Efficient, Cross-Fitting Estimation of Semiparametric Spatial Point Processes

Xindi Lin
Department of Statistics, University of Wisconsin-Madison
and
Hyunseung Kang
Department of Statistics, University of Wisconsin-Madison

August 4, 2024

Abstract

The paper studies the efficient estimation of a class of models called semiparametric spatial point processes where the first-order intensity function contains both a parametric and a nonparametric component. We generalize the notion of semiparametric efficiency lower bound to spatial point processes and propose a novel, spatial cross-fitting estimator based on thinning, a common technique used in simulating markov chain monte carlo. The proposed estimator is shown to be consistent and in many cases, asymptotically Normal. The estimator also achieves the semiparametric efficiency lower bound if the process is Poisson. We conclude with a demonstration of the proposed method in a small simulation study and a re-analysis of the spatial distribution of rainforest trees.

Keywords: Inhomogeneous point processes, partially linear models, semiparametric efficiency lower bound, thinning.

1 Introduction

Parametric spatial point processes are widely to measure the relationship between spatial distribution of events, for instance, the location of trees, violent crime, gene expressions in a cell, stars in a galaxy, and relevant spatial covariates, such as elevation, socioeconomic characteristics of neighborhoods, cellular environment, and physical features of the early universe (e.g., Gatrell et al. [1996], Haberman [2017], Rao et al. [2021]). Specifically, the intensity function $\lambda(\cdot)$, which governs the spatial distribution of events, is a parametric function of the spatial covariates, for example, $\lambda(\mathbf{u};\boldsymbol{\theta}) = \exp(\boldsymbol{\theta}^\top \mathbf{y}(\mathbf{u}))$, where $\mathbf{u}$ is the spatial location, $\mathbf{y}(\mathbf{u})$ are the values of the spatial covariates at location $\mathbf{u}$, and $\boldsymbol{\theta}$ is the parameter of interest. Rathbun and Cressie [1994] showed that the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$ is consistent, asymptotically Normal, and statistically efficient when the point process is Poisson, i.e., the number of events observed in non-overlapping regions of an observation window is independent. Later, Schoenberg [2005], Guan and Loh [2007], and Waagepetersen and Guan [2009] showed that the MLE in Rathbun and Cressie [1994] remains consistent and asymptotically Normal even if the process is not Poisson, but with a loss in statistical efficiency. Also, Berman and Turner [1992], Baddeley and Turner [2000], and Baddeley et al. [2014] proposed computationally efficient approaches to compute the MLE. These methods for parametric spatial point processes are widely accessible with the R package `spatstat` [Baddeley and Turner, 2005].

The main theme of the paper is to propose a more general class of models that study the same scientific questions as above (i.e., the relationship between spatial events and spatial covariates), but with less restrictive modeling assumptions and critically, without sacrificing the aforementioned statistical and computational properties (i.e., consistency, asymptotic Normality, statistical and computational efficiency). Formally, suppose we partition the

set of spatial covariates into two types, $\mathbf{y}(\mathbf{u})$ and $\mathbf{z}(\mathbf{u})$, and the intensity function takes the form[1]:

$$\lambda(\mathbf{u};\boldsymbol{\theta},\eta) = \lambda(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u})),\eta(\mathbf{z}(\mathbf{u}))), \quad \boldsymbol{\theta} \in \Theta,\ \eta \in \mathcal{H}. \tag{1}$$

The term $\tau_{\boldsymbol{\theta}}(\cdot)$ is an user-specified function of $\mathbf{y}(\mathbf{u})$ and is parameterized by the finite-dimensional parameter $\boldsymbol{\theta}$ in some Euclidean space Θ, for instance $\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u})) = \boldsymbol{\theta}^\top \mathbf{y}(\mathbf{u})$ and $\boldsymbol{\theta}$ is a vector of real numbers. The term $\eta(\cdot)$ is an unknown, infinite-dimensional function of $\mathbf{z}(\mathbf{u})$ in a linear space $\mathcal{H}$, for instance the space of smooth functions in a Hilbert space. The class of spatial point processes with an intensity function in (1), which we refer to as *a semiparametric spatial point process*, generalizes or includes many existing models, including: (a) the parametric spatial point process mentioned above, (b) spatial point processes of the form $\lambda(\mathbf{u}) = \lambda_0(\mathbf{u})\exp(\boldsymbol{\theta}^\top \mathbf{y}(\mathbf{u}))$ where $\lambda_0(\mathbf{u})$ is a nonparametric, "baseline" rate/intensity function (e.g., Diggle [1990], Chu et al. [2022], Hessellund et al. [2022]), and (c) recently, spatial point process models for causal inference where $\mathbf{y}(\mathbf{u})$ represents a spatial treatment/exposure variable and $\mathbf{z}(\mathbf{u})$ represents measured, spatial confounders (e.g., Papadogeorgou et al. [2022], Jiang et al. [2023]). More broadly, model (1) generalizes partially linear models in i.i.d. settings (e.g., Robinson [1988], Severini and Staniswalis [1994], Härdle et al. [2000]) and similar to the motivation behind such models, (1) is ideal to address questions about estimating and testing the partial effects of a small set of spatial covariates $\mathbf{y}(\mathbf{u})$ on the spatial distribution of events without imposing parametric assumptions on all available spatial covariates. We remark that there is considerable work on nonparametric estimation of λ (e.g., Berman and Diggle [1989], van Lieshout [2012], Baddeley et al. [2012], Tang and Li [2023]). But, these methods generally do not enable statistical inference in the form of hypothesis testing or constructing confidence intervals

[1]For notational brevity, we overload the notation $\lambda(\cdot)$ on the right-hand side of equation (1). We only do this for Section 1; the rest of the paper uses $\lambda(\mathbf{u})$ as defined on the left-hand side of equation (1).

based on asymptotically Normal limits.

Under model (1), we propose a novel estimator of $\boldsymbol{\theta}$ that is consistent, asymptotically Normal, and in some cases, semiparametrically efficient. Our estimator is inspired by three ideas: (a) cross-fitting [Chernozhukov et al., 2018] in non-spatial, i.i.d. settings, (b) thinning, a common technique to simulate point processes, and (c) the pseudo-likelihood method of Waagepetersen and Guan [2009]. Specifically, we use the independent random thinning method described in Chiu et al. [2013], as a generalization of sample splitting to "correctly" split spatial point processes into subprocesses. Computationally, we extend the numerical approximation techniques in Berman and Turner [1992] and Baddeley et al. [2014] and show that computing our estimator is equivalent to estimating a finite-dimensional parameter in a generalized partial linear regression with carefully chosen weights, for which existing software is available (e.g., the R package `mgcv`).

To the best of our knowledge, we are the first to develop a framework to study semiparametric efficiency of spatial point processes where the data is not i.i.d. and the intensity function is not parametric. Our results extends the classical result in spatial point processes about MLEs [Rathbun and Cressie, 1994] and the results for partially linear models in i.i.d. settings [Robinson, 1988, Severini and Wong, 1992]. We remark that our theoretical analysis uses empirical process theory to study spatial point processes, which may be of independent theoretical interest. We are also unaware of existing works that present a single, universal recipe to construct consistent and asymptotically Normal estimator of $\boldsymbol{\theta}$ for spatial point processes where the intensity function can be written as (1); note that this recipe can lead to a semiparametrically efficient estimator under additional assumptions. Finally, unlike the classical i.i.d. setting in Robinson [1988], our paper also highlights some challenges of constructing asymptotically Normal estimators under spatial dependence, some

Point Process	Log-Linear $\lambda(\cdot)$	Properties of Estimator $\hat{\boldsymbol{\theta}}$			
		Consistent?	Asymptotically Normal?	Consistent SE Estimator?	Efficient?
Poisson	Yes	✓	✓	✓	✓
	No	✓	✓	✓	✓
Non-Poisson	Yes	✓	✓	✓	?
	No	✓	?	?	?

Table 1: Summary of Our Contributions for Semiparametric Point Processes (in ✓) and Direction for Future Work (in ?). SE stands for standard error and $\lambda(\cdot)$ represents the first-order intensity function.

of which leave as future work. A summary of our contributions and future work is in Table 1.

2 Semiparametric Spatial Point Processes

2.1 Setup

Let X denote a spatial point process in $\mathbb{R}^2$ and d$\mathbf{u}$ denote the Lebesgue measure of a point $\mathbf{u} \in \mathbb{R}^2$. For a bounded subset $A \subset \mathbb{R}^2$, the first-order intensity function, often referred to as "the" intensity function, is a function $\lambda : \mathbb{R}^2 \to [0, \infty)$ where $\mathbb{E}[\sum_{\mathbf{u} \in X} I(\mathbf{u} \in A)] = \int_A \lambda(\mathbf{u}) \mathrm{d}\mathbf{u}$ and $I(\cdot)$ is an indicator function. Also, for two bounded subsets $A, B \subset \mathbb{R}^2$, the second-order intensity function is a function $\lambda_2 : A \otimes B \to [0, \infty)$ where $\mathbb{E}[\sum_{\mathbf{u} \in X, \mathbf{v} \in X}^{\mathbf{u} \neq \mathbf{v}} I(\mathbf{u} \in A, \mathbf{v} \in B)] = \int_A \int_B \lambda_2(\mathbf{u}, \mathbf{v}) \mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}$. The pair correlation function (PCF) of X is defined as the normalized second-order intensity function, i.e., $g(\mathbf{u}, \mathbf{v}) = \lambda_2(\mathbf{u}, \mathbf{v}) / \lambda(\mathbf{u})\lambda(\mathbf{v})$. In this paper, we make a common assumption that the PCF of a process X is reweighted, isotropic, and parametric (i.e., $g(\mathbf{u}, \mathbf{v}; \psi) = g(0, \mathbf{u} - \mathbf{v}; \psi)$ for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ and ψ is a finite-dimensional parameter); see Waagepetersen and Guan [2009], Guan et al. [2015], Hessellund et al. [2022] for details.

