Electronic Journal of Statistics
Vol. 14 (2020) 3730–3765
ISSN: 1935-7524
https://doi.org/10.1214/20-EJS1755

Nonparametric estimation of the pair correlation function of replicated inhomogeneous point processes

Ganggang Xu[*†] **and Chong Zhao**

Department of Management Science, University of Miami, Coral Gables, FL 33146, USA
e-mail: gangxu@bus.miami.edu; czhao@bus.miami.edu

Abdollah Jalilian

Department of Statistics, Razi University, Bagh-e-Abrisham, Kermanshah, 67144, Iran
e-mail: jalilian@razi.ac.ir

Rasmus Waagepetersen[‡]

Department of Mathematical Sciences, Aalborg University, Fredrik Bajersvej 7G, DK 9220 Aalborg, Denmark
e-mail: rw@math.aau.dk

Jingfei Zhang[§] **and Yongtao Guan**[¶]

Department of Management Science, University of Miami, Coral Gables, FL 33146, USA
e-mail: ezhang@bus.miami.edu; yguan@bus.miami.edu

Abstract: We consider the nonparametric estimation of the isotropic pair correlation function (PCF) of inhomogeneous point processes when replicates are available. Based on carefully designed estimating equations, two types of nonparametric estimators, i.e., the local polynomial estimator and the orthogonal series estimator, are proposed and studied. The proposed estimators circumvent the problems caused by the need for estimating the unknown intensity function for kernel smoothed PCF estimators and they are free of edge correction terms. Asymptotic properties are investigated for both estimators and valid point-wise confidence bands are derived. Finite sample performances of the proposed estimators are demonstrated by simulation as well as an application to the Sina Weibo posting data.

MSC2020 subject classifications: Primary 60K35, 60K35; secondary 60K35.

[*]Corresponding author.
[†]Research supported by NSF grant SES-1902195.
[‡]Research 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.
[§]Research supported by NSF grant DMS-2015190.
[¶]Research supported by NSF grant DMS-1810591.

Keywords and phrases: Confidence band, estimating equations, local polynomial estimator, nonparametric estimation, orthogonal series estimator, replicated point patterns.

Received August 2019.

Contents

1. Introduction

The pair correlation function (PCF; Stoyan and Stoyan, 1994) is viewed as one of the most informative second-order summary statistics of temporal or spatial point patterns (Stoyan and Stoyan, 1994; Baddeley et al., 2000; Møller and Waagepetersen, 2003). For example, the PCF takes different values when the point patterns are completely random (i.e., Poisson processes), clustered or inhibitive, typically at small lags. As such, it plays an important role in exploratory

statistical analysis, and in suggesting suitable models for the data. Moreover, the popular second-order characteristics called Ripley's K-function (Ripley, 1976) and Besag's L-function Besag (1977) are in one-to-one correspondence with the PCF.

There is extensive literature on estimating the PCF nonparametrically based on a single observed point pattern, ranging from the kernel estimators (Stoyan and Stoyan, 1994; Baddeley et al., 2000; Illian et al., 2008), the Bayesian estimator (Yue and Loh, 2010) and more recently the orthogonal series estimator (Jalilian et al., 2019). However, all of these methods require that the unknown intensity function $\lambda(\cdot)$ is replaced by an estimate $\widehat{\lambda}(\cdot)$. This can cause major uncertainty in the resulting PCF estimator. For example, if $\lambda(\cdot)$ and hence $\widehat{\lambda}(\cdot)$ are close to 0, all existing methods may result in a poor PCF estimator because they include $\widehat{\lambda}(\cdot)$ in the denominator. In certain applications such as the social media posting data investigated in Section 5, $\lambda(\cdot)$ can be close to zero if one has little posting activities consistently during certain time of the day. Furthermore, in the absence of a suitable parametric model, $\lambda(\cdot)$ is typically estimated using kernel smoothing. This implicitly imposes smoothness assumptions on $\lambda(\cdot)$ which may be problematic in some applications. For example, Xu et al. (2019) studies locations of different types of restaurants whose intensity functions may abruptly change across different areas of a city and hence cannot be viewed as being smooth. Another issue is the evaluation of the uncertainty of the PCF estimators. Although central limit theorems regarding PCF estimators are available (e.g. Heinrich, 1988; Heinrich and Klein, 2014; Jalilian et al., 2019), the practical use of these results seems unexplored, perhaps due to difficulty of estimating the asymptotic variances.

In this paper, we address the aforementioned issues in the context of replicated point pattern data. While the literature on analyzing single point patterns is extensive, less has been done for replicated point patterns. The current work is motivated by a temporal point process dataset obtained from Sina Weibo, the largest social media platform in China. The data consist of posting time stamps collected from independent user accounts whereas the posting time stamps from each user can be viewed as a single point pattern. For this dataset, informative exploratory data analysis is crucial in understanding the mechanism in which users interact with social media websites. For example, what patterns (e.g., random, clustered, inhibitive) do the user events follow? Are there any differences in these patterns and what might have contributed to these differences? Some work has considered the analysis of replicated point pattern data. For example, Diggle et al. (1991) and Baddeley et al. (1993) introduced nonparametric approaches to estimate pooled summary statistics. Using replicated temporal Cox processes, Bouzas et al. (2006) proposed a functional approach to estimate the intensity function and Wu et al. (2013) used kernel smoothing to perform functional data analysis. More recently, Gervini (2016, 2017) developed an independent component model and a multiplicative component model which can be used for both replicated temporal and spatial point processes. Xu et al. (2020) studies multi-level principal component analysis of repeatedly observed

temporal point processes.

Our contribution to the analysis of replicated point processes is to introduce nonparametric estimators of the PCF that do not require estimation of the intensity function. Moreover, we provide a rigorous asymptotic analysis of the proposed estimators including the demonstration of asymptotic normality and construct computationally feasible estimators of the asymptotic variances. Confidence bands can, therefore, be constructed based on the asymptotic normality of the proposed estimators. Finally, our theoretical findings apply to both replicated temporal and spatial point processes.

More specifically, we propose two nonparametric methods to estimate the PCF: the local polynomial estimator and the orthogonal series estimator, both of which are obtained by solving a system of estimating equations. While the former is computationally fast and works well for the PCF at medium to large spatial/temporal lags, it suffers from non-negligible bias when the lag is close to zero, which is a common problem for many kernel estimators (Stoyan and Stoyan, 1994; Møller and Waagepetersen, 2003). The orthogonal series estimator, on the other hand, is computationally more expensive but has a smaller bias when the lag is close to zero, as is demonstrated by Jalilian et al. (2019). Thanks to the careful design of the proposed estimating equations, neither estimator requires any knowledge of the intensity $\lambda(\cdot)$, which is a major advantage over existing methods. For statistical inference, we derive limiting distributions of the proposed estimators under a unified framework that covers two asymptotic scenarios: (1) the number of replicates grows to infinity while the observation window is fixed; (2) the number of replicates is fixed but the observation window expands. While the first scenario is more suitable for temporal processes where replicates are easily available such as in our Sina Weibo posting data, the second scenario may be more preferable for spatial point processes where the number of replicates is limited. In both scenarios, we propose empirical estimators for the asymptotic variances of the estimated PCFs and consequently construct valid confidence bands. Another advantage of the proposed estimators is that they do not require any edge correction terms due to the design of the estimating equations.

2. Methodology

Let X_i, $i = 1, \ldots, m$, be m independent and identically distributed point processes defined on $\mathbb{R}^d$, $d \geq 1$. Let $D \subset \mathbb{R}^d$ be a bounded observation window over which the X_i's are observed. For $B \subseteq \mathbb{R}^d$, let $N_i(B)$ denote the random number of points in $X_i \cap B$. The marginal first- and second-order intensity functions, $\lambda(\mathbf{x})$ and $\lambda^{(2)}(\mathbf{x}, \mathbf{y})$, of the X_i's are defined by

$$\mathbb{E}[N_i(B)] = \int_B \lambda(\mathbf{x}) \mathrm{d}\mathbf{x},$$
$$\mathbb{E}[N_i(A)N_i(B)] = \int_{A \cap B} \lambda(\mathbf{x}) \mathrm{d}\mathbf{x} + \int_A \int_B \lambda^{(2)}(\mathbf{x}, \mathbf{y}) \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y},$$

for bounded $A, B \subseteq \mathbb{R}^d$. The first-order intensity function is referred to as the intensity function. The PCF function is defined as $g(\mathbf{x},\mathbf{y}) = \lambda^{(2)}(\mathbf{x},\mathbf{y})/[\lambda(\mathbf{x})\lambda(\mathbf{y})]$ if $\lambda(\mathbf{x})\lambda(\mathbf{y}) > 0$ and $g(\mathbf{x},\mathbf{y}) = 0$ otherwise. We assume that $g(\cdot,\cdot)$ is isotropic, i.e., $g(\mathbf{x},\mathbf{y}) = g_0(\|\mathbf{x}-\mathbf{y}\|)$ for some function $g_0(\cdot)$ with $\|\mathbf{x}-\mathbf{y}\|$ being the distance between $\mathbf{x},\mathbf{y} \in \mathbb{R}^d$.

For $\mathbf{x} \in \mathbb{R}^d$, $\lambda(\mathbf{x})\mathrm{d}\mathbf{x}$ can be interpreted as the probability of observing a point in a small neighborhood of volume $\mathrm{d}\mathbf{x}$ around $\mathbf{x}$. For $\mathbf{x},\mathbf{y} \in \mathbb{R}^d$, $g(\mathbf{x},\mathbf{y})$ is the ratio $\lambda(\mathbf{y}|\mathbf{x})/\lambda(\mathbf{y})$ where $\lambda(\mathbf{y}|\mathbf{x})$ is the intensity at $\mathbf{y}$ given that a point is already present at $\mathbf{x}$ (Coeurjolly et al., 2017). In other words, the PCF measures the increase/decrease in the intensity at $\mathbf{y}$ caused by the presence of a point at $\mathbf{x}$.

2.1. Estimating functions for the PCF

Our nonparametric estimators of $g(\cdot,\cdot)$ are derived from weighted composite likelihood estimating functions (Guan, 2006). Consider a parametric model where the PCF is governed by a vector of unknown parameters $\boldsymbol{\theta}$, denoted as $g(\cdot,\cdot;\boldsymbol{\theta})$. Under the isotropic assumption, with a slight abuse of notation, we denote $g(\cdot,\cdot;\boldsymbol{\theta})$ by $g(r;\boldsymbol{\theta})$ with r being the lag distance.

Following Guan (2006) and Waagepetersen (2007), an estimator of $\boldsymbol{\theta}$ can be obtained by solving the following estimating equation $\mathbf{U}(\boldsymbol{\theta}) = \mathbf{0}$ with

$$\begin{aligned}\mathbf{U}(\boldsymbol{\theta}) = &\sum_{i=1}^{m} \sum_{\mathbf{u},\mathbf{v}\in X_i}^{\neq} w(\mathbf{u},\mathbf{v}) \frac{g^{(1)}(\|\mathbf{u}-\mathbf{v}\|;\boldsymbol{\theta})}{g(\|\mathbf{u}-\mathbf{v}\|;\boldsymbol{\theta})} \\ &- m \int_D \int_D w(\mathbf{x},\mathbf{y})\lambda(\mathbf{x})\lambda(\mathbf{y}) g^{(1)}(\|\mathbf{x}-\mathbf{y}\|;\boldsymbol{\theta})\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y},\end{aligned} \tag{2.1}$$

where $g^{(1)}(r;\boldsymbol{\theta}) = \partial g(r;\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, $\sum^{\neq}$ denotes summation over all distinct points and $w(\cdot,\cdot)$ is some predefined weight function. Consider for a moment the case $w(\mathbf{u},\mathbf{v}) = 1$. Then (2.1) is formally the sum of m Poisson likelihood score functions for m Poisson processes on $\mathbb{R}^d \times \mathbb{R}^d$ with common intensity function $\lambda(\mathbf{x})\lambda(\mathbf{y})g(\|\mathbf{x}-\mathbf{y}\|;\boldsymbol{\theta})$, $\mathbf{x},\mathbf{y} \in \mathbb{R}^d \times \mathbb{R}^d$. Although the points $(\mathbf{u},\mathbf{v}) \in X_i \times X_i$, $\mathbf{u} \neq \mathbf{v}$ do not form Poisson processes for $i = 1,\ldots,m$, the estimation approach can be justified according to composite likelihood considerations. In particular, regardless of the choice of $w(\mathbf{u},\mathbf{v})$, $\mathbf{U}(\boldsymbol{\theta})$ is an unbiased estimating function in the sense that $\mathbb{E}[\mathbf{U}(\boldsymbol{\theta}_0)] = \mathbf{0}$ where $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$. However, without knowing $\lambda(\cdot)$, the double integrals in (2.1) cannot be computed and thus must be estimated.

Note that for any real function $f(\cdot,\cdot)$, one has that

$$\mathbb{E}\Big[\sum_{\mathbf{u}\in X_i}\sum_{\mathbf{v}\in X_j} f(\mathbf{u},\mathbf{v})\Big] = \int_D \int_D f(\mathbf{x},\mathbf{y})\lambda(\mathbf{x})\lambda(\mathbf{y})\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}, \quad \text{if } i \neq j.$$

We therefore modify $U(\boldsymbol{\theta})$ as

$$\begin{aligned}\tilde{\mathbf{U}}(\boldsymbol{\theta}) = &\sum_{i=1}^{m} \sum_{\mathbf{u},\mathbf{v}\in X_i}^{\neq} \sum w(\mathbf{u},\mathbf{v}) \frac{g^{(1)}(\|\mathbf{u}-\mathbf{v}\|;\boldsymbol{\theta})}{g(\|\mathbf{u}-\mathbf{v}\|;\boldsymbol{\theta})} \\ &- \frac{1}{m-1} \sum_{i\neq j} \sum \sum_{\mathbf{u}\in X_i, \mathbf{v}\in X_j} \sum w(\mathbf{u},\mathbf{v}) g^{(1)}(\|\mathbf{u}-\mathbf{v}\|;\boldsymbol{\theta}).\end{aligned} \tag{2.2}$$

It can be easily seen that $\mathbb{E}[\tilde{\mathbf{U}}(\boldsymbol{\theta}_0)] = \mathbf{0}$ and hence $\tilde{\mathbf{U}}(\boldsymbol{\theta})$ is also an unbiased estimating function as long as $m \geq 2$. Note that neither $\lambda(\cdot)$ nor any edge correction term is needed for (2.2). In the following sections we adapt (2.2) to obtain local polynomial and orthogonal series estimators of the PCF.

2.2. Local polynomial estimator

Consider a local polynomial approximation of $g(t)$ for any lag $t \in [r-h, r+h]$ as follows

$$\tilde{g}_{r,h}(t;\boldsymbol{\theta}) = \exp\left[\theta_0 + \theta_1(t-r) + \cdots + \theta_p(t-r)^p\right], \tag{2.3}$$

where $\theta_0 = \log[g(r)]$, $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_p)^T$ and $p \geq 0$. To estimate $\boldsymbol{\theta}$, we can define a localized version of the estimating function (2.2) as follows

$$\begin{aligned}\tilde{\mathbf{U}}_{r,h}(\boldsymbol{\theta}) = &\sum_{i=1}^{m} \sum_{\mathbf{u},\mathbf{v}\in X_i}^{\neq} \sum w_{r,h}(\|\mathbf{u}-\mathbf{v}\|)\mathbf{G}_r(\|\mathbf{u}-\mathbf{v}\|) \\ &- \frac{1}{m-1} \sum_{i\neq j} \sum \sum_{\mathbf{u}\in X_i, \mathbf{v}\in X_j} \sum w_{r,h}(\|\mathbf{u}-\mathbf{v}\|)\mathbf{G}_r(\|\mathbf{u}-\mathbf{v}\|)\tilde{g}_{r,h}(\|\mathbf{u}-\mathbf{v}\|;\boldsymbol{\theta}),\end{aligned} \tag{2.4}$$

where $w_{r,h}(\|\mathbf{u}-\mathbf{v}\|) = K_h(\|\mathbf{u}-\mathbf{v}\| - r)/|D|$ with $|D|$ denoting the Lebesgue measure of $D \subset \mathbb{R}^d$ and $K_h(x) = K(x/h)/h$ for some kernel function $K(\cdot)$ and a bandwidth $h > 0$, and $\mathbf{G}_r(t) = \partial \log[\tilde{g}_{r,h}(t;\boldsymbol{\theta})]/\partial\boldsymbol{\theta} = [1, t-r, \ldots, (t-r)^p]^T$. Equation (2.4) can be viewed as a version of (2.2) with a weight function $w_{r,h}(\|\mathbf{u}-\mathbf{v}\|)$ and $g(t;\boldsymbol{\theta}) = \tilde{g}_{r,h}(t;\boldsymbol{\theta})$ for $t \in [r-h, r+h]$. Letting $\widehat{\boldsymbol{\theta}} = (\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_p)^T$ be the solution to $\tilde{\mathbf{U}}_{r,h}(\boldsymbol{\theta}) = \mathbf{0}$, we define the local polynomial estimator of $g(r)$ as

$$\hat{g}_h(r) = \exp(\hat{\theta}_0), \tag{2.5}$$

which is nonnegative as desired.

