\cap W_n} K_1\Big)^k \Big(\sum_{l=1}^{p}\sum_{\mathbf{u}\in X_l\cap C(\mathbf{l})\cap W_n} K_1\Big)^{m-k}\Big] \\
&= K_1^m \sum_{k=0}^{m}\binom{m}{k}\mathrm{E}\Big[\Big(\sum_{\mathbf{u}\in X_i\cap C(\mathbf{l})\cap W_n} 1\Big)^k \Big(\sum_{l=1}^{p}\sum_{\mathbf{u}\in X_l\cap C(\mathbf{l})\cap W_n} 1\Big)^{m-k}\Big]
\end{aligned}$$

where the inequality follows from assumption C2.

Regarding the expression inside the expectation,

$$
\begin{aligned}
&\Big(\sum_{\mathbf{u}\in X_i\cap C(\mathbf{l})\cap W_n} 1\Big)^k \Big(\sum_{l=1}^{p} \sum_{\mathbf{u}\in X_l\cap C(\mathbf{l})\cap W_n} 1\Big)^{m-k} \\
&= \sum_{\mathbf{u}_1\in X_i\cap C(\mathbf{l})\cap W_n} \sum_{\mathbf{u}_2\in X_i\cap C(\mathbf{l})\cap W_n} \cdots \sum_{\mathbf{u}_k\in X_i\cap C(\mathbf{l})\cap W_n} \\
&\quad \sum_{l_1=1}^{p} \sum_{\mathbf{v}_{1,l_1}\in X_{l_1}\cap C(\mathbf{l})\cap W_n} \sum_{l_2=1}^{p} \sum_{\mathbf{v}_{2,l_2}\in X_{l_2}\cap C(\mathbf{l})\cap W_n} \cdots \sum_{l_{m-k}=1}^{p} \sum_{\mathbf{v}_{m-k,l_{m-k}}\in X_{l_{m-k}}\cap C(\mathbf{l})\cap W_n} 1
\end{aligned}
$$

The above sum consists of p^{m-k} terms of the form

$$
\begin{aligned}
&\sum_{\mathbf{u}_1\in X_i\cap C(\mathbf{l})\cap W_n} \sum_{\mathbf{u}_2\in X_i\cap C(\mathbf{l})\cap W_n} \cdots \sum_{\mathbf{u}_k\in X_i\cap C(\mathbf{l})\cap W_n} \\
&\quad \sum_{\mathbf{v}_{1,l_1}\in X_{l_1}\cap C(\mathbf{l})\cap W_n} \sum_{\mathbf{v}_{2,l_2}\in X_{l_2}\cap C(\mathbf{l})\cap W_n} \cdots \sum_{\mathbf{v}_{m-k,l_{m-k}}\in X_{l_{m-k}}\cap C(\mathbf{l})\cap W_n} 1
\end{aligned}
$$

which again can be split into a number of terms according to the possible combinations of ties between the summation indices $\mathbf{u}_j$, $j = 1, \ldots, k$ and $\mathbf{v}_{j',l_{j'}}$, $j' = 1, \ldots, m-k$, $l_{j'} = 1, \ldots, p$. By Campbell's formula, the expectations of these terms can be evaluated as integrals with respect to cross joint intensities. For the sum where all indices are distinct, the expectation becomes an integral over $C(\mathbf{l})^m$ with respect to the appropriate joint cross intensity $\lambda^{(k_1,k_2\ldots,k_N)}_{j_1\cdots j_N}$ of total order m, where $j_1 = i$, $N \leq p$, $k_1 \geq k$, $\sum_{l=1}^{N} k_l = m$ and $\{j_2, \ldots, j_N\} \subseteq \{1, 2, \ldots, p\} \setminus \{i\}$. For example, if $l_1 = l_2 = \cdots = l_{m-k} = i$ then $N = 1$ and $k_1 = m$ so that the cross joint intensity becomes the mth order joint intensity $\lambda^{(m)}_i$ of X_i. The joint cross intensity $\lambda^{(k_1,k_2\ldots,k_N)}_{j_1\cdots j_N}$ is bounded above by $K_1^m C_g$ by conditions C2 and N2.

If not all indices are distinct we obtain lower order integrals involving lower order joint cross densities. These integrals are of a smaller magnitude compared to the case where all indices are distinct. Therefore, under conditions C2 and N2, we have that

$$
\begin{aligned}
\mathrm{E}&\Big(\sum_{\mathbf{u}\in X_i\cap C(\mathbf{l})\cap W_n} 1\Big)^k \Big(\sum_{l=1}^{p} \sum_{\mathbf{u}\in X_l\cap C(\mathbf{l})\cap W_n} 1\Big)^{m-k} \\
&\leq p^{m-k} K K_1^m C_g |C(\mathbf{l})|^m \\
&= p^{m-k} K K_1^m C_g
\end{aligned}
$$

where $K < \infty$ is an upper bound for the number of combinations of ties mentioned above. This completes the proof of $\mathcal{H}3$. $\square$

C.3 Proof of Theorem 3

The proof of (4.2) in Theorem 3 can be separated into the following steps

A1. $\hat{g}^{\mathrm{n}}_{ij,kl,n}(r; b_n, \boldsymbol{\beta}^*) \xrightarrow{p} g_{ij,kl}(r)$ as $n \to \infty$.

A2. $\hat{g}^{\mathrm{n}}_{ij,kl,n}(r; b_n, \widehat{\boldsymbol{\beta}}) \xrightarrow{p} \hat{g}^{\mathrm{n}}_{ij,kl,n}(r; b_n, \boldsymbol{\beta}^*)$ as $n \to \infty$.

