Supplementary material for 'Efficient, cross-fitting estimation of semiparametric spatial point processes'

Xindi Lin, Hyunseung Kang

July 2024

1 Guide to Notation

X: spatial point process defined on $\mathbb{R}^2$. A: observational window. $\lambda(\mathbf{u};\boldsymbol{\theta},\eta)$: intensity function of X at location $\mathbf{u}$. $\boldsymbol{\theta} \in \Theta \in \mathbb{R}^k$: target parameter. $\eta \in \mathcal{H}$: nuisance parameter. $g(\mathbf{u},\mathbf{v};\psi)$: pair correlation function of X between two locations $\mathbf{u},\mathbf{v}$. ψ: parameter of the pair correlation function. $\boldsymbol{\theta}^*,\eta^*,\psi^*$: true parameter. $\hat{\boldsymbol{\theta}},\hat{\eta},\hat{\psi}$: estimators.

$\mathbf{y}(\mathbf{u})$: p-dimensional target covariates at location $\mathbf{u}$. $\mathbf{z}(\mathbf{u})$: q-dimensional nuisance covariates at location $\mathbf{u}$. $\mathcal{Y} = \{\mathbf{y}(\mathbf{u}) : \mathbf{u} \in \mathbb{R}^2\}$. $\mathcal{Z} = \{\mathbf{z}(\mathbf{u}) : \mathbf{u} \in \mathbb{R}^2\}$. $\lambda(\mathbf{y},\mathbf{z};\boldsymbol{\theta},\eta)$: intensity function expressed as a function of covariates. When we say intensity function of X is log-linear, it means that $\lambda(\mathbf{u};\boldsymbol{\theta},\eta) = \exp(\boldsymbol{\theta}^\top \mathbf{y}(\mathbf{u}) + \eta(\mathbf{z}(\mathbf{u})))$.

$f(\mathbf{y},\mathbf{z})$: Radon-Nikodym derivative of the mapping $\mathbf{u} \mapsto (\mathbf{u}(\mathbf{u},\mathbf{z}))$ in the observational window A. $\mathbb{E}[\ell(\boldsymbol{\theta},\eta)|\mathbf{z}]$: conditional expectation with respect to fixed $\mathbf{z}$. $\mathbb{E}[\ell(\boldsymbol{\theta},\gamma)|\mathbf{z}]$: conditional expectation with respect to fixed $\mathbf{z}$ and $\eta(\mathbf{z}) = \gamma$. $\hat{\mathbb{E}}[\ell(\boldsymbol{\theta},\gamma)|\mathbf{z}]$, $\hat{\mathbb{E}}[\ell(\boldsymbol{\theta},\eta)|\mathbf{z}]$ are the nonparametric estimators.

$\ell(\boldsymbol{\theta},\eta)$: pseudo-log-likelihood function of X. $\eta^*_{\boldsymbol{\theta}} = \arg\max_{\eta\in\mathcal{H}} \ell(\boldsymbol{\theta},\eta)$. $\boldsymbol{\nu}^*(\mathbf{z}) = \frac{\partial}{\partial\boldsymbol{\theta}}\eta^*_{\boldsymbol{\theta}}(\mathbf{z})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$.

$\boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*) = -\mathbb{E}\left[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\ell(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta}})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right]$. $\boldsymbol{\Sigma}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*, \psi^*) = \mathrm{Var}\left(\frac{\partial}{\partial\boldsymbol{\theta}}\ell(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta}})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right)$. $\hat{\eta}_{\boldsymbol{\theta}}, \hat{\boldsymbol{\nu}}$, $\hat{\boldsymbol{S}}(\cdot,\cdot,\cdot)$, $\hat{\boldsymbol{\Sigma}}(\cdot,\cdot,\cdot,\cdot)$: estimators.

When considering a sequence of observational window $\{A_n\}_{n=1}^\infty$, $\hat{\boldsymbol{\theta}}_n, \hat{\eta}_n, \hat{\psi}_n, \ell_n(\boldsymbol{\theta}, \eta), \boldsymbol{S}_n, \boldsymbol{\Sigma}_n$, $f_n(\mathbf{y}, \mathbf{z}), \eta^*_{\boldsymbol{\theta},n}, \boldsymbol{\nu}^*_n, \hat{\eta}_{\boldsymbol{\theta},n}, \hat{\boldsymbol{\nu}}_n \hat{\boldsymbol{S}}_n, \hat{\boldsymbol{\Sigma}}_n$ indicates the same definition with respect to observational window A_n. $\bar{\boldsymbol{S}}_n = \boldsymbol{S}_n / A_n$. $\bar{\boldsymbol{\Sigma}}_n = \boldsymbol{\Sigma}_n / A_n$.

V: random thinning procedure split the original spatial point process to V subprocesses. V is a fixed integer independent of X. For $v = 1, \cdots, V$, X_v is the v-th subprocess, $X_v^c = \bigcap_{i \neq v} X_j$. $\hat{\boldsymbol{\theta}}^{(v)}$, $\hat{\eta}^{(v)}$, $\hat{\eta}^{(v)}_{\boldsymbol{\theta}}$ are the estimator with respect to the v-th subprocess.

$|\cdot|$: ℓ_2 norm of a vector or the spectral norm of a matrix. $\lambda_{\min}[M]$: the minimal eigenvalue of a matrix M.

2 Algorithms

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 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_v^c:

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

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

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

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

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 $\int_{\mathbb{R}} zk(z)dz=0$.

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 $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}}^{(v)}=\arg\max_{\boldsymbol{\theta}\in\Theta}\ell(\boldsymbol{\theta},\hat{\eta}_{\boldsymbol{\theta}}^{(v)};X_v),\quad \ell(\boldsymbol{\theta},\hat{\eta}_{\boldsymbol{\theta}}^{(v)};X_v)=\sum_{\mathbf{u}\in X_v}\log\left(\frac{1}{V}\lambda(\mathbf{u};\boldsymbol{\theta},\hat{\eta}_{\boldsymbol{\theta}}^{(v)})\right)-\int_A\frac{1}{V}\lambda(\mathbf{u};\boldsymbol{\theta},\hat{\eta}_{\boldsymbol{\theta}}^{(v)})\mathrm{d}\mathbf{u}. \tag{2}$$

3 Assumptions

Assumption 1. Regularity conditions needed for consistency.

1.1 (*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.

1.2 *(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 $|C\cap A_n|=\Theta(|A_n|)$

1.3 (*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.

1.4 (*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.

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

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

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

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:

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

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

5.1 (*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 $\mathbf{z}$

5.2 (*Identification*) $\liminf_n \inf_{\boldsymbol{\theta},\mathbf{z}} |A_n|^{-1} \frac{\partial^2}{\partial \gamma^2} \mathbb{E}[\ell_n(\boldsymbol{\theta}, \gamma)|\mathbf{z}]\Big|_{\gamma = \eta^*_{\boldsymbol{\theta},n}(\mathbf{z})} > 0$

5.3 (*Higher Order Kernel*) Kernel function $K(\cdot)$ in 1 is an l-th order kernel.

5.4 (*Weak Dependence*) There exists positive constant C such that the cumulant functions of X satisfies

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

4 Proof of Proposition 1

Proposition 1 (random thinning)**.** *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) If X is a Poisson spatial point process, then the sub-processes $X_1, .., X_V$ are mutually independent.*

Proof. **Proof of result (i):** For every point $\mathbf{u} \in X$, we let $c(\mathbf{u}) = \sum_{j=1}^{V} j \cdot \mathbf{1}(\mathbf{u} \in X_j)$ indicating which subprocess $\mathbf{u}$ belongs to. Then, for any $A \in \mathbb{R}^2$, $I \in [V]$,

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{\mathbf{u} \in \bigcup X_j} \mathbf{1}(\mathbf{u} \in A)\right] \\
=&\mathbb{E}\left[\sum_{\mathbf{u} \in X} \mathbf{1}(\mathbf{u} \in A)\mathbf{1}(c(\mathbf{u}) \in I)\right] \\
=&\mathbb{E}\left[\mathbb{E}\left[\sum_{\mathbf{u} \in X} \mathbf{1}(\mathbf{u} \in A)\mathbf{1}(c(\mathbf{u}) \in I)\middle| X\right]\right] \\
=&\mathbb{E}\left[\sum_{\mathbf{u} \in X} \mathbb{E}\left[\mathbf{1}(\mathbf{u} \in A)\mathbf{1}(c(\mathbf{u}) \in I)\middle| X\right]\right] \\
=&\mathbb{E}\left[\sum_{\mathbf{u} \in X} \mathbf{1}(\mathbf{u} \in A)\mathbb{E}\left[\mathbf{1}(c(\mathbf{u}) \in I)\middle| X\right]\right] \\
=&\mathbb{E}\left[\sum_{\mathbf{u} \in X} \mathbf{1}(\mathbf{u} \in A)\frac{|I|}{V}\right] \\
=&\frac{|I|}{V}\int_A \lambda(\mathbf{u})\mathrm{d}\mathbf{u}
\end{aligned}
$$

Thus, the intensity function of the superposition point process $\bigcup_{j \in I} X_j$ is $\frac{|I|}{K}\lambda(\mathbf{u})$.

Proof of result (ii):

Thinning the original X into V subprocesses can be implemented recursively by first thinning X into two X_1, X_1^c with retaining probability $\frac{1}{V}$, then thinning X_1^c into X_2, X_2^c with retaining probability $\frac{1}{V-1}$ and repeat this procedure until we get $X_1, \ldots, X_V$.

Thus, it suffices to show that for any $0 < q < 1$, randomly thinning X into two subprocesses X_1, X_2 with retaining probability q gives us two independent Poisson spatial point processes. To do so, we proceed in two steps. Step 1 shows X_1, X_2 are Poisson by deriving their density functions. Step 2 shows their independence by illustrating the joint density of X_1, X_2 equals the multiplication of the density of X_1 and X_2.

Step 1: For brevity, we denote $|X|$ as the cardinality of X, i.e. the number of observed points, denote $X = \{\mathbf{u}_1, \ldots, \mathbf{u}_{|X|}\}$, and denote $\mu(A) = \int_A \lambda(\mathbf{u})d\mathbf{u}$.

When X is Poisson, $|X|$ follows the Poisson distribution with expectation $\mu(A)$ and points $\{\mathbf{u}_1, \ldots, \mathbf{u}_{|X|}\}$ are independently identically distributed with density function $\lambda(\mathbf{u})/\mu(A)$ conditioning on $|X|$. Thus, the probability density function of X is

$$\begin{aligned} p(X) =& p(X \mid |X|)\mathbb{P}(|X|) \\ =& \left(\prod_{i=1}^{|X|} \frac{\lambda(\mathbf{u}_i)}{\mu(A)}\right) \frac{\exp(-\mu(A))\mu(A)^{|X|}}{|X|!} \end{aligned}$$

Then, we will show that the probability density functions of X_1, X_2 are consistent with the above form. First, we consider the probability mass function of $|X_1|$ is

$$\begin{aligned} \mathbb{P}(|X_1|) &= \sum_{n=|X_1|}^{\infty} \mathbb{P}(|X_1| \mid |X| = n)\mathbb{P}(|X| = n) \\ &= \sum_{n=|X_1|}^{\infty} \binom{n}{|X_1|} q^{|X_1|}(1-q)^{n-|X_1|} \frac{\exp(-\mu(A))(\mu(A))^n}{n!} \\ &= \frac{\exp(-\mu(A))}{|X_1|!} (\mu(A)q)^{|X_1|} \sum_{n=|X_1|}^{\infty} \frac{((1-q)\mu(A))^{(n-|X_1|)}}{(n-|X_1|)!} \\ &= \frac{\exp(-q\mu(A))}{|X_1|!} (q\mu(A))^{|X_1|} \end{aligned}$$

By the result in (i), the intensity function of X_1 is $q(\lambda(\mathbf{u}))$. Since the thinning is independent with points in X, the retained points in X_1 are still independent and are identically distributed with density $\lambda(\mathbf{u}_i)/q\mu(A)$ conditioning on X_1. Therefore,

$$
\begin{aligned}
p(X_1) =& p(X_1 \mid |X_1|)\mathbb{P}(|X_1|) \\
=& \left(\prod_{i=1}^{|X_1|} \frac{\lambda(\mathbf{u}_i)}{q\mu(A)}\right) \frac{\exp(-q\mu(A))(q\mu(A))^{|X_1|}}{|X_1|!}
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
p(X_2) =& p(X_2 \mid |X_2|)\mathbb{P}(|X_2|) \\
=& \left(\prod_{i=1}^{|X_2|} \frac{\lambda(\mathbf{u}_i)}{(1-q)\mu(A)}\right) \frac{\exp(-(1-q)\mu(A))((1-q)\mu(A))^{|X_2|}}{|X_2|!}
\end{aligned}
$$

Thus, X_1 and X_2 are Poisson spatial point processes.

Step 2: Then, we consider the joint probability mass function of $|X_1|$ and $|X_2|$,

$$
\begin{aligned}
& \mathbb{P}(|X_1|, |X_2|) \\
=& \mathbb{P}(|X_1|, |X_2| \mid |X_1 \cup X_2|)\mathbb{P}(|X_1 \cup X_2|) \\
=& \mathbb{P}(|X_1| \mid |X_1 \cup X_2|)\mathbb{P}(|X_1 \cup X_2|) \\
=& \binom{|X_1 \cup X_2|}{|X_1|} q^{|X_1|}(1-q)^{|X_2|} \frac{\exp(-\mu(A))(\mu(A))^{|X_1+X_2|}}{|X_1+X_2|!} \\
=& \frac{|X_1 \cup X_2|!}{|X_1|!|X_2|!} q^{|X_1|}(1-q)^{|X_2|} \frac{\exp(-\mu(A))(\mu(A))^{|X_1+X_2|}}{|X_1+X_2|!} \\
=& \frac{\exp(-q\mu(A))(q\mu(A))^{|X_1|}}{|X_1|!} \cdot \frac{\exp(-(1-q)\mu(A))((1-q)\mu(A))^{|X_2|}}{|X_2|!} \\
=& \mathbb{P}(|X_1|)\mathbb{P}(|X_2|)
\end{aligned}
$$

Conditioning on $|X_1|, |X_2|$, the points in X_1 and points in X_2 are independent, thus we have

$$
\begin{aligned}
p(X_1, X_2) =& p(X_1, X_2 \mid |X_1|, |X_2|)\mathbb{P}(|X_1|, |X_2|) \\
=& p(X_1 \mid |X_1|, |X_2|)p(X_2 \mid |X_1|, |X_2|)\mathbb{P}(|X_1|, |X_2|) \\
=& p(X_1 \mid |X_1|)p(X_2 \mid |X_2|)\mathbb{P}(|X_1|)\mathbb{P}(|X_2|) \\
=& p(X_1)p(X_2)
\end{aligned}
$$

Therefore, X_1 and X_2 are independent Poisson spatial point processes.

□

5 Proof of Theorem 5.1

Theorem 5.1 (Semiparametric Efficiency Lower Bound for Semiparametric Poisson Process)**.** *The minimum of the sensitivity matrix*

$$
\boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}) = \int_A \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) \left(\frac{\partial}{\partial \boldsymbol{\theta}} \log \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) + \frac{\partial}{\partial \eta} \log \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) \left[\boldsymbol{\nu}(\mathbf{z}(\mathbf{u}))\right] \right)^{\otimes 2} \mathrm{d}\mathbf{u}, \tag{3}
$$

is attained by $\boldsymbol{\nu}^*$ satisfying

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

Furthermore, if the intensity function is log-linear, i.e., $\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) = \exp(\boldsymbol{\theta}^\top \mathbf{y}(\mathbf{u}) + \eta(\mathbf{z}(\mathbf{u})))$, then $\boldsymbol{\nu}^*$ simplifies to

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

Proof. The proof of the theorem consists of three steps. In Step 1, we construct a linear space with the inner product and the outer product on it such that $\boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu})$ is equal to the outer product of

$$|f\rangle\langle g|_A = \int_A \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) f(\mathbf{u}) g(\mathbf{u})^\top \mathrm{d}\mathbf{u} \tag{4}$$

In step 2, we show that $\boldsymbol{\nu}^*$ minimizes the inner product of (6). In step 3, we show that if $\boldsymbol{\nu}^*$ minimizes the inner product of (6), it minimizes the outer product of (6), i.e., the sensitivity matrix, as well.

Step 1: We denote $L_2(A)$ as the linear space consisting of all bounded functions from A to $\mathbb{R}$. We denote $L_2^{(k)}(A) = L_2(A) \times \ldots L_2(A)$ as the linear space consisting of all bounded functions from A to $\mathbb{R}^k$. We equip the linear space $L_2^{(k)}(A)$ with the inner-product $\langle \cdot, \cdot \rangle_A$ and the outer-product $|\cdot\rangle\langle\cdot|_A$ defined as follows: for any $f, g \in L_2^{(k)}(A)$,

$$\langle f, g\rangle_A = \int_A \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) f(\mathbf{u})^\top g(\mathbf{u}) \mathrm{d}\mathbf{u} \tag{5}$$

$$|f\rangle\langle g|_A = \int_A \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) f(\mathbf{u}) g(\mathbf{u})^\top \mathrm{d}\mathbf{u} \tag{6}$$

$\langle f, f\rangle_A$ is referred to as the inner-product of f. $|f\rangle\langle f|_A$ is referred to as the outer-product of f. Then the sensitivity matrix (3) is the outer-product of the following function

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) + \frac{\partial}{\partial \eta} \log \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) \left[\boldsymbol{\nu}(\mathbf{z}(\mathbf{u}))\right] \tag{7}$$

Since $\boldsymbol{\nu}$ is defined via $\boldsymbol{\nu}(\mathbf{z}) = \frac{\partial}{\partial \boldsymbol{\theta}} \eta_{\boldsymbol{\theta}}(\mathbf{z})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$ where $\boldsymbol{\theta} \to \eta_{\boldsymbol{\theta}}$ is a twice continuously differentiable mapping from $\mathbb{R}^k$ to $\mathcal{H}$, thus $\boldsymbol{\nu}$ belongs to $\mathcal{H}^k = \mathcal{H} \times \ldots \times \mathcal{H}$, which is the product space of k nuisance parameter space. In the following section, we will first show that $\boldsymbol{\nu}^*(\mathbf{z})$ defined in Theorem 5.1 minimize the inner-product of (7) among all $\boldsymbol{\nu} \in \mathcal{H}^k$. Secondly, we will show that $\boldsymbol{\nu}^*(\mathbf{z})$ minimizes the outer product of (7), i.e., the sensitivity matrix, as well.

