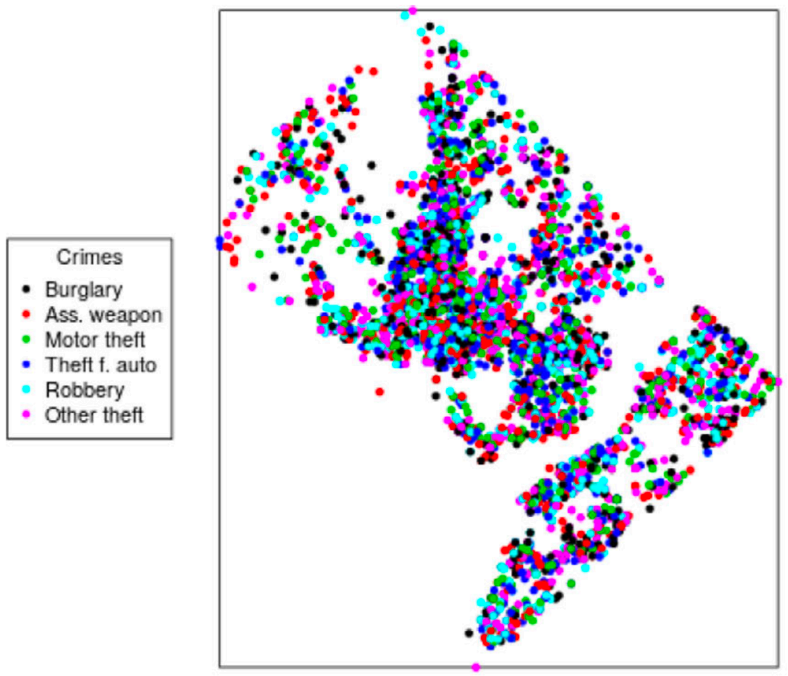
Efficient, Cross-Fitting Estimation of Semiparametric Spatial Point Processes