Proof of A1. In the following, we ease the notation by omitting function arguments. We need to show that

$$\hat{g}^{\mathrm{n}}_{ij,kl,n} = \frac{F_{ij,n}}{F_{kl,n}} = \frac{g_{ij}}{g_{kl}} \frac{g_{kl}|W_n|^{-1}F_{ij,n}/[g_{ij}\mathrm{E}(F_{kl,n})|W_n|^{-1}]}{|W_n|^{-1}F_{kl,n}/[\mathrm{E}(F_{kl,n})|W_n|^{-1}]} \xrightarrow{p} g_{ij,kl},$$

as $n \to \infty$, which is equivalent to showing that as $n \to \infty$

$$|W_n|^{-1}[g_{kl}F_{ij,n} - g_{ij}\mathrm{E}(F_{kl,n})] \xrightarrow{p} 0 \tag{C.4}$$

and

$$|W_n|^{-1}[F_{kl,n} - \mathrm{E}(F_{kl,n})] \xrightarrow{p} 0. \tag{C.5}$$

By rewriting (C.4), we can see that

$$\begin{aligned}\frac{1}{|W_n|}[g_{kl}F_{ij,n} - g_{ij}\mathrm{E}(F_{kl,n})] &= \frac{g_{kl}}{|W_n|}[F_{ij,n} - \mathrm{E}(F_{ij,n})] \\ &\quad + \frac{1}{|W_n|}[g_{kl}\mathrm{E}(F_{ij,n}) - g_{ij}\mathrm{E}(F_{kl,n})].\end{aligned}$$

For the first term on the right hand side it follows from Chebychev's inequality that

$$\mathrm{p}\big(\big|g_{kl}|W_n|^{-1}[F_{ij,n} - \mathrm{E}(F_{ij,n})]\big| > \epsilon\big) < \frac{g_{kl}^2\,\mathrm{Var}(|W_n|^{-1}F_{ij,n})}{\epsilon^2}$$

for any $\epsilon > 0$. Thus the first term converges to zero in probability as $n \to \infty$ due to Lemma 5 in the supplementary material along with condition C2. The second term can be expanded as

$$\begin{aligned}\frac{1}{|W_n|}\int_{W_n}\int_{W_n} &\big[g_{kl}(r)g_{ij}(\|\mathbf{u}-\mathbf{v}\|) - g_{ij}(r)g_{ij}(\|\mathbf{u}-\mathbf{v}\|)\big] \\ &\times \Big[\sum_{i=1}^{p}\lambda_i^*(\mathbf{u})\Big]\Big[\sum_{j=1}^{p}\lambda_j^*(\mathbf{v})\Big]k_{b_n}(\|\mathbf{u}-\mathbf{v}\| - r)\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}.\end{aligned} \tag{C.6}$$

By the continuity of $g_{ij}(\cdot)$ and $g_{kl}(\cdot)$, for any $\delta > 0$, we can choose an $\epsilon > 0$ such that $|g_{kl}(r)g_{ij}(\|\mathbf{u}-\mathbf{v}\|) - g_{kl}(\|\mathbf{u}-\mathbf{v}\|)g_{ij}(r)| < \delta$ provided that $|\|\mathbf{u}-\mathbf{v}\| - r| < \epsilon$. By the definition of the kernel function $k(\cdot)$, a bandwidth $b_n < \epsilon$ implies that the integral is only over $\mathbf{u}$ and $\mathbf{v}$ with $|\|\mathbf{u}-\mathbf{v}\| - r| < \epsilon$. By Lemma 3 in the supplementary material, this further implies that (C.6) is bounded by $\delta p^2 K_1^2 \tilde{C}_1 r^{d-1}$. Since $\delta > 0$ can be arbitrarily chosen and $b_n \to 0$, we have that (C.6) converges to zero as $n \to \infty$. Finally, (C.5) follows from a simple application of Chebyshev's inequality using the conclusion of Lemma 5. □

Proof of A2. We show that $\hat{g}^{\mathrm{n}}_{ij,kl,n}(r; b_n, \widehat{\boldsymbol{\beta}}) \to \hat{g}^{\mathrm{n}}_{ij,kl,n}(r; b_n, \boldsymbol{\beta}^*)$ by using a multivariate Taylor approximation:

$$\hat{g}^{\mathrm{n}}_{ij,kl,n}(r; b_n, \widehat{\boldsymbol{\beta}}) - \hat{g}^{\mathrm{n}}_{ij,kl,n}(r; b_n, \boldsymbol{\beta}^*) = \nabla\hat{g}^{\mathrm{n}}_{ij,kl,n}(r; b_n, \boldsymbol{\beta}')(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^{\mathsf{T}}, \tag{C.7}$$

where $\boldsymbol{\beta}'$ is on the line segment that connects $\boldsymbol{\beta}^*$ and $\widehat{\boldsymbol{\beta}}$. Since the estimator $\widehat{\boldsymbol{\beta}}$ is consistent, i.e. $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^{\mathsf{T}} \xrightarrow{p} 0$ as $n \to \infty$, it suffices to show that $\nabla\hat{g}^{\mathrm{n}}_{ij,kl,n}(r; b_n, \boldsymbol{\beta}')$ is bounded. Some tedious algebra yields that under condition C2,

$$\|\nabla F_{ij,n}(r; b_n, \boldsymbol{\beta}')\|_{\max} \le 2\big[\sup_{\mathbf{u}}\|\mathbf{z}(\mathbf{u})\|_{\max}\big]F_{ij,n}(r; b_n, \boldsymbol{\beta}') \le 2K_1 F_{ij,n}(r; b_n, \boldsymbol{\beta}'),$$

which further gives that

$$
\begin{aligned}
&\|\nabla \hat{g}^{\mathrm{n}}_{ij,kl,n}(r;b_n,\boldsymbol{\beta}')\|_{\max} \\
&\quad = \left\|\nabla \frac{F_{ij,n}(r;b_n,\boldsymbol{\beta}')}{F_{kl,n}(r;b_n,\boldsymbol{\beta}')}\right\|_{\max} \le 4K_1 \frac{F_{ij,n}(r;b_n,\boldsymbol{\beta}')}{F_{kl,n}(r;b_n,\boldsymbol{\beta}')} \\
&\quad = 4K_1 \frac{F_{ij,n}(r;b_n,\boldsymbol{\beta}^*)}{F_{kl,n}(r;b_n,\boldsymbol{\beta}^*)} \frac{F_{ij,n}(r;b_n,\boldsymbol{\beta}')}{F_{ij,n}(r;b_n,\boldsymbol{\beta}^*)} \frac{F_{kl,n}(r;b_n,\boldsymbol{\beta}^*)}{F_{kl,n}(r;b_n,\boldsymbol{\beta}')} \\
&\quad = 4K_1 \hat{g}^{\mathrm{n}}_{ij,kl,n}(r;b_n,\boldsymbol{\beta}^*) \frac{F_{ij,n}(r;b_n,\boldsymbol{\beta}')/|W_n|}{F_{ij,n}(r;b_n,\boldsymbol{\beta}^*)/|W_n|} \frac{F_{kl,n}(r;b_n,\boldsymbol{\beta}^*)/|W_n|}{F_{kl,n}(r;b_n,\boldsymbol{\beta}')/|W_n|}.
\end{aligned}
$$