In contrast to the parametric case (2.2), because $\tilde{g}_{r,h}(t;\boldsymbol{\theta})$ is only a local approximation of $g(t)$ for $t \in [r-h, r+h]$, the estimating function $\tilde{\mathbf{U}}_{r,h}(\boldsymbol{\theta})$ is no longer unbiased, resulting in a biased estimator $\hat{g}_h(r)$ for $g(r) = \exp(\theta_0)$. Using standard theory of estimating equations, one can show that $\mathbb{E}[\hat{g}_h(r)] = \exp(\theta_0^*)$, where $\boldsymbol{\theta}^* = (\theta_0^*, \theta_1^*, \ldots, \theta_p^*)^T$ solves the equation

$$\begin{aligned}\int_D \int_D &w_{r,h}(\|\mathbf{x}-\mathbf{y}\|)\lambda(\mathbf{x})\lambda(\mathbf{y}) \\ &\times \left[g(\|\mathbf{x}-\mathbf{y}\|) - \tilde{g}_{r,h}(\|\mathbf{x}-\mathbf{y}\|;\boldsymbol{\theta})\right] \mathbf{G}_r(\|\mathbf{x}-\mathbf{y}\|)\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} = \mathbf{0}.\end{aligned} \tag{2.6}$$

As we will show in the proof of Lemma A.1, for a given kernel $K(\cdot)$, the estimation bias of $\hat{g}_h(r)$ is determined by the difference between θ_0 and θ_0^*, which is of the order $O(h^{p+1})$.

The local constant estimator is the solution to the equation $\tilde{\mathbf{U}}_{r,h}(\boldsymbol{\theta}) = \mathbf{0}$ when $p = 0$, and can be directly calculated as

$$\widehat{g}_h(r) = \frac{(m-1)\sum_{i=1}^{m}\sum_{\mathbf{u},\mathbf{v}\in X_i}^{\neq} K_h(\|\mathbf{u}-\mathbf{v}\|-r)}{\sum_{i\neq j}\sum_{j}\sum_{\mathbf{u}\in X_i,\mathbf{v}\in X_j}\sum K_h(\|\mathbf{u}-\mathbf{v}\|-r)}, \tag{2.7}$$

which is essentially the classical Nadaraya-Watson estimator extended to the current setting. A data driven method for choosing the bandwidth h will be given in Appendix A.

It is well known in local polynomial regression literature, see, e.g., Fan and Gijbels (1996), that a higher degree p will result in smaller bias but larger variance. In the regression setting, the local linear estimator ($p = 1$) typically outperforms the local constant estimator ($p = 0$) in the sense that it achieves much smaller bias on the boundary without significantly increasing the variance (Fan and Gijbels, 1996). However, it is not a clear cut decision in our setting. See a more detailed discussion on this issue in Section 3.1.1.

2.3. Orthogonal series estimator

It is known that the kernel estimator of PCF suffers from serious biases when the lags are close to zero (Stoyan and Stoyan, 1994; Møller and Waagepetersen, 2003). As a remedy, Jalilian et al. (2019) utilized the orthogonal series density estimators (Hall, 1987; Efromovich, 2010) to estimate the PCF with a single point pattern, which was shown to be less biased for various clustered point processes. In this section, we adopt a different estimating equation approach that can remove these restrictions.

For some predefined distance $R > 0$, the orthogonal series expansion of the square-integrable function $\log[g(r)]$ on $[0, R]$ is given by

$$\log[g(r)] = \sum_{l=1}^{\infty}\theta_{0,l}\phi_l(r), \quad \text{where} \quad \theta_{0,l} = \int_0^R \log[g(r)]\,\phi_l(r)w_o(r)\mathrm{d}r,$$

where the $\phi_l(r)$'s are a set of basis functions defined on $[0, R]$ that are orthogonal with respect to some weight function $w_o(r) \geq 0$, that is, $\int_0^R \phi_l(r)\phi_k(r)w_o(r)\mathrm{d}r = 1$ if $k = l$ and 0 otherwise. One example of such basis functions is the cosine basis functions, i.e., $\phi_1(r) = \frac{1}{\sqrt{R}}$, $\phi_k(r) = \frac{\sqrt{2}}{\sqrt{R}}\cos[(k-1)\pi r/R]$ for $k \geq 2$ and $w_o(r) = 1$ for $r \geq 0$. Although expanding $\log[g(\cdot)]$ may be problematic if $g(r)$ is zero or close to zero for some r, for clustered point patterns where $g(\cdot) \geq 1$, expanding $\log[g(\cdot)]$ seems a reasonable choice.

By the Parseval's identity (Tolstov, 1962, p. 119), we have that

$$\sum_{l=1}^{\infty} \theta_{0,l}^2 = \int_0^R w_o(r) \left\{\log\left[g(r)\right]\right\}^2 \mathrm{d}r,$$

and hence $\sum_{k=l}^{\infty} \theta_{0,k}^2 \to 0$ and $\theta_{0,l} \to 0$, as $l \to \infty$, provided that the right-hand side of the equation is integrable. Therefore, we can consider the truncated orthogonal series representation of $g(r)$ as follows

$$\tilde{g}_L(r;\boldsymbol{\theta}) = \exp\left[\sum_{l=1}^{L} \theta_l \phi_l(r)\right], \tag{2.8}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_L)^T$ and L is a predefined sufficiently large positive integer. Following similar ideas that result in (2.2), the parameter vector $\boldsymbol{\theta}$ can be estimated by solving the estimating equation $\tilde{\mathbf{U}}_L(\boldsymbol{\theta}) = \mathbf{0}$ with

$$\begin{aligned}\tilde{\mathbf{U}}_L(\boldsymbol{\theta}) = &\sum_{i=1}^{m} \sum_{\mathbf{u},\mathbf{v}\in X_i}^{\neq} w_R(\|\mathbf{u}-\mathbf{v}\|)\boldsymbol{\phi}_L(\|\mathbf{u}-\mathbf{v}\|) \\ &- \frac{1}{m-1} \sum_{i\neq j} \sum \sum_{\mathbf{u}\in X_i, \mathbf{v}\in X_j} w_R(\|\mathbf{u}-\mathbf{v}\|)\boldsymbol{\phi}_L(\|\mathbf{u}-\mathbf{v}\|)\tilde{g}_L(\|\mathbf{u}-\mathbf{v}\|;\boldsymbol{\theta}),\end{aligned} \tag{2.9}$$

where the vector-valued function $\boldsymbol{\phi}_L(r) = [\phi_1(r), \ldots, \phi_L(r)]^T$ and the weight function is defined as $w_R(r) = w_0(r)I(r < R)/|D|$ with $I(\cdot)$ being the indicator function. Denoting the solution as $\widehat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_L)^T$, the final orthogonal series estimator of $g(r)$ becomes

$$\hat{g}_L(r) = \exp\left[\widehat{\boldsymbol{\theta}}^T \boldsymbol{\phi}_L(r)\right], \quad \text{for } 0 < r \leq R. \tag{2.10}$$

A data driven method for choosing L will be given in Appendix A.

Remark 2.1. *Jalilian et al. (2019) expanded* $g(\cdot)$ *as* $g(r) = \sum_{l=1}^{\infty} \theta_l \phi_l(r - r_{\min})$, *where* $r_{\min}$ *is a pre-chosen minimum distance. The coefficients* θ_l*'s are then replaced by some moment estimators, which critically depend on the intensity function* $\lambda(\cdot)$. *The final estimator* $\hat{g}(\cdot)$ *is not guaranteed to be nonnegative and a poor estimator of* $\lambda(\cdot)$ *can invalidate the theoretical foundation of Jalilian et al. (2019). In contrast, in our work, we expand* $\log[g(\cdot)]$ *and the expansion coefficient vector* $\boldsymbol{\theta}$ *is estimated using our new estimating functions (2.9), which completely avoid the estimation of* $\lambda(\cdot)$. *Therefore, the newly proposed estimator* $\widehat{\boldsymbol{\theta}}$ *is fundamentally different from the moment estimator used in Jalilian et al. (2019).*

3. Asymptotic properties

In this section, we investigate the asymptotic properties of the proposed local polynomial estimator (2.5) and the orthogonal series estimator (2.10). We assume that there are m independent and identically distributed point processes

being observed over a sequence of observation windows D_n. Define the k'th order joint intensity $\lambda^{(k)}(\cdot)$ by the identity

$$\mathbb{E}\left[\sum_{\mathbf{u}_1\ldots,\mathbf{u}_k\in X}^{\neq} I(\mathbf{u}_1\in A_1,\ldots,\mathbf{u}_k\in A_k)\right]=\int_{A_1\times\cdots\times A_k}\lambda^{(k)}(\mathbf{x}_1,\ldots,\mathbf{x}_k)\mathrm{d}\mathbf{x}_1\ldots\mathrm{d}\mathbf{x}_k$$

for bounded subsets $A_i\subset\mathbb{R}^d$, $i=1,\ldots,k$, where the sum is over distinct $\mathbf{u}_1\ldots,\mathbf{u}_k$. Based on the joint intensities, we can subsequently define higher order normalized joint intensities $g^{(k)}(\mathbf{x}_1,\ldots,\mathbf{x}_k)=\lambda^{(k)}(\mathbf{x}_1,\ldots,\mathbf{x}_k)/[\lambda(\mathbf{x}_1)\ldots\lambda(\mathbf{x}_k)]$, which we further assume to be translation invariant, i.e. $g^{(k)}(\mathbf{x}_1,\ldots,\mathbf{x}_k)=g_0^{(k)}(\mathbf{x}_2-\mathbf{x}_1,\ldots,\mathbf{x}_k-\mathbf{x}_1)$ for some $g_0^{(k)}(\cdot)$ for $k=3,4$.

The asymptotic framework assumes that either m, i.e. the number of independent replicates, or $|D_n|$, i.e. the sequence of the Lebesgue measures of observation windows, or both grow to ∞.

3.1. The local polynomial estimator

To study the asymptotic properties of the local polynomial estimator, we first introduce some new notation. Denote $f(r)=\log[g(r)]$ and assume that its jth derivative $f^{\{j\}}(r)$ exists for $j=1,\ldots,p+1$. Define $(p+1)\times(p+1)$ matrix valued functions for $k=1,2$ as

$$\mathbf{Q}_{n,h}^{(k)}(r)=\frac{1}{|D_n|h}\int_{D_n}\int_{D_n}\lambda(\mathbf{x})\lambda(\mathbf{y})\left[K\left(\frac{\|\mathbf{x}-\mathbf{y}\|-r}{h}\right)\right]^k \times\mathbf{A}_h(\|\mathbf{x}-\mathbf{y}\|-r)\mathbf{A}_h^T(\|\mathbf{x}-\mathbf{y}\|-r)\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}, \tag{3.1}$$

where $\mathbf{A}_h(t)=[1,t/h,\ldots,t^p/h^p]^T$.

The following conditions are sufficient for the asymptotic consistency of $\hat{g}_h(r)$.

[C1] There exists a C_λ such that the intensify function $0\le\lambda(\mathbf{u})\le C_\lambda$ for any $\mathbf{u}\in D_n$.

[C2] There exist positive constants c_g, C_g and C_f such that (a) $c_g\le g(r)\le C_g$; (b) $\max_{1\le j\le p+1}|f^{\{j\}}(r)|\le C_f$ for any $r\ge 0$ and that (c) $\int_0^\infty|g(s)-1|\mathrm{d}s<C_g$.

[C3] It holds that (a) $\left|g^{(k)}(\mathbf{x}_1,\ldots,\mathbf{x}_k)\right|\le C_g$ for any $\mathbf{x}_j\in D_n$, $j=1,\ldots,k$ and $k=3,4,5,6$; (b) $\int_{D_n}|g_0^{(3)}(\mathbf{x},\mathbf{y})-g(\|\mathbf{x}-\mathbf{y}\|)|\mathrm{d}\mathbf{x}\le C_g$; and (c) $\int_{D_n}|g_0^{(4)}(\mathbf{x},\mathbf{y}+\mathbf{w},\mathbf{w})-g(\|\mathbf{x}\|)g(\|\mathbf{y}\|)|\mathrm{d}\mathbf{w}\le C_g$.

[C4] The kernel $K(x)$ has a support $[-1,1]$ such that $\int_{-1}^1K(x)\mathrm{d}x=1$.

[C5] As the bandwidth $h\to 0$ and $m|D_n|h(r+h)^{d-1}\to\infty$, there exists a constant $c_0>0$ such that

$$\eta_{\min}\left[\mathbf{Q}_{n,h}^{(k)}(r)\right](r+h)^{1-d}>c_0,\quad k=1,2,$$

where $\eta_{\min}(\mathbf{Q})$ denotes the smallest eigenvalue of the matrix $\mathbf{Q}$.

Conditions C1-C3 are mild conditions commonly used in point process literature, see, e.g., Dvořák and Prokešová (2016); Coeurjolly et al. (2017). Condition C4 assumes that the kernel function should have a bounded support such as the popular uniform and Epanechnikov kernel functions. The factor $(r+h)^{1-d}$ in the condition C5 only matters when $r \to 0$ and $d > 1$, which will be discussed in more detail in Section 3.1.2. Condition C5 is most apparent when we consider stationary point processes in $\mathbb{R}$ ($d = 1$) with $\lambda(\mathbf{x}) \equiv \lambda$ and using the uniform kernel $K(x) = \frac{1}{2} I(-1 \le x \le 1)$. In this case, we have the following Lemma.

Lemma 3.1. *If the observation window $D_n = [0, T_n] \subset \mathbb{R}$, $\lambda(\mathbf{x}) \equiv \lambda$ and the uniform kernel $K(\cdot)$ is used, we have that*

$$\mathbf{Q}^{(k)}_{n,h}(r) = \frac{\lambda^2}{2^{k-1}} \left(1 - \frac{r}{T_n}\right) \mathbf{B}_1(r) - \frac{\lambda^2}{2^{k-1}} \frac{h}{T_n} \mathbf{B}_2(r), \quad k = 1, 2,$$

where $\mathbf{B}_1(r)$, $\mathbf{B}_2(r)$ are $(p+1) \times (p+1)$ matrices whose (i,j)th elements are $\frac{1}{i+j-1}(q_{up}^{i+j-1} - q_{low}^{i+j-1})$ and $\frac{1}{i+j}(q_{up}^{i+j} - q_{low}^{i+j})$, $i, j = 1, \ldots, p+1$, respectively, with $q_{low} = \max(-r/h, -1)$ and $q_{up} = \min[(T_n - r)/h, 1]$.

The proof is given in the Supplementary Material (Xu et al., 2020).

Under the setting of Lemma 3.1, $\mathbf{Q}^{(k)}_{n,h}(r)$, $k = 1, 2$, for the local constant estimator (i.e. $p = 0$) are identical scalars taking the following simple form for $h \le T_n/2$

$$\mathbf{Q}^{(k)}_{n,h}(r) = \begin{cases} \frac{\lambda^2\left(1-\frac{r}{T_n}\right)\left(1+\frac{r}{h}\right)}{2^{k-1}} - \frac{\lambda^2 h\left(1-\frac{r^2}{h^2}\right)}{2^k T_n}, & \text{if } r \in [0, h], \\ \frac{\lambda^2\left(1-\frac{r}{T_n}\right)}{2^{k-1}}, & \text{if } r \in (h, T_n - h), \\ \frac{\lambda^2\left(1-\frac{r}{T_n}\right)\left(1+\frac{T_n - r}{h}\right)}{2^{k-1}} + \frac{\lambda^2 h\left[1-\frac{(T_n-r)^2}{h^2}\right]}{2^k T_n}, & \text{if } r \in [T_n - h, T_n]. \end{cases}$$

In this homogeneous case, condition C5 essentially requires that the intensity λ is bounded away from 0 and the lag r is relatively small compared to the window size T_n. It is anticipated that, for the inhomogeneous case, if the intensity function $\lambda(\cdot)$ is bounded away from 0 in a sufficiently large area within the observation window, condition C5 should hold for a reasonably small lag r.