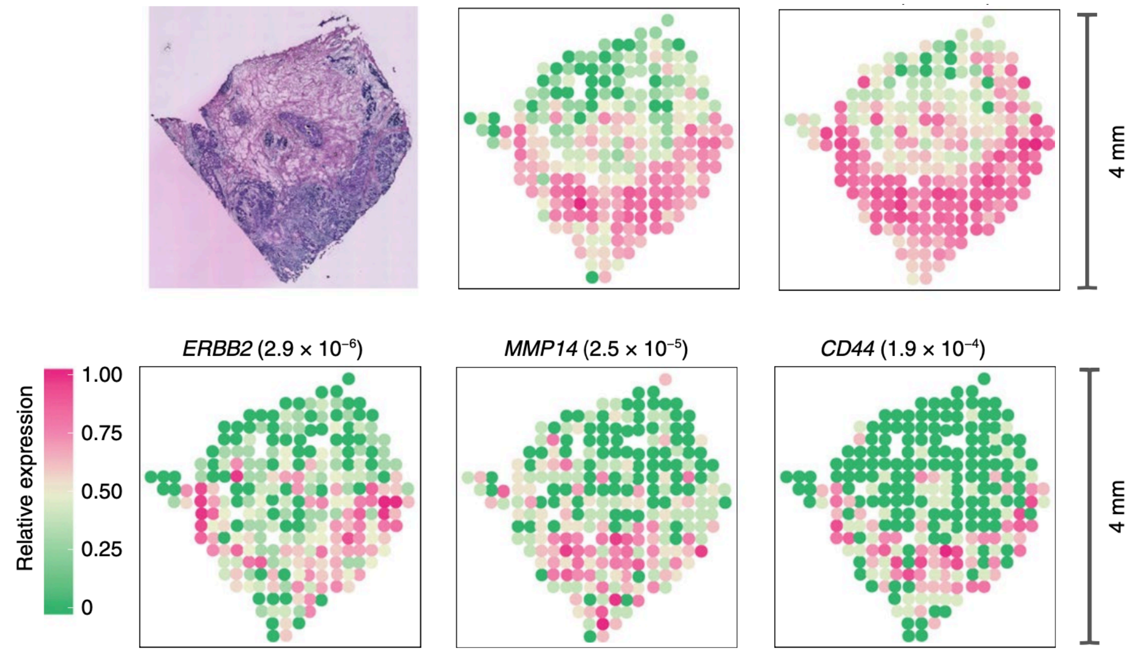
Xindi Lin
Hyunseung Kang

Department of Statistics, University of Wisconsin- Madison

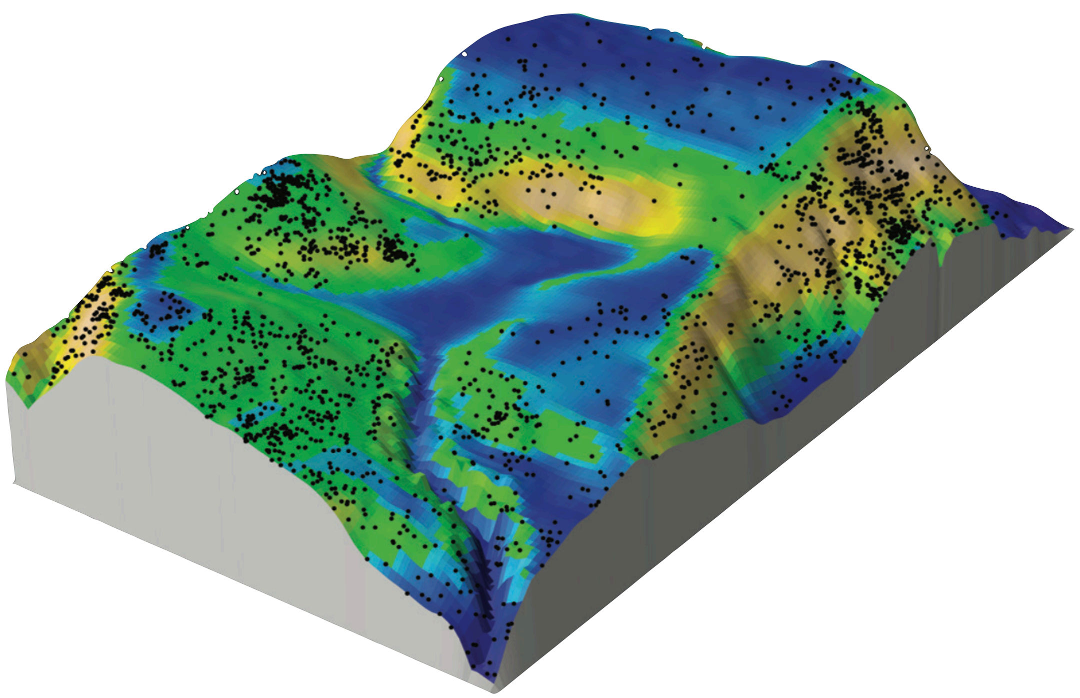
Examples of Spatial Point Processes



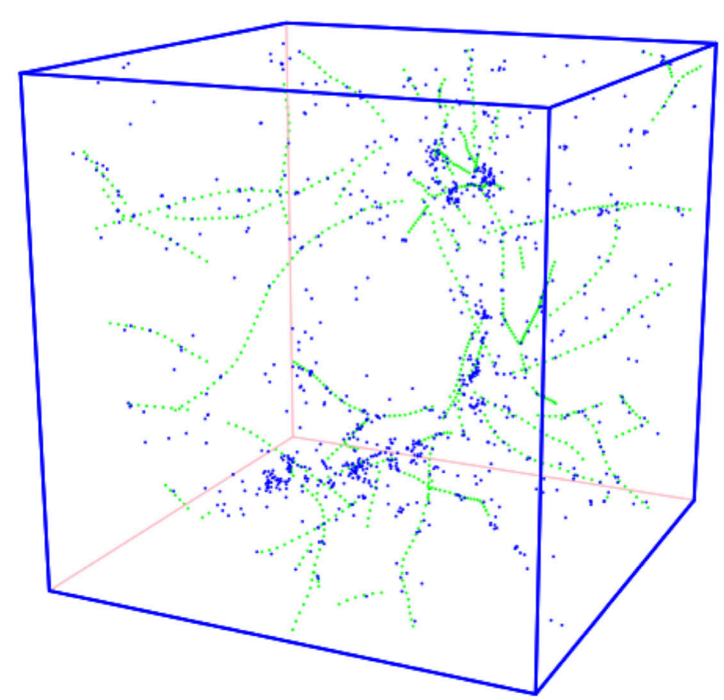
Street crimes locations in Washington D.C from Hessellund et al. (2022). Colors represents different types of crimes.