It follows immediately from A1, that $\hat{g}^{\mathrm{n}}_{ij,kl,n}(r;b_n,\boldsymbol{\beta}^*) \xrightarrow{p} g_{ij,kl}(r)$ as $n \to \infty$. Now it remains to show that $(F_{ij,n}(r;b_n,\boldsymbol{\beta}^*)/|W_n|)/(F_{ij,n}(r;b_n,\boldsymbol{\beta}')/|W_n|) \xrightarrow{p} 1$ as $n \to \infty$ or equivalently,

$$
\frac{1}{|W_n|}|F_{ij,n}(r;b_n,\boldsymbol{\beta}^*) - F_{ij,n}(r;b_n,\boldsymbol{\beta}')| \xrightarrow{p} 0 \qquad \text{as } n \to \infty. \tag{C.8}
$$

Using the inequality $\|\nabla F_{ij,n}(r;b_n,\boldsymbol{\beta}')\|_{\max} \le 2K_1 F_{ij,n}(r;b_n,\boldsymbol{\beta}')$ and a Taylor expansion,

$$
\begin{aligned}
&\frac{1}{|W_n|}|F_{ij,n}(r;b_n,\boldsymbol{\beta}^*) - F_{ij,n}(r;b_n,\boldsymbol{\beta}')| \\
&\qquad = \frac{1}{|W_n|}|\nabla F_{ij,n}(r;b_n,\boldsymbol{\beta}'')^{\mathsf{T}}(\boldsymbol{\beta}^* - \boldsymbol{\beta}')| \\
&\qquad \le \frac{2K_1(p-1)q\|\boldsymbol{\beta}^* - \boldsymbol{\beta}'\|_{\max}}{|W_n|} F_{ij,n}(r;b_n,\boldsymbol{\beta}''),
\end{aligned} \tag{C.9}
$$

where $\boldsymbol{\beta}''$ is between $\boldsymbol{\beta}^*$ and $\boldsymbol{\beta}'$. Recalling the definition of $F_{ij,n}$,

$$
\begin{aligned}
F_{ij,n}(r;b_n,\boldsymbol{\beta}'') &= \sum_{\substack{\mathbf{u}\in X_i\cap W_n \\ \mathbf{v}\in X_j\cap W_n}}^{\neq} \frac{k_{b_n}(\|\mathbf{u}-\mathbf{v}\|-r)}{\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}'')\mathrm{p}_j(\mathbf{v};\boldsymbol{\beta}'')} \\
&= \sum_{\substack{\mathbf{u}\in X_i\cap W_n \\ \mathbf{v}\in X_j\cap W_n}}^{\neq} \frac{k_{b_n}(\|\mathbf{u}-\mathbf{v}\|-r)}{\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)\mathrm{p}_j(\mathbf{v};\boldsymbol{\beta}^*)} \times \frac{\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)\mathrm{p}_j(\mathbf{v};\boldsymbol{\beta}^*)}{\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}'')\mathrm{p}_j(\mathbf{v};\boldsymbol{\beta}'')}.
\end{aligned}
$$

Under condition C2, $\|\mathbf{z}(\mathbf{u})\|_{\max} \le K_1$ ensures that there exists a $c > 0$ such that

$$
\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*) \ge c, \qquad i = 1,\dots,p,
$$

for any $\mathbf{u} \in \{\mathbf{u} \in \bigcup_{l=1}^{\infty} W_l : \lambda_0(\mathbf{u}) > 0\}$. By consistency of $\hat{\boldsymbol{\beta}}_n$ we have that $\boldsymbol{\beta}' \xrightarrow{p} \boldsymbol{\beta}^*$ and hence $\boldsymbol{\beta}'' \xrightarrow{p} \boldsymbol{\beta}^*$, which implies that with a probability tending to 1,

$$
\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}'') \ge c, \qquad i = 1,\dots,p,
$$

for any $\mathbf{u} \in \{\mathbf{u} \in \bigcup_{l=1}^{\infty} W_l : \lambda_0(\mathbf{u}) > 0\}$. Consequently, with a probability tending to 1, we can bound $F_{ij,n}(r;b_n,\boldsymbol{\beta}'')$ as

$$
F_{ij,n}(r;b_n,\boldsymbol{\beta}'') \le \frac{1}{c^2} \sum_{\substack{\mathbf{u}\in X_i\cap W_n \\ \mathbf{v}\in X_j\cap W_n}}^{\neq} \frac{k_{b_n}(\|\mathbf{u}-\mathbf{v}\|-r)}{\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)\mathrm{p}_j(\mathbf{v};\boldsymbol{\beta}^*)} = \frac{1}{c^2} F_{ij,n}(r;b_n,\boldsymbol{\beta}^*). \tag{C.10}
$$

Using equality (C.5),

$$\begin{aligned}|W_n|^{-1}F_{ij,n}(r;b_n,\boldsymbol{\beta}^*) &\xrightarrow{p} |W_n|^{-1}\mathrm{E}[F_{ij,n}(r;b_n,\boldsymbol{\beta}^*)]\\ &= |W_n|^{-1}\int_{W_n^2}\lambda^{\mathrm{pl}}(\mathbf{u})\lambda^{\mathrm{pl}}(\mathbf{v})g_{ij}(\mathbf{u},\mathbf{v})k_b(\|\mathbf{u}-\mathbf{v}\|-r)\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}\\ &\le \frac{p^2K_1^3}{|W_n|}\int_{W_n^2}k_{b_n}(\|\mathbf{u}-\mathbf{v}\|-r)\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}\\ &\le p^2K_1^3\int_{\mathbb{R}^{\mathrm{d}}}k_{b_n}(\|\mathbf{x}\|-r)\mathrm{d}\mathbf{x},\end{aligned}$$

where the first inequality follows from condition C2. Combining the above inequality with Lemma 3 gives that, with a probability tending to 1, there exists a constant C_F such that

$$|W_n|^{-1}F_{ij,n}(r;b_n,\boldsymbol{\beta}^*) \le C_F. \tag{C.11}$$

Finally, for any $\epsilon > 0$ and letting $\bar{F}^*_{ij,n} = F_{ij,n}(r;b_n,\boldsymbol{\beta}^*)/|W_n|$,

$$P(\|\boldsymbol{\beta}'-\boldsymbol{\beta}^*\|_{\max}\bar{F}^*_{ij,n} > \epsilon) \le P(\|\boldsymbol{\beta}'-\boldsymbol{\beta}^*\|_{\max}C_F > \epsilon) + P(\bar{F}^*_{ij,n} > C_F)$$

Thus, since $\boldsymbol{\beta}' \xrightarrow{p} \boldsymbol{\beta}^*$ as $n\to\infty$, (C.8) immediately follows from inequalities (C.9)–(C.11), which completes the proof of A2. □

Therefore, the proof of (4.2) in Theorem 3 is finished.