2.2 Review: Parametric Spatial Point Processes

We briefly review some key results from the literature on parametric spatial point processes. Consider a spatial point process X with a parametric intensity function $\lambda(\mathbf{u};\boldsymbol{\theta})$ that is parametrized by the parameter $\boldsymbol{\theta} \in \Theta$. From Waagepetersen and Guan [2009], the corresponding pseudo-log-likelihood function of X is

$$\ell(\boldsymbol{\theta}) = \sum_{\mathbf{u}\in X\cap A} \log \lambda(\mathbf{u};\boldsymbol{\theta}) - \int_A \lambda(\mathbf{u};\boldsymbol{\theta})\mathrm{d}\mathbf{u} \tag{2}$$

Also, under mild conditions, Waagepetersen and Guan [2009] showed that the solution to the score equation, i.e., $0 = \frac{\partial}{\partial\boldsymbol{\theta}}\ell(\hat{\boldsymbol{\theta}})$, is an asymptotically Normal estimator of $\boldsymbol{\theta}$ with an asymptotic covariance $\boldsymbol{S}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}(\boldsymbol{\theta},g)\boldsymbol{S}^{-1}(\boldsymbol{\theta})$ where

$$\boldsymbol{S}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2}{\partial\boldsymbol{\theta}^\top\partial\boldsymbol{\theta}}\ell(\boldsymbol{\theta})\right] = \int_A \left(\frac{\partial}{\partial\boldsymbol{\theta}}\log\lambda(\mathbf{u};\boldsymbol{\theta})\right)^{\otimes 2}\lambda(\mathbf{u};\boldsymbol{\theta})\mathrm{d}\mathbf{u}, \tag{3}$$

$$\boldsymbol{\Sigma}(\boldsymbol{\theta},g) = \mathrm{Var}\left[\frac{\partial}{\partial\boldsymbol{\theta}}\ell(\boldsymbol{\theta})\right] = \mathbf{S}(\boldsymbol{\theta}) + \int_A\int_A \frac{\partial}{\partial\boldsymbol{\theta}^\top}\log\lambda(\mathbf{u};\boldsymbol{\theta})\frac{\partial}{\partial\boldsymbol{\theta}}\log\lambda(\mathbf{v};\boldsymbol{\theta})(g(\mathbf{u},\mathbf{v})-1)\lambda(\mathbf{u};\boldsymbol{\theta})\lambda(\mathbf{v};\boldsymbol{\theta})\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}. \tag{4}$$

The term $\mathbf{x}^{\otimes 2}$ represents the outer product of $\mathbf{x}$, i.e., $\mathbf{x}^{\otimes 2} = \mathbf{x}\mathbf{x}^\top$. In the literature, $\boldsymbol{S}(\boldsymbol{\theta})$ is called the sensitivity matrix and $\boldsymbol{\Sigma}(\boldsymbol{\theta},g)$ is the covariance matrix of the score function. We remark that the covariance matrix depends on the PCF g.

We mention some important special cases of the above result. First, if the intensity function is log-linear with respect to $y(\mathbf{u})$, i.e., $\lambda(\mathbf{u};\boldsymbol{\theta}) = \exp(\boldsymbol{\theta}^\top\mathbf{y}(\mathbf{u})))$, the sensitivity matrix $\boldsymbol{S}(\boldsymbol{\theta})$ and the covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta},g)$ simplify to:

$$\boldsymbol{S}(\boldsymbol{\theta}) = \int_A \mathbf{y}(\mathbf{u})\mathbf{y}(\mathbf{u})^\top\lambda(\mathbf{u};\boldsymbol{\theta})\mathrm{d}\mathbf{u},\ \boldsymbol{\Sigma}(\boldsymbol{\theta},g) = \mathbf{S}(\boldsymbol{\theta}) + \int_A\int_A \mathbf{y}(\mathbf{u})\mathbf{y}(\mathbf{v})^\top(g(\mathbf{u},\mathbf{v})-1)\lambda(\mathbf{u};\boldsymbol{\theta})\lambda(\mathbf{v};\boldsymbol{\theta})\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}$$

Second, if X is a Poisson spatial point process, $\boldsymbol{\Sigma}(\boldsymbol{\theta}, g) = \mathbf{S}(\boldsymbol{\theta})$ and $\hat{\boldsymbol{\theta}}$ is asymptotically efficient [Rathbun and Cressie, 1994] with an efficiency lower bound given by $\boldsymbol{S}^{-1}(\boldsymbol{\theta})$ (i.e., the Cramér-Rao bound).

Computationally, the score equation cannot be solved directly because of the integral inside the log-likelihood function in Equation 2 and numerical approximation is necessary. Here, we discuss the quadrature method [Berman and Turner, 1992], which we also utilize in our own work. First, a homogeneous Poisson spatial point process on A is generated and combined with the observed spatial points to form the set of quadrature points $\{\mathbf{u}_j\}_{j=1}^m$; typically, a larger m yields more accurate approximations [Baddeley and Turner, 2000]. Second, for each quadrature point, a quadrature weight $\{w_j\}_{j=1}^m$ is computed based on a Dirichlet tessellation [Berman and Diggle, 1989], which involves partitioning a plane into regions close to each point. Then the pseudo-log-likelihood in (2) is approximated as:

$$\ell(\boldsymbol{\theta}) \approx \sum_{\mathbf{u} \in X \cap A} \log \lambda(\mathbf{u}; \boldsymbol{\theta}) - \sum_{j=1}^{m} w_j \lambda(\mathbf{u}_j; \boldsymbol{\theta}) = \sum_{j=1}^{m} w_j (y_j \log \lambda_j - \lambda_j) \tag{5}$$
$$\lambda_j = \lambda(\mathbf{u}_j; \boldsymbol{\theta}),\ y_j = \frac{I(\mathbf{u}_j \in X \cap A)}{w_j}$$

The right-hand side of equation (5) is equivalent to the weighted log-likelihood of independent Poisson random variables $Y_j \sim \text{Poisson}(\lambda_j)$ with weights w_j. This implies that the score equation that solves for $\hat{\boldsymbol{\theta}}$ can be computed with the familiar `glm` function in `R`. Similarly, $\boldsymbol{S}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta}, g)$ cannot be computed directly and can be approximated with quadrature points by

$$\boldsymbol{S}(\boldsymbol{\theta}) \approx \sum_{j=1}^{m} w_j \mathbf{y}(\mathbf{u}_j)^{\otimes 2} \lambda(\mathbf{u}_j; \boldsymbol{\theta}), \tag{6}$$

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}, g) \approx \sum_{j=1}^{m} w_j \mathbf{y}(\mathbf{u}_j)^{\otimes 2} \lambda(\mathbf{u}_j; \boldsymbol{\theta}) + \sum_{i=1}^{m} \sum_{j=1}^{m} w_i w_j \mathbf{y}(\mathbf{u}_i)\mathbf{y}(\mathbf{u}_j)^\top \lambda(\mathbf{u}_i; \boldsymbol{\theta})\lambda(\mathbf{u}_j; \boldsymbol{\theta})(g(\mathbf{u}_i, \mathbf{u}_j) - 1). \tag{7}$$

In practice, $\boldsymbol{\theta}$ and g in equations (6) and (7) are replaced by $\hat{\boldsymbol{\theta}}$ and an estimator of ψ that parametrizes g. A common approach to estimate ψ is using the K-function, which measures the expected number of points within a distance r of another point; for the interested reader, see Waagepetersen and Guan [2009] and Chapter 7 of Baddeley et al. [2015] for details. To estimate g, if the PCF is assumed to be parametric where $g(\mathbf{u}, \mathbf{v}; \psi)$ and ψ is the parameter that parameterizes the PCF [Waagepetersen and Guan, 2009], an estimate of ψ is obtained with the K-function, which measures the expected number of points observed within a distance r of another point; see Chapter 7 of Baddeley et al. [2015] for more details. Specifically, for some r_1, r_2 where $0 < r_1 < r_2$, Waagepetersen and Guan [2009] proposed the following estimator of ψ:

$$\hat{\psi} = \arg\min_{\psi \in \mathbb{R}} \int_{r_1}^{r_2} \left(K(r; \psi) - \hat{K}(r) \right)^2 dr, \quad \hat{K}(r) = \sum_{\mathbf{u}, \mathbf{v} \in X \cap A}^{\neq} \frac{I(\|\mathbf{u} - \mathbf{v}\| \leq r)}{\lambda(\mathbf{u})\lambda(\mathbf{u})|A|} \tag{8}$$

Here, $\hat{K}$ is the empirical K function defined with the ℓ_2 norm $\|\cdot\|$, $|A|$ is the size of the observation window, $K(r; \psi)$ is the true K function parameterized by ψ , and $\sum^{\neq}$ is the sum over different points.

2.3 Semiparametric Spatial Point Process

We are now ready to formally define a semiparametric spatial point process. Consider a spatial point process X in $\mathbb{R}^2$. For every spatial point $\mathbf{u} \in \mathbb{R}^2$, consider two sets of bounded, spatial covariates, $\mathbf{y}(\mathbf{u}) \in \mathcal{Y} \subset \mathbb{R}^p$ and $\mathbf{z}(\mathbf{u}) \in \mathcal{Z} \subset \mathbb{R}^q$. We consider a process X with an intensity function in equation (1) where the target parameter of interest is $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^k$

in a compact set Θ and $\eta(\cdot) \in \mathcal{H}$ is an infinite-dimensional nuisance parameter. For this reason, we refer to $\mathbf{y}(\mathbf{u})$ as *target* covariates and $\mathbf{z}(\mathbf{u})$ as *nuisance* covariates. Let $\boldsymbol{\theta}^*$, η^*, and ψ^* denote the true values of $\boldsymbol{\theta}, \eta$, and ψ, respectively.

We make a couple of remarks about the semiparametric spatial point process. First, the dimension of the parameter $\boldsymbol{\theta}$ (i.e., k) does not necessarily have to equal to the dimension of the target spatial covariates (i.e., p). Second, we say that a semiparametric spatial point process has a log-linear intensity function if

$$\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) = \exp(\boldsymbol{\theta}^\top \mathbf{u} + \eta(\mathbf{u})) \tag{9}$$

Log-linear intensity functions are commonly used in practice Hessellund et al. [2022] , for both theoretical convenience and interpertability. I greatly simplifies some of the estimation procedures

Given an observational window $A \subset \mathbb{R}^2$, the pseudo-log-likelihood function of X is

$$\ell(\boldsymbol{\theta}, \eta) = \sum_{\mathbf{u} \in X \cap A} \log \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) - \int_A \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) \mathrm{d}\mathbf{u} \tag{10}$$

For every $\boldsymbol{\theta} \in \Theta$, let $\eta^*_{\boldsymbol{\theta}} = \arg\max_{\eta \in \mathcal{H}} \mathbb{E}\left[\ell(\boldsymbol{\theta}, \eta)\right]$.

3 Estimation with Spatial Cross-Fitting

3.1 Overview

Algorithm 1 lays out the proposed estimation procedure of $\boldsymbol{\theta}$ and the algorithm has two steps. In the first step, we use *V-fold random thinning* defined in Definition 3.1 to partition the spatial point process X into V sub-processes, denoted as $X_1, X_2..., X_V$. In the second

step, for every $v \in \{1, \ldots, V\}$, we obtain an estimate of $\eta(\cdot)$, denoted as $\hat{\eta}^{(v)}_{\boldsymbol{\theta}}(\cdot)$, based on maximizing the pseudo-log-likelihood with the sub-process $X^c_v = \bigcup_{j \neq v} X_j$ (i.e, all $V-1$ sub-processes except X_v). Then, we estimate $\hat{\boldsymbol{\theta}}^{(v)}$ by maximizing the pseudo-log-likelihood with the estimate $\hat{\eta}^{(v)}_{\boldsymbol{\theta}}(\cdot)$, all evaluated under the sub-process X_v. Finally, we take an average of $\hat{\boldsymbol{\theta}}^{(1)}, ..., \hat{\boldsymbol{\theta}}^{(V)}$ to obtain the final estimate of $\boldsymbol{\theta}$, denoted as $\hat{\boldsymbol{\theta}}$. Subsequent sections discuss each step in detail.