Spatial expression pattern for five known tumor genes in breast cancer from Sun et al. (2020)



Locations of Beilschmiedia trees (in black) in Barro Colorado Island from Baddeley et al. (2022).



Galaxy catalogue from Tempel et al. (2012) (in blue) with the filamentary structure from Tempel et al. (2014a) (in green)

Semiparametric Spatial Point Process

- $X \subset \mathbb{R}^2$: spatial point process, $A \subset \mathbb{R}^2$: observational window
- $\boldsymbol{y}(\mathbf{u}) \in \mathbb{R}^p$: target covariate (e.g. elevation) at location $\mathbf{u} \in \mathbb{R}^2$
- $\mathbf{z}(\mathbf{u}) \in \mathbb{R}^q$: nuisance covariate (e.g. gradient) at location $\mathbf{u} \in \mathbb{R}^2$
- **Our model**: intensity function $\lambda(\mathbf{u})$, the chance of observing at $\mathbf{u}$

$$\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) = \rho(\tau_{\boldsymbol{\theta}}(\boldsymbol{y}(\mathbf{u})), \eta(\mathbf{z}(\mathbf{u}))), \boldsymbol{\theta} \in \mathbb{R}^k, \eta \in H$$

  - Example: $\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) = \exp(\boldsymbol{\theta}^\top \boldsymbol{y}(\mathbf{u}) + \eta(\mathbf{z}(\mathbf{u})))$
  - The goal is to estimate $\boldsymbol{\theta}$ (e.g. effect of elevation on X)

Remark:

- Fithian, et al. (2015), Waagepetersen at al. (2009) assumed parametric η
- Partial linear model [Robinson (1988)] + cross-fitting [Chernozhukov et al (2018)] assumed i.i.d. data

Traditional Cross-Fitting

- Require: $W_i = (X_i, Y_i, Z_i), i = 1,..,n$ be i.i.d. sample from (X, Y, Z)

$$X_i = \theta^\top Y_i + \eta(Z_i) + \epsilon_i$$

- **Step 1**: Sample splitting: take a V-fold random partition of $\{W_i\}_{i=1}^n$
- **Step 2**: For $v \in [V]$,
  - Step 2a: Fix $\boldsymbol{\theta}$, $\hat{\eta}_{\boldsymbol{\theta}}^{(v)} = \arg\max_{\eta} \ell(\boldsymbol{\theta}, \eta; W_v^c)$
  - Step 2b: $\hat{\boldsymbol{\theta}}^{(v)} = \arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta}}^{(v)}; W_v)$
- **Step 3**: Aggregate: $\hat{\boldsymbol{\theta}} = V^{-1} \sum_{v=1}^{V} \hat{\boldsymbol{\theta}}^{(v)}$

Remark:

- No Donsker conditions and full efficiency

Spatial Cross-Fitting

- Require: spatial point process X, observed in $A \subset \mathbb{R}^2$, $\boldsymbol{y}(\mathbf{u})$, $\mathbf{z}(\mathbf{u})$

$$\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) = \rho(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u})), \eta(\mathbf{z}(\mathbf{u})))$$

- **Step 1**: Sample splitting: take a V-fold random partition of $\{W_i\}_{i=1}^n$

1. How to split spatial point process?

- **Step 2**: For $v \in [V]$,
  - Step 2a: Fix $\boldsymbol{\theta}$, $\hat{\eta}_{\boldsymbol{\theta}}^{(v)} = \arg\max_{\eta} \ell(\boldsymbol{\theta}, \eta; W_v^c)$

2. What is ℓ? How to estimate η?

  - Step 2b: $\hat{\boldsymbol{\theta}}^{(v)} = \arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta}}^{(v)}; W_v)$

3. How to compute $\boldsymbol{\theta}$?

- **Step 3**: Aggregate: $\hat{\boldsymbol{\theta}} = V^{-1} \sum_{v=1}^{V} \hat{\boldsymbol{\theta}}^{(v)}$

4. Asymptotics, efficiency, SE estimator

Spatial Cross-Fitting

- Require: spatial point process X, observed in $A \subset \mathbb{R}^2$, $\boldsymbol{y}(\mathbf{u})$, $\mathbf{z}(\mathbf{u})$

$$\lambda(\mathbf{u};\boldsymbol{\theta},\eta) = \rho(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u})),\eta(\mathbf{z}(\mathbf{u})))$$