To show (4.3) in Theorem 3, note that

$$\begin{aligned}&|\hat{g}^{\mathrm{r}}_{ij,kl,n}(r;b_n,\widehat{\boldsymbol{\beta}}_n)-g_{ij,kl}(r)|\\ &\quad\le |\hat{g}^{\mathrm{r}}_{ij,kl,n}(r;b_n,\widehat{\boldsymbol{\beta}}_n)-\hat{g}^{\mathrm{n}}_{ij,kl,n}(r;b_n,\widehat{\boldsymbol{\beta}}_n)|\\ &\qquad+|\hat{g}^{\mathrm{n}}_{ij,kl,n}(r;b_n,\widehat{\boldsymbol{\beta}}_n)-g_{ij,kl}(r)|,\end{aligned}$$

so that it is enough to show that

$$|\hat{g}^{\mathrm{r}}_{ij,kl,n}(r;b_n,\widehat{\boldsymbol{\beta}}_n)-\hat{g}^{\mathrm{n}}_{ij,kl,n}(r;b_n,\widehat{\boldsymbol{\beta}}_n)| \xrightarrow{p} 0 \qquad \text{for } r\ge R^* \text{ as } n\to\infty. \tag{C.12}$$

Combining restriction (3.16) and the convergence $\hat{g}^{\mathrm{n}}_{ij,kl,n}(r;b_n,\widehat{\boldsymbol{\beta}}_n) \xrightarrow{p} g_{ij,kl}(r)$, as $n\to\infty$, we obtain that with a probability tending to 1, $\hat{g}^{\mathrm{n}}_{ij,kl,n}(r;b_n,\widehat{\boldsymbol{\beta}}_n)$ satisfies restriction

$$\hat{g}_{ij,kl,n}(r;b_n,\widehat{\boldsymbol{\beta}}_n) \le \sqrt{\hat{g}_{ii,kl,n}(r;b_n,\widehat{\boldsymbol{\beta}}_n)\hat{g}_{jj,kl,n}(r;b_n,\widehat{\boldsymbol{\beta}}_n)}$$

for a sufficiently large n, in which case $|\hat{g}^{\mathrm{r}}_{ij,kl,n}(r;b_n,\widehat{\boldsymbol{\beta}}_n)-\hat{g}^{\mathrm{n}}_{ij,kl,n}(r;b_n,\widehat{\boldsymbol{\beta}}_n)| = 0$ by the design of algorithm (3.17). Therefore, (C.12) follows, which completes the proof of (4.3) in Theorem 3.

D Auxiliary Lemmas

Lemma 1. *Define* $\mathbf{A}(\mathbf{u})$ *as in* (C.1). *Then for any* $\mathbf{b}=(b_1,\ldots,b_{p-1})^\mathsf{T}\in\mathbb{R}^{p-1}$,

$$\mathbf{b}^\mathsf{T}\mathbf{A}(\mathbf{u})\mathbf{b} \ge \mathrm{p}^*_p(\mathbf{u})\sum_{i=1}^{p-1}b_i^2\mathrm{p}^*_i(\mathbf{u}).$$

Proof. Using the Cauchy-Schwartz inequality,

$$\begin{aligned}
\mathbf{b}^\intercal \mathbf{A}(\mathbf{u})\mathbf{b} &= \sum_{i=1}^{p-1} b_i^2 \mathrm{p}_i^*(\mathbf{u}) - \Big[\sum_{i=1}^{p-1} b_i \mathrm{p}_i^*(\mathbf{u})\Big]^2 \\
&= \sum_{i=1}^{p-1} b_i^2 \mathrm{p}_i^*(\mathbf{u}) - \Big[\sum_{i=1}^{p-1} b_i \frac{\mathrm{p}_i^*(\mathbf{u})}{1-\mathrm{p}_p^*(\mathbf{u})}\Big]^2 [1-\mathrm{p}_p^*(u)]^2 \\
&\geq \sum_{i=1}^{p-1} b_i^2 \mathrm{p}_i^*(\mathbf{u}) - \Big[\sum_{i=1}^{p-1} b_i^2 \mathrm{p}_i^*(\mathbf{u})\Big][1-\mathrm{p}_p^*(\mathbf{u})] = \mathrm{p}_p^*(\mathbf{u}) \sum_{i=1}^{p-1} b_i^2 \mathrm{p}_i^*(\mathbf{u}).
\end{aligned}$$

□

Lemma 2. *Assume C1–C3 holds true. Then as* $n \to \infty$, $|W_n|^{-1} \operatorname{Var}^{\odot}[\mathbf{J}_n(\boldsymbol{\beta})_{ij}] \leq C$ *for some constant* $0 < C < \infty$ *for any* $i, j = 1, \dots, p-1$. *Here, for a random matrix* $\mathbf{A}$, $\operatorname{Var}^{\odot}(\mathbf{A})$ *denotes the element-wise variance of* $\mathbf{A}$.

Proof. Denote by $\mathbf{A}^{\odot 2}$ the element-wise square of the matrix $\mathbf{A}$ and by $\mathbf{A} \odot \mathbf{B}$ the element-wise product of matrices $\mathbf{A}$ and $\mathbf{B}$. Recall that $\mathrm{E}[\mathbf{J}_n(\boldsymbol{\beta})_{ij}] = -\mathbf{S}_n(\boldsymbol{\beta})_{ij}$, $i, j = 1, \dots, p-1$. Clearly, $\operatorname{Var}^{\odot}[\mathbf{J}_n(\boldsymbol{\beta})_{ij}] = \mathrm{E}[\mathbf{J}_n^{\odot 2}(\boldsymbol{\beta})_{ij}] - \mathbf{S}_n^{\odot 2}(\boldsymbol{\beta})_{ij}$. Let

$$\mathbf{H}_{ii}(\mathbf{u}) = \mathbf{Z}(\mathbf{u}, \mathbf{u}) \mathrm{p}_i(\mathbf{u}; \boldsymbol{\beta})[1 - \mathrm{p}_i(\mathbf{u}; \boldsymbol{\beta})]$$

and

$$\mathbf{H}_{ij}(\mathbf{u}) = \mathbf{Z}(\mathbf{u}, \mathbf{u}) \mathrm{p}_i(\mathbf{u}; \boldsymbol{\beta}) \mathrm{p}_j(\mathbf{u}; \boldsymbol{\beta}).$$