Remark 3.1. *The uniform kernel is only used in Lemma 3.1 so that $\mathbf{Q}^{(k)}_{n,h}(r)$'s have closed-form expressions that can be used for a theoretical comparison of the variances of the local polynomial estimators in Section 3.1.1. However, for all numerical examples in this paper, we use the Epanechnikov kernel to achieve smoother PCF estimators.*

Lemma 3.2. *Under conditions C1-C5, as $h \to 0$ and $m|D_n|h(r+h)^{d-1} \to \infty$, we have that*

$$|\hat{g}_h(r) - g(r)| = O_p\left[\frac{1}{\sqrt{m|D_n|h(r+h)^{d-1}}} + h^{p+1}\right]. \tag{3.2}$$

The proof is given in Appendix C and the Supplementary Material.

Lemma 3.2 ensures that the local polynomial estimator for $g(r)$ is consistent. The convergence rate can be decomposed into two parts, with the $O_p[(m|D_n|h(r+h)^{d-1})^{-1/2}]$ due to the estimation variance and $O(h^{p+1})$ resulting from the estimation bias. As the order p of the local polynomials increases, the order of the estimation bias decreases, which is consistent with the local polynomial regression literature. It is also worth pointing out that when $r \to 0$ and $d > 1$, $\hat{g}_h(r)$ has a slower convergence rate as the dimension d increases. See a more detailed discussion in Section 3.1.2.

To make the statistical inference, we next proceed to establish the asymptotic distribution of $\hat{g}_h(r) - g(r)$. To do so, we first introduce the definition of α-mixing coefficient for point processes following Biscio and Waagepetersen (2019). Let $\alpha(\mathcal{F}, \mathcal{G})$ denote the α-mixing coefficient of two σ-algebras $\mathcal{F}$ and $\mathcal{G}$ defined as

$$\alpha(\mathcal{F}, \mathcal{G}) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}, B \in \mathcal{G}\}.$$

Then the α-mixing coefficient of a point process X is defined as

$$\alpha_X(s; a_1, a_2) = \sup\big\{\alpha[\sigma(X \cap E_1), \sigma(X \cap E_2)] : E_k \subset \mathbb{R}^d, |E_k| \le a_k, \\ k = 1, 2, \mathrm{d}(E_1, E_2) \ge s\big\}, \tag{3.3}$$

where $\sigma(X \cap E_i)$ is the σ-algebra generated by $X \cap E_i$, $i = 1, 2$, and $\mathrm{d}(E_1, E_2) = \inf\{\max_{1 \le i \le d} |u_i - v_i| : \mathbf{u} = (u_1, \ldots, u_d)^T \in E_1, \mathbf{v} = (v_1, \ldots, v_d)^T \in E_2\}$. We need to make the following two additional assumptions:

[N1] Either one of the following conditions are true (a) $m \to \infty$; or (b) the mixing coefficient satisfies $\alpha_X(s; h^{-1}, \infty) = O(s^{-d-\varepsilon})$ for some $\varepsilon > 0$.

[N2] There exists $\delta > 2d/\varepsilon$ such that $\left|g^{(k)}(\mathbf{x}_1, \ldots, \mathbf{x}_k)\right| \le C_g$ for any $\mathbf{x}_j \in D_n$, $j = 1, \ldots, k$, $k = 2, \ldots, 2(2 + \lceil\delta\rceil)$, where $\lceil\delta\rceil$ is the smallest integer greater than δ.

Conditions N1 and N2 are commonly used in the literature of point processes, see, e.g., Coeurjolly and Møller (2014). In the case when the number of replicates $m \to \infty$, conditions N1(b) and N2 are not needed to show asymptotic normality of $\hat{g}_h(r)$. However, when m is finite, conditions N1(b) and N2 need to be imposed to control the strength of dependence among the event points of each point pattern within the observation window D_n. Condition N2 is slightly more restrictive than C3(a) in then sense that a higher order of $g^{(k)}(\cdot)$ needs to be bounded since $2(2 + \lceil\delta\rceil) \ge 6$.

Theorem 3.1. *Under C1-C5, N1-N2, as $h \to 0$ and $m|D_n|h(r+h)^{d-1} \to \infty$, we have that*

$$\frac{\sqrt{m|D_n|h}\,[\hat{g}_h(r) - g(r) + b_{n,h}]}{\sigma_{m,n,h}(r)} \xrightarrow{\mathcal{D}} N(0, 1),$$

where $\sigma^2_{m,n,h}(r) = \frac{2g(r)[m-1+g(r)]}{m-1} \boldsymbol{e}^T \left[\mathbf{Q}^{(1)}_{n,h}(r)\right]^{-1} \mathbf{Q}^{(2)}_{n,h}(r) \left[\mathbf{Q}^{(1)}_{n,h}(r)\right]^{-1} \boldsymbol{e}$, with $\boldsymbol{e} = (1, 0, \ldots, 0)^T_{p+1}$ and $\mathbf{Q}^{(k)}_{n,h}(r)$, $k = 1, 2$, are defined in (3.1), and $b_{n,h} = O(h^{p+1})$.

The proof is given in Appendix C and the Supplementary Material.

Theorem 3.1 echoes with Lemma 3.2 in that special attention needs to be paid to $\hat{g}_h(r)$ for $r \approx 0$ when $d > 1$. See a more detailed discussion in Section 3.1.2 on this issue.

3.1.1. Local constant or local linear estimator?

One remaining important question is which order of polynomial one should use in practice. To reduce the asymptotic bias, increasing p is probably a good idea. However, when p increases, the asymptotic variance may also increase. Therefore, to answer that question, we need to consider the actual inflation in the asymptotic variance of $\hat{g}_h(r)$ as p increases. To shed some light on this issue, we consider the homogeneous point processes under Lemma 3.1, in which case the exact asymptotic variance can be computed following Theorem 3.1. Let $\sigma^{(0)}_{m,n,h}(r)$ and $\sigma^{(1)}_{m,n,h}(r)$ be the asymptotic standard deviation for the local constant estimator $p = 0$ and the local linear estimator $p = 1$, respectively. A particularly interesting case is when r is close to 0 where the kernel estimator of the PCF $g(\cdot)$ typically has a large variance (in addition to the large bias discussed in Section 1). Using Lemma 3.1 and Theorem 3.1, we can derive the closed forms of $\sigma^{(p)}_{m,n,h}(r)$ for $p = 0, 1$, based on which we can see that

$$\left[\frac{\sigma^{(1)}_{m,n,h}(0)}{\sigma^{(0)}_{m,n,h}(0)}\right]^2 = \frac{(4 - 3h/T_n)[1 - h/(2T_n)]}{1 - h/T_n + h^2/(6T_n^2)} \to 4 \text{ as } h/T_n \to 0,$$

which indicates that for a small h (compared to the observation window $D_n = [0, T_n]$), the local linear estimator will have a four times larger variance than the local constant estimator. This is in contrast with the local polynomial estimator for regression models, where the asymptotic variance stays the same by using $p = 1$ instead of $p = 0$, see Chapter 3.3 of Fan and Gijbels (1996) for more details. To paint a more complete picture of the variance inflation going from $p = 0$ to $p = 1$, Figure 1 gives the ratio of variances for the local linear estimator and the local constant estimators as r increases with the observation window fixed at $T_n = 1$. This clearly demonstrates that the variance inflation for small r can be rather serious.

The bias term in Theorem 3.1 for the local constant estimator is easy to derive, owing to the closed-form solution (2.7). Specifically, some straightforward algebra gives that

$$b^o_{n,h}(r) = g_0'(r)\frac{\int_{D_n}\int_{D_n}\lambda(\mathbf{x})\lambda(\mathbf{y})K_h(\|\mathbf{x}-\mathbf{y}\| - r)(\|\mathbf{x}-\mathbf{y}\| - r)\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}}{\int_{D_n}\int_{D_n}\lambda(\mathbf{x})\lambda(\mathbf{y})K_h(\|\mathbf{x}-\mathbf{y}\| - r)\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}} + O_p(h^2).$$

The above bias term suggests that when $g_0'(r)$ is small for $r \approx 0$ (e.g., the Thomas process in (4.1)), the local constant estimator has a small bias and thus may be preferable to the local linear estimator due to its smaller variance. On the contrary, when $g_0(r)$ is steep, which is the case for the Variance Gamma

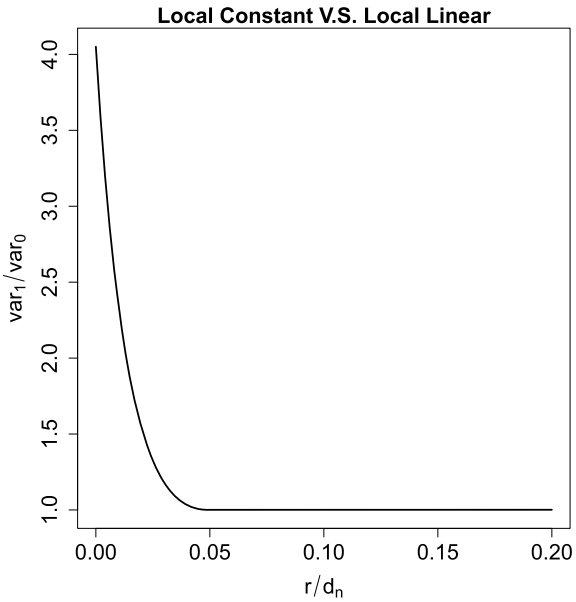

FIG 1. *Ratios of asymptotic variances of local linear estimator* (var_1) *v.s. local constant estimator* (var_0) *with* $h = 0.05$ *and* $T_n = 1$.

process in (4.1) around $r \approx 0$, the local constant estimator may suffer from severe biases, making it less attractive than the local linear estimator for both parameter estimation and statistical inference. Both phenomena are observed in our simulation studies in Section 4.

In summary, the above discussion suggests that it is not a clear-cut decision to prefer the local linear estimator as in the regression models. In practice, if making valid statistical inferences is the primary goal, the local linear estimator appears to be more reliable despite its potentially larger variance.

3.1.2. Impact of dimensionality

We now discuss the impact of dimensionality d on the local polynomial estimator. Since we assume that the true PCF is isotropic, it appears that the dimensionality should have little impact. While this is mostly true, it is not the case for $\hat{g}_h(r)$ when $r \approx 0$. To see this, consider the $\mathbf{Q}_{n,h}^{(k)}(r)$, $k = 1, 2$, defined in (3.1). Under conditions C1-C4, a straightforward application of polar coordinate transformation yields that

$$\|\mathbf{Q}_{n,h}^{(k)}(r)\|_{\max} \leq C_q \int_{\mathbb{R}} [K(s)]^k (sh + r)^{d-1} \mathrm{d}s = O\left((r + h)^{d-1}\right),$$

where $C_q > 0$ is some constant, and $\|\cdot\|_{\max}$ is the maximum norm of a matrix. Consequently, when $r \to 0$ and $d > 1$, we have that $\|\mathbf{Q}_{n,h}^{(k)}(r)\|_{\max} \to 0$. Note

that the asymptotic variance $\sigma^2_{m,n,h}(r)$ in Theorem 3.1 depends critically on the $\mathbf{Q}^{(k)}_{n,h}(r)$'s and $\|\mathbf{Q}^{(k)}_{n,h}(r)\|_{\max} \to 0$ can result in inflated $\sigma^2_{m,n,h}(r)$ for $r \approx 0$ when $d > 1$. This intuitively explains the slower convergence rate of $\hat{g}_h(r)$ in Lemma 3.2 when $r \to 0$ and $d > 1$. On the other hand, if r is an internal point in $(0, R]$, asymptotic properties of $\hat{g}_h(r)$ are not impacted by the dimension d.

3.2. The orthogonal series estimator

In this subsection, we investigate asymptotic properties of the orthogonal series estimator proposed in Section 2.3. We start by defining the truncation error of using a finite number of components in the orthogonal series as

$$\tilde{\zeta}_L(r;\boldsymbol{\theta}_0) = \log[g(r)] - \log[\tilde{g}_L(r;\boldsymbol{\theta}_0)] = \sum_{l=L+1}^{\infty} \theta_{0,l}\phi_l(r). \tag{3.4}$$

Define the $L \times L$ matrix $\mathbf{Q}_L$ as

$$\mathbf{Q}_L = \int_{D_n^2} w_R(\|\mathbf{x}-\mathbf{y}\|)\lambda(\mathbf{x})\lambda(\mathbf{y})g(\|\mathbf{x}-\mathbf{y}\|)\boldsymbol{\phi}_L(\|\mathbf{x}-\mathbf{y}\|)\boldsymbol{\phi}_L^T(\|\mathbf{x}-\mathbf{y}\|)\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}, \tag{3.5}$$

where $w_R(r) = w_0(r)I(r < R)/|D_n|$ as defined in (2.9).

Following conditions are needed for consistency of the orthogonal series estimator (2.10).

[C4'] For some $\nu_1 > 0$, the error term (3.4) satisfies (a) $\int_0^R w_o(r)\tilde{\zeta}_L^2(r;\boldsymbol{\theta}_0)\mathrm{d}r = \sum_{l=L+1}^{\infty}\theta_{0,l}^2 = O\left(L^{-2\nu_1}\right)$; (b) $\sup_{0<r\le R}|\tilde{\zeta}_L(r;\boldsymbol{\theta}_0)| = O\left(L^{-\nu_1+\tau_1}\right)$ for some $0 < \tau_1 < \nu_1$; (c) $\sup_{0<r\le R}\|\boldsymbol{\phi}_L(r)\| = O(L^{\nu_2})$ for some $0 \le \nu_2 < \nu_1$; and (d) the weight function is uniformly bounded, i.e., $w_o(r) \le C_w$ for any $0 < r \le R$.

[C5'] As $L \to \infty$, there exist c_0, ν_0 where $0 \le 2\nu_0 < \nu_1 - \nu_2$, such that

$$\eta_{\min}(\mathbf{Q}_L) > c_0 L^{-\nu_0},$$

where $\eta_{\min}(\mathbf{Q})$ denotes the smallest eigenvalue of a matrix $\mathbf{Q}$.

Condition C4' quantifies how fast the approximation error (3.4) decays using some constants ν_0, ν_1, τ_1 and τ_2, which will be reflected by the convergence rate of $\hat{g}_L(r)$. Condition C5' is a similar to C5 and ensures that the smallest eigenvalue of the sensitivity matrix of the estimating function (2.9) does not approach 0 too fast. The following lemma describes the convergence rate of $\hat{g}_L(r)$.

Lemma 3.3. *Under conditions C1-C3 and C4'-C5', as $L \to \infty$ and that $L^{4\nu_0+2\nu_2}/(m|D_n|) \to 0$, we have that*

$$\sup_{0<r<R}|g(r) - \hat{g}_L(r)| = O_p\left(L^{-\nu_1+\max\{\tau_1,\nu_0+\nu_2\}} + \frac{L^{\nu_0+\nu_2}}{\sqrt{m|D_n|}}\right). \tag{3.6}$$

The proof is given in Appendix C and the Supplementary Material.

Lemma 3.3 essentially states that the convergence rate of $\hat{g}_L(r)$ are determined by the number of basis terms L as well as the quantities $\nu_0, \nu_1, \nu_2, \tau_1$ that govern the approximation errors of the orthogonal series representation and the sensitivity of the estimating function (2.9). The first term $L^{-\nu_1+\max\{\tau_1,\nu_0+\nu_2\}}$ in (3.6) quantifies the magnitude of the bias introduced by the truncated orthogonal series approximation to $g(r)$, which is a smooth function of r and the second term $\frac{L^{\nu_0+\nu_2}}{\sqrt{m|D_n|}}$ measures the estimation standard error.

To further establish asymptotic normality of $\hat{g}_L(r)$, suppose that there exists a vector $\boldsymbol{\theta}^* = (\theta_1^*, \ldots, \theta_L^*)^T$ such that

$$\int_{D_n^2} w_R(\|\mathbf{x}-\mathbf{y}\|)\lambda(\mathbf{x})\lambda(\mathbf{y})\left[g(\|\mathbf{x}-\mathbf{y}\|) - \tilde{g}_L(\|\mathbf{x}-\mathbf{y}\|;\boldsymbol{\theta}^*)\right] \times \boldsymbol{\phi}_L(\|\mathbf{x}-\mathbf{y}|)\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} = \mathbf{0}. \tag{3.7}$$

The following additional conditions are needed for asymptotic normality.