Algorithm 1 V-Fold Spatial Cross-Fitting For Spatial Point Processes

Require: A spatial point process X, target covariates $\mathbf{y}(\mathbf{u})$, nuisance covariates $\mathbf{z}(\mathbf{u})$, and number of folds V.

Step 1: Use random thinning defined in Definition 3.1 to partition the point process X into V sub-processes $X_1, \ldots, X_V$.

for $v \in [V]$ **do**

Step 2a: Estimate the nuisance parameter η with subprocesses that are not in X_v, denoted as X^c_v:

$$\hat{\eta}^{(v)}_{\boldsymbol{\theta}}(\mathbf{z}) = \arg\max_{\gamma \in \mathbb{R}} \hat{\mathbb{E}}\left[\ell(\boldsymbol{\theta}, \gamma; X^c_v) | \mathbf{z}\right]$$

Step 2b: Estimate the target parameter θ with the subprocess X_v:

$$\hat{\boldsymbol{\theta}}^{(v)} = \arg\max_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta}, \hat{\eta}^{(v)}_{\boldsymbol{\theta}}; X_v)$$

end for

return The aggregated estimator $\hat{\boldsymbol{\theta}} = V^{-1} \sum_{v=1}^{V} \hat{\boldsymbol{\theta}}^{(v)}$, $\hat{\eta} = V^{-1} \sum_{v=1}^{V} \hat{\eta}^{(v)}_{\boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$

3.2 Details of The First Step: Random Thinning

Historically, thinning has been used in the point process literature (e.g., Møller and Schoenberg [2010]) or in Bayesian inference to draw samples from complex, often analytically intractable distributions. In many instances, thinning is used to discard samples and the "thinned" samples are only used for subsequent inference. A novel insight in our algorithm is to repurpose thinning as a way to prevent overfitting of $\hat{\boldsymbol{\theta}}^{(v)}$ and $\eta^{(v)}_{\boldsymbol{\theta}}$, akin to how sample splitting is used in Chernozhukov et al. [2018] to prevent overfitting. Notably, unlike

traditional thinning, we use thinning several times to obtain sub-processes $X_1, ..., X_V$ and we use all samples from the thinning process. Also, by the properties of thinning, each sub-process $X_1, ..., X_V$ inherit the same first-order properties of the original process X. Furthermore, if X is a Poisson spatial point process, the subprocesses $X_1, ..., X_V$ are i.i.d. Poisson spatial point processes [Chandramohan et al., 1985, Alie et al., 2022]). We call this approach V-fold random thinning, and it is formally defined below.

Definition 3.1 (*V-fold Random Thinning*)**.** Suppose X is a spatial point process and $V \geq 2$ is an integer. For each point in X, we uniformly sample a number v from $\{1, \cdots, V\}$, and assign the point to the subprocess X_v.

Proposition 1. *The V thinned spatial point process $X_1, \cdots, X_V$ satisfying the following properties: (i) For any set $I \subset [V] = \{1, 2, .., V\}$, the intensity function of the superposition point process $\bigcup_{j \in I} X_j$ is $\frac{|I|}{K}\lambda(\mathbf{u})$ where $|I|$ is the cardinality of I. (ii) For any set $I \subset [V] = \{1, 2, .., V\}$, $(\boldsymbol{\theta}^*, \eta^*) = \arg\max \ell(\boldsymbol{\theta}, \eta; \cup_{j \in I} X_j)$. For every fixed $\boldsymbol{\theta}$, $\eta^*_{\boldsymbol{\theta}} = \arg\max \mathbb{E}[\ell(\boldsymbol{\theta}, \eta; \cup_{j \in I} X_j)]$. (iii) If X is a Poisson spatial point process, then the sub-processes $X_1, .., X_V$ are mutually independent.*

While not discussed further, we believe thinning can have broader applications in semiparametric inference, especially in the presence of infinite-dimensional nuisance functions.

3.3 Details of The Second Step: Semiparametric Estimation

Step 2 consists of two estimators. In step 2a, the objective function $\hat{\mathbb{E}}\left[\ell(\boldsymbol{\theta}, \gamma; X_v^c)|\mathbf{z}\right]$ of nuisance parameter is defined as follows

$$\sum_{\mathbf{u} \in X_v^c} K_h\left(\mathbf{z}(\mathbf{u}) - \mathbf{z}\right) \log\left(\frac{V-1}{V}\lambda(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u})), \gamma)\right) - \int_A K_h\left(\mathbf{z}(\mathbf{u}) - \mathbf{z}\right) \frac{V-1}{V}\lambda(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u})), \gamma)\mathrm{d}\mathbf{u}. \tag{11}$$

which is an extension of the Nadaraya-Watson estimator initially developed for i.i.d. data (e.g., Watson [1964], Robinson [1988]). Denote $\mathbf{z} = (z_1, \ldots, z_q)$. $K_h(\mathbf{z}) = h^{-q}K(\mathbf{z}/h)$ is a kernel function with bandwidth h and $K(\mathbf{z}) = \prod_{i=1}^{q} k(z_i)$ where $k : \mathbb{R} \to \mathbb{R}$ is a bounded even function satisfies $\int_{\mathbb{R}} k(z)dz = 1$ and $\int_{\mathbb{R}} zk(z)dz = 0$. The kernel regression estimator is trained on the subprocesses X_v^c. For a fixed $\boldsymbol{\theta}$, the estimator $\hat{\eta}_{\boldsymbol{\theta}}^{(v)}(\cdot)$ is the result of evaluating (11) for every $\mathbf{z} \in \mathcal{Z}$. We remark that conceptually, investigators can use other estimators of $\eta(\cdot)$ that is not necessarily based on kernel regression, such as the recent spatial random forest estimator in Saha et al. [2023]; this is akin to cross-fitting in the i.i.d. setting [Chernozhukov et al., 2018] where investigators can use machine learning estimators for the nuisance parameters. Here, we provide one specific estimator of the nuisance parameter η that can provably achieve the desired rates of convergence under a spatial point process model.

In step 2b, we estimate the target parameter $\boldsymbol{\theta}$ by plugging in $\hat{\eta}_{\boldsymbol{\theta}}^{(v)}(\cdot)$ into the pseudo-log-likelihood and maximizing it with respect to $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}}^{(v)} = \arg\max_{\boldsymbol{\theta}\in\Theta} \ell(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta}}^{(v)}; X_v), \quad \ell(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta}}^{(v)}; X_v) = \sum_{\mathbf{u}\in X_v} \log\left(\frac{1}{V}\lambda(\mathbf{u}; \boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta}}^{(v)})\right) - \int_A \frac{1}{V}\lambda(\mathbf{u}; \boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta}}^{(v)})\mathrm{d}\mathbf{u}. \tag{12}$$

We reverse the role of the subprocesses for estimating the nuisance parameter and the target parameter and obtain V estimates of $\boldsymbol{\theta}$. The final step is to aggregate V estimates of $\boldsymbol{\theta}$ by averaging, which in turn regains the full efficiency.

We make two important remarks about Algorithm 1. First, throughout the discussion of the two steps, we did not discuss about actually solving the integrals in (11) and (12). Section 3.4 provides details on dealing with this problem. Second, an interesting phenomena we found in our theoretical derivations is that if the intensity function is log-linear (i.e. equation (9)), step 1 of Algorithm 1 is unnecessary to achieve the desired statistical

properties of $\hat{\boldsymbol{\theta}}$, notably asymptotic Normality. In other words, we can use the full sample to estimate both the nuisance parameter in step 2a and the target parameter in step 2b of Algorithm 1. For more detailed discussion of this, see Section 4.3.

3.4 Computation

In practice, the integral $\int_A \lambda(\mathbf{u};\boldsymbol{\theta},\eta)\mathrm{d}\mathbf{u}$ in the pseudo-log-likelihood (10) and in Algorithm 1 needs to be evaluated to compute an estimator for $\boldsymbol{\theta}$. This section provides two popular solutions to this problem in the context of semiparametric point processes. We remark that this problem does not arise in existing works in partially linear models under i.i.d. settings.

The first approach is based on the quadrature method discussed in Section 2.2. Let $\{\mathbf{u}_j, j = 1,\ldots,m\}$ be the quadrature points containing the observed points in the process X and the corresponding quadrature weights $\{w_j, j = 1,\ldots,m\}$ generated by the Dirichlet tessellation described in Berman and Turner [1992]. Then, the pseudo-log-likelihood in (10) can be approximated as

$$\begin{aligned}\ell(\boldsymbol{\theta},\eta) &\approx \sum_{i=1}^{|X\cap A|} \log\lambda(\mathbf{u}_i;\boldsymbol{\theta},\eta) - \sum_{j=1}^{m}\lambda(\mathbf{u}_j;\boldsymbol{\theta},\eta)w_j \qquad (13)\\ &= \sum_{j=1}^{m}(y_j\log\lambda_j - \lambda_j w_j), \quad \lambda_j = \lambda(\mathbf{u}_j;\boldsymbol{\theta},\eta), y_j = I(\mathbf{u}_j\in X\cap A)/w_j\end{aligned}$$

The right-hand side of (13) is equivalent to the weighted log-likelihood of independent Poisson variables $Y_j \sim \text{Poisson}(\lambda_j)$ with weights w_j, $j = 1\ldots,m$. This approximation allows us to use existing software for generalized partial linear regression such as `mgcv` (Wood and Wood [2015]) to implement the second step in Algorithm 1.

The second approach is based on an approximation to logistic regression discussed in Baddeley et al. [2014]. Let D be a dummy point pattern generated by a spatial point

process that is independent of X and has intensity function $\delta(\mathbf{u})$. Then, conditional on the observed points in $X \cup D$, each point $\mathbf{u}$ in $X \cup D$ belongs to X with probability $\frac{\lambda(\mathbf{u};\boldsymbol{\theta},\eta)}{\lambda(\mathbf{u};\boldsymbol{\theta},\eta)+\delta(\mathbf{u})}$ and belongs to D with probability $\frac{\delta(\mathbf{u})}{\lambda(\mathbf{u};\boldsymbol{\theta},\eta)+\delta(\mathbf{u})}$. Furthermore, $\boldsymbol{\theta}, \eta$ can be estimated by maximizing the logistic log-likelihood function $\ell_{\text{logistic}}(\theta, \eta)$ defined as

$$\ell_{\text{logistic}}(\theta,\eta) = \sum_{\mathbf{u}\in X} \log\left\{\frac{\lambda(\mathbf{u};\boldsymbol{\theta},\eta)}{\lambda(\mathbf{u};\boldsymbol{\theta},\eta)+\delta(\mathbf{u})}\right\} + \sum_{\mathbf{u}\in D} \log\left\{\frac{\delta(\mathbf{u})}{\lambda(\mathbf{u};\boldsymbol{\theta},\eta)+\delta(\mathbf{u})}\right\}. \tag{14}$$

which can be solved by existing software for logistic partially linear models (i.e. `mgcv`).