The block elements in $\mathbf{J}_n^{\odot 2}(\boldsymbol{\beta})$ are of the form

$$\mathbf{J}_n^{\odot 2}(\boldsymbol{\beta})_{ij} = \sum_{l,l'=1}^{p}\sum \sum_{\substack{\mathbf{u} \in X_l \cap W_n \\ \mathbf{v} \in X_{l'} \cap W_n}} \mathbf{H}_{ij}(\mathbf{u}) \odot \mathbf{H}_{ij}(\mathbf{v}),$$

for $i, j = 1, \dots, p-1$, whose expectations are of the form

$$\begin{aligned}
\mathrm{E}[\mathbf{J}_n^{\odot 2}(\boldsymbol{\beta})_{ij}] = &\sum_{l,l'=1}^{p}\sum \int_{W_n} \int_{W_n} \mathbf{H}_{ij}(\mathbf{u}) \odot \mathbf{H}_{ij}(\mathbf{v}) \lambda_{ll'}(\mathbf{u}, \mathbf{v}) \mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v} \\
&+ \sum_{l=1}^{p} \int_{W_n} \mathbf{H}_{ij}(\mathbf{u}) \odot \mathbf{H}_{ij}(\mathbf{u}) \lambda_l^*(\mathbf{u}) \mathrm{d}\mathbf{u}.
\end{aligned}$$

Using Theorem 1, it follows that the squared sensitivity is given by

$$\mathbf{S}_n^{\odot 2}(\boldsymbol{\beta})_{ij} = \sum_{l,l'=1}^{p}\sum \int_{W_n} \int_{W_n} \mathbf{H}_{ij}(\mathbf{u}) \odot \mathbf{H}_{ij}(\mathbf{v}) \lambda_l^*(\mathbf{u}) \lambda_l^*(\mathbf{v}) \mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}$$

for $i, j = 1, \dots, p-1$. By condition C2, $\|\mathbf{H}_{ij}\|_{\max} \leq K_1^2$ for any $i, j = 1, \dots, p-1$, which, together with condition C3 and isotropy of $g_{ll'}$, further implies that

$$
\begin{aligned}
&\operatorname{Var}^{\odot}[\mathbf{J}_n(\boldsymbol{\beta})_{ij}] \\
&= \int_{W_n^2} \mathbf{H}_{ij}(\mathbf{u}) \odot \mathbf{H}_{ij}(\mathbf{v}) \sum_{l,l'=1}^{p}\sum [\lambda_{ll'}(\mathbf{u},\mathbf{v}) - \lambda_l^*(\mathbf{u})\lambda_{l'}(\mathbf{v})]\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v} \\
&\quad + \int_{W_n} \mathbf{H}_{ij}(\mathbf{u}) \odot \mathbf{H}_{ij}(\mathbf{u})\lambda^{\mathrm{pl}}(\mathbf{u})\mathrm{d}\mathbf{u} \\
&\leq \int_{W_n^2} K_1^4 \sum_{l,l'=1}^{p}\sum |\lambda_{ll'}(\mathbf{u},\mathbf{v}) - \lambda_l^*(\mathbf{u})\lambda_{l'}^*(\mathbf{v})|\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v} + \int_{W_n} K_1^4 \lambda^{\mathrm{pl}}(\mathbf{u})\mathrm{d}\mathbf{u} \\
&\leq \sum_{l,l'=1}^{p}\sum \int_{W_n^2} K_1^6 |g_{ll'}(\mathbf{u},\mathbf{v}) - 1|\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v} + \sum_{l=1}^{p} \int_{W_n} K_1^5 \mathrm{d}\mathbf{u} \\
&\leq \sum_{l,l'=1}^{p}\sum \int_{W_n} K_1^6 K_2 \mathrm{d}\mathbf{u} + pK_1^5|W_n| \\
&= |W_n|(p^2K_1^6K_2 + pK_1^5),
\end{aligned}
$$

which yields that $|W_n|^{-1}\operatorname{Var}^{\odot}[\mathbf{J}_n(\boldsymbol{\beta})]_{ij} \leq C$ with $C = p^2K_1^6K_2 + pK_1^5$, for any $i, j = 1, \ldots, p-1$. □

Lemma 3. *Let $b > 0$ be a bandwidth and $k_b(\cdot) = k(\cdot/b)/b$ with a kernel function $k(\cdot)$ defined on a bounded support in $\mathbb{R}$. Then for b small enough, we have that*

$$
\begin{aligned}
&\int_{\mathbb{R}^d} k_b(\|\mathbf{u}\| - r)\mathrm{d}\mathbf{u} \leq \tilde{C}_1 r^{d-1} \\
&\int_{\mathbb{R}^d} k_b^2(\|\mathbf{u}\| - r)\mathrm{d}\mathbf{u} \leq \tilde{C}_2 \frac{1}{b} r^{d-1},
\end{aligned}
$$

where $\tilde{C}_1$ and $\tilde{C}_2$ are some positive constants.

Proof. Without loss of generality we assume that the kernel function $k(\cdot)$ has a bounded support $[-1, 1]$. Using the polar coordinates transformation

$$\int_{\mathbb{R}^d} f(\mathbf{u})\mathrm{d}\mathbf{u} = \int_{\mathbb{S}^{\mathrm{d}-1}} \int_0^\infty f(t\mathbf{v})t^{\mathrm{d}-1}\mathrm{d}t\nu_d(\mathrm{d}\mathbf{v}),$$

where $\nu_d(\cdot)$ is surface measure on the unit sphere $\mathbb{S}^{\mathrm{d}-1}$ in $\mathbb{R}^d$, we have that

$$
\begin{aligned}
I(r;b) \equiv \int_{\mathbb{R}^d} k_b(\|\mathbf{u}\| - r)\mathrm{d}\mathbf{u} &= \frac{1}{b}\int_{\mathbb{R}^d} k\left(\frac{\|\mathbf{u}\| - r}{b}\right)\mathrm{d}\mathbf{u} \\
&= \frac{1}{b}\int_{\mathbb{S}^{\mathrm{d}-1}} \int_0^\infty k\left(\frac{t-r}{b}\right) t^{d-1}\mathrm{d}t\nu_d(\mathrm{d}\mathbf{v}) \\
&= \frac{\nu_d\left(\mathbb{S}^{\mathrm{d}-1}\right)}{b}\int_0^\infty k\left(\frac{t-r}{b}\right) t^{d-1}\mathrm{d}t.
\end{aligned}
$$