- **Step 1**: random thinning: take a V-fold random thinning of X
    - For every $\mathbf{u} \in X \cap A$, randomly assign to one of V subprocesses
    - We show every subprocess inherit the first order property of X
    - We show subprocessess are i.i.d. if X is Poisson

Spatial Cross-Fitting

- Require: spatial point process X, observed in $A \subset \mathbb{R}^2$, $\boldsymbol{y}(\mathbf{u})$, $\mathbf{z}(\mathbf{u})$

$$\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) = \rho(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u})), \eta(\mathbf{z}(\mathbf{u})))$$

- **Step 1**: random thinning: take a V-fold random thinning of X
- **Step 2**: For $v \in [V]$,
  - Step 2a: Fix $\boldsymbol{\theta}$ and $\mathbf{z}$, $\hat{\eta}_{\boldsymbol{\theta}}^{(v)}(\mathbf{z}) = \arg\max_{\eta} \hat{\mathbb{E}}[\ell(\boldsymbol{\theta}, \eta; X_v^c) \,|\, \mathbf{z}]$
    - Pseudo-likelihood [Waagepetersen at al. (2009)]
    - We develop a new spatial kernel regression estimator and its convergence rate

Spatial Cross-Fitting

- Require: spatial point process X, observed in $A \subset \mathbb{R}^2$, $\boldsymbol{y}(\mathbf{u})$, $\mathbf{z}(\mathbf{u})$

$$\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) = \rho(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u})), \eta(\mathbf{z}(\mathbf{u})))$$

- **Step 1**: random thinning: take a V-fold random thinning of X
- **Step 2**: For $v \in [V]$,
  - Step 2a: Fix $\boldsymbol{\theta}$ and $\mathbf{z}$, $\hat{\eta}_{\boldsymbol{\theta}}^{(v)}(\mathbf{z}) = \arg\max_{\eta} \hat{\mathbb{E}}[\ell(\boldsymbol{\theta}, \eta; X_v^c) \,|\, \mathbf{z}]$
  - Step 2b: $\hat{\boldsymbol{\theta}}^{(v)} = \arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}, \hat{\eta}_{\boldsymbol{\theta}}^{(v)}; X_v)$
  - $\ell(\boldsymbol{\theta}, \eta; X) = \sum_{\mathbf{u} \in A \cap X} \log \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) - \int_A \lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) \mathrm{d}\mathbf{u}$
  - Numerical approximation of integral [Berman at el. (1992)]
  - We show the numerical equivalence of solving 2b to solving a generalized partially linear model in i.i.d. settings [mgcv]

Spatial Cross-Fitting

- Require: spatial point process X, observed in $A \subset \mathbb{R}^2$, $\boldsymbol{y}(\mathbf{u})$, $\mathbf{z}(\mathbf{u})$

$$\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) = \rho(\tau_{\boldsymbol{\theta}}(\mathbf{y}(\mathbf{u})), \eta(\mathbf{z}(\mathbf{u})))$$

- **Step 1**: random thinning: take a V-fold random thinning of X
- **Step 2**: For $v \in [V]$,
  - Step 2a: Fix $\boldsymbol{\theta}$ and $\mathbf{z}$, $\hat{\eta}^{(v)}_{\boldsymbol{\theta}}(\mathbf{z}) = \arg\max_{\eta} \hat{\mathbb{E}}[\ell(\boldsymbol{\theta}, \eta; X_v^c) \,|\, \mathbf{z}]$
  - Step 2b: $\hat{\boldsymbol{\theta}}^{(v)} = \arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}, \hat{\eta}^{(v)}_{\boldsymbol{\theta}}; X_v)$
- **Step 3**: Aggregate: $\hat{\boldsymbol{\theta}} = V^{-1} \sum_{v=1}^{V} \hat{\boldsymbol{\theta}}^{(v)}$

Asymptotic Theory

- **Poisson:** No interactions between points
- **Asymptotic Scheme:** Consider a sequence of expanding $\{A_n\}_{n=1}^{\infty}$

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

Consistency

Assumption 1.