As discussed by Baddeley and Turner [2000], the quadrature approximation is not generally unbiased. In contrast, while the logistic method maintains unbiasedness, the estimation has additional uncertainty due to the use of the random dummy points in D. We derive the adjusted variance estimator for semiparametric spatial process under the logistic method for approximation. Our adjusted variance estimator is a generalization of the variance estimator in Baddeley et al. [2014] to semiparametric spatial processes.

3.5 Estimation of the Asymptotic Variance of $\hat{\boldsymbol{\theta}}$

This section discusses the estimation of the variance of $\hat{\boldsymbol{\theta}}$. Assuming the PCF of X has a known parametric form $g(\cdot,\cdot;\psi)$ and the true values of $\boldsymbol{\theta}$, η, and ψ are given by $\boldsymbol{\theta}^*$, η^*, and ψ^* respectively, we define the mapping $\boldsymbol{\theta} \mapsto \eta^*_{\boldsymbol{\theta}}$ as $\eta^*_{\boldsymbol{\theta}} = \arg\max_{\eta\in\mathcal{H}} \mathbb{E}\left[\ell(\boldsymbol{\theta},\eta)\right]$ for every $\boldsymbol{\theta} \in \Theta$. We denote $\boldsymbol{\nu}^*(\mathbf{z}) = \frac{\partial}{\partial\boldsymbol{\theta}}\eta^*_{\boldsymbol{\theta}}(\mathbf{z})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$ as the derivative of the mapping $\eta^*_{\boldsymbol{\theta}}$ at $\boldsymbol{\theta}^*$. The asymptotic variance of $\hat{\boldsymbol{\theta}}$ is equivalent to the asymptotic variance of estimating $\boldsymbol{\theta}$ in a parametric spatial point process X' with intensity function $\lambda(\tau(\boldsymbol{\theta}(\mathbf{u})), \eta^*_{\boldsymbol{\theta}}(\mathbf{u})), \boldsymbol{\theta} \in \mathbb{R}^k$ where $\eta^*_{\boldsymbol{\theta}}$ is known. We denote $\ell(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta}})$ as the pseudo-log-likelihood of X'. The sensitivity matrix of X' in 3 is denoted as $\boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*)$, and the covariance matrix of X' in 4 is denoted as $\boldsymbol{\Sigma}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*, g)$. The notation $\boldsymbol{\nu}^*$ inside the notations indicates that the sensitivity matrix

and covariance matrix only depend on $\eta^*_{\boldsymbol{\theta}}$ via its derivative at $\boldsymbol{\theta}^*$. The asymptotic variance of estimating $\boldsymbol{\theta}^*$ in X' is equal to $[\boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*)]^{-1} \boldsymbol{\Sigma}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*, g) [\boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*]^{-1}$. If the intensity function function of X' is log-linear(but not necessarily Poisson), the sensitivity matrix $\boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*)$ and the covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*, g)$ simplify to:

$$\boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*) = \int_A \left(\mathbf{y}(\mathbf{u}) + \boldsymbol{\nu}^*(\mathbf{z}(\mathbf{u}))\right)^{\otimes 2} \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) \mathrm{d}\mathbf{u} \tag{15}$$

$$\begin{aligned}\boldsymbol{\Sigma}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*, \psi^*) =& \boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*) + \int_A \int_A \left(\mathbf{y}(\mathbf{u}) + \boldsymbol{\nu}^*(\mathbf{z}(\mathbf{u}))\right) \left(\mathbf{y}(\mathbf{v}) + \boldsymbol{\nu}^*(\mathbf{z}(\mathbf{v}))\right)^\top \\ & \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) \lambda(\mathbf{v}; \boldsymbol{\theta}^*, \eta^*) (g(\mathbf{u}, \mathbf{v}; \psi^*) - 1) \mathrm{d}\mathbf{u} \mathrm{d}\mathbf{v}\end{aligned} \tag{16}$$

The quadrature approximation of $\boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*)$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*, \psi^*)$ are given by

$$\widehat{\boldsymbol{S}}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*) = \frac{|A|}{m} \sum_{j=1}^{m} \left(\mathbf{y}(\mathbf{u}_j) + \boldsymbol{\nu}^*(\mathbf{z}(\mathbf{u}_j))\right)^{\otimes 2} \lambda(\mathbf{u}_j; \boldsymbol{\theta}^*, \eta^*) \tag{17}$$

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*, \psi^*) = \widehat{\boldsymbol{S}}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*) + \frac{|A|^2}{m^2} \sum_{i,j=1}^{m} & \left(\mathbf{y}(\mathbf{u}_i) + \boldsymbol{\nu}^*(\mathbf{z}(\mathbf{u}_i))\right) \left(\mathbf{y}(\mathbf{u}_j) + \boldsymbol{\nu}^*(\mathbf{z}(\mathbf{u}_j))\right)^\top \\ & \lambda(\mathbf{u}_i; \boldsymbol{\theta}^*, \eta^*) \lambda(\mathbf{u}_j; \boldsymbol{\theta}^*, \eta^*) (g(\mathbf{u}_i, \mathbf{u}_j; \psi^*) - 1)\end{aligned} \tag{18}$$

Here, $\{\mathbf{u}_j\}_{j=1}^m$ represents the quadrature points, and $\{w_j\}_{j=1}^m$ denotes the quadrature weights. We replace $\boldsymbol{\theta}, \eta$ with the aggregated estimator $\hat{\boldsymbol{\theta}}, \hat{\eta}$ in Algorithm 1. The PCF g is assumed to have a parametric form and is substituted by the estimation proposed in Waagepetersen and Guan [2009] (See section 2.2 for review). $\boldsymbol{\nu}^*$ will be replaced by the nonparametric estimators discussed in the next section.

3.5.1 Estimation of $\boldsymbol{\nu}^*$

The R package `mgcv` employs a backfitting algorithm, iteratively refining the parametric estimation of the target parameter $\boldsymbol{\theta}(\cdot)$ and the nonparametric estimation of the nuisance parameter η until convergence. This iterative process does not provide a full path from $\boldsymbol{\theta}$ to $\hat{\eta}_{\boldsymbol{\theta}}$, making it unfeasible to estimate $\boldsymbol{\nu}^*$ directly through $\boldsymbol{\nu}^*(\mathbf{z}) = \frac{\partial}{\partial \boldsymbol{\theta}} \eta^*_{\boldsymbol{\theta}}(\mathbf{z}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$. However, denoting $\mathcal{H}^k = \mathcal{H} \times \cdots \times \mathcal{H}$, our analysis reveals that when intensity function is log-linear, $\boldsymbol{\nu}^*$ satisfies

$$\boldsymbol{\nu}^* = \arg\min_{\boldsymbol{\nu} \in \mathcal{H}^k} \int_A \left(\mathbf{y}(\mathbf{u}) + \boldsymbol{\nu}(\mathbf{z}(\mathbf{u}))\right)^\top \left(\mathbf{y}(\mathbf{u}) + \boldsymbol{\nu}(\mathbf{z}(\mathbf{u}))\right) \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) \mathrm{d}\mathbf{u} \tag{19}$$

indicating that $-\boldsymbol{\nu}^*$ is the weighted linear projection of $\mathbf{y}(\mathbf{u})$ onto $H \circ \mathbf{z}(\mathbf{u})$ with weight $\lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*)$. We therefore proposed the weighted kernel regression estimator of $\boldsymbol{\nu}^*$ as follows:

$$\hat{\boldsymbol{\nu}}(\mathbf{z}) = \frac{\sum_{j=1}^m K_h\left(\mathbf{z}(\mathbf{u}_i) - \mathbf{z}\right) \lambda(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u}_i)), \eta(\mathbf{z})) \mathbf{y}(\mathbf{u}_i)}{\sum_{i=1}^m K_h\left(\mathbf{z}(\mathbf{u}_i) - \mathbf{z}\right) \lambda(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u}_i)), \eta(\mathbf{z}))}, \tag{20}$$

where K_h is the scaled kernel function defined in section 3.3. $\hat{\boldsymbol{\nu}}(\mathbf{z})$ is obtained using merely the quadrature points without the observed points. This characteristic plays a pivotal role in our asymptotic theory, as elaborated in section 4.3.

4 Theory

4.1 Overview and Definitions

In subsequent sections, we show that the proposed estimator $\hat{\boldsymbol{\theta}}$ exhibits: (a) consistency for general spatial point processes, (b) asymptotic Normality under conditions where the

process is Poisson or the intensity function is log-linear, provided that the estimation of $\eta(\cdot)$ converges rapidly enough and (c) semiparametric efficiency in the case of a Poisson process. Additionally, we establish consistent estimation of the asymptotic variance of $\hat{\boldsymbol{\theta}}$ in scenarios where condition (b) is met; see Table 1 for a summary of the theoretical results.

First, we introduce two key technical concepts pivotal for developing the theory, leveraging the push-forward measure (see Bogachev and Ruas [2007]) and the Gateaux derivative (see Frigyik et al. [2008]).