Step 2: The Gateaux derivative

$$\frac{\partial}{\partial \eta}\log\lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*)\left[\boldsymbol{\nu}(\mathbf{z}(\mathbf{u}))\right]=\frac{\partial}{\partial \eta(\mathbf{z}(\mathbf{u}))}\log\lambda(\tau_{\boldsymbol{\theta}^*}(\mathbf{y}(\mathbf{u})),\eta(\mathbf{z}(\mathbf{u})))\cdot\boldsymbol{\nu}(\mathbf{z}(\mathbf{u}))$$

Since $\mathcal{H}^k$ is linear, the following set

$$\left\{\frac{\partial}{\partial \eta}\log\lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*)\left[\boldsymbol{\nu}(\mathbf{z}(\mathbf{u}))\right]:\boldsymbol{\nu}\in\mathcal{H}^k\right\}$$

constitutes a sub-linear space of $L_2^{(k)}(A)$. Therefore, by the projection Theorem in Hilbert space (Theorem 2.1 in Tsiatis [2006]), there exists a unique $\boldsymbol{\nu}^*\in\mathcal{H}^k$ minimizing the inner-product of (7) which satisfies the orthogonal condition as follows: for any $\boldsymbol{\nu}\in\mathcal{H}^k$,

$$\left\langle\frac{\partial}{\partial\boldsymbol{\theta}}\log\lambda(\cdot;\boldsymbol{\theta}^*,\eta^*)+\frac{\partial}{\partial\eta}\log\lambda(\cdot;\boldsymbol{\theta}^*,\eta^*)\left[\boldsymbol{\nu}^*(\mathbf{z}(\cdot))\right],\frac{\partial}{\partial\eta}\log\lambda(\cdot;\boldsymbol{\theta}^*,\eta^*)\left[\boldsymbol{\nu}(\mathbf{z}(\cdot))\right]\right\rangle_A=0$$

We rearrange the orthogonal condition as follows:

$$\begin{aligned}0&=\left\langle\frac{\partial}{\partial\boldsymbol{\theta}}\log\lambda(\cdot;\boldsymbol{\theta}^*,\eta^*)+\frac{\partial}{\partial\eta}\log\lambda(\cdot;\boldsymbol{\theta}^*,\eta^*)\left[\boldsymbol{\nu}^*(\mathbf{z}(\cdot))\right],\frac{\partial}{\partial\eta}\log\lambda(\cdot;\boldsymbol{\theta}^*,\eta^*)\left[\boldsymbol{\nu}(\mathbf{z}(\cdot))\right]\right\rangle_A \qquad (8)\\&=\int_A\left(\frac{\partial}{\partial\boldsymbol{\theta}}\log\lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*)+\frac{\partial}{\partial\eta}\log\lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*)\left[\boldsymbol{\nu}^*(\mathbf{z}(\mathbf{u}))\right]\right)^\top\frac{\partial}{\partial\eta}\log\lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*)\left[\boldsymbol{\nu}(\mathbf{z}(\mathbf{u}))\right]\lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*)\mathrm{d}\mathbf{u}\\&=\int_{\mathcal{Z}}\int_{\mathcal{Y}}\left(\frac{\partial}{\partial\boldsymbol{\theta}}\log\lambda(\mathbf{y},\mathbf{z};\boldsymbol{\theta}^*,\eta^*)+\frac{\partial}{\partial\eta}\log\lambda(\mathbf{y},\mathbf{z};\boldsymbol{\theta}^*,\eta^*)\boldsymbol{\nu}^*(\mathbf{z})\right)^\top\\&\quad\frac{\partial}{\partial\eta}\log\lambda(\mathbf{y},\mathbf{z};\boldsymbol{\theta}^*,\eta^*)\boldsymbol{\nu}(\mathbf{z})\lambda(\mathbf{y},\mathbf{z};\boldsymbol{\theta}^*,\eta^*)f(\mathbf{y},\mathbf{z})\mathrm{d}\mathbf{y}\mathrm{d}\mathbf{z}\end{aligned}$$

The above orthogonality condition will be satisfied if for every $\mathbf{z}\in\mathcal{Z}$,

$$\int_{\mathcal{Y}}\left(\frac{\partial}{\partial\boldsymbol{\theta}}\log\lambda(\mathbf{y},\mathbf{z};\boldsymbol{\theta}^*,\eta^*)+\frac{\partial}{\partial\eta}\log\lambda(\mathbf{y},\mathbf{z};\boldsymbol{\theta}^*,\eta^*)\boldsymbol{\nu}^*(\mathbf{z})\right)^\top\frac{\partial}{\partial\eta}\log\lambda(\mathbf{y},\mathbf{z};\boldsymbol{\theta}^*,\eta^*)f(\mathbf{y},\mathbf{z})\lambda(\mathbf{y},\mathbf{z};\boldsymbol{\theta}^*,\eta^*)\mathrm{d}\mathbf{y}=0$$

which is equivalent to

$$\int_{\mathcal{Y}} \frac{\partial}{\partial \boldsymbol{\theta}^\top} \log \lambda(\mathbf{y}, \mathbf{z}; \boldsymbol{\theta}^*, \eta^*) \frac{\partial}{\partial \eta} \log \lambda(\mathbf{y}, \mathbf{z}; \boldsymbol{\theta}^*, \eta^*) \lambda(\mathbf{y}, \mathbf{z}; \boldsymbol{\theta}^*, \eta^*) f(\mathbf{y}, \mathbf{z}) \mathrm{d}\mathbf{y}$$
$$= - \int_{\mathcal{Y}} \boldsymbol{\nu}^{\boldsymbol{*}}(\mathbf{z})^\top \frac{\partial}{\partial \eta} \log \lambda(\mathbf{y}, \mathbf{z}; \boldsymbol{\theta}^*, \eta^*) \frac{\partial}{\partial \eta} \log \lambda(\mathbf{y}, \mathbf{z}; \boldsymbol{\theta}^*, \eta^*) \lambda(\mathbf{y}, \mathbf{z}; \boldsymbol{\theta}^*, \eta^*) f(\mathbf{y}, \mathbf{z}) \mathrm{d}\mathbf{y}$$

Wen can thus derive the formula of $\boldsymbol{\nu}^*(\mathbf{z})$ as follows

$$\begin{aligned} \boldsymbol{\nu}^*(\mathbf{z}) &= - \frac{\int_{\mathcal{Y}} \frac{\partial}{\partial \boldsymbol{\theta}} \log \lambda(\mathbf{y}, \mathbf{z}; \boldsymbol{\theta}^*, \eta^*) \frac{\partial}{\partial \eta} \log \lambda(\mathbf{y}, \mathbf{z}; \boldsymbol{\theta}^*, \eta^*) \lambda(\mathbf{y}, \mathbf{z}; \boldsymbol{\theta}^*, \eta^*) f(\mathbf{y}, \mathbf{z}) \mathrm{d}\mathbf{y}}{\int_{\mathcal{Y}} \frac{\partial}{\partial \eta} \log \lambda(\mathbf{y}, \mathbf{z}; \boldsymbol{\theta}^*, \eta^*) \frac{\partial}{\partial \eta} \log \lambda(\mathbf{y}, \mathbf{z}; \boldsymbol{\theta}^*, \eta^*) \lambda(\mathbf{y}, \mathbf{z}; \boldsymbol{\theta}^*, \eta^*) f(\mathbf{y}, \mathbf{z}) \mathrm{d}\mathbf{y}} \\ &= - \frac{\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \eta} \mathbb{E}[\ell(\boldsymbol{\theta}^*, \eta^*(\mathbf{z}))]}{\frac{\partial^2}{\partial \eta^2} \mathbb{E}[\ell(\boldsymbol{\theta}^*, \eta^*(\mathbf{z}))]} \end{aligned}$$

Thus, the $\boldsymbol{\nu}^*(\mathbf{z})$ defined in Theorem 5.1 minimize the inner-product.

Step 3: To show the equivalence between the minimization of inner-product and outer-product, we will first prove the Multivariate Pythagorean theorem in Hilbert space $L_2(A)$, which is a generalization of Theorem 3.3 in Tsiatis [2006].

Theorem 5.2 (Multivariate Pythagorean theorem)**.** *Let $\mathcal{U}$ be a sub-linear space of $L_2(A)$, and $\mathcal{U}^k$ will be a sub-linear space of $L_2^{(k)}(A)$. If $f \in \mathcal{U}^k$, and $g \in L_2^{(k)}(A)$ is orthogonal to $\mathcal{U}^k$, we have*

$$|f+g\rangle \langle f+g| = |f\rangle \langle f| + |g\rangle \langle g|$$

Proof. We denote the k-dimensional element g as $g = (g_1, \ldots, g_k)$, $f = (f_1, \ldots, f_k)$. g is orthogonal to $\mathcal{U}^k$ if and only if its each element $g_j, j = 1, \ldots, k$ is orthogonal to $\mathcal{U}$. Consequently, for any $i, j \in [k]$, we have

$$\int_A \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) f_i(\mathbf{u}) g_j(\mathbf{u}) \mathrm{d}\mathbf{u} = 0$$

and thus

$$|f\rangle\langle g| = \int_A \lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*) f(\mathbf{u}) g(\mathbf{u})^\top \mathrm{d}\mathbf{u} = 0^{k\times k}$$

therefore

$$\begin{aligned}|f+g\rangle\langle f+g| &= |f\rangle\langle f| + |g\rangle\langle g| + 2|f\rangle\langle g| \\ &= |f\rangle\langle f| + |g\rangle\langle g|\end{aligned}$$

$\square$

Denote $\mathcal{U} = \left\{\frac{\partial}{\partial\eta}\log\lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*)\left[\nu(\mathbf{z}(\mathbf{u}))\right] : \nu\in\mathcal{H}\right\}$, then we have

$$\left\{\frac{\partial}{\partial\eta}\log\lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*)\left[\boldsymbol{\nu}(\mathbf{z}(\mathbf{u}))\right] : \boldsymbol{\nu}\in\mathcal{H}^k\right\} = \mathcal{U}^k \subset L_2^{(k)}(A)$$

For any $\boldsymbol{\nu}\in\mathcal{H}^k$, we define

$$f = \frac{\partial}{\partial\eta}\log\lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*)\left[\boldsymbol{\nu}(\mathbf{z}(\mathbf{u})) - \boldsymbol{\nu}^*(\mathbf{z}(\mathbf{u}))\right]$$

$$g = \frac{\partial}{\partial\boldsymbol{\theta}}\log\lambda(\cdot;\boldsymbol{\theta}^*,\eta^*) + \frac{\partial}{\partial\eta}\log\lambda(\cdot;\boldsymbol{\theta}^*,\eta^*)\left[\boldsymbol{\nu}^*(\mathbf{z}(\cdot))\right]$$

$f\in\mathcal{U}^k$, $g\in L_2^{(k)}(A)$, f,g are orthogonal according the orthogonality condition in (8). Then, we apply the Multivariate Pythagorean theorem to f and g and obtain

$$\begin{aligned}&\boldsymbol{S}(\boldsymbol{\theta}^*,\eta^*,\boldsymbol{\nu}) - \boldsymbol{S}(\boldsymbol{\theta}^*,\eta^*,\boldsymbol{\nu}^*) = \\ &\quad\left|\frac{\partial}{\partial\eta}\log\lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*)\left[\boldsymbol{\nu}(\mathbf{z}(\mathbf{u})) - \boldsymbol{\nu}^*(\mathbf{z}(\mathbf{u}))\right]\right\rangle\left\langle\frac{\partial}{\partial\eta}\log\lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*)\left[\boldsymbol{\nu}(\mathbf{z}(\mathbf{u})) - \boldsymbol{\nu}^*(\mathbf{z}(\mathbf{u}))\right]\right|_A\end{aligned}$$

where the right-hand-side is always definitely positive when $\boldsymbol{\nu}^*$ is not equal to $\boldsymbol{\nu}$. Therefore,

$\boldsymbol{\nu}^*$ minimizes $\boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu})$. $\square$

6 Proof of Theorem 6.1

Theorem 6.1 (Consistency of $\hat{\boldsymbol{\theta}}$)**.** *Suppose conditions 1.1-1.4 in Assumption 1 and Assumption 3 hold. Then $\hat{\boldsymbol{\theta}}_n$ is consistent for $\boldsymbol{\theta}^*$, i.e., $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^* \to_p \mathbf{0}$.*

Proof. Since

$$\hat{\boldsymbol{\theta}}_n = V^{-1} \sum_{v=1}^{V} \hat{\boldsymbol{\theta}}_n^{(v)}$$

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

where $\ell_n(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta},n}^{(v)}; X_v)$ is the log-pseudo-likelihood of the subprocess X_v and $\hat{\eta}_{\boldsymbol{\theta}}^{(v)}$ is an estimator of $\eta_{\boldsymbol{\theta},n}^*$ satisfies Assumption 3. If we can show that $\hat{\boldsymbol{\theta}}_n^{(v)}$ is a consistent estimator of $\boldsymbol{\theta}^*$ for every $v \in [V]$, $\hat{\boldsymbol{\theta}}_n$ will be a consistent estimator of $\boldsymbol{\theta}^*$ as well.

For the convenience of the proof, we redefine $\hat{\boldsymbol{\theta}}_n$ as

$$\hat{\boldsymbol{\theta}}_n := \arg\max_{\boldsymbol{\theta} \in \Theta} \ell_n(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta},n}) \tag{9}$$

in the following proof where $\hat{\eta}_{\boldsymbol{\theta},n}$ is the estimators of $\eta_{\boldsymbol{\theta},n}^*$.

Since X_v is a spatial point process with intensity function $V^{-1}\lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*)$ and $\eta_{\boldsymbol{\theta},n}^* = \arg\max \ell_n(\boldsymbol{\theta}, \eta; X_v)$ by Proposition 1, it suffices to show that the estimator $\hat{\boldsymbol{\theta}}$ defined in (9) is a consistent estimator of $\boldsymbol{\theta}^*$ if $\hat{\eta}_{\boldsymbol{\theta},n}$ satisfies the condition in Assumption 3.

To do so, we proceed in five steps. Step 1 uses the sufficient separation Lemma 2 to bound the estimation error $|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*|$ by the plug-in error $\mathbb{E}\left[\ell(\boldsymbol{\theta}^*, \eta_{\boldsymbol{\theta}^*,n}^*)\right] - \mathbb{E}\left[\ell(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_{\hat{\boldsymbol{\theta}}_n,n})\right]$ and decompose the the plug-in error into three remainder terms. Step 2 bound the remainder terms with different auxiliary lemmas.

Step 1: When condition 1.1, 1.2, 1.3 in Assumption 1 are satisfied, Lemma 2 states that for any $(\boldsymbol{\theta}, \eta) \neq (\boldsymbol{\theta}^*, \eta^*)$,

$$\Theta(|A_n|) \cdot \min\{|\boldsymbol{\theta} - \boldsymbol{\theta}^*|, 1\} = \mathbb{E}[\ell_n(\boldsymbol{\theta}^*, \eta^*)] - \mathbb{E}[\ell_n(\boldsymbol{\theta}, \eta)]$$

Since $\eta^*_{\boldsymbol{\theta}^*,n} = \eta^*$, we apply Lemma 2 for $(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_{\hat{\boldsymbol{\theta}}_n,n}))$ and we will have

$$\Theta(|A_n|) \min(1, |\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*|) = \mathbb{E}\left[\ell(\boldsymbol{\theta}^*, \eta^*_{\boldsymbol{\theta}^*,n})\right] - \mathbb{E}\left[\ell(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_{\hat{\boldsymbol{\theta}}_n,n})\right]$$

We decompose and upper bound the plug-in error in the right-hand side as follows:

$$\begin{aligned}\mathbb{E}\left[\ell(\boldsymbol{\theta}^*, \eta^*_{\boldsymbol{\theta}^*,n})\right] - \mathbb{E}\left[\ell(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_{\hat{\boldsymbol{\theta}}_n,n})\right] \leq \left|\mathbb{E}\left[\ell(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_{\hat{\boldsymbol{\theta}}_n,n})\right] - \ell(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_{\hat{\boldsymbol{\theta}}_n,n})\right| + \left|\ell(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_{\hat{\boldsymbol{\theta}}_n,n}) - \sup_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n})\right| + \\ + \left|\mathbb{E}\left[\ell(\boldsymbol{\theta}^*, \eta^*_{\boldsymbol{\theta}^*,n})\right] - \sup_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n})\right|\end{aligned}$$

We denote

$$\begin{aligned}R_1 &:= \left|\mathbb{E}\left[\ell(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_{\hat{\boldsymbol{\theta}}_n,n})\right] - \ell(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_{\hat{\boldsymbol{\theta}}_n,n})\right| \\ R_2 &:= \left|\ell(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_{\hat{\boldsymbol{\theta}}_n,n}) - \sup_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n})\right| \\ R_3 &:= \left|\mathbb{E}\left[\ell(\boldsymbol{\theta}^*, \eta^*_{\boldsymbol{\theta}^*,n})\right] - \sup_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n})\right|\end{aligned}$$

To show the consistency of $\hat{\boldsymbol{\theta}}_n$, it suffices to show that $R_i = o_p(|A_n|), i = 1, 2, 3$.

Step 2: When condition 1.1, 1.3, 1.4 in Assumption 1 are satisfied, equation (49) in

Corollary 1 states that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta},\eta\in\mathcal{H}} |\ell_n(\boldsymbol{\theta},\eta) - \mathbb{E}\left[\ell_n(\boldsymbol{\theta},\eta)\right]| = O_p(|A_n|^{\frac{1}{2}})$$

Thus,

$$\begin{aligned} R_1 :=& \left|\mathbb{E}\left[\ell(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_{\hat{\boldsymbol{\theta}}_n,n})\right] - \ell(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_{\hat{\boldsymbol{\theta}}_n,n})\right| \\ \leq& \sup_{\boldsymbol{\theta}\in\boldsymbol{\theta},\eta\in\mathcal{H}} |\ell_n(\boldsymbol{\theta},\eta) - \mathbb{E}\left[\ell_n(\boldsymbol{\theta},\eta)\right]| \\ =& O_p(|A_n|^{\frac{1}{2}}) = o_p(|A_n|) \end{aligned}$$

Step 3: When condition 1.1, 1.3, 1.4 in Assumption 1 are satisfied, Lemma 3 states that

$$\sup \left|\ell_n(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta},n}) - \ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n})\right| = O_p(|A_n| e_n^{(1)})$$

Moreover, when $\hat{\eta}_{\boldsymbol{\theta},n}$ satisfies Assumption 3,

$$e_n^{(1)} = o(1)$$

Since $\ell(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_{\hat{\boldsymbol{\theta}}_n,n}) = \sup_{\boldsymbol{\theta}\in\Theta} \ell(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta},n})$, we thus have

$$\begin{aligned} R_2 :=& \left|\ell(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_{\hat{\boldsymbol{\theta}}_n,n}) - \sup_{\boldsymbol{\theta}\in\Theta} \ell(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n})\right| \\ =& \left|\sup_{\boldsymbol{\theta}\in\Theta} \ell(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta},n}) - \sup_{\boldsymbol{\theta}\in\Theta} \ell(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n})\right| \\ \leq& \sup_{\boldsymbol{\theta}\in\Theta} \left|\ell(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta},n}) - \ell(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n})\right| \\ =& O_p(|A_n| e_n^{(1)}) = o_p(|A_n|) \end{aligned}$$

Step 4: When condition 1.1, 1.3, 1.4 in Assumption 1 are satisfied, equation (49) in Corollary 1 states that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta},\eta\in\mathcal{H}} \left|\ell_n(\boldsymbol{\theta},\eta) - \mathbb{E}\left[\ell_n(\boldsymbol{\theta},\eta)\right]\right| = O_p(|A_n|^{\frac{1}{2}})$$

Since $\mathbb{E}\left[\ell(\boldsymbol{\theta}^*, \eta^*_{\boldsymbol{\theta}^*,n})\right] = \sup_{\boldsymbol{\theta}\in\Theta} \ell(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n})$, we thus have

$$\begin{aligned}
R_3 :=& \left|\mathbb{E}\left[\ell(\boldsymbol{\theta}^*, \eta^*_{\boldsymbol{\theta}^*,n})\right] - \sup_{\boldsymbol{\theta}\in\Theta} \ell(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n})\right| \\
=& \left|\sup_{\boldsymbol{\theta}\in\Theta} \mathbb{E}\left[\ell(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n})\right] - \sup_{\boldsymbol{\theta}\in\Theta} \ell(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n})\right| \\
\leq& \sup_{\boldsymbol{\theta}\in\Theta} \left|\mathbb{E}\left[\ell(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n})\right] - \ell(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n})\right| \\
\leq& \sup_{\boldsymbol{\theta}\in\boldsymbol{\theta},\eta\in\mathcal{H}} \left|\ell_n(\boldsymbol{\theta},\eta) - \mathbb{E}\left[\ell_n(\boldsymbol{\theta},\eta)\right]\right| \\
=& O_p(|A_n|^{\frac{1}{2}}) = o_p(|A_n|)
\end{aligned}$$

Therefore, when Assumption 1 and Assumption 3 are satisfied, $\hat{\boldsymbol{\theta}}_n$ in Algorithm 1 is a consistent estimator of $\boldsymbol{\theta}^*$. □

7 Proof of Theorem 7.1

Theorem 7.1 (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.*

Proof. Central Limit Theorem Lemma 6 states that under condition 1.1, 2.1, 2.2, 2.3,

$$|A_n|^{-\frac{1}{2}} \bar{\boldsymbol{\Sigma}}_n^{-\frac{1}{2}}(\boldsymbol{\theta}^*, \eta^*, \nu^*, \psi^*) \frac{\partial}{\partial \boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}) \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \to_d N(\mathbf{0}, \mathbf{I}_k)$$

Therefore, it suffices to show that

$$|A_n|^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) = -\left(\bar{\boldsymbol{S}}_n(\boldsymbol{\theta}^*, \eta^*, \nu^*)\right)^{-1} |A_n|^{-\frac{1}{2}} \frac{\partial}{\partial \boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n})\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} + o_p(1)$$

To do so, we proceed in five steps. Step 1 shows the main argument, and Steps 2–5 present auxiliary calculations.