[N1'] Either one of the following conditions are true (a) $m \to \infty$; or (b) the mixing coefficient satisfies $\alpha_X(s;2,\infty) = O(s^{-d-\varepsilon'})$ for some $\varepsilon' > 0$.

[N2'] There exists $\delta' > 2d/\varepsilon'$ such that $\left|g^{(k)}(\mathbf{x}_1,\ldots,\mathbf{x}_k)\right| \le C_g$ for any $\mathbf{x}_j \in D_n$, $j = 1,\ldots,k$, $k = 2,\ldots,2(2+\lceil\delta'\rceil)$.

[N3] For $r \in [0,R]$, define the vector $\boldsymbol{\ell}(r) = (\mathbf{Q}_L)^{-1}\boldsymbol{\phi}_L^T(r)$ and its standardized version $\boldsymbol{\ell}_0(r) = \|\boldsymbol{\ell}(r)\|^{-1}\boldsymbol{\ell}(r)$. Assume that as $m|D_n| \to \infty$ and $L \to \infty$, (a) there exists some $c_u > 0$ such that $\boldsymbol{\ell}_0^T(r)\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\boldsymbol{\ell}_0(r) \ge c_u$ with $\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*) = \mathrm{Var}\left[\sqrt{m|D_n|}\tilde{\mathbf{U}}_L(\boldsymbol{\theta}^*)\right]$; and (b) the basis functions satisfy $\int_0^R \left[w_o(s)|\boldsymbol{\ell}_0^T(r)\boldsymbol{\phi}_L(s)|\right]^{2+\lceil\delta'\rceil}\mathrm{d}s \le C_\phi$, for some $C_\phi > 0$.

Theorem 3.2. *Under conditions C1-C3, C4'-C5', N1', N2' and N3, we have that, as $L \to \infty$ and $L^{4\nu_0+2\nu_2}/(m|D_n|) \to 0$,*

$$\frac{\sqrt{m|D_n|}\left[\hat{g}_L(r) - g(r) + b_{n,L}\right]}{g(r)\sqrt{\boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\phi}_L(r)}} \xrightarrow{\mathcal{D}} N(0,1),$$

where $b_{n,L} = O\left(L^{-\nu_1+\max\{\tau_1,\nu_0+\nu_2\}}\right)$, $\mathbf{Q}_L$ is defined in (3.5) and $\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)$ in condition N3.

The proof is given in Appendix C and the Supplementary Material.

3.3. Empirical variance estimation

The asymptotic variances derived in Theorems 3.1 and 3.2 are needed to construct valid confidence bands for $g(r)$ in practice. However, direct calculations of the asymptotic variances are difficult, especially for the orthogonal series estimator. In this section, we propose a straightforward approach to estimate the

asymptotic variances for both the local polynomial estimator and the orthogonal series estimator. We shall illustrate the estimation process for the local polynomial estimator only. The extension to the orthogonal series estimator is straightforward and can be found in the Supplementary Material.

3.3.1. Preliminaries

Define two random vectors

$$\mathbf{Z}_1 = \frac{1}{m}\sum_{i=1}^{m}\sum_{\mathbf{u},\mathbf{v}\in X_i}^{\neq} w_{r,h}(\|\mathbf{u}-\mathbf{v}\|)\mathbf{A}_h(\|\mathbf{u}-\mathbf{v}\|-r), \tag{3.8}$$

$$\mathbf{Z}_2(\boldsymbol{\theta}^*) = \sum_{i\neq j}\sum_{\mathbf{u}\in X_i}\sum_{\mathbf{v}\in X_j}\frac{w_{r,h}(\|\mathbf{u}-\mathbf{v}\|)}{m(m-1)}\tilde{g}_{r,h}(\|\mathbf{u}-\mathbf{v}\|;\boldsymbol{\theta}^*)\mathbf{A}_h(\|\mathbf{u}-\mathbf{v}\|-r), \tag{3.9}$$

where $\boldsymbol{\theta}^*$ is defined in (2.6), $\mathbf{A}_h(\cdot)$ is as defined in (3.1) and $\tilde{g}_{r,h}(\cdot;\boldsymbol{\theta})$ is defined in (2.3). By the definition of $\boldsymbol{\theta}^*$, it is straightforward to show that

$$\mathbb{E}\mathbf{Z}_1 = \mathbb{E}\mathbf{Z}_2(\boldsymbol{\theta}^*) = \int_{D_n^2} w_{r,h}(\|\mathbf{x}-\mathbf{y}\|)\lambda(\mathbf{x})\lambda(\mathbf{y})g(\|\mathbf{x}-\mathbf{y}\|) \times \mathbf{A}_h(\|\mathbf{x}-\mathbf{y}\|-r)\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}. \tag{3.10}$$

Using Lemma S.5 in the Supplementary Material together with a simple application of the Delta method, we can show that the asymptotic variance of $\hat{g}_h(r)$ is of the form

$$\mathrm{Var}^{\mathrm{asy}}[\hat{g}_h(r)] = \boldsymbol{e}^T[\mathbf{Q}_{n,h}^{(1)}(r)]^{-1}\mathrm{Var}[\mathbf{Z}_1-\mathbf{Z}_2(\boldsymbol{\theta}^*)][\mathbf{Q}_{n,h}^{(1)}(r)]^{-1}\boldsymbol{e},$$

where $\boldsymbol{e}$ is defined in Theorem 3.1 and the matrix $\mathbf{Q}_{n,h}^{(1)}(r)$ can be estimated simply as

$$\widehat{\mathbf{Q}}_{n,h}^{(1)}(r) = \sum_{i\neq j}\sum\sum_{\mathbf{u}\in X_i, v\in X_j}\frac{w_{r,h}(\|\mathbf{u}-\mathbf{v}\|)}{m(m-1)}\mathbf{G}_r(\|\mathbf{u}-\mathbf{v}\|)\mathbf{G}_r^T(\|\mathbf{u}-\mathbf{v}\|). \tag{3.11}$$

Therefore, to estimate $\mathrm{Var}^{\mathrm{asy}}[\hat{g}_h(r)]$, it suffices to estimate $\mathrm{Var}[\mathbf{Z}_1-\mathbf{Z}_2(\boldsymbol{\theta}^*)]$.

In Appendix B, we show that $\mathrm{Var}[\mathbf{Z}_1-\mathbf{Z}_2(\boldsymbol{\theta}^*)]$ is asymptotically equivalent to $\mathrm{Var}[\mathbf{Z}_1-\tilde{\mathbf{Z}}_2(\boldsymbol{\theta}^*)]$ where

$$\mathbf{Z}_2(\boldsymbol{\theta}^*) \approx \tilde{\mathbf{Z}}_2(\boldsymbol{\theta}^*) = \frac{2}{m}\sum_{i=1}^{m}\sum_{\mathbf{u}\in X_i} q(\mathbf{u}) - \mathbb{E}\mathbf{Z}_2(\boldsymbol{\theta}^*),$$

and $q(\mathbf{u}) = \int_{D_n}\lambda(\mathbf{v})w_{r,h}(\|\mathbf{u}-\mathbf{v}\|)\tilde{g}_{r,h}(\|\mathbf{u}-\mathbf{v}\|;\boldsymbol{\theta}^*)\mathbf{A}_h(\|\mathbf{u}-\mathbf{v}\|-r)\mathrm{d}\mathbf{v}$. Moreover,

$$\mathbf{Z}_1-\tilde{\mathbf{Z}}_2(\boldsymbol{\theta}^*) = \frac{1}{m}\sum_{i=1}^{m}\tilde{\xi}(X_i) + \mathbb{E}\mathbf{Z}_2(\boldsymbol{\theta}^*),$$

where the $\tilde{\xi}(X_i) = \sum_{u,v\in X_i}\sum w_{r,h}(\|\mathbf{u}-\mathbf{v}\|)\mathbf{A}_h(\|\mathbf{u}-\mathbf{v}\|-r)-2\sum_{\mathbf{u}\in X_i} q(\mathbf{u}), i = 1,\ldots,p,$ are independent and identically distributed. It thus just remains to estimate $\mathrm{Var}\left[\tilde{\xi}(X_1)\right]$.

3.3.2. *Estimation of* $\mathrm{Var}\left[\tilde{\xi}(X_1)\right]$

To estimate $\mathrm{Var}\left[\tilde{\xi}(X_1)\right]$, define a partition $\{\Delta_k\}_{k=1}^{n_p}$ of the observation domain D_n and let

$$\mathbf{Y}_{ik} = \sum_{\mathbf{u}\in\Delta_k\cap X_i}\left[\sum_{\mathbf{v}\in X_i} w_{r,h}(\|\mathbf{u}-\mathbf{v}\|)\mathbf{A}_h(\|\mathbf{u}-\mathbf{v}\|-r) - 2q(\mathbf{u})\right].$$

Under similar conditions such as N1 where the α-mixing coefficient of the point process quickly decays, the correlations among $\mathbf{Y}_{ik_1}$'s and $\mathbf{Y}_{ik_2}$'s can be ignored for any $k_1 \neq k_2$ and $i = 1,\ldots,m$. It follows that

$$\mathrm{Var}\left[\tilde{\xi}(X_1)\right] \approx \sum_{k=1}^{n_p}\mathrm{Var}\left(\mathbf{Y}_{ik}\right).$$

Since $\mathbf{Y}_{ik}$ depend on the unknown $q(\mathbf{u})$, we obtain an approximation $\widehat{\mathbf{Y}}_{ik}$ by replacing the unknown $q(\mathbf{u})$ by $\hat{q}_i(\mathbf{u}) = \frac{1}{m-1}\sum_{j\neq i}\sum_{\mathbf{v}\in X_j}\sum w_{r,h}(\|\mathbf{u}-\mathbf{v}\|)\hat{g}_h(\|\mathbf{u}-\mathbf{v}\|)\mathbf{A}_h(\|\mathbf{u}-\mathbf{v}\|-r)$. The leave-one-out estimator $\hat{q}_i(\mathbf{u})$ is used in $\widehat{\mathbf{Y}}_{ik}$ for each i to avoid dependence between $\hat{q}_i(\mathbf{u})$ and X_i, which gave better finite sample performance in simulation studies (not shown). Our estimator of $\mathrm{Var}\left[\tilde{\xi}(X_1)\right]$ then becomes

$$\widehat{\mathrm{Var}}\left[\tilde{\xi}(X_1)\right] = \sum_{k=1}^{n_p}\frac{1}{m}\sum_{i=1}^{m}(\widehat{\mathbf{Y}}_{ik}-\overline{\mathbf{Y}}_k)(\widehat{\mathbf{Y}}_{ik}-\overline{\mathbf{Y}}_k)^T,$$

where $\overline{\mathbf{Y}}_k = \frac{1}{m}\sum_{i=1}^{m}\widehat{\mathbf{Y}}_{ik}$. It is worth pointing out that if m is relatively large, no partition of D_n is needed, which results in the case with $n_p = 1$.

3.3.3. *Final variance estimator*

By the previous considerations we arrive at

$$\widehat{\mathrm{Var}}\left[\mathbf{Z}_1 - \mathbf{Z}_2(\boldsymbol{\theta}^*)\right] = \frac{1}{m^2}\sum_{k=1}^{n_p}\sum_{i=1}^{m}(\widehat{\mathbf{Y}}_{ik}-\overline{\mathbf{Y}}_k)(\widehat{\mathbf{Y}}_{ik}-\overline{\mathbf{Y}}_k)^T. \tag{3.12}$$

The empirical variance estimator for the local polynomial estimator is then defined as

$$\widehat{\mathrm{Var}}[\hat{g}_h(r)] = \boldsymbol{e}^T[\widehat{\mathbf{Q}}_{n,h}^{(1)}(r)]^{-1}\widehat{\mathrm{Var}}[\mathbf{Z}_1 - \mathbf{Z}_2(\boldsymbol{\theta}^*)][\widehat{\mathbf{Q}}_{n,h}^{(1)}(r)]^{-1}\boldsymbol{e}, \tag{3.13}$$

where $\widehat{\mathbf{Q}}_{n,h}^{(1)}(r)$ is defined in (3.11) and $\widehat{\mathrm{Var}}[\mathbf{Z}_1 - \mathbf{Z}_2(\boldsymbol{\theta}^*)]$ in (3.12).

A similar variance estimator can be obtained for the orthogonal series estimator with $\boldsymbol{e}$ replaced by $\boldsymbol{\phi}_L(r)$ and $\widehat{\mathbf{Q}}^{(1)}_{n,h}(r)$ replaced by a consistent estimator of $\mathbf{Q}_L$. More details can be found in the Supplementary Material. Formal justifications of the proposed variance estimator (3.12) involve tedious lengthy proofs and a few more technical conditions. For the ease of presentation, we shall not pursue this direction. Instead, we will illustrate the effectiveness of these variance estimators through extensive simulation studies in the next section, where the main message is that when m is sufficiently large, the proposed variance estimator (3.13) is unbiased with a diminishing variance when m or $|D_n|$ grows. In the case when m is moderate to small, (3.13) tends to over-estimate $\mathrm{Var}[\hat{g}_h(r)]$ for small lags r in our simulation studies, resulting in a wider confidence band.

4. Simulation study

We conduct a simulation study to evaluate the finite sample performance of the proposed PCF estimators. We generate m replicated realizations of inhomogeneous Thomas and Variance Gamma (VarGamma) processes (Jalilian et al., 2013) on $[0, T]$, where $m = 30, 50, 100$ and $T = 30, 60$. Let ρ, μ and σ denote the parent intensity of the process, the mean number of offspring generated per parent, and the standard deviation of an offspring's position relative to its parent, respectively. We set $\mu = 6$, i.e., each parent generates on average six offspring events. Each simulated offspring event is retained randomly with a probability equal to $p(x) = 0.28[\sin(2\pi x) + \sin(4\pi x) + 1.811256]$; this probability function is used such that the average number of events per realization when $T = 30$ is similar to the pattern observed in the Sina Weibo data analyzed in Section 5. The resulting intensity function is of the form $\lambda(x) = \rho\mu p(x)$ and the PCF is

$$g(r) = \begin{cases} 1 + \frac{1}{2\sqrt{\pi}\rho\sigma} \exp\left(-\frac{r^2}{4\sigma^2}\right), & \text{for the Thomas process,} \\ 1 + \frac{1}{4\rho\sigma} \exp\left(-\frac{r}{2\sigma}\right), & \text{for the Variance Gamma process.} \end{cases} \tag{4.1}$$

Note that a smaller ρ and a smaller σ both lead to larger $g(r)$ values at small lags. We set $\rho = 1.0$ and $\sigma = 0.025, 0.030$ to reflect different strengths of clustering. A selection of simulated point patterns are shown in the top panel of Figure 2.

For each simulation, we estimate the pair correlation function using the three proposed nonparametric estimators: the local constant estimator ($\hat{g}_c$), the local linear estimator ($\hat{g}_l$) and the orthogonal series estimator ($\hat{g}_o$). For the orthogonal series estimator, the cosine basis functions as described in Section 2.3 are used. The bottom panel in Figure 2 shows that the PCF of the Thomas process can be easily approximated by the orthogonal series with only $L = 6$ basis functions, but the PCF of the Variance Gamma process requires more than $L = 20$ basis functions to avoid large approximation error around $r \approx 0$. In other words, the PCF of the Variance Gamma process is much more difficult to approximate using orthogonal series.

For all simulation studies, the tuning parameters h and L are chosen through 5-fold cross validation as suggested in Appendix A. Specifically, h is selected

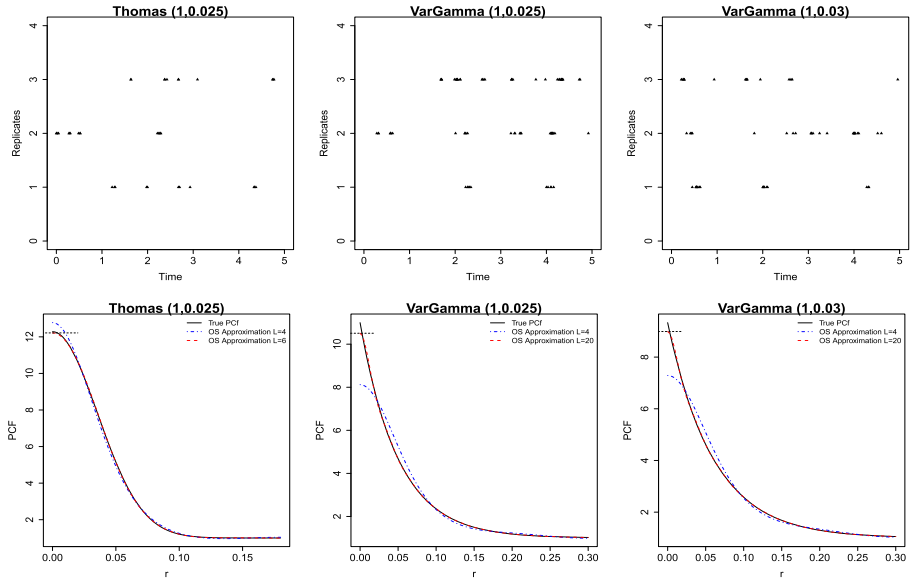

FIG 2. *Top panel: simulated replicated point process with* $T = 5$. *Bottom panel: Orthogonal series approximation to the true PCF with* L *basis functions,* $L = 4, 6,$ *or* 20.

from 50 equally spaced grid points between $[0.001, R/5]$, where the upper bound R is fixed at 0.18 for the Thomas process and 0.30 for the Variance Gamma process. The L is chosen from the integers 4–50.