When substituting the variable $s = (t-r)/b$, we have that

$$
\begin{aligned}
I(r;b) &= \nu_d(\mathbb{S}^{\mathrm{d}-1}) \int_{-r/b}^{\infty} k(s)(bs+r)^{d-1} \mathrm{d}s \\
&\leq \nu_d(\mathbb{S}^{\mathrm{d}-1}) \big[\sup_{s\in[-1,1]} k(s) \big] \int_{\max\{-1,-r/b\}}^{1} (bs+r)^{d-1} \mathrm{d}s. \\
&= \nu_d(\mathbb{S}^{\mathrm{d}-1}) \big[\sup_{s\in[-1,1]} k(s) \big] \frac{1}{db} \big[(b+r)^d - (b\max\{-1,-r/b\}+r)^d \big].
\end{aligned}
$$

Applying L'Hospital's rule,

$$
\lim_{b\to 0} \frac{1}{b} \left[(b+r)^d - (b\max\{-1,-r/b\}+r)^d \right] = 2dr^{d-1}.
$$

Hence, for b small enough, there exists a constant $\tilde{C}_1$ so that

$$
I(r;b) \leq \tilde{C}_1 r^{d-1}.
$$

The second inequality follows the same way. $\square$

Lemma 4. *The variance of $F_{ij}(r;b,\boldsymbol{\beta}^*)$ in (3.15) for $i \neq j$ is:*

$$
\begin{aligned}
&\mathrm{Var}[F_{ij}(r;b,\boldsymbol{\beta}^*)] \\
&\quad = \int_{W^4} [g_{ij}^{(2,2)}(\mathbf{u}_1,\mathbf{u}_2,\mathbf{v}_1,\mathbf{v}_2) - g_{ij}(\mathbf{u}_1,\mathbf{v}_1)g_{ij}(\mathbf{u}_2,\mathbf{v}_2)] \\
&\qquad \times t(\mathbf{u}_1,\mathbf{v}_1)t(\mathbf{u}_2,\mathbf{v}_2)\mathrm{d}\mathbf{u}_1\mathrm{d}\mathbf{u}_2\mathrm{d}\mathbf{v}_1\mathrm{d}\mathbf{v}_2 \\
&\qquad + \int_{W^3} g_{ij}^{(1,2)}(\mathbf{u},\mathbf{v}_1,\mathbf{v}_2)t(\mathbf{u},\mathbf{v}_1)\frac{\lambda^{\mathrm{pl}}(\mathbf{v}_2)}{\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)}k_b(\|\mathbf{u}-\mathbf{v}_2\|-r)\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{u}_1\mathrm{d}\mathbf{v}_2 \\
&\qquad + \int_{W^3} g_{ij}^{(2,1)}(\mathbf{u}_1,\mathbf{u}_2,\mathbf{v})t(\mathbf{u}_1,\mathbf{v})\frac{\lambda^{\mathrm{pl}}(\mathbf{u}_2)}{\mathrm{p}_j(\mathbf{v};\boldsymbol{\beta}^*)}k_b(\|\mathbf{u}_2-\mathbf{v}\|-r)\mathrm{d}\mathbf{u}_1\mathrm{d}\mathbf{u}_2\mathrm{d}\mathbf{v} \\
&\qquad + \int_{W^2} g_{ij}^{(1,1)}(\mathbf{u},\mathbf{v})t(\mathbf{u},\mathbf{v})[\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)\mathrm{p}_j(\mathbf{v};\boldsymbol{\beta}^*)]^{-1}k_b(\|\mathbf{u}-\mathbf{v}\|-r)\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}
\end{aligned}
$$

and for $i = j$:

$$
\begin{aligned}
&\mathrm{Var}[F_{ii}(r;b,\boldsymbol{\beta}^*)] \\
&\quad = \int_{W^4} [g_i^{(4)}(\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3,\mathbf{u}_4) - g_i(\mathbf{u}_1,\mathbf{u}_2)g_i(\mathbf{u}_3,\mathbf{u}_4)] \\
&\qquad \times t(\mathbf{u}_1,\mathbf{u}_2)t(\mathbf{u}_3,\mathbf{u}_4)\mathrm{d}\mathbf{u}_1\mathrm{d}\mathbf{u}_2\mathrm{d}\mathbf{u}_3\mathrm{d}\mathbf{u}_4 \\
&\qquad + 4\int_{W^3} g_i^{(3)}(\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3)t(\mathbf{u}_1,\mathbf{u}_2)\frac{\lambda^{\mathrm{pl}}(\mathbf{u}_3)}{\mathrm{p}_i(\mathbf{u}_1;\boldsymbol{\beta}^*)}k_b(\|\mathbf{u}_1-\mathbf{u}_3\|-r)\mathrm{d}\mathbf{u}_1\mathrm{d}\mathbf{u}_2\mathrm{d}\mathbf{u}_3 \\
&\qquad + 2\int_{W^2} g_i(\mathbf{u}_1,\mathbf{u}_2)t(\mathbf{u}_1,\mathbf{u}_2)[\mathrm{p}_i(\mathbf{u}_1;\boldsymbol{\beta}^*)\mathrm{p}_i(\mathbf{u}_2;\boldsymbol{\beta}^*)]^{-1}k_b(\|\mathbf{u}_1-\mathbf{u}_2\|-r)\mathrm{d}\mathbf{u}_1\mathrm{d}\mathbf{u}_2,
\end{aligned}
$$

where $t(\mathbf{u},\mathbf{v}) = k_b(\|\mathbf{u}-\mathbf{v}\|-r)\lambda^{\mathrm{pl}}(\mathbf{u})\lambda^{\mathrm{pl}}(\mathbf{v})$.