Step 1: Algorithm 1 states that

$$\hat{\boldsymbol{\theta}}_n = V^{-1} \sum_{v=1}^{V} \hat{\boldsymbol{\theta}}_n^{(v)}$$
$$\hat{\boldsymbol{\theta}}_n^{(v)} = \arg\max_{\boldsymbol{\theta}\in\Theta} \ell_n(\boldsymbol{\theta}, \hat{\eta}^{(v)}_{\boldsymbol{\theta},n}; X_v)$$

where $\ell_n(\boldsymbol{\theta}, \hat{\eta}^{(v)}_{\boldsymbol{\theta},n}; X_v)$ is the log-pseudo-likelihood of the subprocess X_v and $\hat{\eta}^{(v)}_{\boldsymbol{\theta}}$ is an estimator of $\eta^*_{\boldsymbol{\theta},n}$ satisfies Assumption 3 and Assumption 4. For every $v \in [V]$, by Taylor Theorem, there exists a sequence $\tilde{\boldsymbol{\theta}}_n^{(v)} \in [\boldsymbol{\theta}^*, \hat{\boldsymbol{\theta}}_n^{(v)}]$ that

$$\begin{aligned} 0 &= \frac{\partial}{\partial \boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}, \hat{\eta}^{(v)}_{\boldsymbol{\theta},n}; X_v)\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n^{(v)}} \\ &= \frac{\partial}{\partial \boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}, \hat{\eta}^{(v)}_{\boldsymbol{\theta},n}; X_v)\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} + \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ell_n(\boldsymbol{\theta}, \hat{\eta}^{(v)}_{\boldsymbol{\theta},n}; X_v)\bigg|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_n^{(v)}} (\hat{\boldsymbol{\theta}}_n^{(v)} - \boldsymbol{\theta}^*) \end{aligned} \tag{10}$$

Denote

$$\hat{\boldsymbol{S}}_n^{(v)} := |A_n|^{-1} \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ell_n(\boldsymbol{\theta}, \hat{\eta}^{(v)}_{\boldsymbol{\theta},n}; X_v)\bigg|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_n^{(v)}}$$
$$\bar{\boldsymbol{S}}_n^{(v)} := |A_n|^{-1} \mathbb{E}\left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}; X_v)\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right]$$
$$R_{1,n}^{(v)} := \hat{\boldsymbol{S}}_n^{(v)} - \bar{\boldsymbol{S}}_n^{(v)}$$
$$R_{2,n}^{(v)} := \frac{\partial}{\partial \boldsymbol{\theta}} \left(\ell_n(\boldsymbol{\theta}, \hat{\eta}^{(v)}_{\boldsymbol{\theta},n}; X_v) - \ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}; X_v)\right)\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$$

In Steps 2-5 respectively, we will show that

$$|R_{1,n}^{(v)}| = o_p(1) \tag{11}$$

$$|R_{2,n}^{(v)}| = o_p(|A_n|^{\frac{1}{2}}) \tag{12}$$

$$\left| |A_n|^{-\frac{1}{2}} \frac{\partial}{\partial \boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta},n}^{(v)}; X_v) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right| = O_p(1) \tag{13}$$

$$\left(\bar{\boldsymbol{S}}_n^{(v)}\right)^{-1} = O(1) \tag{14}$$

Because the infimum limit of the smallest singular values of $\bar{\boldsymbol{S}}_n(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*)$ is bounded below from zero, it follows from (11) that with probability $1 - o(1)$, the infimum limit of the smallest singular values of $|A_n|^{-1} \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ell_n(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta},n}^{(v)}; X_v)\Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_n^{(v)}}$ is bounded below from zero as well. Then with the same $1 - o(1)$ probability, we rearrange (7) and have

$$\begin{aligned} |A_n|^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n^{(v)} - \boldsymbol{\theta}^*) &= -\left\{ |A_n|^{-1} \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ell_n(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta},n}^{(v)}; X_v)\Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_n^{(v)}} \right\}^{-1} \frac{\partial}{\partial \boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta},n}^{(v)}; X_v)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \\ &= -\left(\bar{\boldsymbol{S}}_n^{(v)} + R_{1,n}^{(v)}\right)^{-1} |A_n|^{-\frac{1}{2}} \left(\frac{\partial}{\partial \boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}, \eta_{\boldsymbol{\theta},n}^*; X_v)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} + R_{2,n}^{(v)} \right) \end{aligned} \tag{15}$$

In addition, by Woodbury matrix identity,

$$\left(\bar{\boldsymbol{S}}_n^{(v)} + R_{1,n}^{(v)}\right)^{-1} - \left(\bar{\boldsymbol{S}}_n^{(v)}\right)^{-1} = -\left(\bar{\boldsymbol{S}}_n^{(v)} + R_{1,n}^{(v)}\right)^{-1} R_{1,n}^{(v)} \left(\bar{\boldsymbol{S}}_n^{(v)}\right)^{-1}$$

Then it follows from (11) that

$$\begin{aligned} &\left| \left(\bar{\boldsymbol{S}}_n^{(v)} + R_{1,n}^{(v)}\right)^{-1} - \left(\bar{\boldsymbol{S}}_n^{(v)}\right)^{-1} \right| \\ \le &\left| \left(\bar{\boldsymbol{S}}_n^{(v)} + R_{1,n}^{(v)}\right)^{-1} \right| \times \left| R_{1,n}^{(v)} \right| \times \left| \left(\bar{\boldsymbol{S}}_n^{(v)}\right)^{-1} \right| \\ = &O_p(1) \times o_p(1) \times O(1) = o_p(1) \end{aligned} \tag{16}$$

It follows from (12) and (13) that

$$|A_n|^{-\frac{1}{2}} \left| \frac{\partial}{\partial \boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}; X_v) \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} + R^{(v)}_{2,n} \Bigg| \leq |A_n|^{-\frac{1}{2}} \left| \frac{\partial}{\partial \boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right| + |A_n|^{-\frac{1}{2}} |R_2|$$
$$= O_p(1) + o_p(1) = O_p(1) \tag{17}$$

Combine (16) and (17) gives

$$\begin{aligned}
&\left| \left(\left(\bar{\boldsymbol{S}}^{(v)}_n + R^{(v)}_{1,n} \right)^{-1} - \left(\bar{\boldsymbol{S}}^{(v)}_n \right)^{-1} \right) |A_n|^{-\frac{1}{2}} \left(\frac{\partial}{\partial \boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}; X_v) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} + R^{(v)}_{2,n} \right) \right| \\
\leq & \left| \left(\bar{\boldsymbol{S}}^{(v)}_n + R^{(v)}_{1,n} \right)^{-1} - \left(\bar{\boldsymbol{S}}^{(v)}_n \right)^{-1} \right| \times |A_n|^{-\frac{1}{2}} \times \left| \frac{\partial}{\partial \boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}; X_v) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} + R^{(v)}_{2,n} \right| \\
= & o_p(1) \times O_p(1) = o_p(1)
\end{aligned}$$

Now, substituting the last bound into (15) yields

$$\begin{aligned}
|A_n|^{\frac{1}{2}} (\hat{\boldsymbol{\theta}}^{(v)}_n - \boldsymbol{\theta}^*) = & - \left(\bar{\boldsymbol{S}}^{(v)}_n \right)^{-1} |A_n|^{-\frac{1}{2}} \left(\frac{\partial}{\partial \boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}; X_v) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} + R^{(v)}_{2,n} \right) + o_p(1) \\
= & \left(\bar{\boldsymbol{S}}^{(v)}_n \right)^{-1} |A_n|^{-\frac{1}{2}} \frac{\partial}{\partial \boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}; X_v) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} + \left(\bar{\boldsymbol{S}}^{(v)}_n \right)^{-1} |A_n|^{-\frac{1}{2}} R^{(v)}_{2,n} + o_p(1) \\
= & \left(\bar{\boldsymbol{S}}^{(v)}_n \right)^{-1} |A_n|^{-\frac{1}{2}} \frac{\partial}{\partial \boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}; X_v) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} + o_p(1)
\end{aligned}$$

By proposition 1, the intensity function of X_v is $V^{-1}\lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*)$. If we replace $\lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*)$ by $V^{-1}\lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*)$ in the following formula of sensitivity matrix $\boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*)$

$$\boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*) = \int_A \left(\frac{\partial}{\partial \boldsymbol{\theta}} \log \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta}}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right)^{\otimes 2} \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) \mathrm{d}\mathbf{u}$$

we will obtain that

$$\bar{\boldsymbol{S}}^{(v)}_n = V^{-1} \bar{\boldsymbol{S}}_n(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*)$$

Moreover,

$$\sum_{v=1}^{V} \ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}; X_v) = \ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n})$$

Therefore,

$$\begin{aligned}
|A_n|^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) =& |A_n|^{\frac{1}{2}}\left(V^{-1}\sum_{v=1}^{V}\hat{\boldsymbol{\theta}}_n^{(v)} - \boldsymbol{\theta}^*\right) = V^{-1}\sum_{v=1}^{V}|A_n|^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n^{(v)} - \boldsymbol{\theta}^*) \\
&= \left(V\bar{\boldsymbol{S}}_n^{(v)}\right)^{-1}|A_n|^{-\frac{1}{2}}\frac{\partial}{\partial\boldsymbol{\theta}}\left(\sum_{v=1}^{V}\ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}; X_v)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right) + o_p(1) \\
&= \left(\bar{\boldsymbol{S}}_n(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*)\right)^{-1}|A_n|^{-\frac{1}{2}}\frac{\partial}{\partial\boldsymbol{\theta}}\ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} + o_p(1)
\end{aligned}$$

Step 2: In this step, we establish (11). Observe that by the triangle inequality,

$$|R_{1,n}^{(v)}| \leq \mathcal{I}_{1,n}^v + \mathcal{I}_{2,n}^v + \mathcal{I}_{3,n}^v \tag{18}$$

where

$$\begin{aligned}
\mathcal{I}_{1,n}^v &:= |A_n|^{-1} \times \left|\frac{\partial^2}{\partial\boldsymbol{\theta}^2}\ell_n(\boldsymbol{\theta}, \hat{\eta}^{(v)}_{\boldsymbol{\theta},n}; X_v)\Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_n^{(v)}} - \mathbb{E}\left[\frac{\partial^2}{\partial\boldsymbol{\theta}^2}\ell_n(\boldsymbol{\theta}, \hat{\eta}^{(v)}_{\boldsymbol{\theta},n}; X_v)\Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_n^{(v)}}\right]\right| \\
\mathcal{I}_{2,n}^v &:= |A_n|^{-1} \times \left|\mathbb{E}\left[\frac{\partial^2}{\partial\boldsymbol{\theta}^2}\ell_n(\boldsymbol{\theta}, \hat{\eta}^{(v)}_{\boldsymbol{\theta},n}; X_v)\Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_n^{(v)}}\right] - \mathbb{E}\left[\frac{\partial^2}{\partial\boldsymbol{\theta}^2}\ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}; X_v)\Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_n^{(v)}}\right]\right| \\
\mathcal{I}_{3,n}^v &:= |A_n|^{-1} \times \left|\mathbb{E}\left[\frac{\partial^2}{\partial\boldsymbol{\theta}^2}\ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}; X_v)\Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_n^{(v)}}\right] - \mathbb{E}\left[\frac{\partial^2}{\partial\boldsymbol{\theta}^2}\ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}; X_v)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right]\right|
\end{aligned}$$

To bound $\mathcal{I}_{1,n}^v$, we further denote

$$\hat{\eta}_n^{(v)} := \hat{\eta}_{\boldsymbol{\theta},n}^{(v)}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n^{(v)}}$$
$$\hat{\boldsymbol{\nu}}_n^{(v)} := \frac{\partial}{\partial \boldsymbol{\theta}}\hat{\eta}_{\boldsymbol{\theta},n}^{(v)}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n^{(v)}}$$
$$\hat{\boldsymbol{\nu}}_n^{(v)\prime} := \frac{\partial^2}{\partial \boldsymbol{\theta}^2}\hat{\eta}_{\boldsymbol{\theta},n}^{(v)}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n^{(v)}}$$

Note that by the chain rule,

$$\begin{aligned}\frac{\partial^2}{\partial \boldsymbol{\theta}^2}\ell_n(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta},n}^{(v)}; X_v)\Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_n^{(v)}} &= \frac{\partial^2}{\partial \boldsymbol{\theta}^2}\ell_n(\boldsymbol{\theta}, \hat{\eta}_n^{(v)}; X_v)\Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_n^{(v)}} \\ &\quad + \frac{\partial^2}{\partial \boldsymbol{\theta}\partial\eta}\ell_n(\boldsymbol{\theta}, \hat{\eta}_n^{(v)}; X_v)\Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_n^{(v)}}\hat{\boldsymbol{\nu}}_n^{(v)} + \frac{\partial}{\partial\eta}\ell_n(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta},n}^{(v)}; X_v)\Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_n^{(v)}}\hat{\boldsymbol{\nu}}_n^{(v)\prime} \quad (19)\end{aligned}$$

If we substitute (19) into $\mathcal{I}_{1,n}^v$, then by the trianglar inequality and taking the supremum with respect to $\hat{\boldsymbol{\theta}}^{(v)}, \hat{\eta}_n^{(v)}, \hat{\boldsymbol{\nu}}_n^{(v)}, \hat{\boldsymbol{\nu}}_n^{(v)\prime}$,

$$\begin{aligned}\mathcal{I}_{1,n}^v \leq & |A_n|^{-1} \times \sup_{\boldsymbol{\theta}\in\Theta,\eta\in\mathcal{H}}\left|\frac{\partial^2}{\partial \boldsymbol{\theta}^2}\ell_n(\boldsymbol{\theta}, \eta; X_v) - \mathbb{E}\left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2}\ell_n(\boldsymbol{\theta}, \eta; X_v)\right]\right| \\ &+ |A_n|^{-1} \times \sup_{\boldsymbol{\theta}\in\Theta,\eta\in\mathcal{H}}\left|\frac{\partial^2}{\partial \boldsymbol{\theta}\partial\eta}\ell_n(\boldsymbol{\theta}, \eta; X_v) - \mathbb{E}\left[\frac{\partial^2}{\partial \boldsymbol{\theta}\partial\eta}\ell_n(\boldsymbol{\theta}, \eta; X_v)\right]\right| \times \sup_{\boldsymbol{\nu}\in\mathcal{H}^k}|\boldsymbol{\nu}| \\ &+ |A_n|^{-1} \times \sup_{\boldsymbol{\theta}\in\Theta,\eta\in\mathcal{H}}\left|\frac{\partial}{\partial\eta}\ell_n(\boldsymbol{\theta}, \eta; X_v) - \mathbb{E}\left[\frac{\partial}{\partial\eta}\ell_n(\boldsymbol{\theta}, \eta; X_v)\right]\right| \times \sup_{\boldsymbol{\nu}'\in\mathcal{H}^k\times\mathcal{H}^k}|\boldsymbol{\nu}'|\end{aligned}$$

Following from Corollary 1 and the fact that all functions in $\mathcal{H}$ are bounded, $\mathcal{I}_{1,n}^v = O_p(|A_n|^{-\frac{1}{2}})$.

To bound $\mathcal{I}_{2,n}^v$, it follows from the first order plug-in error Lemma 3 that when $\hat{\eta}_{\boldsymbol{\theta},n}^{(v)}$ satisfies Assumption 3,

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\mathbb{E}\left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2}\ell_n(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta},n}^{(v)}; X_v)\right] - \mathbb{E}\left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2}\ell_n(\boldsymbol{\theta}, \eta_{\boldsymbol{\theta},n}^{*}; X_v)\right]\right| = o(|A_n|)$$

Thus, $\mathcal{I}_{2,n}^{v} = o(1)$.

To bound $\mathcal{I}_{3,n}^{v}$, we note that $\frac{\partial^2}{\partial \boldsymbol{\theta}^2}\ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n})$ is continuous with respect to $\boldsymbol{\theta} \in \Theta$ where Θ is compact, thus for any $\boldsymbol{\theta} \in \Theta$, there exists a positive constant $C <$ such that the third term satisfies

$$\mathcal{I}_{3,n}^{v} \leq |A_n|^{-1} \times \left| \mathbb{E}\left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2}\ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}; X_v)\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right] \right| \times C \left| \boldsymbol{\theta}^* - \tilde{\boldsymbol{\theta}}_n^{(v)} \right|$$

Thus, $\mathcal{I}_{3,n}^{v} = o_p(1)$ follows from the consistency of $\tilde{\boldsymbol{\theta}}_n^{(v)}$ and equation (49) in in Corollary 1. Combining the bounds $\mathcal{I}_{1,n}^{v} = O_p(|A_n|^{-\frac{1}{2}})$, $\mathcal{I}_{2,n}^{v} = o(1)$ and $\mathcal{I}_{3,n}^{v} = o_p(1)$ with (18) gives us 11

Step 3: In this step, we establish (12). To bound $R_{2,n}^{(v)}$, We further denote

$$\begin{aligned} r_n^{(2,v)}(\boldsymbol{\theta}) &:= \ell_n(\boldsymbol{\theta}, \hat{\eta}^{(v)}_{\boldsymbol{\theta},n}; X_v) - \ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}) - \frac{\partial}{\partial \eta}\ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}; X_v))[\hat{\eta}^{(v)}_{\boldsymbol{\theta},n} - \eta^*_{\boldsymbol{\theta},n}] \\ \mathcal{I}_{4,n}^{v} &:= \frac{\partial^2}{\partial \boldsymbol{\theta}\partial \eta}\ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}; X_v)\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}[\hat{\eta}^{(v)}_{\boldsymbol{\theta}^*,n} - \eta^*_{\boldsymbol{\theta}^*,n}] \\ \mathcal{I}_{5,n}^{v} &:= \frac{\partial}{\partial \eta}\ell_n(\boldsymbol{\theta}^*, \eta^*_{\boldsymbol{\theta}^*,n}; X_v)\left[\frac{\partial}{\partial \boldsymbol{\theta}}\hat{\eta}^{(v)}_{\boldsymbol{\theta}^*,n} - \frac{\partial}{\partial \boldsymbol{\theta}}\eta^*_{\boldsymbol{\theta}^*,n}\right] \end{aligned}$$

Then by the chain rule,

$$\begin{aligned} R_{2,n}^{(v)} =& \frac{\partial}{\partial \boldsymbol{\theta}}\left(\frac{\partial}{\partial \eta}\ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}; X_v))[\hat{\eta}^{(v)}_{\boldsymbol{\theta},n} - \eta^*_{\boldsymbol{\theta},n}] + r_n^{(2,v)}(\boldsymbol{\theta})\right)\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \\ =& \frac{\partial}{\partial \boldsymbol{\theta}}\left(\frac{\partial}{\partial \eta}\ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}; X_v))[\hat{\eta}^{(v)}_{\boldsymbol{\theta},n} - \eta^*_{\boldsymbol{\theta},n}]\right)\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} + \frac{\partial}{\partial \boldsymbol{\theta}} r_n^{(2,v)}(\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \\ =& \mathcal{I}_{4,n}^{v} + \mathcal{I}_{5,n}^{v} + \frac{\partial}{\partial \boldsymbol{\theta}} r_n^{(2,v)}(\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \end{aligned} \tag{20}$$

We call $r_n^{(2,v)}(\boldsymbol{\theta})$ the second order plug-in error rate because it is the second order approximation error of $\ell(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}; X_c)$ when $\eta^*_{\boldsymbol{\theta},n}$ is replaced by $\hat{\eta}^{(v)}_{\boldsymbol{\theta},n}$. Lemma 4 shows that the rate of second order plug-in error satisfies

$$\frac{\partial}{\partial \boldsymbol{\theta}} r_n^{(2,v)}(\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = o_p(|A_n^{\frac{1}{2}}|) \tag{21}$$

We call $\mathcal{I}_{4,n}^v$ and $\mathcal{I}_{5,n}^v$ spatial empirical processes because it corresponds to the empirical processes in classic *i.i.d.* setting in semiparametric statistics (See Wellner et al. [2013]). However, the Maximal Inequality (See Lemma 6.2 in Chernozhukov et al. [2018]) can not be generalized directly to bound $\mathcal{I}_{4,n}^v$ and $\mathcal{I}_{5,n}^v$. In the proof of Lemma 5, we utilize the special structure of the pseudo-likelihood of spatial point process to show that

$$\mathcal{I}_{4,5}^{(v)} = o_p(|A_n^{\frac{1}{2}}|), \quad \mathcal{I}_{4,5}^{(v)} = o_p(|A_n^{\frac{1}{2}}|) \tag{22}$$

if either the intensity function is log-linear or $\hat{\eta}_{\boldsymbol{\theta},n}^{(v)}$ is independent of X_v. Combining the bounds gives us (21) and (22) with (20) gives us (12)

Step 4: In this step, we establish (13). We use the same notation when we bound $\mathcal{I}_{1,n}^v$ that

$$\hat{\eta}_n^{(v)} := \hat{\eta}_{\boldsymbol{\theta},n}^{(v)}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n^{(v)}}$$
$$\hat{\boldsymbol{\nu}}_n^{(v)} := \frac{\partial}{\partial \boldsymbol{\theta}} \hat{\eta}_{\boldsymbol{\theta},n}^{(v)}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n^{(v)}}$$

Then by the chain rule and taking supremum over $\boldsymbol{\theta}$ and $\hat{\eta}_n^{(v)}$,

$$\begin{aligned}\left| |A_n|^{-\frac{1}{2}} \frac{\partial}{\partial \boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta},n}^{(v)}; X_v)\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right| &\leq \left| \frac{\partial}{\partial \boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}, \hat{\eta}_n^{(v)}; X_v)\bigg|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_n^{(v)}} \right| + \left| \frac{\partial}{\partial \eta} \ell_n(\boldsymbol{\theta}, \hat{\eta}_n^{(v)}; X_v)\bigg|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_n^{(v)}} \hat{\boldsymbol{\nu}}_n^{(v)} \right| \\ &\leq \sup_{\boldsymbol{\theta}\in\Theta, \eta\in\mathcal{H}} \left| \frac{\partial}{\partial \boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}, \eta; X_v) \right| + \sup_{\boldsymbol{\theta}\in\Theta, \eta\in\mathcal{H}} \left| \frac{\partial}{\partial \eta} \ell_n(\boldsymbol{\theta}, \eta; X_v) \right| \times \sup_{\boldsymbol{\nu}\in\mathcal{H}^k} |\boldsymbol{\nu}|\end{aligned}$$

Then (13) follows from Corollary 1 and the fact that all functions in $\mathcal{H}$ is bounded.