4.1. Estimation accuracy

We first show the computational cost of the proposed nonparametric PCF estimators in Figure 3. All simulations are conducted in the software R on a cluster of 200 Linux machines with a total of 200 CPU cores, each of which runs at approximately 2 GFLOPS. Each simulation run is carried out independently using a CPU core without parallel computing. The average CPU times for all three estimators are less than half a second for a given h or L, indicating that the proposed estimators are computationally efficient.

Next, the estimation accuracy of the proposed estimators are evaluated by the mean integrated squared errors (MISEs), i.e., $\int_0^U [\hat{g}_k(r) - g(r)]^2 dt$, $k = c, l, o$, for $U = 0.06, 0.12, 0.18$ based on 1,000 simulations runs. As a comparison benchmark, we introduce the following naive PCF estimator

$$\hat{g}_{naive}(r) = \frac{1}{2mc(r)} \sum_{i=1}^{m} \sum_{u,v \in X_i}^{\neq} \frac{K_h(|u-v| - r)}{\lambda(u)\lambda(v)|T - |u-v||},$$

where $|T - |u - v||$ is the 1-dimensional translation edge correction weight for $u, v \in [0, T]$, $c(r) = \int_0^T K_h(s - r)\mathrm{d}s$ is a boundary correction term for $r < h$

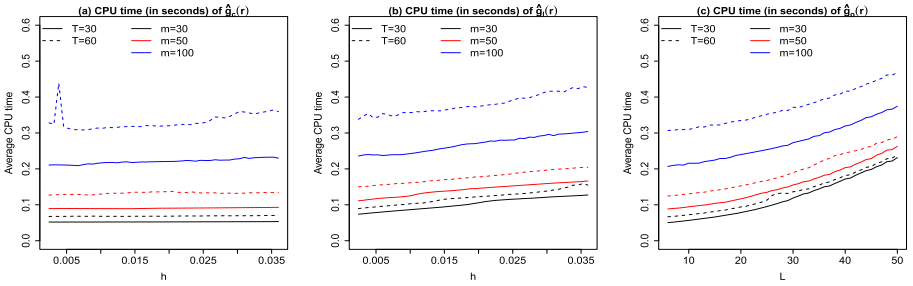

FIG 3. *Computation times for* $\hat{g}_c(\cdot)$, $\hat{g}_l(\cdot)$ *and* $\hat{g}_o(\cdot)$ *for the Thomas process.*

and $\lambda(\cdot)$ is the true first-order intensity function. It is straightforward to see that $\hat{g}_{naive}(\cdot)$ is an average of popular kernel PCF estimators for individual temporal point process X_i's. Since our goal is simply to provide a benchmark to evaluate performances of the proposed PCF estimators, we use the true intensity $\lambda(\cdot)$ instead of an estimated version in order to avoid further complications arising from the estimation of $\lambda(\cdot)$. If an estimated $\lambda(\cdot)$ is used, our unreported simulation results suggest that the performance of $\hat{g}_{naive}(\cdot)$ will become worse.

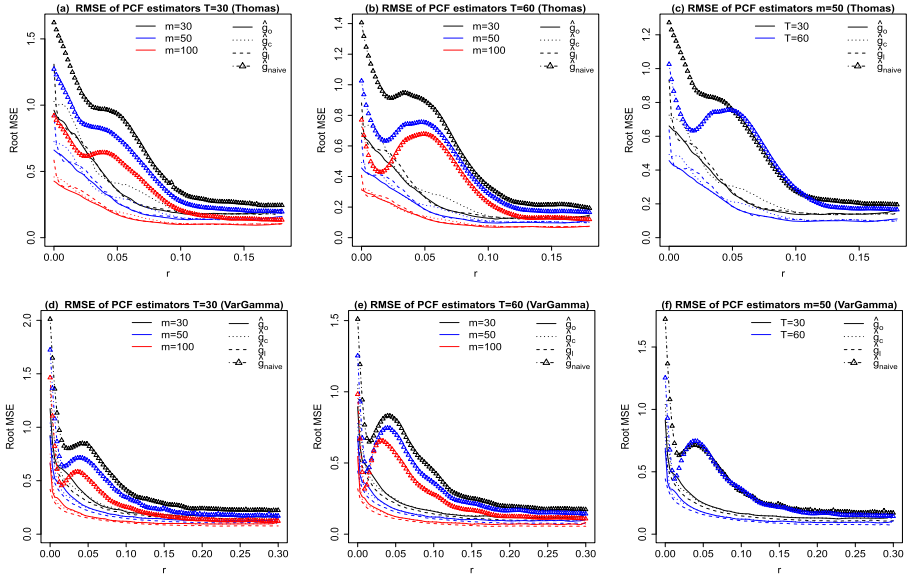

FIG 4. *The point-wise root-mean squared errors (RMSE) for all four PCF estimators with* $\rho = 1$ *and* $\sigma = 0.025$.

Figure 4 summarizes the point-wise root mean squared errors (RMSE) for the four estimators under various scenarios. In all case scenarios, the three proposed nonparametric PCF estimators outperform the naive PCF estimator $\hat{g}_{naive}(\cdot)$.

For the Thomas process, it is apparent that when $r \approx 0$, the RMSE of the local linear estimator $\hat{g}_l(r)$ inflates, resulting in its inferior performance. The orthogonal series estimator $\hat{g}_o(r)$ outperforms the other two almost uniformly for all $r \in [0, 0.18]$. However, the message is rather different from the Variance Gamma process, in which case the local linear estimator $\hat{g}_l(r)$ achieves the best performance when r is small. The same conclusion can be reached from the MISEs summarized in Table 1, which shows that for every simulation scenario with the Thomas process, the orthogonal series estimator always yields the smallest MISE values. In contrast, for all cases that involve the Variance Gamma process, the local linear estimator consistently outperforms the other two estimators.

TABLE 1
MISEs (10^{-2}) *of proposed nonparametric PCF estimators over intervals* $[0, U]$ *with* $U = 0.06, 0.12, 0.18$.

			Thomas			VarGamma		
	σ	Method	$[0, 0.06]$	$[0, 0.12]$	$[0, 0.18]$	$[0, 0.06]$	$[0, 0.12]$	$[0, 0.18]$
$m = 50$ $T = 30$	$\sigma = 0.025$	$\hat{g}_c$	1.444	1.659	1.776	1.181	1.379	1.501
		$\hat{g}_l$	1.410	1.570	1.686	0.740	0.896	0.986
		$\hat{g}_o$	1.229	1.376	1.501	1.085	1.287	1.406
	$\sigma = 0.03$	$\hat{g}_c$	1.032	1.249	1.349	0.884	1.06	1.17
		$\hat{g}_l$	1.060	1.220	1.319	0.554	0.692	0.774
		$\hat{g}_o$	0.859	1.009	1.104	0.817	0.999	1.109
$m = 50$ $T = 60$	$\sigma = 0.025$	$\hat{g}_c$	0.705	0.814	0.877	0.577	0.683	0.748
		$\hat{g}_l$	0.702	0.783	0.845	0.346	0.432	0.479
		$\hat{g}_o$	0.579	0.656	0.716	0.545	0.649	0.713
	$\sigma = 0.03$	$\hat{g}_c$	0.512	0.626	0.679	0.470	0.564	0.626
		$\hat{g}_l$	0.530	0.612	0.664	0.286	0.359	0.403
		$\hat{g}_o$	0.448	0.526	0.576	0.443	0.541	0.601
$m = 100$ $T = 30$	$\sigma = 0.025$	$\hat{g}_c$	0.588	0.682	0.745	0.513	0.611	0.675
		$\hat{g}_l$	0.616	0.694	0.756	0.302	0.383	0.429
		$\hat{g}_o$	0.502	0.576	0.635	0.493	0.593	0.656
	$\sigma = 0.03$	$\hat{g}_c$	0.452	0.548	0.601	0.404	0.491	0.545
		$\hat{g}_l$	0.485	0.563	0.615	0.250	0.318	0.356
		$\hat{g}_o$	0.393	0.465	0.515	0.397	0.485	0.539
$m = 100$ $T = 60$	$\sigma = 0.025$	$\hat{g}_c$	0.284	0.333	0.365	0.294	0.347	0.381
		$\hat{g}_l$	0.301	0.342	0.374	0.167	0.207	0.229
		$\hat{g}_o$	0.271	0.308	0.338	0.285	0.337	0.37
	$\sigma = 0.03$	$\hat{g}_c$	0.215	0.265	0.292	0.219	0.268	0.299
		$\hat{g}_l$	0.231	0.272	0.298	0.118	0.154	0.175
		$\hat{g}_o$	0.212	0.251	0.276	0.211	0.261	0.291

To explain such performance differences of different estimators between the Thomas process and the Variance Gamma process, Figure 5 shows the relative biases of the three estimators under various case scenarios. We can observe that for the Thomas process, all three estimators have biases with similar magnitudes, and are relatively small compared to $g_0(r)$. In this case, the orthogonal series estimator has the best estimation performance due to its low variance, see also Figure 6 for variance comparisons. In contrast, for the Variance Gamma process, when $r \approx 0$, both $\hat{g}_c(r)$ and $\hat{g}_o(r)$ have appreciably larger biases than those of $\hat{g}_l(r)$. This observation echoes with Figure 2 in that the PCFs of the Variance Gamma process need more than $L = 20$ basis functions for a decent orthogonal series approximation. However, the proposed tuning parameter selection method in Appendix A tends not to choose a sufficiently large L, which results in the observed biases for $\hat{g}_o(r)$ in Figure 2 (d)-(f). The biases for $\hat{g}_c(r)$

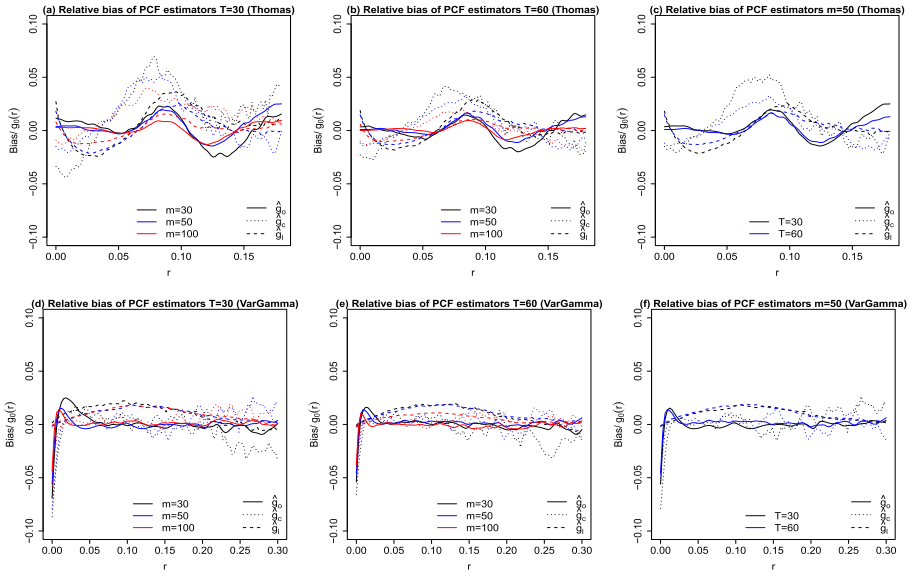

FIG 5. *The point-wise relative bias for all three PCF estimators with* $\rho = 1$ *and* $\sigma = 0.025$.

can be similarly explained by the relatively large bandwidth h chosen by our cross-validation approach. On the contrary, the local linear estimator $\hat{g}_l(\cdot)$ with h chosen by cross-validation achieves small biases for both the Thomas and Variance Gamma processes. In particular, for the Variance Gamma process, although $\hat{g}_l(\cdot)$ still has larger variances (see Figure 7), the much smaller biases lead to its overall superior performance in terms of RMSE, compared to the other two PCF estimators.

4.2. Performances of empirical variance estimators

In this subsection, we evaluate performances of the empirical variance estimators proposed in Section 3.3 for both $\hat{g}_l(\cdot)$ and $\hat{g}_o(\cdot)$. The performance of the variance estimator for $\hat{g}_c(\cdot)$ is rather similar to that of $\hat{g}_o(\cdot)$, and hence is omitted. In this simulation study, we fix $\mu = 6$, $\rho = 1$ and $\sigma = 0.025$ and consider increasing the number of replicated point patterns (i.e., m) and the length of the observation windows (i.e., T). For $\hat{g}_l(r)$, we denote the empirical variance estimator proposed in Section 3.3 as $\hat{\sigma}_l^2(r)$. Similarly, we denote by $\hat{\sigma}_o^2(r)$ the estimated empirical variance for $\hat{g}_o(r)$. The partition used for both estimators is obtained by dividing $[0, T]$ into 10 equally spaced subintervals. Summary statistics based on $1,000$ simulation runs are illustrated in Figures 6–7. For the Thomas process, Figure 6(a)-(b) and (d)-(e) show that when $m = 30$, both $\hat{\sigma}_l(r)$ and $\hat{\sigma}_o(r)$ tend to slightly overestimate the truth for small lags r. This observation is expected because the proposal in Section 3.3 is based on the assumption that m is large. As we can see, when m increases, the bias quickly decreases. Figure 6(c) and 6(f)

illustrate that both $\hat{\sigma}_l(r)$ and $\hat{\sigma}_o(r)$ have very small standard deviations compared to their true values, and hence can be reliably used in practice. Similar observations can also be made on Figure 7 for the Variance Gamma process.

4.3. Confidence band coverage probabilities

In this subsection, we study the coverage probabilities of the point-wise confidence band defined as

$$\mathrm{CI}_k: \quad \hat{g}_k(r) \pm z_{1-\alpha/2}\hat{\sigma}_k(r), \quad k = c, l, o, \text{ and } r \in [0, R],$$

where $\hat{\sigma}_k(\cdot)$'s are as defined in Section 4.2. In this simulation study, we fix $\mu = 6$, $\rho = 1$ and $\sigma = 0.025$ and consider increasing the number of replicated point patterns (i.e., m) and the length of the observation windows (i.e., T). The empirical coverage probabilities of the 95% confidence bands based on 1,000 simulation runs are illustrated in Figures 8.