1.1 Smoothness (2nd order), boundedness, identification

1.2 Weak pairwise dependence of X : $\int_{\mathbb{R}^2} |\, g(0,\mathbf{u}) - 1 \,|\, \mathrm{d}\mathbf{u} < C$

Remark: $g(\,\cdot\,,\cdot\,)$ is the pairwise correlation function. $g = 1$ indicate no correlation.

Theorem 1. If the nuisance estimator is uniformly consistent (same as classic semiparametric setting), $\hat{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta}$

Asymptotic Normality

Assumption 2

2.1 Nonsingularity of asymptotic variance

2.2 α-mixing condition $\alpha_{2,\infty}^{X}(r) = O(r^{-(2+\epsilon)})$

Theorem 2. Under Assumptions 1 and 2, when the nuisance error rate is $o_p(|A_n|^{-\frac{1}{4}})$, (similar to $o_p(n^{-\frac{1}{4}})$ in classic semiparametric setting), $\hat{\boldsymbol{\theta}}_n$ is asymptotically Normal with an asymptotic variance $\boldsymbol{\Sigma}_n$ if either X is Poisson or $\lambda(\cdot)$ is log-linear

Theorem 3. $\hat{\boldsymbol{\theta}}_n$ is a semiparametric efficient estimator of $\boldsymbol{\theta}$ if X is Poisson

Theorem 4. Under Assumptions 1 and 2, we can consistently estimate $\boldsymbol{\Sigma}_n$

A Key Step in the Proof

Classic Empirical Process

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

Spatial Empirical Process

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

- Maximal inequality fails because of higher-order dependence in X
- $\mathbb{G}_n^s[V_n^{(1)}(\hat{\eta}_{\theta*,n} - \eta^*_{\theta*,n}) \circ \mathbf{z}]$, $\mathbb{G}_n^s[V_n^{(2)}(\hat{\nu}_n - \nu_n^*) \circ \mathbf{z}]$
- Poisson $\rightarrow \hat{\eta}_{\theta,n}, \hat{\nu}_n \perp \mathbb{G}_n^s$ and Log-linear $\lambda(\cdot) \rightarrow V_n^{(1)} = 0, \hat{\nu}_n \perp \mathbb{G}_n^s$

Rate of Convergence of Nuisance Estimator

Assumption 3

3.1 Smoothness (l-th order)

3.2 m-th order weak dependence of X

Theorem 5. Under Assumptions 1 and 3, $\hat{\eta}_{\boldsymbol{\theta}}$ converge at a rate