Step 5: In this step, we establish (14). By proposition 1, the intensity function of X_v is $V^{-1}\lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*)$. If we replace $\lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*)$ by $V^{-1}\lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*)$ in the formula of sensitivity matrix $\boldsymbol{S}(\boldsymbol{\theta}^*,\eta^*,\boldsymbol{\nu}^*)$ in the following formula of sensitivity matrix $\boldsymbol{S}(\boldsymbol{\theta}^*,\eta^*,\boldsymbol{\nu}^*)$

$$\boldsymbol{S}(\boldsymbol{\theta}^*,\eta^*,\boldsymbol{\nu}^*) = \int_A \left(\frac{\partial}{\partial\boldsymbol{\theta}} \log\lambda(\mathbf{u};\boldsymbol{\theta},\eta^*_{\boldsymbol{\theta}}) \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right)^{\otimes 2} \lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*)\mathrm{d}\mathbf{u}$$

, we will obtain that

$$\bar{\boldsymbol{S}}_n^{(v)} = V^{-1}\bar{\boldsymbol{S}}_n(\boldsymbol{\theta}^*,\eta^*,\boldsymbol{\nu}^*)$$

Then, (14) follows from condition 2.1 which states that the limit infimum of the smallest eigne value of $\bar{\boldsymbol{S}}_n(\boldsymbol{\theta}^*,\eta^*,\boldsymbol{\nu}^*)$ is positive.

□

8 Proof of Theorem 8.1

Theorem 8.1 (Consistent variance estimator)**.** *$\hat{\boldsymbol{\nu}}_n$ defined as*

$$\hat{\boldsymbol{\nu}}_n(\mathbf{z}) = -\frac{\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\eta}\hat{\mathbb{E}}[\ell_n(\hat{\boldsymbol{\theta}}_n,\hat{\eta}_n(\mathbf{z}))]}{\frac{\partial^2}{\partial\eta^2}\hat{\mathbb{E}}[\ell_n(\hat{\boldsymbol{\theta}}_n,\hat{\eta}_n(\mathbf{z}))]} \tag{23}$$

where m_n is the number of samples uniformly drawn from the observational window A_n. Under the same assumptions in Theorem 7.1, $\hat{\boldsymbol{\nu}}_n(\mathbf{z})$ is a consistent estimator of $\boldsymbol{\nu}_n^*$, i.e. $\sup_{\mathbf{z}\in\mathcal{Z}} |\hat{\boldsymbol{\nu}}_n(\mathbf{z}) - \boldsymbol{\nu}^*(\mathbf{z})| \to_p 0$*. Under the 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

$$\left| |A_n|^{-1} \bar{\boldsymbol{S}}_n(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_n, \hat{\boldsymbol{\nu}}_n) - \bar{\boldsymbol{S}}_n(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*) \right| \to_p 0$$
$$\left| |A_n|^{-1} \widehat{\boldsymbol{\Sigma}}_n^{-\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n, \hat{\eta}_n, \hat{\boldsymbol{\nu}}_n, \hat{\psi}_n) - \bar{\boldsymbol{\Sigma}}_n^{-\frac{1}{2}}(\boldsymbol{\theta}^*, \eta^*, \boldsymbol{\nu}^*, \psi^*) \right| \to_p 1$$

Proof. To start with, note that $\widehat{\boldsymbol{S}}_n(\boldsymbol{\theta}, \eta, \boldsymbol{\nu})$ is absolutely continuous with respect to $\boldsymbol{\theta}, \eta, \boldsymbol{\nu}$, $\widehat{\boldsymbol{\Sigma}}_n(\boldsymbol{\theta}, \eta, \boldsymbol{\nu}, \psi)$ is absolutely continuous with respect to $\boldsymbol{\theta}, \eta, \boldsymbol{\nu}, \psi$, and $\hat{\boldsymbol{\theta}}_n, \hat{\eta}_n, \hat{\psi}_n$ are consistent estimators of $\boldsymbol{\theta}^*, \eta^*, \psi^*$. Therefore, it suffices to show that for any $\boldsymbol{\theta}, \eta, \boldsymbol{\nu}, \psi$,

$$\left| \bar{\boldsymbol{S}}_n(\boldsymbol{\theta}, \eta, \boldsymbol{\nu}) - |A_n|^{-1} \widehat{\boldsymbol{S}}_n(\boldsymbol{\theta}, \eta, \boldsymbol{\nu}) \right| \to_p 0 \tag{24}$$

$$\left| \bar{\boldsymbol{\Sigma}}_n(\boldsymbol{\theta}, \eta, \boldsymbol{\nu}, \psi) - |A_n|^{-1} \widehat{\boldsymbol{\Sigma}}_n(\boldsymbol{\theta}, \eta, \boldsymbol{\nu}, \psi) \right| \to_p 0 \tag{25}$$

and

$$\sup_{\mathbf{z} \in \mathcal{Z}} |\hat{\boldsymbol{\nu}}_n(\mathbf{z}) - \boldsymbol{\nu}^*(\mathbf{z})| \to_p 0 \tag{26}$$

To do so, we proceed in six steps. Step 1 shows (24) and (25). Step 2 shows the main argument of proving (26), and Steps 3–5 present auxiliary calculations of proving (26).

Step 1: In this step, we establish (24) and (25). For any fixed $\boldsymbol{\theta}, \eta, \boldsymbol{\nu}, \psi$, denote

$$\phi_1(\mathbf{u}) := \frac{\partial}{\partial \boldsymbol{\theta}} \log \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) + \frac{\partial}{\partial \eta} \log \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta)[\boldsymbol{\nu}]$$
$$\phi_2(\mathbf{u}) := \left(\frac{\partial}{\partial \boldsymbol{\theta}} \log \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) + \frac{\partial}{\partial \eta} \log \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta)[\boldsymbol{\nu}] \right) \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*)$$
$$\bar{g}(\mathbf{u}, \mathbf{v}) := (g(\mathbf{u}, \mathbf{v}; \psi) - 1)$$

Then the formula of $\bar{\boldsymbol{S}}_n(\boldsymbol{\theta}, \eta, \boldsymbol{\nu})$ and $\bar{\boldsymbol{\Sigma}}_n(\boldsymbol{\theta}, \eta, \boldsymbol{\nu}, \psi)$ are given by

$$
\begin{aligned}
\bar{\boldsymbol{S}}_n(\boldsymbol{\theta}, \eta, \boldsymbol{\nu}) &= |A_n|^{-1} \int_{A_n} \phi_1(\mathbf{u})^{\otimes 2} \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) \mathrm{d}\mathbf{u} \\
\bar{\boldsymbol{\Sigma}}_n(\boldsymbol{\theta}, \eta, \boldsymbol{\nu}, \psi) &= \bar{\boldsymbol{S}}_n(\boldsymbol{\theta}, \eta, \boldsymbol{\nu}) + |A_n|^{-1} \int_{A_n} \int_{A_n} \phi_1(\mathbf{u}) \phi_2(\mathbf{v})^\top \bar{g}(\mathbf{u}, \mathbf{v}) \mathrm{d}\mathbf{u} \mathrm{d}\mathbf{v}
\end{aligned}
$$

The formula of $\widehat{\boldsymbol{S}}_n(\boldsymbol{\theta}, \eta, \boldsymbol{\nu})$ and $\widehat{\boldsymbol{\Sigma}}_n(\boldsymbol{\theta}, \eta, \boldsymbol{\nu}, \psi)$ are given by

$$
\begin{aligned}
\widehat{\boldsymbol{S}}_n(\boldsymbol{\theta}, \eta, \boldsymbol{\nu}) &= \frac{|A_n|}{m_n} \sum_{j=1}^{m_n} \phi_1(\mathbf{u}_j)^{\otimes 2} \lambda(\mathbf{u}_j; \boldsymbol{\theta}, \eta) \\
\widehat{\boldsymbol{\Sigma}}_n(\boldsymbol{\theta}, \eta, \boldsymbol{\nu}, \psi) &= \widehat{\boldsymbol{S}}_n(\boldsymbol{\theta}, \eta, \boldsymbol{\nu}) + \frac{|A_n|^2}{m_n^2} \sum_{i,j=1}^{m} \phi_2(\mathbf{u}_i) \phi_2(\mathbf{u}_j)^\top \bar{g}(\mathbf{u}_i, \mathbf{u}_j)
\end{aligned}
$$

where the set $\{\mathbf{u}_j\}_{j=1}^{m_n}$ points randomly drawn from the observational window A_n.

Note that the set $\{\mathbf{u}_j\}_{j=1}^{m_n}$ can be regarded as a realization of a homogeneous Poisson process with intensity $|A_n|^{-1} m_n$. Additionally, h_1 is a uniformly bounded function. Thus it followed from Lemma 1 that

$$
\begin{aligned}
\left|\bar{\boldsymbol{S}}_n(\boldsymbol{\theta}, \eta, \boldsymbol{\nu})\right| &= O(1) \\
\left| |A_n|^{-1} \widehat{\boldsymbol{S}}(\boldsymbol{\theta}, \eta, \boldsymbol{\nu}) - \bar{\boldsymbol{S}}_n(\boldsymbol{\theta}, \eta, \boldsymbol{\nu}) \right| &= O_p(|A_n|^{-\frac{1}{2}})
\end{aligned}
$$

To establish 25, we further denote that for any fixed $\mathbf{v} \in \mathbb{R}^2$,

$$
R_n(\mathbf{v}) := \int_{A_n} \phi_2(\mathbf{u}) \bar{g}(\mathbf{u}, \mathbf{v}) \mathrm{d}\mathbf{u} - \frac{|A_n|}{m_n} \sum_{i=1}^{m_n} \phi_2(\mathbf{u}_i) \bar{g}(\mathbf{u}_i, \mathbf{v})
$$

By the triangular inequality,

$$\left|\bar{\boldsymbol{\Sigma}}_n(\boldsymbol{\theta},\eta,\boldsymbol{\nu},\psi)-|A_n|^{-1}\widehat{\boldsymbol{\Sigma}}_n(\boldsymbol{\theta},\eta,\boldsymbol{\nu},\psi)\right|\le\left|\bar{\boldsymbol{S}}_n(\boldsymbol{\theta},\eta,\boldsymbol{\nu})-|A_n|^{-1}\widehat{\boldsymbol{S}}_n(\boldsymbol{\theta},\eta,\boldsymbol{\nu})\right|+$$
$$\left||A_n|^{-1}\int_{A_n}\int_{A_n}\phi_2(\mathbf{u})\phi_2(\mathbf{v})^\top\bar{g}(\mathbf{u},\mathbf{v})\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}-\frac{|A_n|}{m_n^2}\sum_{i,j=1}^{m}\phi_2(\mathbf{u}_i)\phi_2(\mathbf{u}_j)\bar{g}(\mathbf{u}_i,\mathbf{u}_j)\right|$$

Given (24), it suffices to show that

$$\begin{aligned}&|A_n|^{-1}\left|\int_{A_n}\int_{A_n}\phi_2(\mathbf{u})\phi_2(\mathbf{v})^\top\bar{g}(\mathbf{u},\mathbf{v})\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}-\frac{|A_n|^2}{m_n^2}\sum_{i,j=1}^{m}\phi_2(\mathbf{u}_i)\phi_2(\mathbf{u}_j)\bar{g}(\mathbf{u}_i,\mathbf{u}_j)\right|\\=&|A_n|^{-1}\left|\int_{A_n}\phi_2(\mathbf{v})R_n(\mathbf{v})\mathrm{d}\mathbf{v}-\frac{|A_n|}{m_n}\sum_{i=1}^{m_n}\phi_2(\mathbf{v}_i)R_n(\mathbf{v}_i)\right|=o_p(1)\end{aligned}$$

Then by Lemma 1, it suffices to show that $R_n(\mathbf{v})$ is uniformly bounded.

Since $\{\mathbf{u}_j\}_{j=1}^{m_n}$ is a realization of a homogeneous Poisson process with intensity $|A_n|^{-1}m_n$, it follows from condition 1.4 that

$$\left|\mathbb{E}\left[\frac{|A_n|}{m_n}\sum_{i=1}^{m_n}\phi_2(\mathbf{u}_i)\bar{g}(\mathbf{u}_i,\mathbf{v})\right]\right|=\left|\int_{A_n}\phi_2(\mathbf{u})\bar{g}(\mathbf{u},\mathbf{v})\mathrm{d}\mathbf{u}\right|=O(1)$$

$$\left|\mathrm{Var}\left(\frac{|A_n|}{m_n}\sum_{i=1}^{m_n}\phi_2(\mathbf{u}_i)\bar{g}(\mathbf{u}_i,\mathbf{v})\right)\right|=\left|\int_{A_n}\int_{A_n}\phi_2(\mathbf{u}_1)\bar{g}(\mathbf{u}_1,\mathbf{v})\phi_2(\mathbf{u}_2)^\top\bar{g}(\mathbf{u}_2,\mathbf{v})\mathrm{d}\mathbf{u}_1\mathrm{d}\mathbf{u}_2\right|=O(1)$$

Therefore, $\sup_{\mathbf{v}}|R_n(\mathbf{v})|=O(1)$.

Step 3: In this step, we establish the main argument of proving (26). Denote

$$\begin{aligned}\phi_3(\mathbf{y},\mathbf{z};\boldsymbol{\theta},\eta)&:=\frac{\partial}{\partial\boldsymbol{\theta}}\log\lambda(\mathbf{y},\mathbf{z};\boldsymbol{\theta},\eta)\frac{\partial}{\partial\eta}\log\lambda(\mathbf{y},\mathbf{z};\boldsymbol{\theta},\eta)\lambda(\mathbf{y},\mathbf{z};\boldsymbol{\theta},\eta)\\\phi_4(\mathbf{y},\mathbf{z};\boldsymbol{\theta},\eta)&:=\frac{\partial}{\partial\eta}\log\lambda(\mathbf{y},\mathbf{z};\boldsymbol{\theta},\eta)\frac{\partial}{\partial\eta}\log\lambda(\mathbf{y},\mathbf{z};\boldsymbol{\theta},\eta)\lambda(\mathbf{y},\mathbf{z};\boldsymbol{\theta},\eta)\end{aligned}$$

For any $\phi(\mathbf{y}, \mathbf{z}; \boldsymbol{\theta}, \eta)$, we denote

$$
F_n(\mathbf{z}, \phi, \boldsymbol{\theta}, \eta) := \int_{\mathcal{Y}} \phi(\mathbf{y}, \mathbf{z}; \boldsymbol{\theta}, \eta) f_n(\mathbf{y}, \mathbf{z}) \mathrm{d}\mathbf{y}
$$
$$
\hat{F}_n(\mathbf{z}, \phi, \boldsymbol{\theta}, \eta) := \frac{|A_n|}{m_n} \sum_{j=1}^{m_n} K_h\left(\mathbf{z}(\mathbf{u}_j) - \mathbf{z}\right) \phi(\mathbf{y}(\mathbf{u}_j), \mathbf{z}(\mathbf{u}_j); \boldsymbol{\theta}, \eta)
$$

Then by triangular inequality,

$$
|\hat{\boldsymbol{\nu}}_n(\mathbf{z}) - \boldsymbol{\nu}_n^*(\mathbf{z})| \leq \left| \frac{\hat{F}_n(\mathbf{z}, \phi_3, \boldsymbol{\theta}^*, \eta^*)}{\hat{F}_n(\mathbf{z}, \phi_4, \boldsymbol{\theta}^*, \eta^*)} - \frac{F_n(\mathbf{z}, \phi_3, \boldsymbol{\theta}^*, \eta^*)}{F_n(\mathbf{z}, \phi_4, \boldsymbol{\theta}^*, \eta^*)} \right| + \left| \frac{\hat{F}_n(\mathbf{z}, \phi_3, \boldsymbol{\theta}^*, \eta^*)}{\hat{F}_n(\mathbf{z}, \phi_4, \boldsymbol{\theta}^*, \eta^*)} - \frac{\hat{F}_n(\mathbf{z}, \phi_3, \hat{\boldsymbol{\theta}}_n, \hat{\eta}_n)}{\hat{F}_n(\mathbf{z}, \phi_4, \hat{\boldsymbol{\theta}}_n, \hat{\eta}_n)} \right|
$$

Since $\hat{\boldsymbol{\theta}}_n, \hat{\eta}_n$ are consistent estimators of $\boldsymbol{\theta}^*, \eta^*$, and $F_n(\mathbf{z}; \phi_3, \boldsymbol{\theta}, \eta), F_n(\mathbf{z}; \phi_4, \boldsymbol{\theta}, \eta)$ are twice continuously differentiable with respect to $\boldsymbol{\theta}, \eta$ by condition 1.1, it suffices to show that

$$
\sup_{\mathbf{z} \in \mathcal{Z}} \left| \hat{F}_n(\mathbf{z}, \phi_3, \boldsymbol{\theta}^*, \eta^*) - F_n(\mathbf{z}, \phi_3, \boldsymbol{\theta}^*, \eta^*) \right| = o_p(|A_n|) \tag{27}
$$
$$
\sup_{\mathbf{z} \in \mathcal{Z}} \left| \hat{F}_n(\mathbf{z}, \phi_4, \boldsymbol{\theta}^*, \eta^*) - F_n(\mathbf{z}, \phi_4, \boldsymbol{\theta}^*, \eta^*) \right| = o_p(|A_n|) \tag{28}
$$
$$
|F_n(\mathbf{z}, \phi_4, \boldsymbol{\theta}^*, \eta^*) = \Omega(|A_n|)| \tag{29}
$$
$$
|F_n(\mathbf{z}, \phi_3, \boldsymbol{\theta}^*, \eta^*) = O(|A_n|)| \tag{30}
$$

We will proved (27) and (28) in Step 4, prove (29), (30) in Step 5- 6 respectively.

Step 4: We will establish (27), (28) in this step. First, we note that by triangular inequality, for $\phi = \phi_3, \phi_4$,

$$
\begin{aligned}
\sup_{\mathbf{z} \in \mathcal{Z}} \left| \hat{F}_n(\mathbf{z}, \phi, \boldsymbol{\theta}^*, \eta^*) - F_n(\mathbf{z}, \phi, \boldsymbol{\theta}^*, \eta^*) \right| \leq& \sup_{\mathbf{z} \in \mathcal{Z}} \left| \mathbb{E}\left[\hat{F}_n(\mathbf{z}, \phi, \boldsymbol{\theta}^*, \eta^*) \right] - F_n(\mathbf{z}, \phi, \boldsymbol{\theta}^*, \eta^*) \right| \\
&+ \sup_{\mathbf{z} \in \mathcal{Z}} \left| \hat{F}_n(\mathbf{z}, \phi, \boldsymbol{\theta}^*, \eta^*) - \mathbb{E}\left[\hat{F}_n(\mathbf{z}, \phi, \boldsymbol{\theta}^*, \eta^*) \right] \right| \qquad (31)
\end{aligned}
$$

Then we obtain upper bounds of the two terms in the right-hand. Since the order of Kernel function $K(\cdot)$ is at least two, and ϕ_3, ϕ_4 are uniformly bounded, it followed from Lemma 7 that for $\phi = \phi_3, \phi_4$

$$\sup_{\mathbf{z}\in\mathcal{Z}} \left| \mathbb{E}\left[\hat{F}_n(\mathbf{z}, \phi, \boldsymbol{\theta}^*, \eta^*)\right] - F_n(\mathbf{z}, \phi, \boldsymbol{\theta}^*, \eta^*) \right| = O(|A_n| h^2) \tag{32}$$

Furthermore, since $\boldsymbol{\theta}^*$ and η^* are fixed, it followed from Lemma 9 that for $\phi = \phi_3, \phi_4$,

$$\sup_{\mathbf{z}\in\mathcal{Z}} \left| \hat{F}_n(\mathbf{z}, \phi, \boldsymbol{\theta}^*, \eta^*) - \mathbb{E}\left[\hat{F}_n(\mathbf{z}, \phi, \boldsymbol{\theta}^*, \eta^*)\right] \right| = O_p(|A_n|^{\frac{1+q}{q+m}} h^{-q-\frac{q}{q+m}}) \tag{33}$$

Substitute (32), (33) into (31), we have

$$\begin{aligned}\sup_{\mathbf{z}\in\mathcal{Z}} \left| \hat{F}_n(\mathbf{z}, \phi, \boldsymbol{\theta}^*, \eta^*) - F_n(\mathbf{z}, \phi, \boldsymbol{\theta}^*, \eta^*) \right| =& O_p(|A_n|^{\frac{1+q}{q+m}} h^{-q-\frac{q}{q+m}}) + O(|A_n| h^2) \\ =& O_p(|A_n|^{1+\frac{2(1-m)}{(2+q)(q+m)+q}})\end{aligned}$$

If we let the bandwidth, we have $h = \Theta(|A_n|^{\frac{1-m}{(2+q)(q+m)}+q})$,

$$\sup_{\mathbf{z}\in\mathcal{Z}} \left| \hat{F}_n(\mathbf{z}, \phi, \boldsymbol{\theta}^*, \eta^*) - F_n(\mathbf{z}, \phi, \boldsymbol{\theta}^*, \eta^*) \right| = O_p(|A_n|^{1+\frac{2(1-m)}{(2+q)(q+m)+q}})$$

Since $m \geq 2$ is implied by condition 1.4 (decay rate of PCF), we thus have (27) and (28).