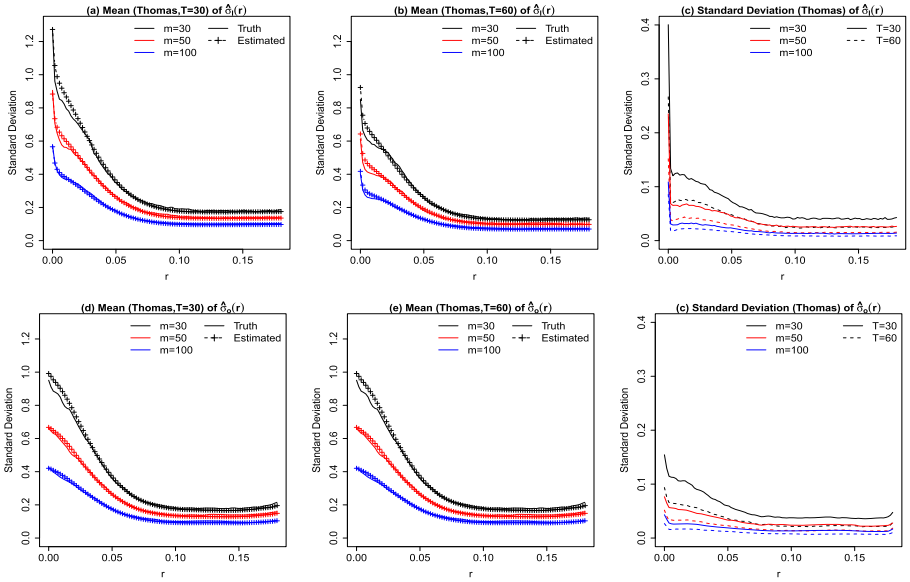

FIG 6. *The means and standard deviations of $\hat{\sigma}_l(r)$ and $\hat{\sigma}_o(r)$ for the Thomas process.*

In Figures 8(a)-(c), we can see that for the Thomas process, the empirical coverage probabilities of all three types of confidence intervals are reasonably close to the nominal level. This is expected based on our observations from Figures 5(a)-(c) and Figures 6, where the estimation biases are relatively small and the variance estimators are close to the truth. However, for the Variance Gamma process, the coverage probabilities of CI_c and CI_o are far away from 0.95 when $r \approx 0$, which can be explained by the large estimation biases we observed from Figures 5(d) and (f) for $\hat{g}_c(r)$ and $\hat{g}_o(r)$ when $r \approx 0$. On the contrary, the

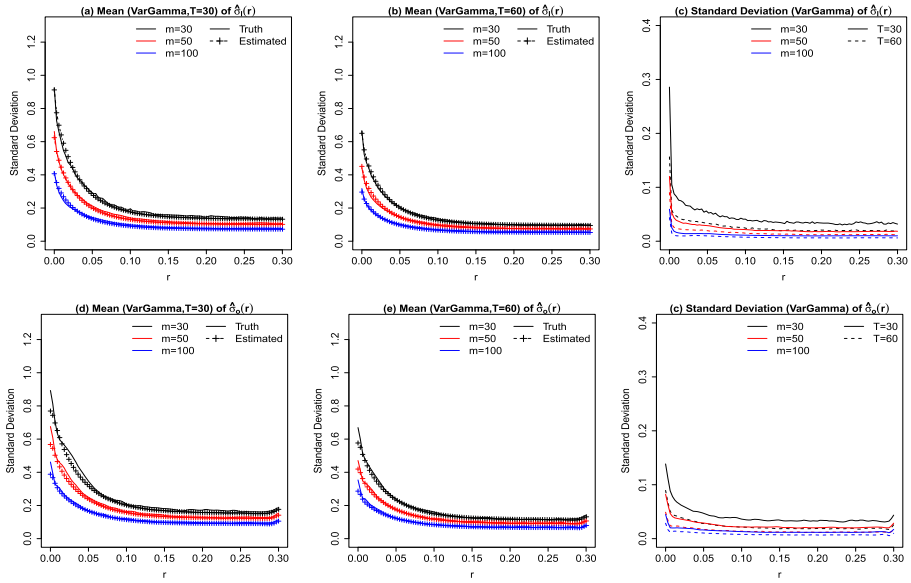

FIG 7. *The means and standard deviations of $\hat{\sigma}_I(r)$ and $\hat{\sigma}_o(r)$ for the Variance Gamma process.*

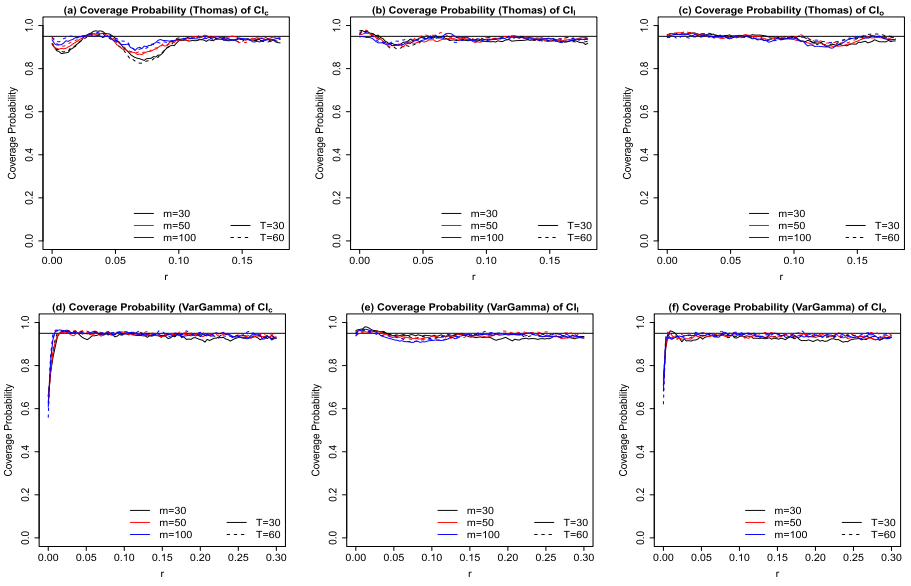

FIG 8. *Coverage probabilities of point-wise confidence bands.*

CI_l still has coverage probabilities that are rather close to 0.95. To sum up, the local linear PCF estimator appears to be more reliable for statistical inferences than the other two estimators, despite its potentially larger variance. The local constant estimator seems to be consistently inferior to the orthogonal series estimator in terms of both estimation accuracies and the coverage probabilities.

5. Sina Weibo data analysis

Sina Weibo (www.weibo.com) is the largest Twitter-type social media in China. The data set being studied contains $m = 1,695$ active followers of the official Weibo account of an MBA program, whose tweeting time stamps are recorded for a total of $T = 30$ days. The objective is to study the posting activity patterns of the Sina Weibo users. The users are categorized according to the following criteria: 'Male' (1221) vs. 'Female' (474); 'Celebrity' (424) vs. 'nonCelebrity' (1271); and 'Member' (107) vs. 'nonMember' (1588). A 'Member' is defined as a user who pays a monthly fee to Sina Weibo. Plots of posting activities for subsets of users within the different categories are shown in the top panel of Figure 9.

The bottom panel in Figure 9 shows the kernel smoothed intensity functions of posting activities for different groups of users over a period of 24 hours which is rescaled to the unit interval. It can be seen that male users tend to have a larger intensity than female users with a similar shape and celebrity users have a larger intensity than the nonCelebrity users. The last plot shows that the member users' intensity function seems to be much greater than that of nonMember users. In all these groups, we can see that there exists a period during a day with intensities approximately equal to 0, which may cause problems for existing nonparametric PCF estimators as we argued in Section 1.

Next, we use the local linear constant and orthogonal series estimators to estimate the PCFs for different groups of users and over a time lag range $[0, 0.12]$. For the orthogonal series estimator, the cosine basis functions as described in Section 2.3 are used. We also compute 95% point-wise confidence intervals based on the empirical variance estimators proposed in Section 3.3. The tuning parameters h and L are chosen through 5-fold cross validation as suggested in Appendix A. The results are summarized in Figures 10 for all user groups.

The left panel in Figure 10 shows that the female users have a slightly larger PCF than the male users for short time lags. Thus, male users tend to post more Weibo in a day (cf. Figure 9) while the posting timestamps have a more clustered pattern for the female users. The middle panel shows that the nonCelebrity users have markedly larger PCF for small time-lags relative to the celebrity users, while the celebrity users have a notably larger intensity function than nonCelebrity users, as illustrated in Figure 9. This suggests that users who have more followers tend to post more frequently with a less clustered tweeting pattern than users who have fewer followers. Finally, the Member vs. nonMember comparison based on the right panel suggests that Member users tend to not only post much more than nonMember users but also have more clustered posting patterns.

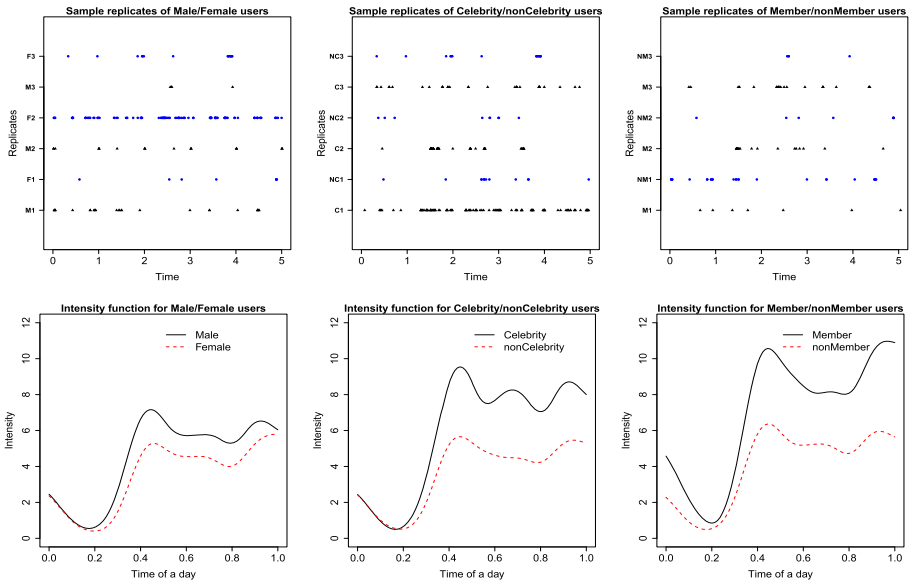

FIG 9. *Top Panels: Sample replicates of different users (first five days). Bottom Panels: Intensity functions for various user groups.*

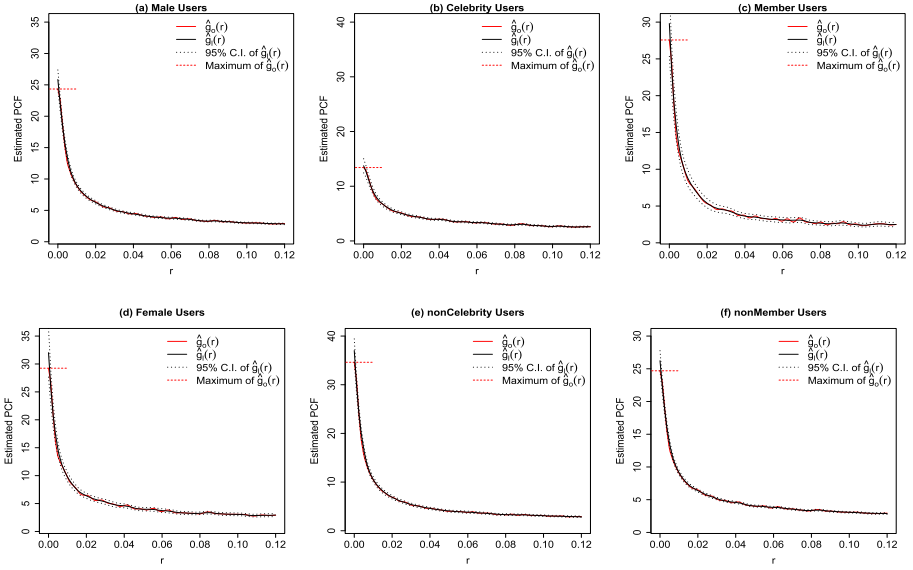

FIG 10. *Orthogonal series estimators and local linear estimators of PCF functions for various user groups.*

6. Discussion

We proposed two types of nonparametric estimators for the PCF of inhomogeneous point processes when replicates are available. Their asymptotic properties and numerical performances are carefully studied and empirical variance estimators are proposed to construct point-wise confidence bands for the PCF for practical analysis. The orthogonal series estimator appears to have lower variance than the local polynomial estimators, and therefore is preferred when the true PCF can be well approximated by a small number of basis functions, in which case the estimation bias is well controlled. However, when a large number of basis functions are needed to approximate the true PCF, the orthogonal series estimator suffers from severe estimation bias. On the contrary, the local linear estimator can avoid large estimation biases and thus is more reliable for statistical inferences. Therefore, in practice, we recommend to use the local linear estimator, although the orthogonal series estimator potentially has better performances in some case scenarios.

The current work assumes that all replicated point patterns share the same first-order intensity, which may be restrictive for some applications. An immediate extension is to follow the work of Xu et al. (2019) to model the intensity function semi-parametrically utilizing some covariates related to the observed point patterns. For example, one can assume that the intensity function for each point process X_i is of the form $\lambda_i(\mathbf{x};\boldsymbol{\beta}) = \lambda_0(\mathbf{x})\exp\left[\boldsymbol{\beta}^T\mathbf{z}_i(\mathbf{x})\right]$, where $\mathbf{z}_i(\mathbf{x})$ is a vector of covariates at location $\mathbf{x}$ for X_i and $\lambda_0(\mathbf{x})$ is a shared background intensity for all X_i's, $i = 1,\ldots,m$. An estimator $\widehat{\boldsymbol{\beta}}$ of the regression parameter $\boldsymbol{\beta}$ can be obtained using the partial likelihood estimation approach (Cook and Lawless, 2007). Next the estimating function (2.4) can be modified by replacing the weight function $w_{r,h}(\|\mathbf{u}-\mathbf{v}\|)$ with $w_{r,h}(\|\mathbf{u}-\mathbf{v}\|)\exp\left[-\widehat{\boldsymbol{\beta}}^T\mathbf{z}_i(\mathbf{u}) - \widehat{\boldsymbol{\beta}}^T\mathbf{z}_i(\mathbf{v})\right]$. We anticipate our theories to hold as long as $\widehat{\boldsymbol{\beta}}$ converges to $\boldsymbol{\beta}$ at a sufficiently fast rate.

Another direction of extension is to relax the assumption of independence among the X_i's. One way to deal with this issue is to model X_i's as a sequence of point processes. For example, X_i can be a point process observed on day i, $i = 1,\ldots,m$ and one can assume that X_i is only correlated with $X_{i-1},\ldots,X_{i-k}$ from the past k days. When k is negligible compared to m as m increases, the proposed nonparametric PCF estimators should still be consistent. However, the asymptotic distributions of the proposed estimators will be different and depend on the specific correlation structure among the X_i's.

Finally, throughout the paper, we have assumed the true PCF to be isotropic. However, this assumption can be removed when there are a large number of replicates. For example, the local constant estimator (2.7) can be readily modified using a product kernel as

$$\widehat{g}_{\mathbf{h}}(\mathbf{x},\mathbf{y}) = \frac{(m-1)\sum_{i=1}^{m}\sum_{\mathbf{u},\mathbf{v}\in X_i}^{\neq} K_{h_1}(\|\mathbf{u}-\mathbf{x}\|)K_{h_2}(\|\mathbf{v}-\mathbf{y}\|)}{\sum\sum_{i\neq j}\sum_{\mathbf{u}\in X_i,\mathbf{v}\in X_j}\sum K_{h_1}(\|\mathbf{u}-\mathbf{x}\|)K_{h_2}(\|\mathbf{v}-\mathbf{y}\|)},$$

for $\mathbf{h} = (h_1, h_2)^T$, and $\mathbf{x}, \mathbf{y} \in D_n$. The bandwidth vector $\mathbf{h}$ can be selected by a small modification to the definitions of $\widehat{M}_1(h)$ and $\widehat{M}_2(h)$ in (A.2). Specifically, one can simply remove the term $\frac{I(\|\mathbf{u}-\mathbf{v}\|\le R)}{\|\mathbf{u}-\mathbf{v}\|^{d-1}}$ from the definitions of $\widehat{M}_1(h)$ and $\widehat{M}_2(h)$. A local linear estimator and an orthogonal series estimator for $g(\mathbf{x}, \mathbf{y})$ without the isotropy assumption can also be developed, but the derivations will be less straightforward and we therefore omit a detailed discussion.