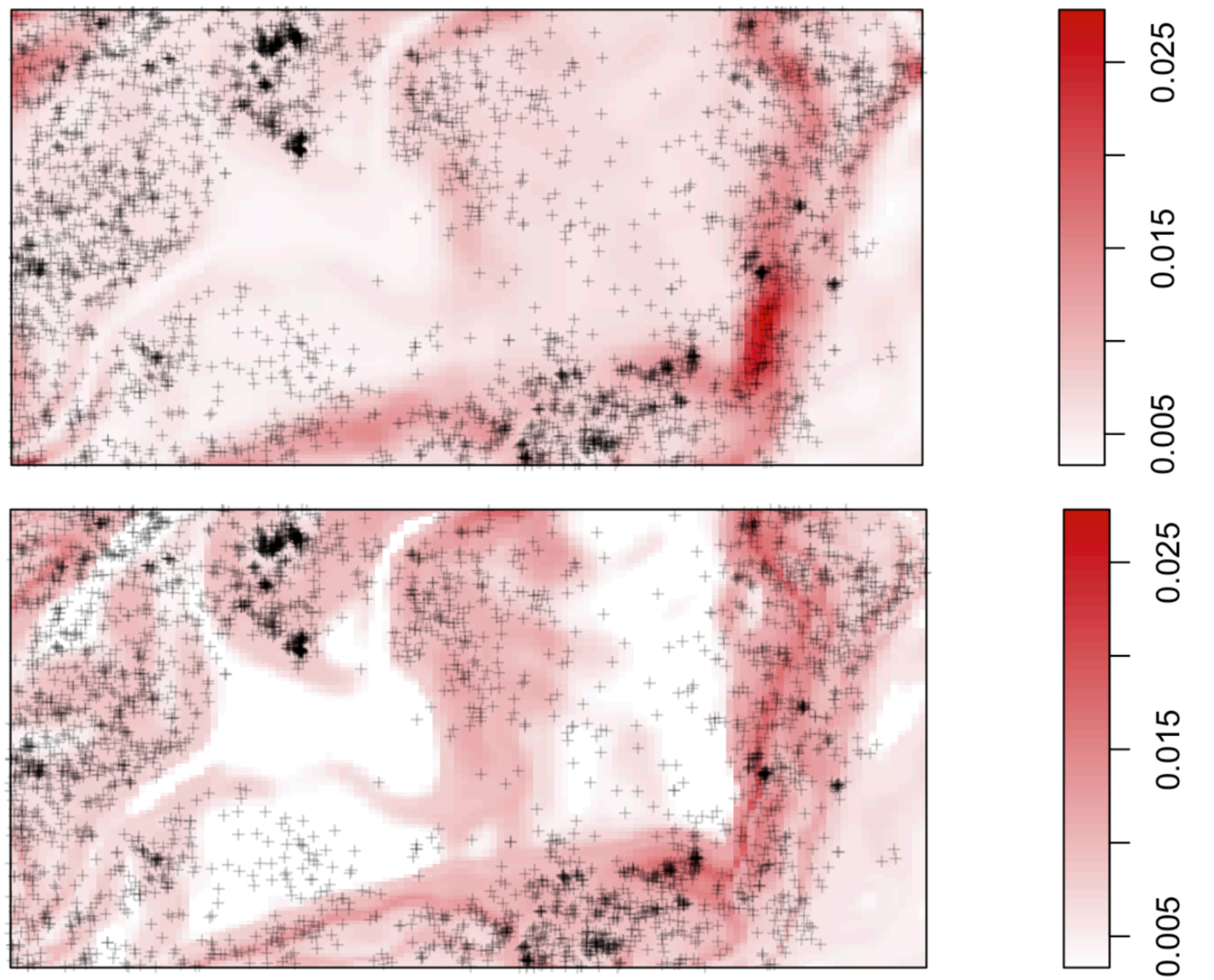
$$o_p\left(|A_n|^{-\frac{m-1}{m+k+q+1}\cdot\frac{l}{l+q+1}}\right)$$

where k is the dimension of $\boldsymbol{\theta}$ and q is the dimension of nuisance covariate

Remark: When X is Poisson, the rate can be close to $O_p(|A_n|^{-1})$ if l is large; faster than $O_p(n^{-\frac{1}{2}})$ in i.i.d; $o_p(|A_n|^{-\frac{1}{4}})$ can be satisfied