Step 5: In this step, we will establish (29). Note that $f(\mathbf{y}, \mathbf{z})$ satisfies

$$\int_{\mathcal{Y}} \int_{\mathcal{Z}} f(\mathbf{y}, \mathbf{z}) \mathrm{d}\mathbf{y} \mathrm{d}\mathbf{z} = |A_n|$$

Thus almost surely in $\mathbf{z} \in \mathcal{Z}$,

$$\int_{\mathcal{Y}} f_n(\mathbf{y}, \mathbf{z}) \mathrm{d}\mathbf{y} = O(|A_n|)$$

Moreover, ϕ_3 is uniformly bounded, we thus have

$$|F_n(\mathbf{z}, \phi_3, \boldsymbol{\theta}^*, \eta^*)| \leq \sup_{\mathbf{y}, \mathbf{z}} |\phi_3(\mathbf{y}, \mathbf{z}; \boldsymbol{\theta}^*, \eta^*)| \int_{\mathcal{Y}} f_n(\mathbf{y}, \mathbf{z}) \mathrm{d}\mathbf{y} = O(|A_n|)$$

Step 6: By our definition,

$$F_n(\mathbf{z}, \phi_4, \boldsymbol{\theta}^*, \eta^*) = \frac{\partial^2}{\partial \gamma} \mathbb{E}[\ell(\boldsymbol{\theta}^*, \gamma) | \mathbf{z}] \bigg|_{\gamma = \eta^*(\mathbf{z})}$$

Then 29 followed from the nuisance identification condition 5.2.

□

9 Proof of Theorem 9.1

Theorem 9.1. *Suppose Assumptions 1, 5 hold. Then, with appropriately chosen bandwidth, for $j = 0, 1, 2$, the nuisance estimation satisfies*

$$\sup_{\mathbf{z} \in \mathcal{Z}} \left| \frac{\partial^j}{\partial \boldsymbol{\theta}^j} (\eta^*_{\boldsymbol{\theta}, n}(\mathbf{z}) - \hat{\eta}^{(v)}_{\boldsymbol{\theta}, n}(\mathbf{z})) \right| = o_p \left(|A_n|^{-\frac{m-1}{m+k+q+1} \cdot \frac{l}{l+q+1}} \right) \tag{34}$$

Proof. $\hat{\eta}^{(v)}_{\boldsymbol{\theta}}(\mathbf{z})$ is obtained

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

using the thinned process $X^{(v)}_c$. By proposition 1, it suffices to show that $\hat{\eta}_{\boldsymbol{\theta}, n}(\mathbf{z})$ obtained by

$$\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}) := \arg\max_{\gamma\in\mathbb{R}} \hat{\mathbb{E}}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}]$$

using the full process X satisfies the error rate in (34). We will use the approximation error of $\hat{\mathbb{E}}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}]$ to show the results. We proceed in four steps. Step 1 shows that $\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z})$ is a uniformly consistent estimator of $\eta^*_{\boldsymbol{\theta},n}(\mathbf{z})$. Step 2 shows (34).

Step 1: In this step, we show that $\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z})$ is a uniformly consistent estimator of

$$\eta^*_{\boldsymbol{\theta},n}(\mathbf{z}) := \arg\max_{\gamma\in\mathbb{R}} \mathbb{E}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}]$$

To do this, note that under condition 5.2, it followed from the Implicit Function Theorem 11.1 that $\eta^*_{\boldsymbol{\theta},n}(\mathbf{z})$ is the unique solution of $\frac{\partial}{\partial\gamma}\mathbb{E}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}] = 0$ for every $\boldsymbol{\theta},\mathbf{z}$ when n is sufficiently large, and there exists a positive constant c such that

$$\left| |A_n|^{-1}\frac{\partial}{\partial\gamma}\mathbb{E}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}]\bigg|_{\gamma=\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z})}\right| \geq c\left|\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z})\right|$$

Thus, for any $\epsilon > 0$, there exists $\delta > 0$ such that

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\Theta,\mathbf{z}\in\mathcal{Z}}\left|\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z})\right| > \epsilon\right)\\
\leq&\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\Theta,\mathbf{z}\in\mathcal{Z}}\left| |A_n|^{-1}\frac{\partial}{\partial\gamma}\mathbb{E}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}]\bigg|_{\gamma=\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z})}\right| > \delta\right)\\
\leq&\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\Theta,\mathbf{z}\in\mathcal{Z}}|A_n|^{-1}\left|\frac{\partial}{\partial\gamma}\hat{\mathbb{E}}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}]\bigg|_{\gamma=\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z})} - \frac{\partial}{\partial\gamma}\mathbb{E}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}]\bigg|_{\gamma=\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z})}\right| > \delta\right)\\
\leq&\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\Theta,\mathbf{z}\in\mathcal{Z},\gamma\in H}|A_n|^{-1}\left|\frac{\partial}{\partial\gamma}\hat{\mathbb{E}}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}] - \frac{\partial}{\partial\gamma}\mathbb{E}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}]\right| > \delta\right) \qquad (35)
\end{aligned}$$

Moreover, Lemma 10 indicates that

$$\sup_{\boldsymbol{\theta}\in\Theta,\mathbf{z}\in\mathcal{Z},\gamma\in H} |A_n|^{-1} \left| \frac{\partial}{\partial\gamma}\hat{\mathbb{E}}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}] - \frac{\partial}{\partial\gamma}\mathbb{E}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}] \right| = o_p(1) \tag{36}$$

Combine (35) and (36), it follows that for any $\epsilon > 0$,

$$\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\Theta,\mathbf{z}\in\mathcal{Z}} \left|\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z})\right| > \epsilon \right) \to 0$$

Thus, we have the uniform consistency of nuisance estimation:

$$\sup_{\boldsymbol{\theta}\in\Theta,\mathbf{z}\in\mathcal{Z}} \left|\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z})\right| = o_p(1) \tag{37}$$

Step 3: In this step, we establish (34). To do this, we first note that that

$$\begin{aligned} \frac{\partial}{\partial\gamma}\hat{\mathbb{E}}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}]\bigg|_{\gamma=\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z})} &= 0 \\ \frac{\partial}{\partial\gamma}\mathbb{E}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}]\bigg|_{\gamma=\eta^*_{\boldsymbol{\theta},n}(\mathbf{z})} &= 0 \end{aligned}$$

Denote

$$\begin{aligned} r_n(\boldsymbol{\theta},\mathbf{z}) &= \frac{\partial}{\partial\gamma}\hat{\mathbb{E}}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}]\bigg|_{\gamma=\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z})} - \frac{\partial}{\partial\gamma}\mathbb{E}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}]\bigg|_{\gamma=\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z})} \\ d_n(\boldsymbol{\theta},\mathbf{z}) &= \int_0^1 \frac{\partial^2}{\partial\gamma^2}\mathbb{E}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}]\bigg|_{\gamma=t\hat{\eta}_{\boldsymbol{\theta},n}+(1-t)\eta^*_{\boldsymbol{\theta},n}} \mathrm{d}t \end{aligned}$$

Then, by Taylor Theorem,

$$
\begin{aligned}
0 =& \frac{\partial}{\partial \gamma} \hat{\mathbb{E}}[\ell_n(\boldsymbol{\theta}, \gamma)|\mathbf{z}]\Big|_{\gamma=\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z})} - \frac{\partial}{\partial \gamma} \mathbb{E}[\ell_n(\boldsymbol{\theta}, \gamma)|\mathbf{z}]\Big|_{\gamma=\eta^*_{\boldsymbol{\theta},n}(\mathbf{z})} \\
=& r_n(\boldsymbol{\theta}, \mathbf{z}) + d_n(\boldsymbol{\theta}, \mathbf{z})(\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z})) \qquad (38)
\end{aligned}
$$

Combine the identification condition 5.2 and the uniform consistency of $\hat{\eta}_{\boldsymbol{\theta},n}$ in (37) gives

$$
\begin{aligned}
\liminf_n \inf_{\boldsymbol{\theta},\mathbf{z}} d_n(\boldsymbol{\theta}, \mathbf{z}) &= \liminf_n \inf_{\boldsymbol{\theta},\mathbf{z}} \int_0^1 \frac{\partial^2}{\partial \gamma^2} \mathbb{E}[\ell_n(\boldsymbol{\theta}, \gamma)|\mathbf{z}]\Big|_{\gamma=t\hat{\eta}_{\boldsymbol{\theta},n}+(1-t)\eta^*_{\boldsymbol{\theta},n}} \mathrm{d}t \\
&= \inf_{\boldsymbol{\theta},\mathbf{z}} \int_0^1 \liminf_n \frac{\partial^2}{\partial \gamma^2} \mathbb{E}[\ell_n(\boldsymbol{\theta}, \gamma)|\mathbf{z}]\Big|_{\gamma=t\hat{\eta}_{\boldsymbol{\theta},n}+(1-t)\eta^*_{\boldsymbol{\theta},n}} \mathrm{d}t \\
&= \inf_{\boldsymbol{\theta},\mathbf{z}} \frac{\partial^2}{\partial \gamma^2} \mathbb{E}[\ell_n(\boldsymbol{\theta}, \gamma)|\mathbf{z}]\Big|_{\gamma=\eta^*_{\boldsymbol{\theta},n}} = \Omega(|A_n|) \qquad (39)
\end{aligned}
$$

Moreover, Lemma 10 shows that

$$
\sup_{\boldsymbol{\theta},\mathbf{z}} |r_n(\boldsymbol{\theta}, \mathbf{z})| = o_p\left(|A_n|^{1-\frac{m-1}{m+k+q+1}\cdot\frac{l}{l+q+1}}\right) \qquad (40)
$$

Substitute (39), (40) into (38) gives us

$$
\sup_{\boldsymbol{\theta}\in\Theta,\mathbf{z}\in\mathcal{Z}} \left|\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z})\right| = o_p\left(|A_n|^{-\frac{m-1}{m+k+q+1}\cdot\frac{l}{l+q+1}}\right) \qquad (41)
$$

To further derive the approximation error of the derivative of $\hat{\eta}_{\boldsymbol{\theta},n}$, we differentiating equation (38) with respect to $\boldsymbol{\theta}$ and yield

$$
0 = \frac{\partial}{\partial \boldsymbol{\theta}} r_n(\boldsymbol{\theta}, \mathbf{z}) + \frac{\partial}{\partial \boldsymbol{\theta}} d_n(\boldsymbol{\theta}, \mathbf{z})(\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z})) + d_n(\boldsymbol{\theta}, \mathbf{z})\left(\frac{\partial}{\partial \boldsymbol{\theta}} \hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}) - \frac{\partial}{\partial \boldsymbol{\theta}} \eta^*_{\boldsymbol{\theta},n}(\mathbf{z})\right) \qquad (42)
$$

It followed from Corollary 1 that

$$\sup_{\boldsymbol{\theta},\gamma,\mathbf{z}} \left| \frac{\partial^2}{\partial \gamma^2 \partial \boldsymbol{\theta}} \mathbb{E}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}] \right| = O(|A_n|)$$

We thus have

$$\sup_{\boldsymbol{\theta},\mathbf{z}} \left| \frac{\partial}{\partial \boldsymbol{\theta}} d_n(\boldsymbol{\theta},\mathbf{z}) \right| = O(|A_n|) \tag{43}$$

Moreover, Lemma 10 shows that

$$\sup_{\boldsymbol{\theta},\mathbf{z}} \left| \frac{\partial}{\partial \boldsymbol{\theta}} r_n(\boldsymbol{\theta},\mathbf{z}) \right| = o_p\left(|A_n|^{1-\frac{m-1}{m+k+q+1}\cdot\frac{l}{l+q+1}} \right) \tag{44}$$

Substitute (41), (43), (44) into (42) gives us

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

Differentiating (38) twice with respect to $\boldsymbol{\theta}$ and use similar approach and we will obtain

$$\sup_{\boldsymbol{\theta},\mathbf{z}} \left| \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}) - \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \eta^*_{\boldsymbol{\theta},n}(\mathbf{z}) \right| = o_p\left(|A_n|^{-\frac{m-1}{m+k+q+1}\cdot\frac{l}{l+q+1}} \right)$$

$\square$

10 Auxiliary Lemmas

Lemma 1 (Spatial Summation)**.** *Let X be a spatial point process with intensity function $\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta)$ for some $\boldsymbol{\theta} \in \Theta, \eta \in \mathcal{H}$. Let $\{A_n\}_{n=1}^{\infty}$ be an expanding sequence of region in $\mathbb{R}^2$ that $|A_n| \to \infty$. Let $f(\mathbf{u})$ be a function defined on $\mathbb{R}^2$ that $|f(\mathbf{u})| < C$ on $\cup_{n=1}^{\infty} A_n$ for some $0 < C < \infty$. Under condition 1.1, 1.3, 1.4, we have*

$$\mathbb{E}\left[\sum_{\mathbf{u} \in X \cap A_n} f(\mathbf{u})\right] = O(|A_n|), \qquad \sum_{\mathbf{u} \in X \cap A_n} f(\mathbf{u}) - \mathbb{E}\left[\sum_{\mathbf{u} \in X \cap A_n} f(\mathbf{u})\right] = O_p(|A_n|^{\frac{1}{2}})$$

Furthermore, if there exists $0 < c$ that $|f(\mathbf{u})| > c$ on $\cup_{n=1}^{\infty} A_n$, we have

$$\mathbb{E}\left[\sum_{\mathbf{u} \in X \cap A_n} f(\mathbf{u})\right] = \Theta(|A_n|)$$

Proof. By Campbell's theorem, we have

$$\sum_{\mathbf{u} \in X \cap A_n} f(\mathbf{u}) = \int_{A_n} f(\mathbf{u}) \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) \mathrm{d}\mathbf{u}$$

By condition 1.1, the intensity function is uniformly bounded. So there exists $0 < C_0 < \infty$ such that $\sup_{\mathbf{u} \in \mathbb{R}^2, \boldsymbol{\theta} \in \Theta, \eta \in \mathcal{H}} \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) < C_0$. We let $B = \{\mathbf{u} : \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) \leq c_2\}$ where c_2 is defined in condition 1.3. Then the we decompose $\mathbb{E}\left[\sum_{\mathbf{u} \in X \cap A_n} f(\mathbf{u})\right]$ as follows:

$$\mathbb{E}\left[\sum_{\mathbf{u} \in X \cap A_n} f(\mathbf{u})\right] = \int_{A_n \cap B^c} f(\mathbf{u}) \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) \mathrm{d}\mathbf{u} + \int_{A_n \cap B} f(\mathbf{u}) \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) \mathrm{d}\mathbf{u} \tag{45}$$

Since $c < f(\mathbf{u}) < C$ on $\cup_{n=1}^{\infty} A_n$, we have

$$c \cdot c_2(|A_n| - |B|) \leq \int_{A_n \cap B^c} f(\mathbf{u})\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta)\mathrm{d}\mathbf{u} \leq C \cdot C_0|A_n| \tag{46}$$

$$0 \leq \int_{A_n \cap B} f(\mathbf{u})\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta)\mathrm{d}\mathbf{u} \leq C \cdot c_2|B| \tag{47}$$

Since B is a bounded subset of $\mathbb{R}^2$, substituting (46), (47) into (45) and we will have

$$\mathbb{E}\left[\sum_{\mathbf{u} \in X \cap A_n} f(\mathbf{u})\right] = \Theta(|A_n|)$$

If we only have $f(\mathbf{u}) < C$ on $\cup_{n=1}^{\infty} A_n$, then only the upper bounds in (46) and (47) hold and we will obtain

$$\mathbb{E}\left[\sum_{\mathbf{u} \in X \cap A_n} f(\mathbf{u})\right] = O(|A_n|)$$

Then, the variance of $\sum_{\mathbf{u} \in X \cap A_n} f(\mathbf{u})$ can be decomposed as follows:

$$\mathrm{Var}\left(\sum_{\mathbf{u} \in X \cap A_n} f(\mathbf{u})\right) = \underbrace{\int_{A_n}\int_{A_n} f(\mathbf{u})f(\mathbf{v})\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta)\lambda(\mathbf{v}; \boldsymbol{\theta}, \eta)(g(\mathbf{u}, \mathbf{v}) - 1)\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}}_{=R_1} + \underbrace{\int_{A_n} f(\mathbf{u})^2\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta)\mathrm{d}\mathbf{u}}_{=R_2}$$

Since both $f(\mathbf{u})$ and $\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta)$ are upper bounded, we have $R_2 = O(|A_n|)$. Then we combine substitute the bound of $f(\mathbf{u})$ and $\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta)$, the bound C_2 in condition 1.4 into R_1, and obtain

$$\begin{aligned} R_1 &\leq C^2 C_0^2 \int_{A_n}\int_{A_n} (g(\mathbf{u}, \mathbf{v}) - 1)\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v} \\ &= C^2 C_0^2 \int_{A_n}\int_{\mathbb{R}^2} (g(\mathbf{u}, \mathbf{v}) - 1)\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v} \\ &= C^2 C_0^2 C_2 \int_{A_n} \mathrm{d}\mathbf{v} = O(|A_n|) \end{aligned}$$

Therefore,

$$\mathrm{Var}\left(\sum_{\mathbf{u}\in X\cap A_n} f(\mathbf{u})\right) = O(|A_n|)$$

For any $M > 0$, by Markov inequality,

$$\begin{aligned}
&\mathbb{P}\left(\sum_{\mathbf{u}\in X\cap A_n} f(\mathbf{u}) - \mathbb{E}\left[\sum_{\mathbf{u}\in X\cap A_n} f(\mathbf{u})\right] \geq M|A_n|^{\frac{1}{2}}\right)\\
\leq&\frac{\mathrm{Var}\left(\sum_{\mathbf{u}\in X\cap A_n} f(\mathbf{u})\right)}{|A_n|M^2}\\
=&\frac{O(|A_n|)}{M^2|A_n|} = O(1)
\end{aligned}$$

Therefore,

$$\sum_{\mathbf{u}\in X\cap A_n} f(\mathbf{u}) - \mathbb{E}\left[\sum_{\mathbf{u}\in X\cap A_n} f(\mathbf{u})\right] = O_p(|A_n|^{\frac{1}{2}})$$

□

Corollary 1 (Spatial Summation)**.** *Let X be the spatial point process with intensity function $\lambda(\mathbf{u};\boldsymbol{\theta},\eta)$. Let $\{A_n\}_{n=1}^{\infty}$ be an expanding sequence of region in $\mathbb{R}^2$. Under condition 1.1, 1.3, 1.4, we have*

$$\mathbb{E}\left[|X\cap A_n|\right] = \Theta(|A_n|), \quad \left||X\cap A_n| - \mathbb{E}\left[|X\cap A_n|\right]\right| = O_p(|A_n|^{\frac{1}{2}}) \tag{48}$$

For $i,j\in\{0,1,2\}$, denote $\ell_n^{(i,j)}(\boldsymbol{\theta},\eta) = \frac{\partial\boldsymbol{\theta}^{i+j}}{\partial\boldsymbol{\theta}^i\partial\eta^j}\ell_n(\boldsymbol{\theta},\eta)$, we have

$$\sup_{\boldsymbol{\theta},\eta\in\mathcal{H}} \mathbb{E}\left[\ell_n^{(i,j)}(\boldsymbol{\theta},\eta)\right] = O(|A_n|), \qquad \sup_{\boldsymbol{\theta}\in\boldsymbol{\theta},\eta\in\mathcal{H}} \left|\ell_n^{(i,j)}(\boldsymbol{\theta},\eta) - \mathbb{E}\left[\ell_n^{(i,j)}(\boldsymbol{\theta},\eta)\right]\right| = O_p(|A_n|^{\frac{1}{2}}) \tag{49}$$

Proof. If we let $f(\mathbf{u}) = 1$ in Lemma 1, equation (48) follows.