Proof. The variance is

$$\mathrm{Var}[F_{ij}(r;b,\boldsymbol{\beta}^*)] = \mathrm{E}[F_{ij}(r;b,\boldsymbol{\beta}^*)^2] - \mathrm{E}[F_{ij}(r;b,\boldsymbol{\beta}^*)]^2,$$

where $\mathrm{E}[F_{ij}(r;b,\boldsymbol{\beta}^*)]^2 = \left[\int_{W^2} g_{ij}(\mathbf{u},\mathbf{v})t(\mathbf{u},\mathbf{v})\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}\right]^2$. In the following we denote by $s_{ij}(\mathbf{u},\mathbf{v}) = [\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)\mathrm{p}_j(\mathbf{v};\boldsymbol{\beta}^*)]^{-1}$ and suppose first that $i \neq j$. Then we have that

$$\begin{aligned}
&\mathrm{E}[F_{ij}(r;b,\boldsymbol{\beta}^*)^2] \\
&\quad = \mathrm{E}\sum^{\neq}_{\substack{\mathbf{u}_1\in X_i\cap W\\ \mathbf{v}_1\in X_j\cap W}} s_{ij}(\mathbf{u}_1,\mathbf{v}_1)k_b(\|\mathbf{u}_1-\mathbf{v}_1\|-r)\sum^{\neq}_{\substack{\mathbf{u}_2\in X_i\cap W\\ \mathbf{v}_2\in X_j\cap W}} s_{ij}(\mathbf{u}_2,\mathbf{v}_2)k_b(\|\mathbf{u}_2-\mathbf{v}_2\|-r) \\
&\quad = \mathrm{E}\sum^{\neq}_{\substack{\mathbf{u}_1,\mathbf{u}_2\in X_i\cap W\\ \mathbf{v}_1,\mathbf{v}_2\in X_j\cap W}} s_{ij}(\mathbf{u}_1,\mathbf{v}_1)s_{ij}(\mathbf{u}_2,\mathbf{v}_2)k_b(\|\mathbf{u}_1-\mathbf{v}_1\|-r)k_b(\|\mathbf{u}_2-\mathbf{v}_2\|-r) \\
&\qquad + \mathrm{E}\sum^{\neq}_{\substack{\mathbf{u}\in X_j\cap W\\ \mathbf{v}_1,\mathbf{v}_2\in X_j\cap W}} s_{ij}(\mathbf{u},\mathbf{v}_1)s_{ij}(\mathbf{u},\mathbf{v}_2)k_b(\|\mathbf{u}-\mathbf{v}_1\|-r)k_b(\|\mathbf{u}-\mathbf{v}_2\|-r) \\
&\qquad + \mathrm{E}\sum^{\neq}_{\substack{\mathbf{u}_1,\mathbf{u}_2\in X_i\cap W\\ \mathbf{v}\in X_j\cap W}} s_{ij}(\mathbf{u}_1,\mathbf{v})s_{ij}(\mathbf{u}_2,\mathbf{v})k_b(\|\mathbf{u}_1-\mathbf{v}\|-r)k_b(\|\mathbf{u}_2-\mathbf{v}\|-r) \\
&\qquad + \mathrm{E}\sum^{\neq}_{\substack{\mathbf{u}\in X_i\cap W\\ \mathbf{v}\in X_j\cap W}} s_{ij}(\mathbf{u},\mathbf{v})^2 k_b(\|\mathbf{u}-\mathbf{v}\|-r)^2,
\end{aligned}$$

where we recall that $\sum^{\neq}$ means summation over distinct points. If $i = j$ we can rename the indices and have that

$$\begin{aligned}
&\mathrm{E}[F_{ii}(r;b,\boldsymbol{\beta}^*)^2] \\
&\quad = \mathrm{E}\sum^{\neq}_{\mathbf{u}_1,\mathbf{u}_2\in X_i\cap W} s_{ii}(\mathbf{u}_1,\mathbf{u}_2)k_b(\|\mathbf{u}_1-\mathbf{u}_2\|-r)\sum^{\neq}_{\mathbf{u}_3,\mathbf{u}_4\in X_i\cap W} s_{ii}(\mathbf{u}_3,\mathbf{u}_4)k_b(\|\mathbf{u}_3-\mathbf{u}_4\|-r) \\
&\quad = \mathrm{E}\sum^{\neq}_{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3,\mathbf{u}_4\in X_i\cap W} s_{ii}(\mathbf{u}_1,\mathbf{u}_2)s_{ii}(\mathbf{u}_3,\mathbf{u}_4)k_b(\|\mathbf{u}_1-\mathbf{u}_2\|-r)k_b(\|\mathbf{u}_3-\mathbf{u}_4\|-r) \\
&\qquad + 4\mathrm{E}\sum^{\neq}_{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3\in X_i\cap W} s_{ii}(\mathbf{u}_1,\mathbf{u}_2)s_{ii}(\mathbf{u}_1,\mathbf{u}_3)k_b(\|\mathbf{u}_1-\mathbf{u}_2\|-r)k_b(\|\mathbf{u}_1-\mathbf{u}_3\|-r) \\
&\qquad + 2\mathrm{E}\sum^{\neq}_{\mathbf{u}_1,\mathbf{u}_2\in X_i\cap W} s_{ii}(\mathbf{u}_1,\mathbf{u}_2)^2 k_b(\|\mathbf{u}_1-\mathbf{u}_2\|-r)^2.
\end{aligned}$$

Lemma 4 then follows directly from applying Campbell's formula to each expectation. □

Lemma 5. *Under conditions C2, K1–K3, the variance of* $|W_n|^{-1}F_{ij,n}(r;b_n,\boldsymbol{\beta}^*_{ij})$ *converges to zero in probability as* $n\to\infty$.

Proof. Following Lemma 4 we have two cases for $\mathrm{Var}[F_{ij,n}(r; b_n, \boldsymbol{\beta}^*_{ij})]$ when $i \neq j$ and $i = j$. For $i \neq j$ we write $\mathrm{Var}[F_{ij,n}(r; b_n, \boldsymbol{\beta}^*_{ij})]$ as a sum of four terms $T_{1,n}, \dots, T_{4,n}$ and for $i = j$ we write $\mathrm{Var}[F_{ii,n}(r; b_n, \boldsymbol{\beta}^*_{ii})]$ as a sum of three terms $T'_{1,n}, \dots, T'_{3,n}$. First we consider $i \neq j$ and applying condition C2, translation invariance of $g^{(2,2)}_{ij}$ (condition K1) and a change of variable, it follows that $T_{1,n}$ is bounded as

$$T_{1,n} \leq p^4 K_1^4 \int_{W_n^4} \left| g^{(2,2)}_{ij}(\mathbf{0}, \mathbf{u}, \mathbf{v}, \mathbf{w} + \mathbf{u}) - g_{ij}(\mathbf{0}, \mathbf{v}) g_{ij}(\mathbf{0}, \mathbf{w}) \right| \\ k_{b_n}(\|\mathbf{v}\| - r) k_{b_n}(\|\mathbf{w}\| - r) \mathrm{d}\mathbf{u}_1 \mathrm{d}\mathbf{u} \mathrm{d}\mathbf{v} \mathrm{d}\mathbf{w}.$$