Definition 4.1 (Conditional Expectation of Pseudo-Log-Likelihood)**.** Suppose the push-forward measure of the Lebesgue measure on $\mathbb{R}^2$ induced by $\mathbf{u} \mapsto (\mathbf{y}(\mathbf{u}), \mathbf{z}(\mathbf{u}))$ has the "joint" Radon-Nikodym derivative $f(\mathbf{y}, \mathbf{z})$, and the "marginal" Radon-Nikodym derivative $f(\mathbf{z})$. Then, $\mathbb{E}[\ell(\boldsymbol{\theta}, \eta)|\mathbf{z}]$, i.e. the conditional expectation of pseudo-log-likelihood function given the value of nuisance covariates, can be written as

$$\mathbb{E}[\ell(\boldsymbol{\theta}, \eta)|\mathbf{z}] = \int_{\mathcal{Y}} (\log \lambda(\tau_{\boldsymbol{\theta}}(\mathbf{y}), \eta(\mathbf{z}))\lambda(\tau_{\boldsymbol{\theta}^*}(\mathbf{y}), \eta^*(\mathbf{z})) - \lambda(\tau_{\boldsymbol{\theta}}(\mathbf{y}), \eta(\mathbf{z})))f(\mathbf{y}, \mathbf{z})\mathrm{d}\mathbf{y} \tag{21}$$

Definition 4.2 (Gateaux Derivative for Semiparametric Intensity Functions)**.** The first-order Gateaux derivative of $\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) = \lambda(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u})), \eta(\mathbf{z}(\mathbf{u})))$ with respect to η along the direction $\bar{\eta} \in H$ is defined as

$$\frac{\partial}{\partial \eta}\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta)[\bar{\eta}] = \lim_{h \to 0} \frac{\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta + h\bar{\eta}) - \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta)}{h} = \frac{\partial}{\partial h}\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta + h\bar{\eta})\bigg|_{h=0} \tag{22}$$

Also, the m-th order Gateaux derivative along the directions $(\bar{\eta}_1, \ldots, \bar{\eta}_m)$ is

$$\frac{\partial}{\partial \eta}\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta)[\bar{\eta}_1, \ldots, \bar{\eta}_m] = \frac{\partial^m}{\partial \left(\eta(\mathbf{z}(\mathbf{u}))\right)^m}\lambda(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u})), \eta(\mathbf{z}(\mathbf{u}))) \cdot \bar{\eta}_1(\mathbf{z}(\mathbf{u})) \cdot \bar{\eta}_2(\mathbf{z}(\mathbf{u})) \cdots \bar{\eta}_m(\mathbf{z}(\mathbf{u}))$$

We remark that we use $\mathbb{E}[\ell(\boldsymbol{\theta}, \eta(\mathbf{z}))]$ and $\mathbb{E}[\ell(\boldsymbol{\theta}, \eta)|\mathbf{z}]$ interchangeably. Also, the m-th order Gateaux derivative is essentially the product of real-valued derivatives of $\lambda(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u})), \eta(\mathbf{z}(\mathbf{u})))$ with respect to $\eta(\mathbf{z})$.

4.2 Semiparametric Efficiency Lower Bound for Poisson Spatial Point Process

This section derives the semiparametric efficiency lower bound of estimating $\boldsymbol{\theta}$ in 1 when the point process X is Poisson. The quantities $\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*, \eta^*_{\boldsymbol{\theta}}$ are defined in section 3.5. Consider all the twice continuously differentiable mapping $\eta_{\boldsymbol{\theta}}$ from Θ to $\mathcal{H}$ satisfying $\eta_{\boldsymbol{\theta}^*} = \eta^*$, and let $\boldsymbol{\nu} = \frac{\partial}{\partial \boldsymbol{\theta}} \eta_{\boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$. Each $\eta_{\boldsymbol{\theta}}$ corresponds to a parametric spatial point process, i.e. a submodel of X, with intensity function $\lambda(\tau(\boldsymbol{\theta}(\mathbf{u})), \eta_{\boldsymbol{\theta}}(\mathbf{u})), \boldsymbol{\theta} \in \mathbb{R}^k$, whose Cramer-Rao lower bound of estimating $\boldsymbol{\theta}$ is $[\boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu})]^{-1}$ where $\boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu})$ is given by

$$\boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}) = \int_A \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) \left(\frac{\partial}{\partial \boldsymbol{\theta}} \log \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) + \frac{\partial}{\partial \eta} \log \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) \left[\boldsymbol{\nu}(\mathbf{z}(\mathbf{u})) \right] \right)^{\otimes 2} \mathrm{d}\mathbf{u}, \tag{23}$$

Theorem 4.1 shows that, the supremum of the Cramer-Rao lower bound is attained by X' defined in 3.5 among all the parametric submodel associated with $\eta_{\boldsymbol{\theta}}$, i.e. $\boldsymbol{\nu}^* = \arg\max_{\boldsymbol{\nu} \in \mathcal{H}} [\boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*)]^{-1}$. The explicit form of $\boldsymbol{\nu}^*$ is derived in Theorem 4.1 as well. In section 4.3, we show that that asymptotic variance of $\boldsymbol{\theta}$ in Algorithm 1 is equal to $[\boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu})]^{-1}$ when X is Poisson. Since the variance of all asymptotically Normal estimator of $\boldsymbol{\theta}$ in X must be greater than the supremum of the Cramer-Rao lower bounds of parametric submodels. (See chapter 4.3 in Tsiatis [2006]), $\hat{\boldsymbol{\theta}}$ is thus semiparametrically efficient when X is Poisson.

Theorem 4.1 (Semiparametric Efficiency Lower Bound for Semiparametric Poisson Pro-

ces). *The minimum of 23 is attained by* $\boldsymbol{\nu}^*$ *defined in 3.5 and* $\boldsymbol{\nu}^*$ *satisfies*

$$\boldsymbol{\nu}^*(\mathbf{z}) = -\frac{\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\eta}\mathbb{E}[\ell(\boldsymbol{\theta}^*, \eta^*(\mathbf{z}))]}{\frac{\partial^2}{\partial\eta^2}\mathbb{E}[\ell(\boldsymbol{\theta}^*, \eta^*(\mathbf{z}))]}$$

and $\mathbb{E}[\ell(\boldsymbol{\theta}, \eta(\mathbf{z}))]$ *is the conditional expectation defined in Definition 21. Furthermore, if the intensity function is log-linear as in equation* (9)*,* $\boldsymbol{\nu}^*$ *simplify to*

$$\boldsymbol{\nu}^*(\mathbf{z}) = -\frac{\int_{\mathcal{Y}} \exp(\boldsymbol{\theta}^{*\top}\mathbf{y})\mathbf{y} f(\mathbf{y}, \mathbf{z})\mathrm{d}\mathbf{y}}{\int_{\mathcal{Y}} \exp(\boldsymbol{\theta}^{*\top}\mathbf{y}) f(\mathbf{y}, \mathbf{z})\mathrm{d}\mathbf{y}}.$$

Particularly, Theorem 4.1 is derived based on the fact that $\boldsymbol{\nu}(\mathbf{z})$ belongs to a k-replicating linear space $\mathcal{H}^k$ (See Definition 6 in Tsiatis [2006]). Thus, the infimum of $\boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu})$ is attained when $-\frac{\partial}{\partial\eta}\log\lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*)\,[\boldsymbol{\nu}(\mathbf{z}(\mathbf{u}))]$ is the weighted linear projection of $\frac{\partial}{\partial\boldsymbol{\theta}}\log\lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*)$ onto the linear space $\frac{\partial}{\partial\eta}\log\lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) \circ \mathcal{H}^k$ with weights equal to $\lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*)$.

4.3 Consistency and Asymptotic Normality

Similar to other works' asymptotic framework in spatial point processes (e.g., Rathbun and Cressie [1994]), we consider a sequence of expanding observation windows A_n in $\mathbb{R}^2$ where $A_1 \subset A_2 \cdots \subset A_n$ and the area $|A_n|$ goes to infinity. Denote by $\ell_n(\boldsymbol{\theta}, \eta)$ the psuedo-likelihood in 10 evaluated on A_n and by $\hat{\boldsymbol{\theta}}_n, \hat{\eta}_n, \hat{\eta}^{(v)}_{\boldsymbol{\theta},n}, \hat{\psi}_n$ the sequence of estimators in Algorithm 1. Quantities $\boldsymbol{\theta}^*, \eta^*, \psi^*, \eta^*_{\boldsymbol{\theta},n}, \boldsymbol{\nu}^*_n, \boldsymbol{S}_n(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*), \boldsymbol{\Sigma}_n(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*, \psi^*)$ are defined as in section 3.5 with $A = A_n$ for the last four. Radon-Nikodym derivatives $f_n(\mathbf{y}, \mathbf{z}), f_n(\mathbf{y})$ are defined in 4.1 with $A = A_n$. We also define "averaged" versions $\bar{\boldsymbol{S}}_n(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*) = \boldsymbol{S}_n(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*)/|A_n|$ and $\bar{\boldsymbol{\Sigma}}_n(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*, \psi^*) = \boldsymbol{\Sigma}_n(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*, \psi^*)/|A_n|$. Finally, let $|\cdot|$ denote the ℓ_2 norm of a vector or the spectral norm of a matrix. Let $|\cdot|_{\max}$ denote the maximum norm of a vector or the matrix. Let $\lambda_{\min}[M]$ denote the minimal

eigenvalue of a matrix M.

Assumption 1. Regularity conditions needed for consistency.

1. (*Smoothness of* λ) The intensity function $\lambda(\mathbf{u};\boldsymbol{\theta},\eta), \boldsymbol{\theta} \in \Theta, \eta \in \mathcal{H}$ is twice continuously differentiable with respect to $\boldsymbol{\theta}$ and η.

2. *(Sufficient Separation)* There exists $0 < c_0, c_1 < \infty$ that the set $C = \{\mathbf{u} : |\log \lambda(\mathbf{u};\boldsymbol{\theta},\eta) - \log \lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*)| \geq \min(c_0, c_1|\boldsymbol{\theta} - \boldsymbol{\theta}^*|)\}$ satisfies $|C \cap A_n| = \Theta(|A_n|)$

3. (*Boundedness of* λ) There exists $c_2 > 0$ that $B = \{\mathbf{u} : \inf_{\boldsymbol{\theta},\eta} \lambda(\mathbf{u};\boldsymbol{\theta},\eta) < c_2\}$ is bounded.

4. (*Decay Rate of PCF*) There exists an $0 < C_2 < \infty$ so that $\int_{\mathbb{R}^2} |g(0,\mathbf{u})\mathrm{d}\mathbf{u}| < C_2$

Assumption 2. Regularity conditions needed for asymptotic Normality.

1. *(Nonsigularity)* $\liminf_n \lambda_{\min}\left[\bar{\boldsymbol{S}}_n(\boldsymbol{\theta}^*,\eta^*,\boldsymbol{\nu}^*)\right] > 0$

2. *(Nonsigularity)* $\liminf_n \lambda_{\min}\left[\bar{\boldsymbol{\Sigma}}_n(\boldsymbol{\theta}^*,\eta^*,\boldsymbol{\nu}^*,\psi^*)\right] > 0$

3. (α*-mixing rate*) The α-mixing coefficient satisfies $\alpha^X_{2,\infty}(r) = O(r^{-(2+\epsilon)})$ for some $\epsilon > 0$

The above conditions are semiparametric extensions of regularity conditions widely used in spatial point process literature. The sufficient separation condition 2 is the spatial, semiparametric extension of condition (A1) in Schoenberg [2005], or condition (C5) in Rathbun and Cressie [1994]. The nonsigularity condition 1 and 2 are the semiparametric extension of condition (C4) and (N3) in Hessellund et al. [2022]. The boundedness condition 3 is a semiparametric extension of condition 2 in Theorem 11 of Rathbun and Cressie [1994]. PCF decay rate condition 4 is identical to (C3) in Hessellund et al. [2022]. The α-mixing condition 3 is identical to condition (v) in Theorem 1 of Waagepetersen and Guan [2009],

which is a common condition required for asymptotic Normality. Heuristically, α-mixing condition states that for any two fixed sets, the dependence between them must decay to 0 at a polynomial rate of the interset distance r.