By condition 1.1, $\frac{\partial \boldsymbol{\theta}^{i+j}}{\partial \boldsymbol{\theta}^i \partial \eta^j} \log \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta)$ and $\frac{\partial \boldsymbol{\theta}^{i+j}}{\partial \boldsymbol{\theta}^i \partial \eta^j} \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta)$ are upper bounded. Moreover,

$$\frac{\partial \boldsymbol{\theta}^{i+j}}{\partial \boldsymbol{\theta}^i \partial \eta^j} \ell_n(\boldsymbol{\theta}, \eta) = \sum_{\mathbf{u} \in X \cap A_n} \frac{\partial \boldsymbol{\theta}^{i+j}}{\partial \boldsymbol{\theta}^i \partial \eta^j} \log \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) - \int_{A_n} \frac{\partial \boldsymbol{\theta}^{i+j}}{\partial \boldsymbol{\theta}^i \partial \eta^j} \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) \mathrm{d}\mathbf{u}$$

We apply Lemma 1 for $f(\mathbf{u}) = \frac{\partial \boldsymbol{\theta}^{i+j}}{\partial \boldsymbol{\theta}^i \partial \eta^j} \log \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta)$ and $f(\mathbf{u}) = \frac{\partial \boldsymbol{\theta}^{i+j}}{\partial \boldsymbol{\theta}^i \partial \eta^j} \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta)$ then equation (49) follows. □

10.1 Key Lemma for Proof of Theorem 6.1

Lemma 2 (Sufficient Separation)**.** *Let X be the spatial point process with intensity function $\lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*)$ where $\boldsymbol{\theta}^*, \eta^*$ are the true parameters. Let $\{A_n\}_{n=1}^{\infty}$ be an expanding sequence of region. Under condition 1.2, 1.3, for any $(\boldsymbol{\theta}, \eta) \neq (\boldsymbol{\theta}^*, \eta^*)$ we have*

$$\mathbb{E}[\ell_n(\boldsymbol{\theta}^*, \eta^*)] - \mathbb{E}[\ell_n(\boldsymbol{\theta}, \eta)] = \Theta(|A_n|) \cdot \min\{|\boldsymbol{\theta} - \boldsymbol{\theta}^*|, 1\}$$

Proof. For any $(\boldsymbol{\theta}, \eta) \neq (\boldsymbol{\theta}^*, \eta^*)$, denote $\Psi(\mathbf{u}; \boldsymbol{\theta}, \eta) \equiv \log \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) - \log \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta)$. Given the set $C \in \mathbb{R}^2$ and constants c_0, c_1 defined in condition 1.2, we have $|\Psi(\mathbf{u}; \boldsymbol{\theta}, \eta)| \leq \min(c_0, c_1|\boldsymbol{\theta} - \boldsymbol{\theta}^*|)$ in C. The inequality $\exp(x) \geq x + 1$ has equality iff $x = 0$. Thus, there exist constants $0 < c_0', c_1' < \infty$ that

$$\inf_{\mathbf{u} \in C} |\exp(\Psi(\mathbf{u}; \boldsymbol{\theta}, \eta)) - \Psi(\mathbf{u}; \boldsymbol{\theta}, \eta) - 1| \geq \min\{c_0', c_1'|\boldsymbol{\theta} - \boldsymbol{\theta}^*|\} \tag{50}$$

Recall the definition of the set $B \in \mathbb{R}^2$ and constant c_2 defined in condition 1.3, combine with (50), we have

$$
\begin{aligned}
&\mathbb{E}[\ell_n(\boldsymbol{\theta}^*, \eta^*)] - \mathbb{E}[\ell_n(\boldsymbol{\theta}, \eta)] \\
=&\int_{A_n} (\log \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) - \log \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta))\lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*)\mathrm{d}\mathbf{u} - \int_{A_n} (\lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) - \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta))\mathrm{d}\mathbf{u} \\
=&\int_{A_n} \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) \left(\log \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*) - \log \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) - 1 + \frac{\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta)}{\lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*)}\right) \mathrm{d}\mathbf{u} \\
=&\int_{A_n} \lambda(\mathbf{u}; \boldsymbol{\theta}^*, \eta^*)(\exp(\Psi(\mathbf{u}; \boldsymbol{\theta}, \eta)) - \Psi(\mathbf{u}; \boldsymbol{\theta}, \eta) - 1)\mathrm{d}\mathbf{u} \\
\geq& c_2 \int_{A_n \cap B^c} (\exp(\Psi(\mathbf{u}; \boldsymbol{\theta}, \eta)) - \Psi(\mathbf{u}; \boldsymbol{\theta}, \eta) - 1)\mathrm{d}\mathbf{u} \\
\geq& c_2 \int_{A_n \cap B^c \cap C} \min\{c_0', c_1'|\boldsymbol{\theta} - \boldsymbol{\theta}^*|\}\mathrm{d}\mathbf{u} \\
\geq& c_2 \min\{c_0', c_1'|\boldsymbol{\theta} - \boldsymbol{\theta}^*|\}|(|A_n \cap C| - |B|) \qquad (51)
\end{aligned}
$$

By condition 1.2, $|A_n \cap C| = \Theta(A_n)$. By condition 1.3, $|B| < \infty$. Thus,

$$
(|A_n \cap C| - |B|) = \Theta(|A_n|) \qquad (52)
$$

Substitute into , we will obtain

$$
\mathbb{E}[\ell_n(\boldsymbol{\theta}^*, \eta^*)] - \mathbb{E}[\ell_n(\boldsymbol{\theta}, \eta)] = \Theta(|A_n|) \cdot \min\{|\boldsymbol{\theta} - \boldsymbol{\theta}^*|, 1\}
$$

□

10.2 Rate of Plug-in Error

Consider the expanding sequence of observational windows $\{A_n\}$. Let $\hat{\eta}_{\boldsymbol{\theta},n}$ be some nuisance estimation of $\eta^*_{\boldsymbol{\theta},n}$ related to A_n. We will first show how the nuisance estimation error of $\hat{\eta}_{\boldsymbol{\theta},n} - \eta^*_{\boldsymbol{\theta},n}$ affect the first order plug-in error $\ell_n(\boldsymbol{\theta}, \hat{\eta}^*_{\boldsymbol{\theta},n}) - \ell_n(\boldsymbol{\theta}, \eta_{\boldsymbol{\theta},n})$, and the second order plug-in error $\ell_n(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta},n}) - \ell_n(\boldsymbol{\theta}, \eta_{\boldsymbol{\theta},n}) - \frac{\partial}{\partial \eta}\ell_n(\boldsymbol{\theta}, \eta_{\boldsymbol{\theta},n})[\hat{\eta}_{\boldsymbol{\theta},n} - \eta^*_{\boldsymbol{\theta},n}]$

Lemma 3 (First Order Plug-in Error)**.** *Let $\hat{\eta}_{\boldsymbol{\theta},n}$ be a sequence of estimators of $\eta_{\boldsymbol{\theta},n}$. Let*

$$r_n^{(1)}(\boldsymbol{\theta}) = \ell_n(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta},n}) - \ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}) \tag{53}$$

$$e_n^{(1)} = \sup_{j\in\{0,1,2\}, \boldsymbol{\theta}\in\Theta, \mathbf{z}\in\mathcal{Z}} \left| \frac{\partial^j}{\partial \boldsymbol{\theta}^j}(\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z}))\right| \tag{54}$$

Then, under condition 1.1, 1.3, 1.4, we have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}} |r_n^{(1)}(\boldsymbol{\theta})| = O(|A_n| \cdot e_n^{(1)}) \tag{55}$$

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}} \left| \frac{\partial^2}{\partial \boldsymbol{\theta}^2} r_n^{(1)}(\boldsymbol{\theta}) \right| = O(|A_n| \cdot e_n^{(1)}) \tag{56}$$

Proof. For fixed $\boldsymbol{\theta}$, denote

$$Q^{(1)}_{\boldsymbol{\theta},n}(\mathbf{u}) = \int_0^1 \frac{\partial}{\partial \eta(\mathbf{z}(\mathbf{u}))} \lambda(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u})), \eta_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) + t(\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u}))))dt$$

$$Q^{(2)}_{\boldsymbol{\theta},n}(\mathbf{u}) = \int_0^1 \frac{\partial}{\partial \eta(\mathbf{z}(\mathbf{u}))} \log\left\{\lambda(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u})), \eta_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) + t(\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u}))))\right\} dt$$

By the smoothness condition 1.1, there exists $0 < C_4 < \infty$ that uniformly with respect to $\mathbf{u} \in \mathbb{R}^2$, $\boldsymbol{\theta} \in \Theta$ and $\eta \in \mathcal{H}$.

$$\sup_{i,j\in\{0,1,2\}} \left| \frac{\partial^{i+j}}{\partial \eta(\mathbf{z}(\mathbf{u}))^i \partial \boldsymbol{\theta}^j} \lambda(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u})), \eta(\mathbf{z}(\mathbf{u})))\right| < C_4$$

Thus, uniformly with respect to $\mathbf{u} \in \mathbb{R}^2, \boldsymbol{\theta} \in \Theta$,

$$\sup_{i\in\{0,1,2\} j\in\{1,2\}} \left| \frac{\partial^i}{\partial \boldsymbol{\theta}^i} Q^{(j)}_{\boldsymbol{\theta},n}(\mathbf{u})\right| < C_4 \tag{57}$$

We apply first order Taylor Theorem to every point $\mathbf{u}$ to (53) as follows:

$$\begin{aligned} r_n^{(1)}(\boldsymbol{\theta}) &= \sum_{\mathbf{u}\in X\cap A_n} \left\{\log\lambda(\mathbf{u};\boldsymbol{\theta},\hat{\eta}_{\boldsymbol{\theta},n}) - \log\lambda(\mathbf{u};\boldsymbol{\theta},\eta^*_{\boldsymbol{\theta},n})\right\} - \int_{A_n}\left\{\lambda(\mathbf{u};\boldsymbol{\theta},\hat{\eta}_{\boldsymbol{\theta},n}) - \lambda(\mathbf{u};\boldsymbol{\theta},\eta^*_{\boldsymbol{\theta},n})\right\}\mathrm{d}\mathbf{u} \\ &= \sum_{\mathbf{u}\in X\cap A_n} Q^{(2)}_{\boldsymbol{\theta},n}(\mathbf{u})\cdot(\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u}))) - \int_{A_n} Q^{(1)}_{\boldsymbol{\theta},n}(\mathbf{u})\cdot(\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) - \eta_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})))\mathrm{d}\mathbf{u} \end{aligned}$$

Moreover, the second order derivative of (53) is:

$$\begin{aligned} \frac{\partial^2}{\partial\boldsymbol{\theta}^2} r_n^{(1)}(\boldsymbol{\theta}) = \sum_{i=0}^{2}\Bigg\{ &\sum_{\mathbf{u}\in X\cap A_n} \frac{\partial^i}{\partial\boldsymbol{\theta}^i} Q^{(2)}_{\boldsymbol{\theta},n}(\mathbf{u})\cdot\frac{\partial^{2-i}}{\partial\boldsymbol{\theta}^{2-i}}(\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u}))) - \\ &\int_{A_n} \frac{\partial^i}{\partial\boldsymbol{\theta}^i} Q^{(1)}_{\boldsymbol{\theta},n}(\mathbf{u})\cdot\frac{\partial^{2-i}}{\partial\boldsymbol{\theta}^{2-i}}(\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})))\mathrm{d}\mathbf{u}\Bigg\} \end{aligned} \quad (58)$$

Combining (54), (57), and equation (48) in Corallary 1, Cauchy-Schwartz inequality gives us that for $i = 0, 1, 2, j = 0, 1, 2$ satisfying $i + j \leq 2$, we have

$$\begin{aligned} &\left|\sum_{\mathbf{u}\in X\cap A_n} \frac{\partial^i}{\partial\boldsymbol{\theta}^i} Q^{(2)}_{\boldsymbol{\theta},n}(\mathbf{u})\cdot\frac{\partial^j}{\partial\boldsymbol{\theta}^j}(\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})))\right| \\ \leq &\left\{\sum_{\mathbf{u}\in X\cap A_n}\left|\frac{\partial^i}{\partial\boldsymbol{\theta}^i} Q^{(2)}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u}))\right|^2\right\}^{\frac{1}{2}}\cdot\left\{\sum_{\mathbf{u}\in X\cap A_n}\left|\frac{\partial^j}{\partial\boldsymbol{\theta}^j}(\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})))\right|^2\right\}^{\frac{1}{2}} \\ \leq &C_4|X\cap A_n|e_n^{(1)} = O_p(|A_n|\cdot e_n^{(1)}) \end{aligned}$$

Similarly, we have

$$\begin{aligned} &\left|\int_{A_n} \frac{\partial^i}{\partial\boldsymbol{\theta}^i} Q^{(1)}_{\boldsymbol{\theta},n}(\mathbf{u})\cdot\frac{\partial^j}{\partial\boldsymbol{\theta}^j}(\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})))\mathrm{d}\mathbf{u}\right| \\ \leq &\left\{\int_{A_n}\left|\frac{\partial^i}{\partial\boldsymbol{\theta}^i} Q^{(1)}_{\boldsymbol{\theta},n}(\mathbf{u})\right|^2\mathrm{d}\mathbf{u}\right\}^{\frac{1}{2}}\cdot\left\{\int_{A_n}\left|\frac{\partial^j}{\partial\boldsymbol{\theta}^j}(\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})))\right|^2\mathrm{d}\mathbf{u}\right\}^{\frac{1}{2}} \\ \leq &C_4|A_n|e_n^{(1)} = O(|A_n|\cdot e_n^{(1)}) \end{aligned}$$

Substituting the last two bounds into (58) gives us

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}} |r_n^{(1)}(\boldsymbol{\theta})| = O(|A_n| \cdot e_n^{(1)})$$
$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}} \left|\frac{\partial^2}{\partial\boldsymbol{\theta}^2} r_n^{(1)}(\boldsymbol{\theta})\right| = O(|A_n| \cdot e_n^{(1)})$$

□

Lemma 4 (Second Order Plug-in Error)**.** *Let $\hat{\eta}_{\boldsymbol{\theta},n}$ be a sequence of estimators of $\eta_{\boldsymbol{\theta},n}$. Let*

$$r_n^{(2)}(\boldsymbol{\theta}) = \ell_n(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta},n}) - \ell_n(\boldsymbol{\theta}, \eta_{\boldsymbol{\theta},n}) - \frac{\partial}{\partial\eta}\ell_n(\boldsymbol{\theta}, \eta_{\boldsymbol{\theta},n}) \left[\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z})\right] \tag{59}$$
$$e_n^{(2)}(\boldsymbol{\theta}) = \sup_{j\in\{0,1\},\mathbf{z}\in\mathcal{Z}} \left|\frac{\partial^j}{\partial\boldsymbol{\theta}^j}(\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z}))\right| \tag{60}$$

Then, under condition 1.3, 1.4, we have

$$\left|\frac{\partial}{\partial\boldsymbol{\theta}} r_n^{(2)}(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right| = O_p(|A_n| \cdot e_n^{(2)}(\boldsymbol{\theta}^*)) \tag{61}$$

Proof. For fixed $\boldsymbol{\theta}$, we denote

$$Q^{(3)}_{\boldsymbol{\theta},n}(\mathbf{u}) = \frac{1}{2}\int_0^1 \frac{\partial^2}{\partial\eta(\mathbf{z}(\mathbf{u}))^2}\lambda(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u})), \eta_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) + t(\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u}))))dt$$

$$Q^{(4)}_{\boldsymbol{\theta},n}(\mathbf{u}) = \frac{1}{2}\int_0^1 \frac{\partial^2}{\partial\eta(\mathbf{z}(\mathbf{u}))^2}\log\left\{\lambda(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u})), \eta_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) + t(\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u}))))\right\}dt$$

By the smoothness condition 1.1, there exists $0 < C_4 < \infty$ that uniformly with respect to $\mathbf{u} \in \mathbb{R}^2$, $\boldsymbol{\theta} \in \Theta$ and $\eta \in \mathcal{H}$.

$$\sup_{i,j\in\{0,1,2\}} \left|\frac{\partial^{i+j}}{\partial\eta(\mathbf{z}(\mathbf{u}))^i\partial\boldsymbol{\theta}^j}\lambda(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u})), \eta(\mathbf{z}(\mathbf{u})))\right| < C_4$$

Thus, uniformly with respect to $\mathbf{u} \in \mathbb{R}^2, \boldsymbol{\theta} \in \Theta$

$$\sup_{i\in\{0,1\} j\in\{3,4\}} \left| \frac{\partial^i}{\partial \boldsymbol{\theta}^i} Q^{(j)}_{\boldsymbol{\theta},n}(\mathbf{u}) \right| < C_4 \tag{62}$$

Apply the second-order Taylor Theorem at every point $\mathbf{u}$ to (59) gives us

$$r^{(2)}_n(\boldsymbol{\theta}) = \sum_{\mathbf{u}\in X\cap A_n} Q^{(4)}_{\boldsymbol{\theta},n}(\mathbf{u}) \cdot (\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})))^2 - \int_{A_n} Q^{(3)}_{\boldsymbol{\theta},n}(\mathbf{u}) \cdot (\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})))^2 \mathrm{d}\mathbf{u}$$

Moreover, the first-order derivative of $r^{(2)}_n(\boldsymbol{\theta})$ can be expressed as follows:

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\theta}} r^{(2)}_n(\boldsymbol{\theta}) = \sum_{i=0}^{1} \Bigg\{ & \sum_{\mathbf{u}\in X\cap A_n} \frac{\partial^i}{\partial \boldsymbol{\theta}^i} Q^{(4)}_{\boldsymbol{\theta},n}(\mathbf{u}) \cdot \frac{\partial^{1-i}}{\partial \boldsymbol{\theta}^{1-i}} (\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})))^2 \\ & - \int_{A_n} \frac{\partial^i}{\partial \boldsymbol{\theta}^i} Q^{(3)}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) \cdot \frac{\partial^{1-i}}{\partial \boldsymbol{\theta}^{1-i}} (\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) - \eta_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})))^2 \mathrm{d}\mathbf{u} \Bigg\}\end{aligned} \tag{63}$$

Combining (60), (62), and (48), Cauchy-Schwartz inequality gives that for $i = 0, 1, j = 0, 1$ satisfying $i + j \leq 1$

$$\begin{aligned}& \left| \sum_{\mathbf{u}\in X\cap A_n} \frac{\partial^i}{\partial \boldsymbol{\theta}^i} Q^{(4)}_{\boldsymbol{\theta},n}(\mathbf{u}) \cdot \frac{\partial^j}{\partial \boldsymbol{\theta}^j} (\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})))^2 \right| \\ \leq & \left\{ \sum_{\mathbf{u}\in X\cap A_n} \left| \frac{\partial^i}{\partial \boldsymbol{\theta}^i} Q^{(4)}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) \right|^2 \right\}^{\frac{1}{2}} \cdot \left\{ \sum_{\mathbf{u}\in X\cap A_n} \left| \frac{\partial^j}{\partial \boldsymbol{\theta}^j} (\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})) - \eta^*_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})))^2 \right|^2 \right\}^{\frac{1}{2}} \\ \leq & C_4 |X \cap A_n| e^{(2)}_n(\boldsymbol{\theta}) = O_p(|A_n| \cdot e^{(2)}_n(\boldsymbol{\theta}))\end{aligned}$$

Similarly, combining (60), (62),Cauchy-Schwartz inequality gives that for $i = 0, 1, j = 0, 1$

satisfying $i+j\leq 1$

$$
\begin{aligned}
&\left|\int_{A_n}\frac{\partial^i}{\partial\boldsymbol{\theta}^i}Q^{(3)}_{\boldsymbol{\theta},n}(\mathbf{u})\cdot\frac{\partial^j}{\partial\boldsymbol{\theta}^j}(\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u}))-\eta^*_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})))^2\mathrm{d}\mathbf{u}\right|\\
\leq&\left\{\int_{A_n}\left|\frac{\partial^i}{\partial\boldsymbol{\theta}^i}Q^{(3)}_{\boldsymbol{\theta},n}(\mathbf{u})\right|^2\mathrm{d}\mathbf{u}\right\}^{\frac{1}{2}}\cdot\left\{\int_{A_n}\left|\frac{\partial^j}{\partial\boldsymbol{\theta}^j}(\hat{\eta}_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u}))-\eta^*_{\boldsymbol{\theta},n}(\mathbf{z}(\mathbf{u})))^2\right|^2\mathrm{d}\mathbf{u}\right\}^{\frac{1}{2}}\\
\leq& C_4|A_n|e^{(2)}_n(\boldsymbol{\theta})=O(|A_n|\cdot e^{(2)}_n(\boldsymbol{\theta}))
\end{aligned}
$$

Substituting the last two bounds into (63) gives us

$$
\left|\frac{\partial}{\partial\boldsymbol{\theta}}r^{(2)}_n(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right|=O_p(|A_n|\cdot r^{(2)}_n(\boldsymbol{\theta}^*))
$$

□

Lemma 5 (Spatial Empirical Process)**.** *Let X be the spatial point process with intensity function $\lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*)$ where $\boldsymbol{\theta}^*,\eta^*$ are the true parameters. Let $\hat{\eta}_{\boldsymbol{\theta},n}$ be a estimator of $\eta^*_{\boldsymbol{\theta},n}$ satisfies Assumption 3. Then under condition 1.1, 1.3, 1.4, if either $\hat{\eta}_{\boldsymbol{\theta},n}$ is independent with $\ell_n(\boldsymbol{\theta},\eta_{\boldsymbol{\theta},n})$ or the intensity function is log-linear, we have*

$$
\left|\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\eta}\ell_n(\boldsymbol{\theta},\eta^*_{\boldsymbol{\theta},n})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}[\hat{\eta}_{\boldsymbol{\theta}^*,n}-\eta^*_{\boldsymbol{\theta}^*,n}]\right|=o_p(|A_n|^{\frac{1}{2}}) \tag{64}
$$

$$
\left|\frac{\partial}{\partial\eta}\ell_n(\boldsymbol{\theta}^*,\eta^*_{\boldsymbol{\theta}^*,n})\left[\frac{\partial}{\partial\boldsymbol{\theta}}\hat{\eta}_{\boldsymbol{\theta}^*,n}-\frac{\partial}{\partial\boldsymbol{\theta}}\eta^*_{\boldsymbol{\theta}^*,n}\right]\right|=o_p(|A_n|^{\frac{1}{2}}) \tag{65}
$$

Proof. The proof of the lemma consists of three steps. In step 1, we illustrate how the two terms we need to bound can be regarded as the generalization of empirical processes and the challenge of extending the Maximal Inequality (See Lemma 6.2 Chernozhukov et al. [2018]) to the spatial setting. Step 2 shows the result when $\hat{\eta}_{\boldsymbol{\theta},n}$ is independent with $\ell_n(\boldsymbol{\theta},\eta_{\boldsymbol{\theta},n})$. Step 3 shows the result when the intensity function is log-linear.