Appendix A: Tuning parameter selection

As for any nonparametric method, the proposed estimators depend on tuning parameters, i.e., the bandwidth h for the local polynomial estimator and the number of basis functions L for the orthogonal series estimator. In this section we describe a data driven method for the selection of the tuning parameters. We will focus on the local polynomial estimator to illustrate our approach. Let R be the largest lag for which the PCF is to be estimated. Then, inspired by Guan (2007), we choose the optimal bandwidth h as the one that minimizes an estimate of the least-square discrepancy measure

$$M(h) = \int_D \int_D \frac{\lambda(\mathbf{x})\lambda(\mathbf{y}) I(\|\mathbf{x}-\mathbf{y}\| \le R)}{\|\mathbf{x}-\mathbf{y}\|^{d-1}} [\hat{g}_h(\|\mathbf{x}-\mathbf{y}\|) - g(\|\mathbf{x}-\mathbf{y}\|)]^2 \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}. \quad \text{(A.1)}$$

Minimizing (A.1) with respect to h is equivalent to minimizing $M_1(h) - 2M_2(h)$ where

$$M_1(h) = \int_D \int_D \frac{\lambda(\mathbf{x})\lambda(\mathbf{y}) I(\|\mathbf{x}-\mathbf{y}\| \le R)}{\|\mathbf{x}-\mathbf{y}\|^{d-1}} [\hat{g}_h(\|\mathbf{x}-\mathbf{y}\|)]^2 \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}, \quad \text{and}$$

$$M_2(h) = \int_D \int_D \frac{\lambda(\mathbf{x})\lambda(\mathbf{y}) I(\|\mathbf{x}-\mathbf{y}\| \le R)}{\|\mathbf{x}-\mathbf{y}\|^{d-1}} \hat{g}_h(\|\mathbf{x}-\mathbf{y}\|) g(\|\mathbf{x}-\mathbf{y}\|) \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}.$$

We propose to use cross-validation to estimate $M_1(h)$ and $M_2(h)$. Specifically, we randomly divide the m replicates into n_{cv} non-overlapping folds. For each $k = 1, \ldots, n_{cv}$, let $\mathcal{S}_k$ be the collection of replicates used as the test data and denote $\hat{g}_h^{-(k)}(\cdot)$ as the $\hat{g}_h(\cdot)$ obtained by using the training data that do not include $\mathcal{S}_k$. Then the proposed estimators for $M_1(h)$ and $M_2(h)$ are $\widehat{M}_j(h) = \frac{1}{n_{cv}} \sum_{k=1}^{n_{cv}} \widehat{M}_j^{(k)}(h)$, $j = 1, 2$, with

$$\widehat{M}_1^{(k)}(h) = \frac{1}{m_k(m_k-1)} \sum_{i \ne j \in \mathcal{S}_k} \sum \sum_{\mathbf{u} \in X_i, \mathbf{v} \in X_j} \frac{I(\|\mathbf{u}-\mathbf{v}\| \le R)}{\|\mathbf{u}-\mathbf{v}\|^{d-1}} \left[\hat{g}_h^{-(k)}(\mathbf{u}, \mathbf{v})\right]^2, \quad \text{and}$$

$$\widehat{M}_2^{(k)}(h) = \frac{1}{m_k} \sum_{i \in \mathcal{S}_k} \sum \sum^{\ne}_{\mathbf{u}, \mathbf{v} \in X_i} \frac{I(\|\mathbf{u}-\mathbf{v}\| \le R)}{\|\mathbf{u}-\mathbf{v}\|^{d-1}} \hat{g}_h^{-(k)}(\mathbf{u}, \mathbf{v}),$$

where m_k is the cardinality of the set $\mathcal{S}_k$, $k = 1, \ldots, n_{cv}$.

Using the above two estimators, the bandwidth h is chosen as

$$\hat{h} = \arg\min_{h>0} \left[\widehat{M}_1(h) - 2\widehat{M}_2(h)\right]. \quad \text{(A.2)}$$

The tuning parameter L for the orthogonal series estimator (2.10) can be chosen in a similar fashion. In our simulation studies and the real data analysis, we use $n_{cv} = 5$.

Appendix B: More on the empirical variance estimator

B.1. Approximation of $\mathbf{Z}_2(\boldsymbol{\theta}^*)$ in equation (3.9)

To estimate $\mathrm{Var}[\mathbf{Z}_1 - \mathbf{Z}_2(\boldsymbol{\theta}^*)]$ it would be helpful to approximate $\mathbf{Z}_1 - \mathbf{Z}_2(\boldsymbol{\theta}^*)$ by a linear combination of independent random variables. To achieve this goal, we view $\mathbf{Z}_2(\boldsymbol{\theta}^*) = \frac{1}{m(m-1)} \sum_{i \neq j=1}^{m} \sum \boldsymbol{\xi}(X_i, X_j)$ as a U-statistic with $\boldsymbol{\xi}(X_i, X_j) = \sum_{\mathbf{u} \in X_i, \mathbf{v} \in X_j} \sum w_{r,h}(\|\mathbf{u} - \mathbf{v}\|) \tilde{g}_{r,h}(\|\mathbf{u} - \mathbf{v}\|; \boldsymbol{\theta}^*) \mathbf{A}_h(\|\mathbf{u} - \mathbf{v}\| - r)$. Following the standard U-statistic theory, we consider the linear projection of $\mathbf{Z}_2(\boldsymbol{\theta}^*) - \mathbb{E}\mathbf{Z}_2(\boldsymbol{\theta}^*)$ defined as $\sum_{i=1}^{m} \mathbb{E}\left[\mathbf{Z}_2(\boldsymbol{\theta}^*) - \mathbb{E}\mathbf{Z}_2(\boldsymbol{\theta}^*) | X_i\right]$, based on which some straightforward calculations yield the approximation

$$\mathbf{Z}_2(\boldsymbol{\theta}^*) \approx \tilde{\mathbf{Z}}_2(\boldsymbol{\theta}^*) = \frac{2}{m} \sum_{i=1}^{m} \sum_{\mathbf{u} \in X_i} q(\mathbf{u}) - \mathbb{E}\mathbf{Z}_2(\boldsymbol{\theta}^*), \tag{A.3}$$

where $q(\mathbf{u}) = \int_{D_n} \lambda(\mathbf{v}) w_{r,h}(\|\mathbf{u} - \mathbf{v}\|) \tilde{g}_{r,h}(\|\mathbf{u} - \mathbf{v}\|; \boldsymbol{\theta}^*) \mathbf{A}_h(\|\mathbf{u} - \mathbf{v}\| - r) \mathrm{d}\mathbf{v}$. Letting $\|\cdot\|_F$ denote the Frobenius norm of a matrix and following similar arguments as in the proof of Lemma S.8 in the Supplementary material, we can show that under conditions C1-C5,

$$\mathrm{Var}\left[\tilde{\mathbf{Z}}_2(\boldsymbol{\theta}^*) - \mathbf{Z}_2(\boldsymbol{\theta}^*)\right] = o\left(\|\mathrm{Var}\left[\mathbf{Z}_1 - \mathbf{Z}_2(\boldsymbol{\theta}^*)\right]\|_F\right), \text{ as } m \to \infty,$$

which implies that $\mathrm{Var}\left[\mathbf{Z}_1 - \mathbf{Z}_2(\boldsymbol{\theta}^*)\right]$ can be well approximated by the covariance matrix of the following random variable

$$\mathbf{Z}_1 - \tilde{\mathbf{Z}}_2(\boldsymbol{\theta}^*) = \frac{1}{m} \sum_{i=1}^{m} \tilde{\xi}(X_i) + \mathbb{E}\mathbf{Z}_2(\boldsymbol{\theta}^*),$$

where $\tilde{\xi}(X_i) = \sum_{\mathbf{u}, \mathbf{v} \in X_i} \sum w_{r,h}(\|\mathbf{u} - \mathbf{v}\|) \mathbf{A}_h(\|\mathbf{u} - \mathbf{v}\| - r) - 2 \sum_{\mathbf{u} \in X_i} q(\mathbf{u})$.

B.2. Empirical variance estimator: orthogonal series estimator

The algorithm proposed in this section follows exactly the same logic as in Section 3.3, hence detailed reasonings are omitted. Define two random vectors

$$\mathbf{Z}_1^L = \frac{1}{m} \sum_{i=1}^{m} \sum_{\mathbf{u}, \mathbf{v} \in X_i}^{\neq} w_R(\|\mathbf{u} - \mathbf{v}\|) \boldsymbol{\phi}_L(\|\mathbf{u} - \mathbf{v}\|), \tag{A.4}$$

$$\mathbf{Z}_2^L(\boldsymbol{\theta}^*) = \sum_{i\neq j=1}^{m}\sum_{\mathbf{u}\in X_i}\sum_{\mathbf{v}\in X_j}\frac{w_R(\|\mathbf{u}-\mathbf{v}\|)}{m(m-1)}\tilde{g}_L(\|\mathbf{u}-\mathbf{v}\|;\boldsymbol{\theta}^*)\boldsymbol{\phi}_L(\|\mathbf{u}-\mathbf{v}\|). \quad \text{(A.5)}$$

Let $\Delta_k, k = 1,\ldots,n_p$ be a partition of the observation domain D_n and define $\mathbf{Y}_{ik} = \sum_{\mathbf{u}\in\Delta_k\cap X_i}\left[\sum_{\mathbf{v}\in X_i} w_R(\|\mathbf{u}-\mathbf{v}\|)\boldsymbol{\phi}_L(\|\mathbf{u}-\mathbf{v}\|-r) - 2\hat{q}_i(\mathbf{u})\right]$ with random variables $\hat{q}_i(\mathbf{u}) \equiv \frac{1}{m-1}\sum_{j\neq i}\sum_{\mathbf{v}\in X_j} w_R(\|\mathbf{u}-\mathbf{v}\|)\hat{g}_L(\|\mathbf{u}-\mathbf{v}\|)\boldsymbol{\phi}_L(\|\mathbf{u}-\mathbf{v}\|)$. If the partition $\{\Delta_k, k=1,\ldots,n_p\}$'s of D_n can be made so that correlations among $\mathbf{Y}_{ik_1}$'s and $\mathbf{Y}_{ik_2}$'s are not too strong for any $k_1 \neq k_2$, an estimator of $\mathrm{Var}\left[\mathbf{Z}_1^L - \mathbf{Z}_2^L(\boldsymbol{\theta}^*)\right]$ can be obtained as

$$\widehat{\mathrm{Var}}\left[\mathbf{Z}_1^L - \mathbf{Z}_2^L(\boldsymbol{\theta}^*)\right] = \frac{1}{m^2}\sum_{k=1}^{n_p}\sum_{i=1}^{m}(\widehat{\mathbf{Y}}_{ik}-\overline{\mathbf{Y}}_k)(\widehat{\mathbf{Y}}_{ik}-\overline{\mathbf{Y}}_k)^T, \text{ and } \overline{\mathbf{Y}}_k = \frac{1}{m}\sum_{i=1}^{m}\widehat{\mathbf{Y}}_{ik}.$$

Once $\widehat{\mathrm{Var}}\left[\mathbf{Z}_1^L - \mathbf{Z}_2^L(\boldsymbol{\theta}^*)\right]$ is obtained, a variance estimator of $\hat{g}_L(\cdot)$ can be defined as

$$\widehat{\mathrm{Var}}[\hat{g}_L(r)] = [\hat{g}_L(r)]^2\boldsymbol{\phi}_L(r)^T\widehat{\mathbf{Q}}_L^{-1}\widehat{\mathrm{Var}}[\mathbf{Z}_1^L - \mathbf{Z}_2^L(\boldsymbol{\theta}^*)]\widehat{\mathbf{Q}}_L^{-1}\boldsymbol{\phi}_L(r), \quad r\in[0,R].$$

Appendix C: Sketches of technical proofs

In this section, we outline the sketches of all technical proofs. Details on the proof can be found in an online Supplementary material.

Proof of Lemma 1. When $\lambda(\mathbf{s}) \equiv \lambda$ and the observation window is $D_n = [0,T_n] \subset \mathbb{R}$, straightforward calculus gives that

$$\mathbf{Q}_{n,h}^{(k)}(r) = \lambda^2\frac{2}{T_n}\int_{-r/h}^{(T_n-r)/h}(T_n - sh - r)\left[K(s)\right]^k\mathbf{A}_1(s)\mathbf{A}_1^T(s)\mathrm{d}s.$$

If the uniform kernel $K(x) = \frac{1}{2}I(-1\leq x\leq 1)$ is used, it can be further simplified as

$$\mathbf{Q}_{n,h}^{(1)}(r) = \mathbf{Q}_{n,h}^{(2)}(r) = \lambda^2\left(1-\frac{r}{T_n}\right)\mathbf{B}_1(r) - \lambda^2\frac{h}{T_n}\mathbf{B}_2(r),$$

where $\mathbf{B}_1(r)$ is a $(p+1)\times(p+1)$ matrix whose (i,j)th element is $\frac{1}{i+j-1}(q_{up}^{i+j-1} - q_{low}^{i+j-1})$ and $\mathbf{B}_2(r)$ is a $(p+1)\times(p+1)$ matrix whose (i,j)th element is $\frac{1}{i+j}(q_{up}^{i+j} - q_{low}^{i+j})$, with $q_{low} = \max(-r/h,-1)$ and $q_{up} = \min\left[(T_n-r)/h,1\right]$. □

Proof of Lemma 3.2. The proof of Lemma 3.2 is carried out by proving the following two technical lemmas. The Lemma A.1 quantifies the magnitude of the estimation bias and Lemma A.2 gives the convergence rate of $\widehat{\boldsymbol{\theta}}$.

Lemma A.1. *Under conditions C1-C5, we have that as* $h\to 0$,

$$h^j\left[\theta_j^* - f^{\{j\}}(r)/j!\right] = O(h^{p+1}), \quad j = 0,1,\ldots,p, \quad \text{(A.6)}$$

$$|g(t) - \tilde{g}_{r,h}(t;\boldsymbol{\theta}^*)| = O(h^{p+1}), \quad \textit{for } t\in[r-h,r+h]. \quad \text{(A.7)}$$

Lemma A.2. *Under conditions C1-C5, we have that as* $m|D_n|h(r+h)^{d-1} \to \infty$ *and* $h \to 0$,

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_h = O_P\left(\frac{1}{\sqrt{m|D_n|h(r+h)^{d-1}}}\right), \tag{A.8}$$

where the norm $\|\mathbf{x}\|_h^2 = x_0^2 + (hx_1)^2 + \cdots + (h^p x_p)^2$ *for any* $\mathbf{x} = (x_0, x_1, \ldots, x_p)^T \in \mathbb{R}^{p+1}$ *and* $\boldsymbol{\theta}^*$ *is defined in equation* (2.6).

Lemma 3.2 follows immediately from Lemmas A.1–A.2, because

$$|\hat{g}_h(r) - g(r)| \le \exp(\theta_0^*)|\exp(\hat{\theta}_0 - \theta_0^*) - 1| + |\exp(\theta_0^*) - g(r)|. \qquad \square$$

Proof of Theorem 3.1. To show Theorem 3.1, we first establish the next three technical lemmas.

Lemma A.3. *Under C1-C5, as* $h \to 0$ *and* $m|D_n|h(r+h)^{d-1} \to \infty$, *we have that,*

$$(m|D_n|h)\mathrm{Var}\left(\mathbf{Z}_1\right) = 2g(r)\mathbf{Q}_{n,h}^{(2)}(r) + O(h(r+h)^{d-1}), \tag{A.9}$$

$$(m|D_n|h)\mathrm{Var}\left[\mathbf{Z}_2(\boldsymbol{\theta}^*)\right] = \frac{2g^2(r)}{m-1}\mathbf{Q}_{n,h}^{(2)}(r) + O(h(r+h)^{d-1}), \tag{A.10}$$

$$(m|D_n|h)\mathrm{Cov}\left[\mathbf{Z}_1, \mathbf{Z}_2(\boldsymbol{\theta}^*)\right] = O(h(r+h)^{d-1}), \tag{A.11}$$

where $\mathbf{Z}_1$ *and* $\mathbf{Z}_2$ *are defined in* (3.8) *and* (3.9), *and* $\mathbf{Q}_{n,h}^{(2)}(r)$ *is as defined in equation* (3.1) *and the convergence is entry-wise.*

Lemma A.4. *Under C1-C5 and N1-N2, we have that, as* $h \to 0$ *and* $m|D_n|h(r+h)^{d-1} \to \infty$,

$$\sqrt{m|D_n|h}\boldsymbol{\Sigma}_Z^{-1/2}(\boldsymbol{\theta}^*)\left[\mathbf{Z}_1 - \mathbf{Z}_2(\boldsymbol{\theta}^*)\right] \xrightarrow{\mathcal{D}} N\left(\mathbf{0}, \mathbf{I}\right), \tag{A.12}$$

where $\boldsymbol{\Sigma}_Z(\boldsymbol{\theta}^*) = 2(m-1+g(r))/(m-1)g(r)\mathbf{Q}_{n,h}^{(2)}(r)$ *with* $\mathbf{Q}_{n,h}^{(2)}(r)$ *defined in equation (11).*

Lemma A.5. *Denote* $\widehat{\boldsymbol{\theta}}$ *as the solution to estimating equations (6), then under conditions C1-C5, N1-N2, we have that, as* $h \to 0$ *and* $m|D_n|h(r+h)^{d-1} \to \infty$,

$$\begin{bmatrix} (\hat{\theta}_0 - \theta_0^*) \\ h(\hat{\theta}_1 - \theta_1^*) \\ \vdots \\ h^p(\hat{\theta}_p - \theta_p^*) \end{bmatrix} = \frac{\left[\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1}}{g(r)}\left[\mathbf{Z}_1 - \mathbf{Z}_2(\boldsymbol{\theta}^*) + o_p\left(\sqrt{\frac{(r+h)^{1-d}}{m|D_n|h}}\right)\right], \tag{A.13}$$

where $\mathbf{Z}_1$ *and* $\mathbf{Z}_2(\boldsymbol{\theta}^*)$ *are defined in* (3.8) *and* (3.9), *respectively.*

By applying the delta method to $\hat{g}_h(r) = \exp(\hat{\theta}_0) = \exp(\boldsymbol{e}^T\widehat{\boldsymbol{\theta}})$, where $\boldsymbol{e} = (1, 0, \ldots, 0)^T$, with Lemmas A.4 and A.5, we have that

$$\frac{\sqrt{m|D_n|h}\,[\hat{g}_h(r) - \exp(\theta_0^*)]}{\exp(\theta_0^*)\sqrt{\frac{2(m-1+g(r))}{(m-1)g(r)}\boldsymbol{e}^T\left[\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1}\mathbf{Q}_{n,h}^{(2)}(r)\left[\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1}\boldsymbol{e}}} \xrightarrow{\mathcal{D}} N(0,1).$$

By Lemma A.1, one can readily show that $\exp(\theta_0^*) - g(r) = O(h^{p+1})$, which further implies

$$\begin{aligned}
&\frac{\sqrt{m|D_n|h}\,[\hat{g}_h(r) - g(r)]}{\exp(\theta_0^*)\sqrt{\frac{2(m-1+g(r))}{(m-1)g(r)}\boldsymbol{e}^T\left[\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1}\mathbf{Q}_{n,h}^{(2)}(r)\left[\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1}\boldsymbol{e}}} \\
&= \frac{\sqrt{m|D_n|h}\,[\hat{g}_h(r) - \exp(\theta_0^*) + O(h^{p+1})]}{\sqrt{\frac{2(m-1+g(r))}{(m-1)g(r)}\boldsymbol{e}^T\left[\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1}\mathbf{Q}_{n,h}^{(2)}(r)\left[\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1}\boldsymbol{e}}} + o_p(1),
\end{aligned}$$

which completes the proof. □

Proof of Lemma 3.3. The proof of Lemma 3.3 relies on the next two technical Lemmas. The first Lemma A.6 gives the estimation bias by quantifying the distance between $g(r)$ and $\tilde{g}_L(r;\boldsymbol{\theta}^*)$. And the Lemma A.7 gives the convergence rate of $\widehat{\boldsymbol{\theta}}$.