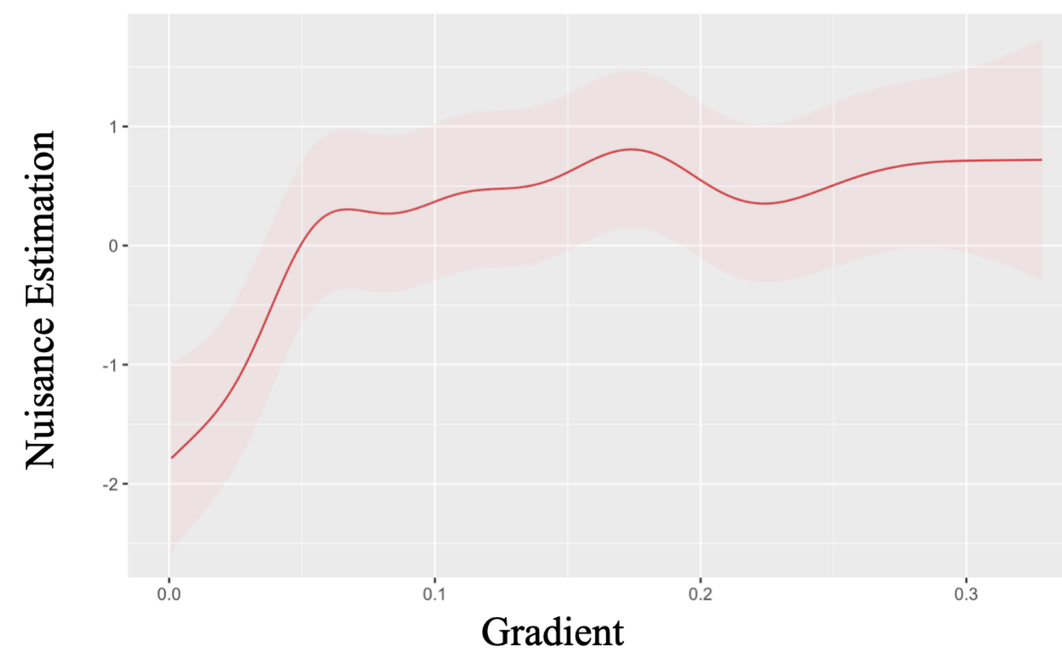
Real Data Analysis

- X be locations of Beilschmiedia trees, $\mathbf{y}(\mathbf{u})$ be elevation, $\mathbf{z}(\mathbf{u})$ be gradient
- Compare our semiparametric method and parametric method



Fitted Intensity. **Top**: parametric model. **Bottom**: semiparametric model

Method	$\hat{\theta}$	SE($\hat{\theta}$)	95%CI
Ours	2.983	2.268	(-1.462,7.428)
Parametric	2.144	2.346	(-2.454,6.742)



Summary

- We propose a semiparametric spatial point process to study relationships between spatial points and spatial covariates.
- We propose a spatial cross-fitting estimation with the following properties:

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

References

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

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

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.

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

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

Simulation Result

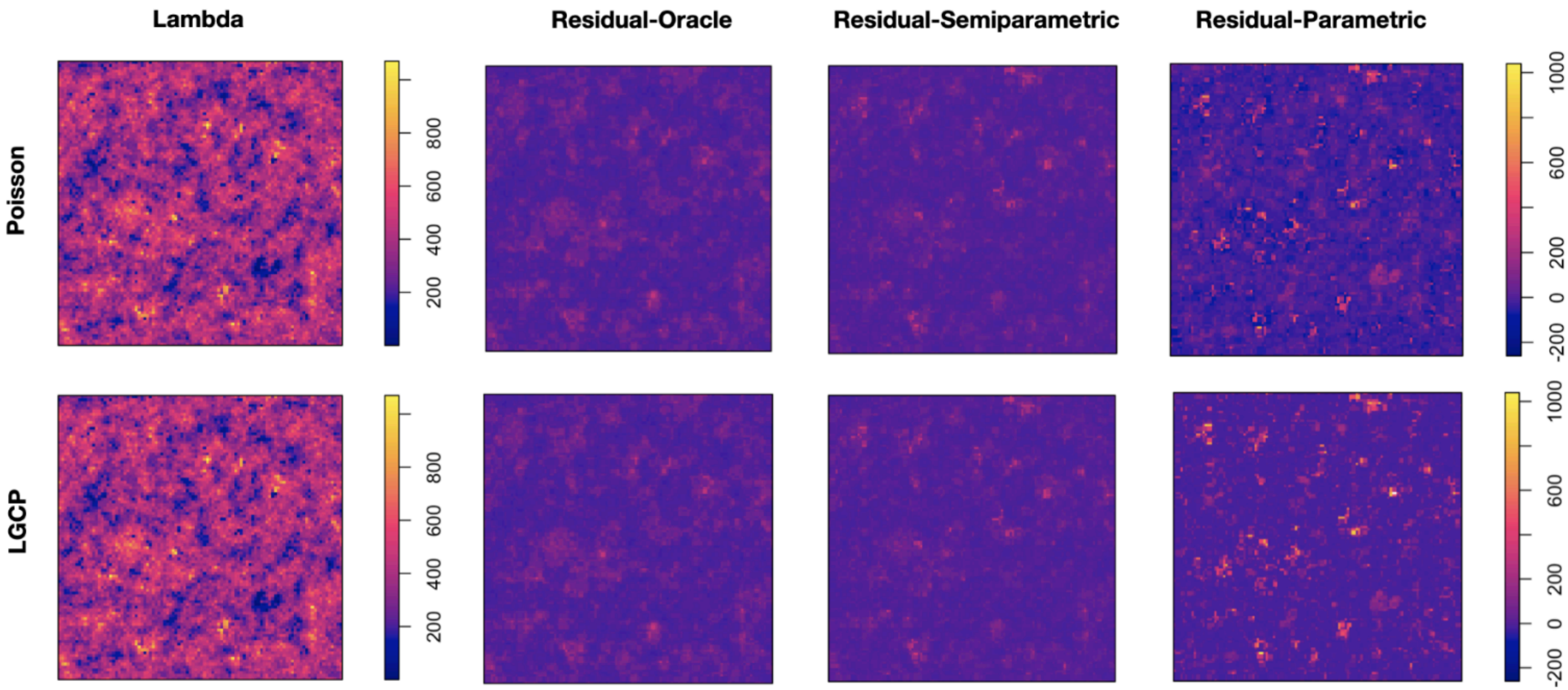
$W_1 : [0,1] \times [0,1]$, $W_2 : [0,2] \times [0,2]$