Step 1: First, we will explain how the two terms we need to bound corresponds to the empirical process in the classic *i.i.d.* setting. Suppose $\{Y_i\}_{i=1}^n$ are n *i.i.d* samples drawn from random variable Y, $\mathcal{F}$ is a class of functional of Y with probability measure $\mathbb{P}$, $\mathbb{G}_n$ be the random measure $n^{-\frac{1}{2}}(\mathbb{P}_n - \mathbb{P})$ Then empirical process $\mathbb{G}_n[f]$ is defined as

$$\mathbb{G}_n[f] := n^{-\frac{1}{2}} \sum_{i=1}^{n} (f(Y_i) - \mathbb{E}[f(Y_i)])$$

Then under some restrction of the complexity of the functional class $\mathcal{F}$ (VC-dimension, packing/covering number), we can use Maximal Inequality (Lemma 6.2 in Chernozhukov et al. [2018]) to bound $\sup_{f\in\mathcal{F}} |\mathbb{G}_n[f]|$ without considering the dependence between $\mathbb{G}_n$ and $\mathcal{F}$.

Now, we consider the generalization of the classic empirical process to the spatial point process. Given a spatial point processes X defined on $\mathbb{R}^2$ with intensity function $\lambda(\mathbf{u})$ and a sequence of expanding region $\{A_n\}_{n=1}^{\infty}$, we let $\mathcal{F}$ be functional class consisting of functions $f(\mathbf{u})$ defined on spatial region $\mathbb{R}^2$. We define the *spatial empirical processes* $\mathbb{G}_n^s[f]$ as

$$\mathbb{G}_n^s[f] := |A_n|^{-\frac{1}{2}} \left\{ \sum_{\mathbf{u}\in X\cap A_n} f(\mathbf{u}) - \int_{A_n} f(\mathbf{u})\lambda(\mathbf{u})\mathrm{d}\mathbf{u} \right\}$$

Note that $|A_n|^{\frac{1}{2}}\mathbb{G}_n^s$ is also referred to as innovation/residual measure in spatial point process literatures (See Baddeley et al. [2008]). We further denote

$$
\begin{aligned}
V_n^{(1)}(f)(\mathbf{u}) &:= \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \eta} \log \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n})\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} [f(\mathbf{u})] \\
V_n^{(2)}(f)(\mathbf{u}) &:= \frac{\partial}{\partial \eta} \log \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n})\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} [f(\mathbf{u})] \\
\mathcal{F}_n^1 &:= \left\{ \left(\hat{\eta}_{\boldsymbol{\theta}^*,n} - \eta^*_{\boldsymbol{\theta}^*,n}\right) \circ \mathbf{z} \right\} \\
\mathcal{F}_n^2 &:= \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \left(\hat{\eta}_{\boldsymbol{\theta}^*,n} - \eta^*_{\boldsymbol{\theta}^*,n}\right) \circ \mathbf{z} \right\} \\
|\mathcal{F}| &:= \sup_{\mathbf{u} \in \mathbb{R}^2} |f(\mathbf{u})|
\end{aligned}
$$

Then the equality (64) is equivalent to

$$
\sup_{f \in \mathcal{F}_n^1} \left| \mathbb{G}_n^s [V_n^{(1)}(f)] \right| = o_p(1) \tag{66}
$$

The equality (65) is equivalent to

$$
\sup_{f \in \mathcal{F}_n^2} \left| \mathbb{G}_n^s [V_n^{(2)}(f)] \right| = o_p(1) \tag{67}
$$

which have the same expression as empirical process in $i.i.d$ case expect that the empirical measure $\mathbb{G}_n$ is replaced by spatial residual measure $\mathbb{G}_n^s$.

Step 2: In this step, we will establish (66) and (67) when $\hat{\eta}_{\boldsymbol{\theta},n}$ is independent with $\ell_n(\boldsymbol{\theta}, \eta_{\boldsymbol{\theta},n})$, i.e. independent with $\mathbb{G}_n^s$. First, we observe that by triangular inequality

$$
\sup_{f \in \mathcal{F}_n^1} \left| \mathbb{G}_n^s [V_n^{(1)}(f)] \right| \leq \sup_{f \in \mathcal{F}_n^1} \left| \mathbb{G}_n^s [V_n^{(1)}(f)] - \mathbb{E}\left[\mathbb{G}_n^s [V_n^{(1)}(f)] \right] \right| + \sup_{f \in \mathcal{F}_n^1} \left| \mathbb{E}\left[\mathbb{G}_n^s [V_n^{(1)}(f)] \right] \right| \tag{68}
$$

Note that for any fixed function f, by Campbell's theorem,

$$\mathbb{E}\left[\mathbb{G}_n^s[f]\right] = 0$$

We thus have

$$\sup_{f \in \mathcal{F}_n^1} \left| \mathbb{E}\left[\mathbb{G}_n^s[V_n^{(1)}(f)]\right] \right| = \sup_{f \in \mathcal{F}_n^1} \left| \mathbb{E}\left[\mathbb{E}\left[\mathbb{G}_n^s[V_n^{(1)}(f)] \mid f\right]\right] \right| = 0 \tag{69}$$

Additionally, since Lemma 1 indicates that $\mathbb{G}_n^s[1] = O_p(1)$, Assumption 3 indicates that $|\mathcal{F}_n^1| = o_p(1)$, condition 1.1 indicates that $|V_n^{(1)}|$ is bounded, we thus have

$$\begin{aligned}
&\sup_{f \in \mathcal{F}_n^1} \left| \mathbb{G}_n^s[V_n^{(1)}(f)] - \mathbb{E}\left[\mathbb{G}_n^s[V_n^{(1)}(f)]\right] \right| \\
&= \sup_{f \in \mathcal{F}_n^1} \left| \mathbb{G}_n^s[V_n^{(1)}(f)] \right| \leq \sup_{f \in \mathcal{F}_n^1} |V_n^{(1)}(f)| \, |\mathbb{G}_n^s[1]| \leq |\mathcal{F}_n^1| |V_n^{(1)}| \, |\mathbb{G}_n^s[1]| = o_p(1)
\end{aligned} \tag{70}$$

Substituting the bound of expectation (69) and the bound of deviation (70) into the triangular inequality (68) gives us (66). Following the exact same argument for $\sup_{f \in \mathcal{F}_n^2} \left| \mathbb{G}_n^s[V_n^{(2)}(f)] \right|$ will give us (67).

Step 3: In this step, we will establish the bound of spatial empirical processes (66) and (67) when the intensity function of X is log-linear, i.e. $\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) = \exp(\boldsymbol{\theta}^\top \mathbf{y}(\mathbf{u}) + \eta(\mathbf{z}(\mathbf{u})))$.

To obtain (66), note that for every $\mathbf{u}$, by the chain rule,

$$
\begin{aligned}
&\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\eta(\mathbf{z})}\log\lambda(\mathbf{u};\boldsymbol{\theta},\eta^*_{\boldsymbol{\theta},n})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\\
=&\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\eta(\mathbf{z})}\log\lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*)+\frac{\partial^2}{\partial\eta(\mathbf{z})^2}\log\lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*)\boldsymbol{\nu}^*_n(\mathbf{z}(\mathbf{u}))\\
=&\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\eta(\mathbf{z})}(\boldsymbol{\theta}^{*\top}\mathbf{y}(\mathbf{u})+\eta^*(\mathbf{z}(\mathbf{u})))+\frac{\partial^2}{\partial\eta(\mathbf{z})^2}(\boldsymbol{\theta}^{*\top}\mathbf{y}(\mathbf{u})+\eta^*(\mathbf{z}(\mathbf{u})))\boldsymbol{\nu}^*_n(\mathbf{z}(\mathbf{u}))\\
=&\mathbf{0}+0\cdot\boldsymbol{\nu}^*_n(\mathbf{z}(\mathbf{u}))=\mathbf{0}
\end{aligned}
$$

Thus, $V_n^{(1)}(f)=0$ for abitrary f so that $\sup_{f\in\mathcal{F}_n^1}\left|\mathbb{G}_n^s[V_n^{(1)}(f)]\right|=0$.

To obtain the second spatial empirical process bound (67), we note that in the notation guideline we have already defined

$$
\hat{\boldsymbol{\nu}}_n(\mathbf{z})=\frac{\partial}{\partial\boldsymbol{\theta}}\hat{\eta}_{\boldsymbol{\theta}^*,n},\quad \boldsymbol{\nu}^*_n(\mathbf{z})=\frac{\partial}{\partial\boldsymbol{\theta}}\eta^*_{\boldsymbol{\theta}^*,n}
$$

When the intensity function is log-linear,

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

Since $\hat{\boldsymbol{\nu}}_n(\mathbf{z})$ is obtained using set $\{\mathbf{u}_i\}_{i=1}^{m_n}$ randomly drawn from A_n, $\mathcal{F}_n^2$ is always independent with $\mathbb{G}_n^s$. Then the second spatial empirical process bound (67) is followed from the argument in Step 2. □

Lemma 6 (Central Limit Theorem)**.** *Under condition 1.1, 2.1, 2.2, 2.3,*

$$
|A_n|^{-\frac{1}{2}}\bar{\boldsymbol{\Sigma}}_n^{-\frac{1}{2}}(\boldsymbol{\theta}^*,\eta^*,\nu^*,\psi^*)\frac{\partial}{\partial\boldsymbol{\theta}}\ell_n(\boldsymbol{\theta},\eta^*_{\boldsymbol{\theta},n})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\to_d N(\mathbf{0},\mathbf{I}_k)
$$

Proof. For every pair of integer $(i,j)\in\mathbb{Z}^2$, we let $C(i,j)$ be the unit volume square centered at (i,j) and let $\mathcal{D}_n = \{(i,j) \in \mathbb{Z}^2 : C(i,j) \cap A_n \neq \varnothing\}$. We denote

$$Z_{i,j} = \sum_{\mathbf{u} \in X \cap C(i,j) \cap A_n} \frac{\partial}{\partial \boldsymbol{\theta}} \log \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} - \int_{C(i,j) \cap A_n} \frac{\partial}{\partial \boldsymbol{\theta}} \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \mathrm{d}\mathbf{u}$$

Then,

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = \sum_{(i,j) \in \mathcal{D}_n} Z_{i,j}$$

By Theorem 1 in Biscio and Waagepetersen [2019],

$$\mathrm{Var}\left(\sum_{(i,j) \in \mathcal{D}_n} Z_{i,j} \right)^{-\frac{1}{2}} \left(\sum_{(i,j) \in \mathcal{D}_n} Z_{i,j} - \mathbb{E}\left[\sum_{(i,j) \in \mathcal{D}_n} Z_{i,j} \right] \right) \to_d N(0, I_k) \tag{71}$$

if the following assumptions are satisfied

1. $A_1 \subset A_2 \subset \ldots$ and $|\bigcup_{n=1}^{\infty} A_n| = \infty$
2. The α-mixing coefficient of X satisfies $\alpha^X_{2,\infty}(r) = O(r^{-(2+\epsilon)})$ for some $\epsilon > 0$
3. There exists $\tau > \frac{4}{\epsilon}$ such that $\sup_{n \in \mathbb{N}} \sup_{(i,j) \in \mathcal{D}_n} \mathbb{E}\left[\left\| Z_{(i,j)} - \mathbb{E}\left[Z_{(i,j)} \right] \right\|^{2+\tau} \right] < \infty$
4. The limit infimum of the smallest eigen value of $\boldsymbol{\Sigma}_n(\boldsymbol{\theta}^*, \eta^*, \nu^*, \psi^*)$ is larger than zero.

where the first condition 1 naturally holds in our setting. The second condition 2 is satisfied by when condition 2.3 holds. The third condition 3 is satisifed because $Z_{i,j}$ are uniformly bounded due to condition 1.1. The fourth condition 4 is satisfied by condition 2.2.

Additionally, since

$$\mathbb{E}\left[\sum_{(i,j) \in \mathcal{D}_n} Z_{i,j} \right] = \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}, \eta^*_{\boldsymbol{\theta},n}) \right] \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$$

$$\operatorname{Var}\left(\sum_{(i,j)\in\mathcal{D}_n} Z_{i,j}\right) = |A_n|\bar{\mathbf{\Sigma}}_n(\boldsymbol{\theta}^*, \eta^*, \nu^*, \psi^*)$$

Equation (71) leads to the result. $\square$

10.3 Lemmas for the proof of Theorem 9.1

Lemma 7 (Kernel Approximation Bias)**.** *Let $h(\mathbf{y}, \mathbf{z})$ be a bounded function defined on the covariance domain $\mathcal{Y} \times \mathcal{Z}$ that is l-th order continuously differentiable with respect to $\mathbf{z}$. Let $K_h(\cdot)$ be an l-th order Kernel function(See definition 11.1) where h is the bandwidth. Then we have*

$$\sup_{\mathbf{z}^*\in\mathcal{Z}}\left|\int_{A_n} h(\mathbf{y}(\mathbf{u}), \mathbf{z}(\mathbf{u}))K_h\left(\mathbf{z}(\mathbf{u}) - \mathbf{z}^*\right)\mathrm{d}\mathbf{u} - \int_{\mathcal{Y}} h(\mathbf{y}, \mathbf{z}^*)f_n(\mathbf{y}, \mathbf{z}^*)\mathrm{d}\mathbf{y}\right| = O(|A_n|h^l) \tag{72}$$

Proof. To make a change of the variable, we let $\mathbf{t} = \frac{\mathbf{z}-\mathbf{z}^*}{h}$ and let $f_n(\mathbf{y}, \mathbf{z}^* + h\mathbf{t}) = 0$ for $\mathbf{t} \notin \frac{\mathcal{Z}-\mathbf{z}^*}{t}$. Then by the change of variable, we have

$$\begin{aligned}
&\int_{A_n} h(\mathbf{y}(\mathbf{u}), \mathbf{z}(\mathbf{u}))K_h\left(\mathbf{z}(\mathbf{u}) - \mathbf{z}^*\right)\mathrm{d}\mathbf{u} \\
=&\int_{\mathcal{Z}}\int_{\mathcal{Y}} h(\mathbf{y}, \mathbf{z})K_h\left(\mathbf{z} - \mathbf{z}^*\right) f_n(\mathbf{y}, \mathbf{z})\mathrm{d}\mathbf{y}\mathrm{d}\mathbf{z} \\
=&\int_{\mathcal{Y}}\int_{\mathbb{R}^q} h(\mathbf{y}, \mathbf{z}^* + h\mathbf{t})K\left(\mathbf{t}\right) f_n(\mathbf{y}, \mathbf{z}^* + h\mathbf{t})\mathrm{d}\mathbf{t}\mathrm{d}\mathbf{y}
\end{aligned} \tag{73}$$

Apply this change of variable to the left-hand-side of 72 as follows

$$\begin{aligned}
&\sup_{\mathbf{z}^*\in\mathcal{Z}}\left|\int_{A_n} h(\mathbf{y}(\mathbf{u}), \mathbf{z}(\mathbf{u}))K_h\left(\mathbf{z}(\mathbf{u}) - \mathbf{z}^*\right)\mathrm{d}\mathbf{u} - \int_{\mathcal{Y}} h(\mathbf{y}, \mathbf{z}^*)f_n(\mathbf{y}, \mathbf{z}^*)\mathrm{d}\mathbf{y}\right| \\
=&\sup_{\mathbf{z}^*\in\mathcal{Z}}\left|\int_{\mathcal{Y}}\int_{\mathbb{R}^q} h(\mathbf{y}, \mathbf{z}^* + h\mathbf{t})K\left(\mathbf{t}\right) f_n(\mathbf{y}, \mathbf{z}^* + h\mathbf{t})\mathrm{d}\mathbf{t}\mathrm{d}\mathbf{y} - \int_{\mathcal{Y}}\int_{\mathbb{R}^q} h(\mathbf{y}, \mathbf{z}^*)K\left(\mathbf{t}\right) f_n(\mathbf{y}, \mathbf{z}^*)\mathrm{d}\mathbf{t}\mathrm{d}\mathbf{y}\right| \\
=&\int_{\mathcal{Y}}\int_{\mathbb{R}^q} K\left(\mathbf{t}\right)\sup_{\mathbf{z}^*\in\mathcal{Z}}\left|h(\mathbf{y}, \mathbf{z}^* + h\mathbf{t})f_n(\mathbf{z}^* + h\mathbf{t}|\mathbf{y}) - h(\mathbf{y}, \mathbf{z}^*)f_n(\mathbf{z}^*|\mathbf{y})\right|\mathrm{d}\mathbf{t}f_n(\mathbf{y})\mathrm{d}\mathbf{y}
\end{aligned} \tag{74}$$

Since $h(\mathbf{y}, \mathbf{z}) f_n(\mathbf{y}, \mathbf{z})$ is l-th order continuously differentiable with respect to $\mathbf{z}$ and $\mathcal{Z}$ is compact, there exists some $0 < c < \infty$ that

$$\sup_{\mathbf{z} \in \mathcal{Z}} |h(\mathbf{y}, \mathbf{z}^* + h\mathbf{t}) f_n(\mathbf{z}^* + h\mathbf{t}|\mathbf{y}) - h(\mathbf{y}, \mathbf{z}^*) f_n(\mathbf{z}^*|\mathbf{y})| \leq c h^l \mathbf{t}^l \tag{75}$$

By definition 11.1,

$$\int_{\mathbb{R}^q} K(\mathbf{t}) \mathbf{t}^l \mathrm{d}\mathbf{t} = \kappa_q \tag{76}$$

Then we substitute (75) and (76) into (74) and obtain

$$\begin{aligned}
&\sup_{\mathbf{z}^* \in \mathcal{Z}} \left| \int_{A_n} h(\mathbf{y}(\mathbf{u}), \mathbf{z}(\mathbf{u})) K_h(\mathbf{z}(\mathbf{u}) - \mathbf{z}^*) \mathrm{d}\mathbf{u} - \int_{\mathcal{Y}} h(\mathbf{y}, \mathbf{z}^*) f_n(\mathbf{y}, \mathbf{z}^*) \mathrm{d}\mathbf{y} \right| \\
\leq & c h^l \int_{\mathcal{Y}} \int_{\mathbb{R}^q} K(\mathbf{t}) \mathbf{t}^l \mathrm{d}\mathbf{t} f_n(\mathbf{y}) \mathrm{d}\mathbf{y} \\
\leq & c h^l \kappa_q \int_{\mathcal{Y}} f_n(\mathbf{y}) \mathrm{d}\mathbf{y} \\
= & c h^l \kappa_q |A_n| = O(|A_n| h^l)
\end{aligned}$$

□

Lemma 8. *Let $f(\mathbf{u})$ be a bounded function defined on $\mathbb{R}^2$. Under condition 5.4, we have*

$$\sum_{\mathbf{u} \in X \cap A_n} f(\mathbf{u}) - \mathbb{E}\left[\sum_{\mathbf{u} \in X \cap A_n} f(\mathbf{u})\right] = O_p(|A_n|^{\frac{1}{m}})$$

Proof. The center order moment and the cumulant functions has the following relationship. We denote $\kappa_{m'}$ to be the integral of m'-th order cumulant function over A_n as follows

$$\kappa_{m'} = \int_{A_n} \cdots \int_{A_n} Q_{m'}(\mathbf{u}_1, \ldots, \mathbf{u}_{m'}) \mathrm{d}\mathbf{u}_1 \ldots \mathrm{d}\mathbf{u}_{m'}$$

Then when condition 5.4 is satisfied,

$$\begin{aligned}\kappa_{m'} &= \int_{A_n} \cdots \int_{A_n} Q_{m'}(\mathbf{u}_1, \ldots, \mathbf{u}_{m'}) \mathrm{d}\mathbf{u}_1 \ldots \mathrm{d}\mathbf{u}_{m'} \\ &\leq \int_{A_n} \left\{ \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} Q_{m'}(\mathbf{u}_1, \ldots, \mathbf{u}_{m'}) \mathrm{d}\mathbf{u}_1 \ldots \mathrm{d}\mathbf{u}_{m'-1} \right\} \mathrm{d}\mathbf{u}_{m'} \\ &= \int_{A_n} C \mathrm{d}\mathbf{u}_{m'} = O(|A_n|)\end{aligned}$$

Thus, $\kappa_{m'} = O(|A_n|)$ for $2 \leq m' \leq m$. The centered moments of spatial point processes relate with cumulant functions as follows

$$\begin{aligned}&\mathbb{E}[|X|] = \kappa_1 \\ &\mathbb{E}[|X - \mathbb{E}[X]|^2] = \kappa_2 + \kappa_1 \\ &\mathbb{E}[|X - \mathbb{E}[X]|^3] = \kappa_3 + 3\kappa_2 + \kappa_1 \\ &\mathbb{E}[|X - \mathbb{E}[X]|^4] = \kappa_4 + 6\kappa_2 + 7\kappa_2 + \kappa_1 \\ &\vdots\end{aligned}$$

Generally, $\mathbb{E}[|X - \mathbb{E}[X]|^{m'}] = O(|A_n|)$ for $2 \leq m' \leq m$. (See chapter 5.2 in Daley and Vere-Jones [2007]). Then we have

$$\mathbb{E}\left[\left|\sum_{\mathbf{u} \in X \cap A_n} f(\mathbf{u}) - \mathbb{E}\left[\sum_{\mathbf{u} \in X \cap A_n} f(\mathbf{u})\right]\right|^m\right] \leq \sup_{\mathbf{u} \in \mathbb{R}^2} |f(\mathbf{u})| \mathbb{E}[|X - \mathbb{E}[X]|^m] = O(|A_n|)$$

Thus, by Markov inequality, we have

$$\sum_{\mathbf{u} \in X \cap A_n} f(\mathbf{u}) - \mathbb{E}\left[\sum_{\mathbf{u} \in X \cap A_n} f(\mathbf{u})\right] = O_p(|A_n|^{\frac{1}{m}})$$

$\square$

Lemma 9 (Kernel Approximation Deviation)**.** *Let $\phi(\mathbf{y}, \mathbf{z}; \pi)$ be a bounded function defined on the covariance domain $\mathcal{Y} \times \mathcal{Z}$ parameterized by $\pi \in \Pi \subset \mathbb{R}^s$. Assume Π be compact, $\phi(\mathbf{y}, \mathbf{z}; \pi)$ be l-th order continuously differentiable with respect to $\mathbf{z}$, continuously differentiable with respect to π. Let X be a spatial point process with intensity function $\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta)$. Let $K_h(\cdot)$ be an l-th order Kernel function(See definition 11.1) where h is the bandwidth. Denote*

$$\hat{F}_n(\mathbf{z}, \pi; \phi) = \sum_{\mathbf{u} \in X \cap A_n} \phi(\mathbf{y}(\mathbf{u}), \mathbf{z}(\mathbf{u}); \pi) K_h\left(\mathbf{z}(\mathbf{u}) - \mathbf{z}\right)$$

Then under Assumption 5, we have

$$\sup_{\mathbf{z} \in \mathcal{Z}, \pi \in \Pi} \left| \hat{F}_n(\mathbf{z}, \pi; \phi) - \mathbb{E}\left[\hat{F}_n(\mathbf{z}, \pi; \phi)\right] \right| = O_p(|A_n|^{\frac{1+q+s}{q+m+s}} h^{-q-\frac{q+s}{q+m+s}}) \tag{77}$$

Proof. To establish (77), we proceed in three steps. Step 1 shows the main argument. Step 2 shows the auxiliary calculations.