Lemma A.6. *Under conditions C1–C3 and C4'–C5', we have that as* $L \to \infty$,

$$\begin{aligned}
\|\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*\| &= O(L^{\nu_0-\nu_1}), &\text{(A.14)}\\
\sup_{0<r<R}|g(r) - \tilde{g}_L(r;\boldsymbol{\theta}^*)| &= O\left(L^{-\nu_1+\max\{\tau_1,\nu_0+\nu_2\}}\right) = o(1), &\text{(A.15)}\\
\sup_{0<r<R}|\tilde{g}_L(r;\boldsymbol{\theta}^*)| &= O(1), &\text{(A.16)}
\end{aligned}$$

where ν_0, ν_1, τ_1 *and* ν_2 *are defined in conditions C4' and C5'.*

Lemma A.7. *Under conditions C1-C3, and C4'-C5', as* $L \to \infty$, *we have that,* $\frac{L^{4\nu_0+2\nu_2}}{m|D_n|} \to 0$,

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| = O_p\left(L^{\nu_0}/\sqrt{m|D_n|}\right), \tag{A.17}$$

where $\boldsymbol{\theta}^*$ *is defined in equation* (3.7).

To show (3.6), using equation (A.15)-(A.16), we have that

$$\begin{aligned}\left|g(r)-\tilde{g}_L(r;\widehat{\boldsymbol{\theta}})\right| &\le |g(r)-\tilde{g}_L(r;\boldsymbol{\theta}^*)| + \left|\tilde{g}_L(r;\boldsymbol{\theta}^*)-\tilde{g}_L(r;\widehat{\boldsymbol{\theta}})\right| \\ &= O\left(L^{-\nu_1+\max\{\tau_1,\nu_0+\nu_2\}}\right) + \tilde{g}_L(r;\boldsymbol{\theta}^*)\left|1-\exp\left[\sum_{l=1}^{L}(\hat{\theta}_l-\theta_l^*)\phi_l(r)\right]\right| \\ &= O\left(L^{-\nu_1+\max\{\tau_1,\nu_0+\nu_2\}}\right) + \tilde{g}_L(r;\boldsymbol{\theta}^*)O\left(\sup_{0<r\le R}\|\boldsymbol{\phi}_L(r)\|\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|\right) \\ &= O\left(L^{-\nu_1+\max\{\tau_1,\nu_0+\nu_2\}}\right) + O_p\left(\frac{L^{\nu_0+\nu_2}}{\sqrt{m|D_n|}}\right) = o(1),\end{aligned}$$

where the upper bounds do not depend on r, which completes the proof. □

Proof of Theorem 3.2. The proof of Theorem 3.2 relies on the next two technical Lemmas.

Lemma A.8. *Let* $\tilde{\sigma}^2_{\boldsymbol{\delta}}(\boldsymbol{\theta}^*) = \boldsymbol{\delta}_L^T\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\boldsymbol{\delta}_L^T$ *with* $\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*) = \mathrm{Var}\left[\sqrt{m|D_n|}\tilde{\mathbf{U}}(\boldsymbol{\theta}^*)\right]$. *If the vector* $\boldsymbol{\delta}_L$ *satisfies (a)* $\|\boldsymbol{\delta}_L\| = 1$*; (b)* $\int_0^R\left[w_o(s)|\boldsymbol{\delta}_L^T\boldsymbol{\phi}_L(s)|\right]^{2+\lceil\delta\rceil}\mathrm{d}s = O(1)$*; and (c)* $\tilde{\sigma}^2_{\boldsymbol{\delta}}(\boldsymbol{\theta}^*) \ge c_u$ *for some constant* $c_u > 0$*, then under conditions C1-C3, C4'-C5' and N1-N2, we have that, as* $L\to\infty$ *and* $m|D_n|\to\infty$*,*

$$\frac{\sqrt{m|D_n|}\boldsymbol{\delta}_L^T\tilde{\mathbf{U}}(\boldsymbol{\theta}^*)}{\tilde{\sigma}_{\boldsymbol{\delta}}(\boldsymbol{\theta}^*)} \xrightarrow{\mathcal{D}} N(0,1). \tag{A.18}$$

Lemma A.9. *Denote* $\widehat{\boldsymbol{\theta}}$ *as the solution to* $\tilde{\mathbf{U}}_L(\boldsymbol{\theta}) = \mathbf{0}$*, then under conditions C1-C3 and C4'-C5', we have that as* $L\to\infty$ *and* $L^{4\nu_0+2\nu_2}/m|D_n|\to 0$*, for any* $0 < r \le R$*,*

$$\begin{aligned}&\sqrt{m|D_n|}\boldsymbol{\phi}_L^T(r)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*) \\ &\qquad= \sqrt{m|D_n|}\boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\tilde{\mathbf{U}}_L(\boldsymbol{\theta}^*) + o_p(1)\left\|\boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\right\|,\end{aligned} \tag{A.19}$$

where $\boldsymbol{\theta}^*$ *and* $\mathbf{Q}_L$ *are defined in (3.7) and (14), respectively. Furthermore, under additional conditions N1-N3, we have that*

$$\frac{\sqrt{m|D_n|}\boldsymbol{\phi}_L^T(r)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)}{\sigma_L(r;\boldsymbol{\theta}^*)} \xrightarrow{\mathcal{D}} N(0,1), \tag{A.20}$$

where $\sigma_L^2(r;\boldsymbol{\theta}^*) = \boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\phi}_L(r)$ *and the covariance matrix* $\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*) = \mathrm{Var}\left[\sqrt{m|D_n|}\tilde{\mathbf{U}}_L(\boldsymbol{\theta}^*)\right]$.

Recall the definition $\tilde{g}_L(r;\boldsymbol{\theta}) = \exp\left[\boldsymbol{\theta}^T\boldsymbol{\phi}_L(r)\right]$ and $\hat{g}_L(r) = \tilde{g}_L(r;\widehat{\boldsymbol{\theta}})$, we apply the delta method to the asymptotic distribution of $\sqrt{m|D_n|}\boldsymbol{\phi}_L^T(r)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)$

from A.9, we have that

$$\frac{\sqrt{m|D_n|}\left[\tilde{g}_L(r;\widehat{\boldsymbol{\theta}})-\tilde{g}_L(r;\boldsymbol{\theta}^*)\right]}{\tilde{g}_L(r;\boldsymbol{\theta}^*)\sqrt{\boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\phi}_L(r)}}\xrightarrow{\mathcal{D}}N(0,1).$$

By equation (A.15) in Lemma A.6, one has that $\sup_{0<r<R}|g(r)-\tilde{g}_L(r;\boldsymbol{\theta}^*)| = O(L^{-\nu_1+\tau_1}+L^{\nu_0-\nu_1+\nu_2}) = o(1)$, it readily follows that

$$\begin{aligned}
&\frac{\sqrt{m|D_n|}\left[\tilde{g}_L(r;\widehat{\boldsymbol{\theta}})-\tilde{g}_L(r;\boldsymbol{\theta}^*)\right]}{\tilde{g}_L(r;\boldsymbol{\theta}^*)\sqrt{\boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\phi}_L(r)}}\\
&=\frac{\sqrt{m|D_n|}\left[\tilde{g}_L(r;\widehat{\boldsymbol{\theta}})-g(r)+g(r)-\tilde{g}_L(r;\boldsymbol{\theta}^*)\right]}{\tilde{g}_L(r;\boldsymbol{\theta}^*)\sqrt{\boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\phi}_L(r)}}\\
&=\frac{\sqrt{m|D_n|}\left[\tilde{g}_L(r;\widehat{\boldsymbol{\theta}})-g(r)+O(L^{-\nu_1+\tau_1}+L^{\nu_0-\nu_1+\nu_2})\right]}{g(r)\sqrt{\boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\phi}_L(r)}}+o_P(1)\\
&\xrightarrow{\mathcal{D}}N(0,1),
\end{aligned}$$

which completes the proof. □

Acknowledgment

Xu's research is supported by NSF grant SES-1902195. Rasmus Waagepetersen was 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. Zhang's research is supported by NSF grant DMS-2015190 and Yongtao Guan is supported by NSF grant DMS-1810591.

Supplementary Material

Supplement Material for "Nonparametric Estimation of the Pair Correlation Function of Replicated Inhomogeneous Point processes"
(doi: 10.1214/20-EJS1755SUPP; .pdf). Relevant R code and detailed technical proofs can be found in an online supplementary material available on Rasmus Waagepetersen's home page.

References

Baddeley, A. J., Møller, J., and Waagepetersen, R. (2000), "Non- and semi-parametric estimation of interaction in inhomogeneous point patterns," *Statistica Neerlandica*, 54, 329–350. MR1804002

Baddeley, A. J., Moyeed, R. A., Howard, C. V., and Boyde, A. (1993), "Analysis of a three-dimensional point pattern with replication," *Applied Statistics*, 42, 641–668. MR1234146

Besag, J. (1977), "Contribution to the discussion on Dr Ripley's paper," *Journal of the Royal Statistical Society. Series B (Methodological)*, 39, 193–195. MR0488279

Biscio, C. A. N. and Waagepetersen, R. (2019), "A general central limit theorem and a subsampling variance estimator for α-mixing point processes," *Scandinavian Journal of Statistics*, 46, 1168–1190. MR4033808

Bouzas, P. R., Valderrama, M. J., Aguilera, A. M., and Ruiz-Fuentes, N. (2006), "Modelling the mean of a doubly stochastic Poisson process by functional data analysis," *Computational Statistics & Data Analysis*, 50, 2655–2667. MR2227341

Coeurjolly, J.-F. and Møller, J. (2014), "Variational approach for spatial point process intensity estimation," *Bernoulli*, 20, 1097–1125. MR3217439

Coeurjolly, J.-F., Møller, J., and Waagepetersen, R. (2017), "A tutorial on Palm distributions for spatial point processes," *International Statistical Review*, 85, 404–420. MR3723609

Cook, R. J. and Lawless, J. (2007), *The Statistical Analysis of Recurrent Events*, Springer Science & Business Media. MR3822124

Diggle, P. J., Lange, N., and Beneš, F. M. (1991), "Analysis of variance for replicated spatial point patterns in clinical neuroanatomy," *Journal of the American Statistical Association*, 86, 618–625. MR0743593

Dvořák, J. and Prokešová, M. (2016), "Asymptotic properties of the minimum contrast estimators for projections of inhomogeneous space-time shot-noise Cox processes," *Applications of Mathematics*, 61, 387–411. MR3532250

Efromovich, S. (2010), "Orthogonal series density estimation," *Wiley Interdisciplinary Reviews: Computational Statistics*, 2, 467–476.

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66*, vol. 66, CRC Press. MR1383587

Gervini, D. (2016), "Independent component models for replicated point processes," *Spatial Statistics*, 18, 474–488. MR3575503

Gervini, D. (2017), "Multiplicative component models for replicated point processes," *arXiv preprint* arXiv:1705.09693. MR3575503

Guan, Y. (2006), "A composite likelihood approach in fitting spatial point process models," *Journal of the American Statistical Association*, 101, 1502–1512. MR2279475

Guan, Y. (2007), "A least-squares cross-validation bandwidth selection approach in pair correlation function estimations," *Statistics & Probability Letters*, 77, 1722–1729. MR2394568

Hall, P. (1987), "Cross-validation and the smoothing of orthogonal series density estimators," *Journal of Multivariate Analysis*, 21, 189–206. MR0884096

Heinrich, L. (1988), "Asymptotic Gaussianity of some estimators for reduced factorial moment measures and product densities of stationary Poisson cluster processes," *Statistics*, 19, 77–86. MR0921628

Heinrich, L. and Klein, S. (2014), "Central limit theorems for empirical product densities of stationary point processes," *Statistical Inference for Stochastic Processes*, 17, 121–138. MR3219525

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008), *Statistical Analysis and Modelling of Spatial Point Patterns*, vol. 76, Wiley. MR2384630

Jalilian, A., Guan, Y., and Waagepetersen, R. (2013), "Decomposition of variance for spatial Cox processes," *Scandinavian Journal of Statistics*, 40, 119–137. MR3024035

Jalilian, A., Guan, Y., and Waagepetersen, R. (2019), "Orthogonal series estimation of the pair correlation function of a spatial point process," *Statistica Sinica*, 29, 769–787. MR3931387

Møller, J. and Waagepetersen, R. P. (2003), *Statistical Inference and Simulation for Spatial Point Processes*, Chapman and Hall/CRC, Boca Raton. MR2004226

Ripley, B. D. (1976), "The second-order analysis of stationary point processes," *Journal of Applied Probability*, 13, 255–266. MR0402918

Stoyan, D. and Stoyan, H. (1994), *Fractals, Random Shapes, and Point Fields: Methods of Geometrical Statistics*, vol. 302, John Wiley & Sons Inc. MR1297125

Tolstov, G. (1962), *Fourier Series*, Dover Publications, New York. MR0425474

Waagepetersen, R. P. (2007), "An estimating function approach to inference for inhomogeneous Neyman–Scott processes," *Biometrics*, 63, 252–258. MR2345595

Wu, S., Müller, H.-G., and Zhang, Z. (2013), "Functional data analysis for point processes with rare events," *Statistica Sinica*, 23, 1–23. MR3076156

Xu, G., Waagepetersen, R., and Guan, Y. (2019), "Stochastic quasi-likelihood for case-control point pattern data," *Journal of the American Statistical Association*, 114, 631–644. MR3963168

Xu, G., Wang, M., Bian, J., Burch, T. R., Andrade, S. C., Huang, H., Zhang, J., and Guan, Y. (2020), "Semi-parametric learning of structured temporal point processes," *Journal of Machine Learning Research*, in press.

Xu, G., Zhao, C., Jalilian, A., Waagepetersen, R., Zhang, J. and Guan, Y. (2020). Supplement to "Nonparametric estimation of the pair correlation function of replicated inhomogeneous point processes" DOI: 10.1214/20-EJS1755SUPP.

Yue, Y. R. and Loh, J. M. (2010), "Bayesian nonparametric estimation of pair correlation function for inhomogeneous spatial point processes," *Journal of Nonparametric Statistics*, 25, 463–474. MR3056096