The next two assumptions are about the estimation of the nuisance parameters.

Assumption 3. [Consistency of Estimated Nuisance Parameter] For every $v \in [V]$ and $i = 0, 1, 2$, the estimated nuisance parameter $\hat{\eta}^{(v)}_{\boldsymbol{\theta},n}$ in Algorithm 1 satisfies:

$$\sup_{\boldsymbol{\theta}\in\Theta,\mathbf{z}\in\mathcal{Z}} \left| \frac{\partial^i}{\partial \boldsymbol{\theta}^i} \hat{\eta}^{(v)}_{\boldsymbol{\theta},n}(\mathbf{z}) - \frac{\partial^i}{\partial \boldsymbol{\theta}^i} \eta^*_{\boldsymbol{\theta},n}(\mathbf{z}) \right| = o_p(1).$$

Assumption 4. [Rates of Convergence of Estimated Nuisance Parameter] For every $v \in [V]$ and $i = 0, 1$, the nuisance parameter estimator $\hat{\eta}^{(v)}_{\boldsymbol{\theta},n}$ in Algorithm 1 satisfies:

$$\sup_{\mathbf{z}\in\mathcal{Z}} \left| \frac{\partial^i}{\partial \boldsymbol{\theta}^i} \hat{\eta}^{(v)}_{\boldsymbol{\theta}^*,n}(\mathbf{z}) - \frac{\partial^i}{\partial \boldsymbol{\theta}^i} \eta^*_{\boldsymbol{\theta}^*,n}(\mathbf{z}) \right| = o_p\left(|A_n|^{-\frac{1}{4}} \right)$$

In Section 4.4, we show that the the spatial kernel regression estimator in Algorithm **??** can satisfy both assumptions in very mild conditions. However, the investigator is free to choose any estimator of the nuisance function that satisfy the two assumptions instead of the spatial kernel regression estimator. Also, because of V-fold thinning, we do not need Donsker-style conditions on top of these assumptions to achieve asymptotic Normality of $\hat{\boldsymbol{\theta}}$ when X is Poisson.

The following theorems establish the consistency and the asymptotic Normality of $\hat{\boldsymbol{\theta}}$.

Theorem 4.2 (Consistency of $\hat{\boldsymbol{\theta}}$)**.** *Suppose conditions (1)-(4) in Assumption 1 and Assumption 3 hold. Then $\hat{\boldsymbol{\theta}}_n$ is consistent for $\boldsymbol{\theta}^*$, i.e., $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^* \to_p \mathbf{0}$.*

Theorem 4.3 (Asymptotic Normality of $\hat{\boldsymbol{\theta}}$)**.** *Suppose Assumptions 1 - 4 hold. Then, $\hat{\boldsymbol{\theta}}$ is asymptotically Normal, i.e., $|A_n|^{\frac{1}{2}} \bar{\boldsymbol{S}}_n(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*) \bar{\boldsymbol{\Sigma}}_n^{-\frac{1}{2}}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*, \psi^*)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \to_d N(\mathbf{0}, \mathbf{I}_k)$ if*

either X is Poisson or the intensity function of X is log-linear as 9.

Assumption 3 is used to show that $\hat{\boldsymbol{\theta}}$ is consistent. For $\hat{\boldsymbol{\theta}}$ to be asymptotically Normal, the estimated nuisance parameter needs to converge at the rate $o_p\left(|A_n|^{-\frac{1}{4}}\right)$, as stipulated in Assumption 4. This rate mirrors the typical rate $o_p\left(N^{-\frac{1}{4}}\right)$ in i.i.d setting where N is the sample size. Furthermore, if the process is Poisson, the PCF $g(\mathbf{u}, \mathbf{v}) = 0$ for all $\mathbf{u}, \mathbf{v}$, resulting in the sensitivity matrix $\boldsymbol{S}_n(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*)$ being equal to the covariance matrix $\boldsymbol{\Sigma}_n(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*, g^*)$. Consequently, the asymptotic variance of Theorem 4.1 simplifies to $\boldsymbol{S}_n^{-\frac{1}{2}}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*)$.

Random thinning becomes unnecessary when the intensity function is log-linear in 9. It's primarily employed to ensure independence between $\ell(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta}})$ and $\hat{\eta}_{\boldsymbol{\theta}}, \hat{\boldsymbol{\nu}}_{\boldsymbol{\theta}}$, which ensures that the expectations of the remainder terms $\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \eta}\ell(\boldsymbol{\theta}, \eta_{\boldsymbol{\theta}})\left[\hat{\eta}_{\boldsymbol{\theta}} - \eta_{\boldsymbol{\theta}}\right]$ and $\frac{\partial}{\partial \eta}\ell(\boldsymbol{\theta}, \eta_{\boldsymbol{\theta}})\left[\hat{\boldsymbol{\nu}}_{\boldsymbol{\theta}} - \boldsymbol{\nu}_{\boldsymbol{\theta}}\right]$ are zero. However, when the intensity function is log-linear, $\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \eta}\ell(\boldsymbol{\theta}, \eta_{\boldsymbol{\theta}})$ becomes non-stochastic, rendering it independent of $\ell(\boldsymbol{\theta}, \eta_{\boldsymbol{\theta}})$. Additionally, $\hat{\boldsymbol{\nu}}_{\boldsymbol{\theta}}$ is estimated merely using quadrature points (see 20), ensuring its independence from $\ell(\boldsymbol{\theta}, \eta_{\boldsymbol{\theta}})$. Hence, random thinning is unnecessary in the log-linear intensity function scenario. In contrast, while sample splitting serves a similar purpose in generalized partial linear models with i.i.d. data, it remains necessary for ensuring asymptotic normality, irrespective of the specified inverse link function.

Our last theorem shows that our variance estimator in Section 3.5 is consistent.

Theorem 4.4 (Consistent variance estimator)**.** *Suppose the same assumptions in Theorem 4.3 hold. $\hat{\boldsymbol{\nu}}_n$ in 20 is a consistent estimator of $\boldsymbol{\nu}^*_n$, i.e. $\sup_{\mathbf{z} \in \mathcal{Z}} |\hat{\boldsymbol{\nu}}_n(\mathbf{z}) - \boldsymbol{\nu}^*(\mathbf{z})| \to_p 0$. Further assume assumptions in Theorem 1 in Waagepetersen and Guan [2009], $\hat{\psi}_n$ is consistent estimator of ψ^*, i.e. $|\hat{\psi}_n - \psi^*|$. Then we have $\widehat{\boldsymbol{S}}_n(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_n, \hat{\boldsymbol{\nu}}_n)/\boldsymbol{S}_n(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*) \to_p 1$ and $\widehat{\boldsymbol{\Sigma}}_n^{-\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_n, \hat{\boldsymbol{\nu}}_n, \hat{\psi}_n)/\boldsymbol{\Sigma}_n^{-\frac{1}{2}}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*, \psi^*) \to_p 1$.*

An implication of Theorem 4.4 is that we can use $\widehat{\boldsymbol{S}}_n^{-1}(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_n, \hat{\boldsymbol{\nu}}_n)\widehat{\boldsymbol{\Sigma}}_n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_n, \hat{\boldsymbol{\nu}}_n, \hat{\psi}_n)$ as the standard error of $\hat{\boldsymbol{\theta}}$ and construct confidence intervals of $\boldsymbol{\theta}$. If X is a Poisson spatial point process, we can simply use $\widehat{\boldsymbol{S}}_n^{-1}(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_n, \hat{\boldsymbol{\nu}}_n)$ as the estimator of the standard error of $\hat{\boldsymbol{\theta}}$, based on the result from Theorem 4.3, without needing an estimate of the PCF.

4.4 Convergence Rate of Estimated Nuisance Parameter $\hat{\eta}^{(v)}_{\boldsymbol{\theta},n}(\mathbf{z})$

In the section will show that when the intensity function and the nuisance are sufficiently smooth, and the dependence between points in X is weak, the nuisance estimation in Algorithm 1 attains the convergence rate specified in Assumptions 3 and 4.

Assumption 5. For some integers $l \geq 2$ and $m \geq 2$,

1. (*Smoothness*) $\lambda(\tau_{\boldsymbol{\theta}^*}(\mathbf{y}), \eta^*(\mathbf{z}))$ and $f_n(\mathbf{y}, \mathbf{z})$ are l-times continuously differentiable with respect to $\mathbf{z}$
2. (*Identification*) $\liminf_n \inf_{\boldsymbol{\theta},\mathbf{z}} |A_n|^{-1} \frac{\partial^2}{\partial \gamma^2} \mathbb{E}[\ell_n(\boldsymbol{\theta}, \gamma)|\mathbf{z}]\Big|_{\gamma=\eta^*_{\boldsymbol{\theta},n}(\mathbf{z})} > 0$
3. (*Higher Order Kernel*) Kernel function $K(\cdot)$ in 11 is an l-th order kernel.
4. (*Weak Dependence*) There exists positive constant C such that the cumulant functions of X satisfies

$$\sup_{\mathbf{u}_1 \in \mathbb{R}^2} \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} |Q_{m'}(\mathbf{u}_1, \ldots, \mathbf{u}_{m'})| \, \mathrm{d}\mathbf{u}_2 \cdots \mathrm{d}\mathbf{u}_{m'} < C, \quad m' = 2, 3, \ldots, m$$

The identification condition guarantees The higher-order kernel 3 is a bias-reduction technique used in kernel regression in i.i.d. setting, which necessitates additional smoothness requirements as we state in condition 1. If the conditional mean is infinitely differentiable with respect to covariates, the convergence rate can be arbitrarily close to $O_p(N^{-\frac{1}{2}})$,

where N denotes the sample size. (see Li and Racine [2023] chapter 1.11 for details). The weak dependence condition in Assumption 5 is identical to condition (6) in Guan and Loh [2007] if $m = 4$. Particularly, $Q_{m'}(\mathbf{u}_1, \ldots, \mathbf{u}_{m'})$ describes the dependence among points $\mathbf{u}_1, \ldots, \mathbf{u}_{m'}$, with values close to zero indicating near independence. In the case of a Poisson process, $Q_{m'}(\mathbf{u}_1, \ldots, \mathbf{u}_{m'}) = 0$ if at least two of $\mathbf{u}_1, \ldots, \mathbf{u}_{m'}$ differ. $Q_{m'}(\mathbf{u}_1, \ldots, \mathbf{u}_{m'})$ can be expressed as a function of intensity functions up to the m'-th order. For instance, $Q_2(\mathbf{u}_1, \mathbf{u}_2) = \lambda_2(\mathbf{u}_1, \mathbf{u}_2) - \lambda(\mathbf{u}_1)\lambda(\mathbf{u}_2)$. The integral of $Q_{m'}(\mathbf{u}_1, \ldots, \mathbf{u}_{m'})$ is equal to the m'-th order centered moment of X. (See Daley and Vere-Jones [2007] chapter 9.5, chapter 12.6 for details.)