Using the second part of condition K1, the above upper bound can be simplied to

$$p^4 K_1^4 K_3 |W_n| \int_{W_n} k_{b_n}(\|\mathbf{v}\| - r) \mathrm{d}\mathbf{v} \int_{W_n} k_{b_n}(\|\mathbf{w}\| - r) \mathrm{d}\mathbf{w}.$$

Consequently, it follows from Lemma 3 and condition K3 that

$$\frac{T_{1,n}}{|W_n|^2} \leq \frac{p^4 K_1^4 K_3 (\tilde{C}_1 r^{d-1})^2}{|W_n|} \to 0, \qquad \text{as } n \to \infty.$$

Using similar arguments for $T'_{1,n}$ is can be shown that $T'_{1,n}$ tend to zero. Regarding $T_{2,n}$, first note that under condition C2, $\|\mathbf{z}(\mathbf{u})\|_{\max} \leq K_1$ ensures that there exists a $c > 0$ such that

$$p_i(\mathbf{u}; \boldsymbol{\beta}^*) \geq c, \qquad i = 1, \dots, p, \tag{D.1}$$

for any $\mathbf{u} \in \{\mathbf{u} \in \bigcup_{l=1}^{\infty} W_l : \lambda_0(\mathbf{u}) > 0\}$. Applying further condition K2, $T_{2,n}$ is bounded as

$$T_{2,n} \leq \frac{K_4 K_1^3}{c} \int_{W_n^3} k_{b_n}(\|\mathbf{u} - \mathbf{v}_1\| - r) k_{b_n}(\|\mathbf{u} - \mathbf{v}_2\| - r) \mathrm{d}\mathbf{u} \mathrm{d}\mathbf{v}_1 \mathrm{d}\mathbf{v}_2$$

By Lemma 3,

$$T_{2,n} \leq \frac{K_4 K_1^3}{c} |W_n| (\tilde{C}_1 r^{d-1})^2.$$

Hence $T_{2,n}/|W_n|^2 \to 0$ as $n \to \infty$. It follows along the same lines that $T_{3,n}$ and $T'_{2,n}$ tend to zero as $n \to \infty$.

Regarding the fourth term $T_{4,n}$, we have

$$\begin{aligned} T_{4,n} &\leq \frac{2 K_4 p^2 K_1^2}{|W_n|^2 c^2} \int_{W_n^2} k^2_{b_n}(\|\mathbf{u} - \mathbf{v}\| - r) \mathrm{d}\mathbf{u} \mathrm{d}\mathbf{v} \\ &\leq \frac{2 K_4 p^2 K_1^2}{|W_n| c^2} \int_{\mathbb{R}^d} k^2_{b_n}(\|\mathbf{u}\| - r) \mathrm{d}\mathbf{u} \end{aligned}$$

Applying Lemma 3 and the last part of condition K3, $T_{4,n}$ tends to zero as $n \to \infty$. Using similar arguments for $T'_{3,n}$ it can be shown that $T'_{3,n}$ tends to zero as $n \to \infty$. Thus $\mathrm{Var}\left[|W_n|^{-1} F_{ij,n}(r; b_n, \boldsymbol{\beta}^*)\right] \to 0$ as $n \to \infty$. □

References

Adrian Baddeley, Aruna Jammalamadaka, and Gopalan Nair. Multitype point process analysis of spines on the dendrite network of a neuron. *Journal of the Royal Statistical Society: Series C*, 63(5):673–694, 2014.

C. A. N. Biscio and R. P. Waagepetersen. A general central limit theorem and subsampling variance estimator for α-mixing point processes. *Scandinavian Journal of Statistics*, 2019. Appeared online, 1-23.

A. Choiruddin, F. Cuevas-Pacheco, J.-F. Coeurjolly, and R. Waagepetersen. Regularized estimation for highly multivariate log Gaussian Cox processes. Under revision, 2019.

D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34:187–220, 1972.

Martin Crowder. On consistency and inconsistency of estimating equations. *Econometric Theory*, 2(3):305–330, 1986.

P. J. Diggle and B. S. Rowlingson. A conditional approach to point process modelling of elevated risk. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 157:433–440, 1994.

Peter Diggle, Pingping Zheng, and Peter Durr. Nonparametric estimation of spatial segregation in a multivariate point process: bovine tuberculosis in Cornwall, UK. *Journal of the Royal Statistical Society: Series C*, 54(3):645–658, 2005. ISSN 1467-9876. doi: 10.1111/j.1467-9876.2005.05373.x. URL `http://dx.doi.org/10.1111/j.1467-9876.2005.05373.x`.

Y. Guan, R. P. Waagepetersen, and C. M. Beale. Second-order analysis of inhomogeneous spatial point processes with proportional intensity functions. *Journal of the American Statistical Association*, 103:769–777, 2008.

K. B. Hessellund, G. Xu, Y. Guan, and R. Waagepetersen. Supplementary material for 'Semi-parametric multinomial logistic regression for multivariate point pattern data'. Supplementary material, 2019.

Abdollah Jalilian, Yongtao Guan, Jorge Mateu, and Rasmus Waagepetersen. Multivariate product-shot-noise Cox point process models. *Biometrics*, 71:1022–1033, 2015.

Tuomas Rajala, D. J. Murrell, and S. C. Olhede. Detecting multivariate interactions in spatial point patterns with Gibbs models and variable selection. *Journal of the Royal Statistical Society: Series C*, 67:1237–1273, 2018.

Alex Reinhart and Joel Greenhouse. Self-exciting point processes with spatial covariates: modelling the dynamics of crime. *Journal of the Royal Statistical Society: Series C*, 67: 1305–1329, 2018. doi: 10.1111/rssc.12277. URL `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssc.12277`.

R. P. Waagepetersen and Y. Guan. Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 65:685–702, 2009.

R. P. Waagepetersen, Y. Guan, A. Jalilian, and J. Mateu. Analysis of multispecies point patterns by using multivariate log-Gaussian Cox processes. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 65:77–96, 2016.

Ganggang Xu, Rasmus Waagepetersen, and Yongtao Guan. Stochastic quasi-likelihood for case-control point pattern data. *Journal of the American Statistical Association*, 114: 631–644, 2019.