Step 1: Let $\mathcal{Z}_\delta$ be a uniform grid on $\mathcal{Z}$, Π_δ be a uniform grid on Π where δ is the spacing. Denote

$$R_{n,1} := \max_{\mathbf{z} \in \mathcal{Z}_\delta, \pi \in \Pi_\delta} \left| \hat{F}_n(\mathbf{z}, \pi; \phi) - \mathbb{E}\left[\hat{F}_n(\mathbf{z}, \pi; \phi)\right] \right|$$
$$R_{n,2} := \sup_{|\mathbf{z} - \mathbf{z}'| \leq \delta, |\pi - \pi'| \leq \delta} \left| \hat{F}_n(\mathbf{z}, \pi; \phi) - \hat{F}_n(\mathbf{z}', \pi'; \phi) \right|$$

Then by the triangular inequality,

$$\sup_{\mathbf{z} \in \mathcal{Z}, \pi \in \Pi} \left| \hat{F}_n(\mathbf{z}, \pi; \phi) - \mathbb{E}\left[\hat{F}_n(\mathbf{z}, \pi; \phi)\right] \right| \leq R_{n,1} + R_{n,2}$$

In Step 2,3 respectively, we will show that

$$R_{n,1} = O_p(\delta^{-\frac{q+s}{m}}|A_n|^{\frac{1}{m}}h^{-q}) \tag{78}$$

$$R_{n,2} = O_p(|A_n|h^{-(q+1)}\delta) \tag{79}$$

If we specify $\delta = \Theta(|A_n|^{\frac{1-m}{q+m+s}}h^{\frac{m}{q+m+s}})$, we will have $R_{n,1} = R_{n,2} = O_p(|A_n|^{\frac{1+q+s}{q+m+s}}h^{-q-\frac{q+s}{q+m+s}})$ and (77) follows.

Step 2: In this step, we establish (78). First, since $\phi(\cdot)$ and the Kernel function $K(\cdot)$ are uniformly bounded,

$$\phi(\mathbf{y}(\mathbf{u}), \mathbf{z}(\mathbf{u}); \pi)K_h\left(\mathbf{z}(\mathbf{u}) - \mathbf{z}\right) = \phi(\mathbf{y}(\mathbf{u}), \mathbf{z}(\mathbf{u}); \pi)h^{-q}K\left(\frac{\mathbf{z}(\mathbf{u}) - \mathbf{z}}{h}\right) = O(h^{-q})$$

It follows from Lemma 8 that for any $\epsilon > 0$,

$$\begin{aligned} &\mathbb{P}\left(\left|\hat{F}_n(\mathbf{z}, \pi; \phi) - \mathbb{E}\left[\hat{F}_n(\mathbf{z}, \pi; \phi)\right]\right| \geq \epsilon\right) \\ =&\mathbb{P}\left(\left|\hat{F}_n(\mathbf{z}, \pi; \phi) - \mathbb{E}\left[\hat{F}_n(\mathbf{z}, \pi; \phi)\right]\right|^m \geq \epsilon^m\right) = O(|A_n|h^{-qm}\epsilon^{-m}) \end{aligned}$$

Then by union bound,

$$\begin{aligned} \mathbb{P}\left(R_{n,1} \geq \epsilon\right) \leq &\sum_{\mathbf{z}\in\mathcal{Z}_\delta, \pi\in\Pi_\delta} \mathbb{P}\left(\left|\hat{F}_n(\mathbf{z}, \pi; \phi) - \mathbb{E}\left[\hat{F}_n(\mathbf{z}, \pi; \phi)\right]\right| \geq \epsilon\right) \\ =&O\left(\delta^{-(q+s)}|A_n|h^{-qm}\epsilon^{-m}\right) \end{aligned}$$

Thus, we have (78)

Step 3: In this step, we establish (79). First, we observe that by triangular inequality,

$$R_{n,2} \leq \sup_{|\mathbf{z}-\mathbf{z}'|\leq\delta} \left|\hat{F}_n(\mathbf{z},\pi;\phi) - \hat{F}_n(\mathbf{z}',\pi;\phi)\right| + \sup_{|\pi-\pi'|\leq\delta} \left|\hat{F}_n(\mathbf{z}',\pi;\phi) - \hat{F}_n(\mathbf{z}',\pi';\phi)\right| \tag{80}$$

By the smoothness of the Kernel function K and ϕ, the first term in the right-hand side of (80) satisfies

$$\begin{aligned}
&\sup_{|\mathbf{z}-\mathbf{z}'|\leq\delta} \left|\hat{F}_n(\mathbf{z},\pi;\phi) - \hat{F}_n(\mathbf{z}',\pi;\phi)\right| \\
=&\sup_{|\mathbf{z}-\mathbf{z}'|\leq\delta} \left|\sum_{\mathbf{u}\in X\cap A_n} \phi(\mathbf{y}(\mathbf{u}),\mathbf{z}(\mathbf{u});\pi)\left(K_h\left(\mathbf{z}(\mathbf{u})-\mathbf{z}\right) - K_h\left(\mathbf{z}(\mathbf{u})-\mathbf{z}'\right)\right)\right| \\
\leq&\sup_{\mathbf{y},\mathbf{z},\pi} |\phi(\mathbf{y},\mathbf{z};\pi)| h^{-q} \sup_{|\mathbf{z}-\mathbf{z}'|\leq\delta} \left|\sum_{\mathbf{u}\in X\cap A_n} K\left(\frac{\mathbf{z}(\mathbf{u})-\mathbf{z}}{h}\right) - K\left(\frac{\mathbf{z}(\mathbf{u})-\mathbf{z}'}{h}\right)\right| \\
\leq&\sup_{\mathbf{y},\mathbf{z},\pi} |\phi(\mathbf{y},\mathbf{z};\pi)| \cdot \sup_{\mathbf{z}} \left|\frac{\partial}{\partial \mathbf{z}} K(\mathbf{z})\right| \cdot |X\cap A_n| \cdot h^{-(q+1)}\delta \\
=&O_p(|A_n| h^{-(q+1)}\delta)
\end{aligned} \tag{81}$$

The second term in the right-hand side of (80) satisfies

$$\begin{aligned}
&\sup_{|\mathbf{z}-\mathbf{z}'|\leq\delta} \left|\hat{F}_n(\mathbf{z}',\pi;\phi) - \hat{F}_n(\mathbf{z}',\pi';\phi)\right| \\
=&\sup_{|\mathbf{z}-\mathbf{z}'|\leq\delta} \left|\sum_{\mathbf{u}\in X\cap A_n} (\phi(\mathbf{y}(\mathbf{u}),\mathbf{z}(\mathbf{u});\pi) - \phi(\mathbf{y}(\mathbf{u}),\mathbf{z}(\mathbf{u});\pi')) K_h\left(\mathbf{z}(\mathbf{u})-\mathbf{z}'\right)\right| \\
\leq&\sup_{\mathbf{z}} |K(\mathbf{z})| h^{-q} \sup_{|\mathbf{z}-\mathbf{z}'|\leq\delta} \left|\sum_{\mathbf{u}\in X\cap A_n} (\phi(\mathbf{y}(\mathbf{u}),\mathbf{z}(\mathbf{u});\pi) - \phi(\mathbf{y}(\mathbf{u}),\mathbf{z}(\mathbf{u});\pi'))\right| \\
\leq&\sup_{\mathbf{z}} |K(\mathbf{z})| \cdot \sup_{\mathbf{y},\mathbf{z},\pi} \left|\frac{\partial}{\partial \pi} \phi(\mathbf{y},\mathbf{z};\pi)\right| \cdot |X\cap A_n| \cdot h^{-q}\delta \\
=&O_p(|A_n| h^{-q}\delta)
\end{aligned} \tag{82}$$

Substitute (81) and (82) into (80) and we will have (79).

$\square$

Lemma 10 (Approximation of Conditioning on Covariates)**.** *Recall that $\boldsymbol{\theta} \in \mathbb{R}^k$, $\mathbf{z}(\mathbf{u}) \in \mathbb{R}^q$, m is the defined in condition 5.4. Denote $\alpha = \frac{m-1}{k+q+m+1}$, $\beta = \frac{k+q+1}{k+q+m+1}$. Under all the conditions in Assumption 5, for $i = 0, 1$ and $j = 0, 1, 2$, we have*

$$\sup_{\boldsymbol{\theta}\in\Theta,\gamma\in H,\mathbf{z}\in\mathcal{Z}} \left| \frac{\partial^{i+j}}{\partial\gamma^i\partial\boldsymbol{\theta}^j}\hat{\mathbb{E}}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}] - \frac{\partial^{i+j}}{\partial\gamma^i\partial\boldsymbol{\theta}^j}\mathbb{E}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}] \right| = O_p(|A_n|^{1-\alpha}h^{-q-\beta} + |A_n|h^l)$$

If we further let $h = \Theta(|A_n|^{-\frac{\alpha}{l+q+\beta}})$, we will have

$$\sup_{\boldsymbol{\theta}\in\Theta,\gamma\in H,\mathbf{z}\in\mathcal{Z}} \left| \frac{\partial^{i+j}}{\partial\gamma^i\partial\boldsymbol{\theta}^j}\hat{\mathbb{E}}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}] - \frac{\partial^{i+j}}{\partial\gamma^i\partial\boldsymbol{\theta}^j}\mathbb{E}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}] \right| = o_p(|A_n|^{1-\frac{\alpha l}{l+q+1}})$$

Proof. We proceed the proof into three steps. Step 1 shows the main argument. Step 2-3 shows the auxiliary calculation.

Step 1: Recall that

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

$$\hat{\mathbb{E}}[\ell(\boldsymbol{\theta},\gamma)|\mathbf{z}] = \sum_{\mathbf{u}\in X} K_h\left(\mathbf{z}(\mathbf{u}) - \mathbf{z}\right)\log\lambda(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u})),\gamma) - \int_{A_n} K_h\left(\mathbf{z}(\mathbf{u}) - \mathbf{z}\right)\lambda(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u})),\gamma)\mathrm{d}\mathbf{u}.$$

Denote

$$\phi_5(\mathbf{y};\boldsymbol{\theta},\gamma) := \log\lambda(\tau_{\boldsymbol{\theta}}(\mathbf{y}),\gamma)$$

$$\phi_6(\mathbf{y};\boldsymbol{\theta},\gamma) := \lambda(\tau_{\boldsymbol{\theta}}(\mathbf{y}),\gamma)$$

$$\phi_5^{i,j}(\mathbf{y};\boldsymbol{\theta},\gamma) := \frac{\partial^{i+j}}{\partial\gamma^i\partial\boldsymbol{\theta}^j}\phi_5(\mathbf{y};\boldsymbol{\theta},\gamma)$$

$$\phi_6^{i,j}(\mathbf{y};\boldsymbol{\theta},\gamma) := \frac{\partial^{i+j}}{\partial\gamma^i\partial\boldsymbol{\theta}^j}\phi_6(\mathbf{y};\boldsymbol{\theta},\gamma)$$

For any $\phi(\mathbf{y};\boldsymbol{\theta},\gamma)$, we denote

$$F_n(\mathbf{z},\boldsymbol{\theta},\gamma;\phi) := \int_{\mathcal{Y}} \phi(\mathbf{y},\mathbf{z};\boldsymbol{\theta},\eta) f_n(\mathbf{y},\mathbf{z})\mathrm{d}\mathbf{y}$$
$$\hat{F}_n(\mathbf{z},\boldsymbol{\theta},\gamma;\phi) := \sum_{\mathbf{u}\in X\cap A_n} K_h\left(\mathbf{z}(\mathbf{u})-\mathbf{z}\right)\phi(\mathbf{y}(\mathbf{u});\boldsymbol{\theta},\gamma)$$

Then by triangular inequality,

$$\begin{aligned}
&\sup_{\boldsymbol{\theta},\gamma,\mathbf{z}}\left|\frac{\partial^{i+j}}{\partial\gamma^i\partial\boldsymbol{\theta}^j}\hat{\mathbb{E}}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}] - \frac{\partial^{i+j}}{\partial\gamma^i\partial\boldsymbol{\theta}^j}\mathbb{E}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}]\right| \\
\leq &\sup_{\boldsymbol{\theta},\gamma,\mathbf{z}}\left|\mathbb{E}\left[\hat{F}_n(\mathbf{z},\boldsymbol{\theta},\gamma;\phi_5^{i,j})\right] - F_n(\mathbf{z},\boldsymbol{\theta},\gamma;\phi_5^{i,j})\right| + \sup_{\boldsymbol{\theta},\gamma,\mathbf{z}}\left|\hat{F}_n(\mathbf{z},\boldsymbol{\theta},\gamma;\phi_5^{i,j}) - \mathbb{E}\left[\hat{F}_n(\mathbf{z},\boldsymbol{\theta},\gamma;\phi_5^{i,j})\right]\right| \\
&+\sup_{\boldsymbol{\theta},\gamma,\mathbf{z}}\left|\mathbb{E}\left[\hat{F}_n(\mathbf{z},\boldsymbol{\theta},\gamma;\phi_6^{i,j})\right] - F_n(\mathbf{z},\boldsymbol{\theta},\gamma;\phi_6^{i,j})\right| \qquad (83)
\end{aligned}$$

By the smoothness condition 1.1, $\phi_5^{i,j}$ and $\phi_6^{i,j}$ are uniformly bounded for every i,j. Thus, it followed from Lemma 7 that under Assumption 5,

$$\sup_{\boldsymbol{\theta},\gamma,\mathbf{z}}\left|\mathbb{E}\left[\hat{F}_n(\mathbf{z},\boldsymbol{\theta},\gamma;\phi_6^{i,j})\right] - F_n(\mathbf{z},\boldsymbol{\theta},\gamma;\phi_6^{i,j})\right| = O(|A_n|h^l) \qquad (84)$$
$$\sup_{\boldsymbol{\theta},\gamma,\mathbf{z}}\left|\mathbb{E}\left[\hat{F}_n(\mathbf{z},\boldsymbol{\theta},\gamma;\phi_5^{i,j})\right] - F_n(\mathbf{z},\boldsymbol{\theta},\gamma;\phi_5^{i,j})\right| = O(|A_n|h^l) \qquad (85)$$

Additionally, since $\phi_5^{i,j}$ are continuously differentiable with respect to $\boldsymbol{\theta}$ and γ, it followed from Lemma 9 that

$$\sup_{\boldsymbol{\theta},\gamma,\mathbf{z}}\left|\hat{F}_n(\mathbf{z},\boldsymbol{\theta},\gamma;\phi_5^{i,j}) - \mathbb{E}\left[\hat{F}_n(\mathbf{z},\boldsymbol{\theta},\gamma;\phi_5^{i,j})\right]\right| = O_p(|A_n|^{\frac{p+q+2}{p+q+m+1}}h^{-q-\frac{p+q+1}{p+q+m+1}}) \qquad (86)$$

Finally, if we substitute the rate of R_1 in (84), (85) and (86) into (83), we have

$$\sup_{\boldsymbol{\theta},\gamma,\mathbf{z}}\left|\frac{\partial^{i+j}}{\partial\gamma^i\partial\boldsymbol{\theta}^j}\hat{\mathbb{E}}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}] - \frac{\partial^{i+j}}{\partial\gamma^i\partial\boldsymbol{\theta}^j}\mathbb{E}[\ell_n(\boldsymbol{\theta},\gamma)|\mathbf{z}]\right| = O_p(|A_n|^{1-\alpha}h^{-q-\beta} + |A_n|h^l)$$

where $\alpha = \frac{m-1}{k+q+m+1}$, $\beta = \frac{k+q+1}{k+q+m+1}$. If we further let $h = \Theta(|A_n|^{-\frac{\alpha}{l+q+\beta}})$, the rate will be $O_p(|A_n|^{1-\frac{\alpha l}{l+q+\beta}}) = o_p(|A_n|^{1-\frac{\alpha l}{l+q+1}})$. □

11 Preliminary

Definition 11.1 (Higher Order Kernel)**.** l is a positive even integer. A l-th order kernel function $k(\cdot) : \mathbb{R} \to \mathbb{R}$ satisfies

$$\int_R k(z)\mathrm{d}z = 1$$
$$\int_R z^i k(z)\mathrm{d}z = 0, \quad (i = 2, \ldots, l-1)$$
$$\int_R z^l k(z)\mathrm{d}z = \kappa_l \neq 0$$

Let $K(\mathbf{z}) = \prod_{j=1}^q k(z_j)$ for $\mathbf{z} = (z_1, \ldots, z_q) \in \mathbb{R}^q$. Then $K(\cdot)$ satisfies

$$\int_{R^q} K(\mathbf{z})\mathrm{d}\mathbf{z} = 1$$
$$\int_{R^q} \mathbf{z}^i K(\mathbf{z})\mathrm{d}\mathbf{z} = 0, \quad (i = 2, \ldots, l-1)$$
$$\int_{R^q} \mathbf{z}^l K(\mathbf{z})\mathrm{d}\mathbf{z} = \kappa_l^q \neq 0$$

Theorem 11.1 (Implicit Function Theorem)**.** *Let $f : \mathbb{R}^{n+m} \to \mathbb{R}^m$ be a k-times continuously differentiable function, and let $\mathbb{R}^{n+m}$ have coordinates $(\mathbf{x}, \mathbf{y})$. Fix a point $(\mathbf{a}, \mathbf{b})$ with $f(\mathbf{a}, \mathbf{b}) = \mathbf{0}$, where $\mathbf{0} \in \mathbb{R}^m$ is the zero vector. If the Jacobian matrix*

$$J_{f,\mathbf{y}}(\mathbf{a}, \mathbf{b}) = \left[\frac{\partial f_i}{\partial y_j}(\mathbf{a}, \mathbf{b})\right]$$

is invertible, then there exists an open set $U \in \mathbb{R}^n$ containing $\mathbf{a}$ such that there exists a *unique function* $g : U \to \mathbb{R}^m$ *such that* $g(\mathbf{a}) = b$*, and* $f(\mathbf{x}, g(\mathbf{x})) = \mathbf{0}$ *for all* $\mathbf{x} \in U$*. Moreover,* g *is also* k*-times continuously differentiable and the Jacobian matrix of partial derivatives of* g *in* U *is given by*

$$\left[\frac{\partial g_i}{\partial x_j}(\mathbf{x})\right]_{m\times n} = -\left[J_{f,\mathbf{y}}(\mathbf{x}, g(\mathbf{x}))\right]^{-1}_{m\times m}\left[J_{f,\mathbf{x}}(\mathbf{x}, g(\mathbf{x}))\right]^{-1}_{m\times n}$$

References

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

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

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

Jon Wellner et al. *Weak convergence and empirical processes: with applications to statistics.* Springer Science & Business Media, 2013.

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.

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.

Adrian Baddeley, Jesper Møller, and Anthony G Pakes. Properties of residuals for spatial point processes. *Annals of the Institute of Statistical Mathematics*, 60(3):627–649, 2008.

Christophe Ange Napoléon Biscio and Rasmus Waagepetersen. A general central limit theorem and a subsampling variance estimator for α-mixing point processes. *Scandinavian Journal of Statistics*, 46(4):1168–1190, 2019.

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.