Theorem 4.5. *Suppose Assumptions 1, 5 hold. Then, with appropriately chosen bandwidth, for $j = 0, 1, 2$, the nuisance estimation satisfies*

$$\sup_{\mathbf{z}\in\mathcal{Z}} \left| \frac{\partial^j}{\partial \boldsymbol{\theta}^j} (\eta^*_{\boldsymbol{\theta},n}(\mathbf{z}) - \hat{\eta}^{(v)}_{\boldsymbol{\theta},n}(\mathbf{z})) \right| = o_p\left(|A_n|^{-\frac{m-1}{m+k+q+1} \cdot \frac{l}{l+q+1}} \right) \tag{24}$$

The convergence rate in equation (24) is determined by the m and l in Assumption 5, the target parameter dimension k, and the nuisance covariates dimension q. When the higher-order dependence of points in X is sufficiently weak enough, i.e. m can be arbitrarily large, the nuisance convergence rate is $O_p\left(|A_n|^{-\frac{l}{l+q+1}}\right)$. Likewise, with adequate smoothness in the intensity function and nuisance covariates, i.e. l can be arbitrarily large, the nuisance convergence is $O_p\left(|A_n|^{-\frac{m-1}{m+k+q+1}}\right)$. However, if the dimensions of the nuisance covariates and the target parameter are large, the nuisance convergence rate will be slower. To ensure that Assumption 3 and 4 hold, m and l must be sufficiently large to satisfy $\frac{m-1}{m+k+q+1} \cdot \frac{l}{l+q+1} \geq \frac{1}{4}$. Interestingly, when both m and l are allowed to grow arbitrarily, the rate can approach $O_p(|A_n|^{-1})$, which differs from kernel regression, where the error rate is always slower than $O_p(N^{-\frac{1}{2}})$ for i.i.d. data, regardless of smoothness.

5 Simulation and Real Data Analysis

5.1 Simulation

In this section, we evaluate the finite-sample performance of our method through a simulation study conducted on an observational window W_a of size $a \times a$, where $a = 1, 2$. We consider Poisson spatial point processess with an intensity function given by

$$\lambda(\mathbf{u}; \theta, \eta) = 400 \exp(\theta \mathbf{y}(\mathbf{u}) + \eta(\mathbf{z}(\mathbf{u}))) \quad \theta \in \mathbb{R}, \eta \in L_2(\mathbb{R}) \tag{25}$$

where the true target parameter $\theta^* = 0.3$, and the true nuisance parameter $\eta^*(z) = -0.09z^2$. Here, $\mathbf{y}(\mathbf{u}), \mathbf{z}(\mathbf{u})$ are dependent Gaussian random fields. Specifically, we first generate two i.i.d. zero-mean Gaussian random field $\mathbf{y}(\mathbf{u}), \bar{\mathbf{z}}(\mathbf{u})$ with covariance function $C(r) = \exp(-r/0.05)$, and then let $\mathbf{z}(\mathbf{u}) = \bar{\mathbf{z}}(\mathbf{u}) \times \mathbf{y}(\mathbf{u})$. We also consider log-Gaussian Cox processes(LGCP) with a conditional intensity function given by

$$\Lambda(\mathbf{u}; \theta, \eta) = 400 \exp\left(\theta \mathbf{y}(\mathbf{u}) + \eta(\mathbf{z}(\mathbf{u})) + G(\mathbf{u}) - \frac{2}{\sigma^2}\right) \quad \theta \in \mathbb{R}, \eta \in L_2(\mathbb{R}) \tag{26}$$

where $\theta^*, \eta^*, \mathbf{y}(\mathbf{u}), \mathbf{z}(\mathbf{u})$ are the same as the Poisson spatial processes. $G(\mathbf{u})$ is a zero-mean Gaussian random field with covariance function $C(r) = \sigma^2 \exp(-r/0.2)$, where $\sigma^2 = 0.2$. We generated 1000 realizations for both the Poisson spatial point process and the log-Gaussian Cox process. For each realization, we fitted (1) a semiparametric spatial point process where $\eta(z)$ is unspecified, (2) a parametric spatial point process where $\eta(z)$ is misspecified as a linear function of z, (3) an oracle parametric spatial where $\eta^*(z) = -0.09z^2$ is known as priori. For semiparametric spatial point process, we use Algorithm 1 with logistic likelihood (Equation (14)) to obtain the $\hat{\theta}$. The square root of variance estimates

of $\hat{\theta}$ in section 3.5 served as the standard error $\mathrm{SE}(\hat{\theta})$. For parametric spatial point process, we use `ppm` and `kppm` in package `spatstat` (Baddeley and Turner [2005]) to obtain $\hat{\theta}$ and $\mathrm{SE}(\hat{\theta})$.

	Process	Model	$\mathrm{Bias}_{\times 100}$	rMSE	meanSE	CP90 (%)	CP95 (%)
W_1	Poisson	Semi	-0.7548	0.0531	0.0532	89.3	95.3
		Para	-4.1990	0.0633	0.0495	78.7	86.3
		Oracle	0.1458	0.0523	0.0522	90.5	94.7
	LGCP	Semi	-0.6000	0.0673	0.0611	86.1	91.9
		Para	-4.3380	0.0726	0.0597	82.2	89.3
		Oracle	-0.2632	0.0644	0.0618	89.2	93.5
W_2	Poisson	Semi	-0.0219	0.0266	0.0275	91.3	96.2
		Para	-2.5624	0.0350	0.0257	75.4	85.8
		Oracle	0.2058	0.0268	0.0275	91.2	95.9
	LGCP	Semi	-0.2485	0.0328	0.0327	89.8	94.3
		Para	-3.5219	0.0457	0.0316	70.5	81.6
		Oracle	-0.6360	0.0331	0.0327	88.8	94.6

Table 2: W_a: observational window $[0, a]\times[0, a]$; Poisson: spatial pattern generated from 25; LGCP: spatial pattern generated from 26; Semi: semiparametric model; Para: misspecified parametric model: Oracle: oracle parametric model

For each method, we present the results in Table 2, including the empirical bias ($\mathrm{Bias}_{\times 100}$), root mean square error (rMSE), empirical mean of $\mathrm{SE}(\hat{\theta})$ (meanSE) and the empirical coverage probability of 90% level and 95% level confidence intervals assuming the Normality of $\hat{\theta}$ (CP90, CP95). Across all simulation settings, our method exhibits close-to-unbiased estimates, whereas the estimates from the misspecified parametric model show significant bias. The variance estimation of our method tends to be under-estimated when the point pattern is generated from an LGCP, primarily due to unsatisfactory estimates of the pair correlation function rather than bias in variance estimates or errors in asymptotic theory. This issue persists even when fitting the Oracle parametric model. However, it diminishes notably as the observational window expands, providing enough points for a satisfactory estimation of the pair correlation function. Furthermore, the rMSE of our method only

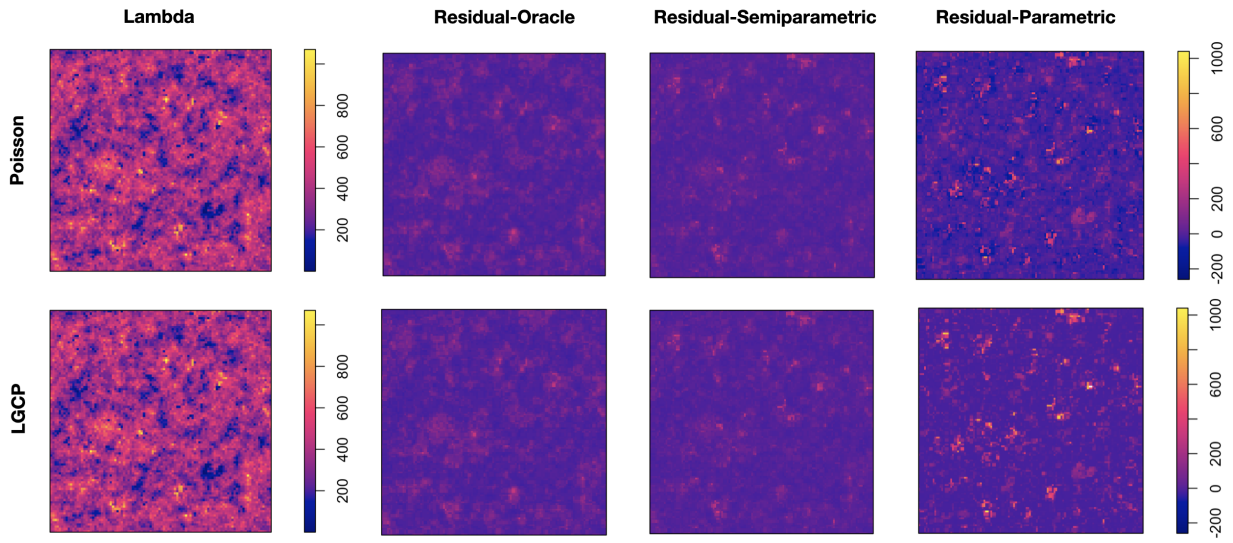

Figure 1: Residual of fitted intensity using different models. Lambda: intensity function used to generate spatial point processes.

slightly exceeds that of the Oracle model, suggesting that our approach maintains high efficiency while accommodating a flexible functional form for the nuisance parameter.

5.2 Real Data Analysis

We applied our method to a tropical rainforest dataset collected in a 1000 $\times$ 500 meters plot on Barro Colorado Island (See Hubbell [1983], Condit et al. [1996], and Condit [1998] for further details). In our analysis, we focus on understanding how elevation influences the growth of *Beilschmiedia pendula* and *Capparis frondosa* trees. Following Waagepetersen and Guan [2009], we assume that the locations of trees follow a Thomas process. The elevation serves as the target covariate $\mathbf{y}(\mathbf{u})$, while the gradient of elevation acts as the nuisance covariate $\mathbf{z}(\mathbf{u})$. The target parameter θ represents the effect of elevation on tree growth. For both tree species, we fit both a semiparametric spatial point process and a parametric spatial point process. In the parametric model, we assume the effect of gradient to be log-linear. For each method, we compute the point estimate $\hat{\theta}$, the standard error

SE($\hat{\theta}$), and plot the fitted intensity map in 2.

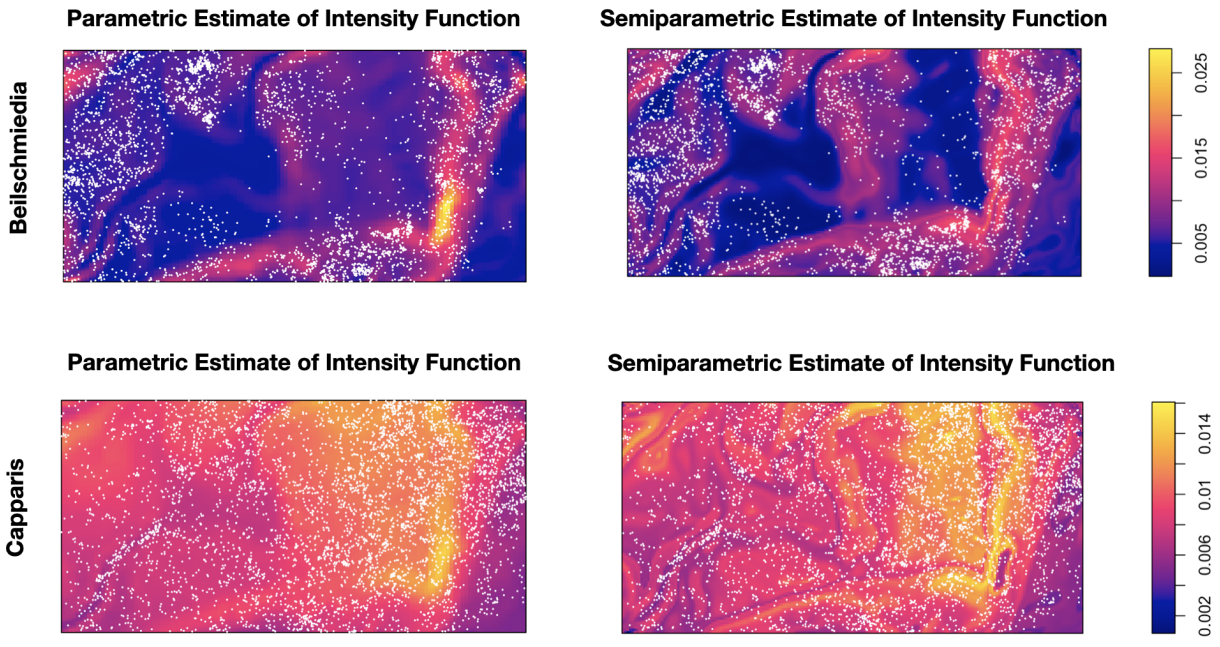

Figure 2: Heat maps: fitted intensity functions; White points: observed trees. First row: Beilschmiedia; Second row: Capparis.

For *Beilschmiedia pendula*, the estimate of θ from the semiparametric model is 3.132 (95% CI: (-1.480, 7.744)), whereas the estimate from the parametric model is 2.144 (95% CI: (-2.454, 6.742)). The nuisance estimation obtained by Algorithm 1 reveals a non-log-linear relationship between the gradient and the spatial distribution of *Beilschmiedia pendula*. Consequently, inference based on the parametric model may suffer from bias due to the mis-specification of the nuisance functional form. Furthermore, the intensity map obtained from the semiparametric model exhibits a better fit compared to the parametric model, effectively capturing dense occurrences in the top left corner.

For *Capparis frondosa*, the estimate of θ from the semiparametric model is 2.565 (95% CI: (-0.032, 5.162)), while the estimate from the parametric model is similar, with a point estimate of 2.692 (95% CI: (0.036, 5.348)). Further examination reveals that the estimated

nuisance covariates has a minor effect on the intensity function. However, our method's fitted intensity still captures slightly more detail than the parametric method, such as a sparse region in the bottom right of the observation window. In summary, because our approach makes fewer assumptions about the parametric form of the nuisance covariate, our estimations of the target parameters are likely to be more robust than those from the parametric model. Our fitted intensity maps are likely to be more detailed and accurate.

6 Discussion

We study the estimation of the semiparametric spatial point process where we are interested in the effect of target covariates on the spatial distribution of points when there are nuisance covariates present. We show the semiparametric efficiency lower bound of estimating this effect and present a cross-fitting estimator that can achieve this lower bound. We demonstrate the performance of our proposed method through a simulation study and a re-analysis of the spatial distribution of tree species, where our approach avoids potential biases from model misspecification compared to a parametric model and retains good efficiency characteristics. More generally, to the best of our knowledge, our work is the first to study a general class of semiparametric intensity functions where parts of the spatial covariates are modeled nonparametrically and present optimal estimators in terms of semiparametric efficiency. However, as laid out in Table 1, there are some remaining challenges, especially when the process is not Poisson without a log-linear intensity function. Regardless, we believe our work presents a new approach to study semiparametric point processes and we hope the techniques in the paper is useful to researchers in point processes and in semiparametric inference.

References

Anthony C Gatrell, Trevor C Bailey, Peter J Diggle, and Barry S Rowlingson. Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British geographers*, pages 256–274, 1996.

Cory P Haberman. Overlapping hot spots? examination of the spatial heterogeneity of hot spots of different crime types. *Criminology & Public Policy*, 16(2):633–660, 2017.

Anjali Rao, Dalia Barkley, Gustavo S França, and Itai Yanai. Exploring tissue architecture using spatial transcriptomics. *Nature*, 596(7871):211–220, 2021.

Stephen L Rathbun and Noel Cressie. Asymptotic properties of estimators for the parameters of spatial inhomogeneous poisson point processes. *Advances in Applied Probability*, 26(1):122–154, 1994.

Frederic Paik Schoenberg. Consistent parametric estimation of the intensity of a spatial–temporal point process. *Journal of Statistical Planning and Inference*, 128(1):79–93, 2005.

Yongtao Guan and Ji Meng Loh. A thinned block bootstrap variance estimation procedure for inhomogeneous spatial point patterns. *Journal of the American Statistical Association*, 102(480):1377–1386, 2007.

Rasmus Waagepetersen and Yongtao Guan. Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(3):685–702, 2009.

Mark Berman and T Rolf Turner. Approximating point process likelihoods with glim. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 41(1):31–38, 1992.

Adrian Baddeley and Rolf Turner. Practical maximum pseudolikelihood for spatial point patterns: (with discussion). *Australian & New Zealand Journal of Statistics*, 42(3):283–322, 2000.

Adrian Baddeley, Jean-François Coeurjolly, Ege Rubak, and Rasmus Waagepetersen. Logistic regression for spatial gibbs point processes. *Biometrika*, 101(2):377–392, 2014.

Adrian Baddeley and Rolf Turner. Spatstat: an r package for analyzing spatial point patterns. *Journal of statistical software*, 12:1–42, 2005.

Peter J. Diggle. A point process modelling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 153(3):349–362, 1990.

Tingjin Chu, Yongtao Guan, Rasmus Waagepetersen, and Ganggang Xu. Quasi-likelihood for multivariate spatial point processes with semiparametric intensity functions. *Spatial Statistics*, 50:100605, 2022.

Kristian Bjørn Hessellund, Ganggang Xu, Yongtao Guan, and Rasmus Waagepetersen. Semiparametric multinomial logistic regression for multivariate point pattern data. *Journal of the American Statistical Association*, 117(539):1500–1515, 2022.

Georgia Papadogeorgou, Kosuke Imai, Jason Lyall, and Fan Li. Causal inference with spatio-temporal data: estimating the effects of airstrikes on insurgent violence in iraq. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(5):1969–1999, 2022.

Zhichao Jiang, Shizhe Chen, and Peng Ding. An instrumental variable method for point

processes: generalized wald estimation based on deconvolution. *Biometrika*, 110(4):989–1008, 2023.

Peter M Robinson. Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, pages 931–954, 1988.

Thomas A Severini and Joan G Staniswalis. Quasi-likelihood estimation in semiparametric models. *Journal of the American statistical Association*, 89(426):501–511, 1994.

Wolfgang Härdle, Hua Liang, and Jiti Gao. *Partially linear models*. Springer Science & Business Media, 2000.

Mark Berman and Peter Diggle. Estimating weighted integrals of the second-order intensity of a spatial point process. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 51(1):81–92, 1989.

Marie-Colette NM van Lieshout. On estimation of the intensity function of a point process. *Methodology and Computing in Applied Probability*, 14:567–578, 2012.

Adrian Baddeley, Ya-Mei Chang, Yong Song, and Rolf Turner. Nonparametric estimation of the dependence of a spatial point process on spatial covariates. *Statistics and Its Interface*, 5(2):221–236, 2012.

Xiwei Tang and Lexin Li. Multivariate temporal point process regression. *Journal of the American Statistical Association*, 118(542):830–845, 2023.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

Sung Nok Chiu, Dietrich Stoyan, Wilfrid S Kendall, and Joseph Mecke. *Stochastic geometry and its applications*. John Wiley & Sons, 2013.

Thomas A Severini and Wing Hung Wong. Profile likelihood and conditionally parametric models. *The Annals of Statistics*, pages 1768–1802, 1992.

Yongtao Guan, Abdollah Jalilian, and Rasmus Waagepetersen. Quasi-likelihood for spatial point processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 77(3):677–697, 2015.

Adrian Baddeley, Ege Rubak, and Rolf Turner. *Spatial Point Patterns: Methodology and Applications with R*. CRC press, 2015.

Jesper Møller and Frederic Paik Schoenberg. Thinning spatial point processes into poisson processes. *Advances in Applied Probability*, 42(2):347–358, 2010.

Jagadeesh Chandramohan, Robert D Foley, and Ralph L Disney. Thinning of point processes—covariance analyses. *Advances in Applied Probability*, 17(1):127–146, 1985.

Renaud Alie, David A Stephens, and Alexandra M Schmidt. On data augmentation in point process models based on thinning procedures. *arXiv preprint arXiv:2203.06743*, 2022.

Geoffrey S Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 26(4):359–372, 1964.

Arkajyoti Saha, Sumanta Basu, and Abhirup Datta. Random forests for spatially dependent data. *Journal of the American Statistical Association*, 118(541):665–683, 2023.

Simon Wood and Maintainer Simon Wood. Package 'mgcv'. *R package version*, 1(29):729, 2015.

Vladimir Igorevich Bogachev and Maria Aparecida Soares Ruas. *Measure theory*, volume 1. Springer, 2007.

Béla A Frigyik, Santosh Srivastava, and Maya R Gupta. An introduction to functional derivatives. *Dept. Electr. Eng., Univ. Washington, Seattle, WA, Tech. Rep*, 1, 2008.

Anastasios A Tsiatis. *Semiparametric theory and missing data*, volume 4. Springer, 2006.

Qi Li and Jeffrey Scott Racine. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, 2023.

Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media, 2007.

Robin B. Hubbell, Stephen P. Foster. Diversity of canopy trees in a neotropical forest and implications for conservation. *Tropical Rain Forest: Ecology and Management*, 1983.

Richard Condit, Stephen P Hubbell, and Robin B Foster. Changes in tree species abundance in a neotropical forest: impact of climate change. *Journal of tropical ecology*, 12(2):231–256, 1996.

Richard Condit. *Tropical forest census plots: methods and results from Barro Colorado Island, Panama and a comparison with other plots*. Springer Science & Business Media, 1998.