$\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) = \exp(0.3\boldsymbol{y}(\boldsymbol{u}) - 0.09(z(\mathbf{u})^2)$, $\boldsymbol{y}(\boldsymbol{u}), z(\boldsymbol{u})$ be dependent GP

Semi: Our model, Para: wrong model, Oracle: η is known

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

Simulation Result



Semiparametric Efficiency

- **Psuedo-log-likelihood**: $\ell(\boldsymbol{\theta}, \eta)$ [Waagepetersen at al. (2009)]
- $\boldsymbol{\theta}^*, \eta^*$: true parameters.
- **Step1**: construct parametric submodel
  - Consider $\boldsymbol{\theta} \mapsto \eta_{\boldsymbol{\theta}}$ s.t. $\eta_{\boldsymbol{\theta}^*} = \eta^*$, $\nu = \partial/\partial\boldsymbol{\theta}\eta_{\boldsymbol{\theta}} \mid_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$
  - $\boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \nu) = \mathbb{E}[-\partial^2/\partial\boldsymbol{\theta}^2\ell(\boldsymbol{\theta}, \eta_{\boldsymbol{\theta}}) \mid_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}]$
- **Step 2**: supremum of Fisher-Rao bound
  - $\nu^* = \arg\max_{\nu} \boldsymbol{S}(\boldsymbol{\theta}^*, \eta^*, \nu)$ [Thm 4.1, Lin, Kang (2024)]
- **Step 3**: bound in Step 2 is attainable
  - $\eta^*_{\boldsymbol{\theta}} = \arg\max_{\eta} \ell(\boldsymbol{\theta}, \eta)$, $\partial/\partial\boldsymbol{\theta}\eta^*_{\boldsymbol{\theta}} \mid_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = \nu^*$

Efficiency Bound

Definition 1 (Conditional on Covariates). Let $f(\mathbf{y}, \mathbf{z})$ joint" Radon-Nikodym derivative of the push-forward measure induced by $\mathbf{y}(\mathbf{u}), \mathbf{z}(\mathbf{u}))$. For any $\phi(\mathbf{u}) = \phi(\mathbf{y}(\mathbf{u}), \mathbf{z}(\mathbf{u}))$

$$\mathbb{E}\left[\sum_{\mathbf{u}\in X\cap A} \phi(\mathbf{u}) \middle| \mathbf{z}\right] = \int_{\mathscr{Y}} \phi(\mathbf{y}, \mathbf{z}) f(\mathbf{y}, \mathbf{z}) \mathrm{d}\mathbf{y}$$

Theorem 1 (Efficiency Bound). The supremum of Cramer-Rao lower bound is attained by

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

If the intensity function is log-linear, we further have

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

Consistency Condition

Assumption 1 (Conditions for Consistency).

1.1 (Smoothness) $\lambda(\mathbf{u};\boldsymbol{\theta},\eta)$ is twice continuously differentiable w.r.t. $\boldsymbol{\theta},\eta$

1.2 (Identification) The area $\{\mathbf{u} : |\lambda(\mathbf{u};\boldsymbol{\theta},\eta) - \lambda(\mathbf{u};\boldsymbol{\theta}^*,\eta^*)| \geq c|\boldsymbol{\theta} - \boldsymbol{\theta}^*|\}$ grow at same rate as $|A_n|$

1.3 (Boundedness) The set $\{\mathbf{u} : \lambda(\mathbf{u};\boldsymbol{\theta},\eta) < c\}$ is bounded

1.4 (weak pairwise dependence) $\int_{\mathbb{R}^2} g(0,\mathbf{u})\mathrm{d}\mathbf{u}| < C$

Theorem 2 (Consistency). Under Assumption 1, when the nuisance estimator is uniformly consistent, $\hat{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta}^*$

Remark: 1.PCF: pairwise dependence of spatial point process). 2. Cross-fitting in classic semiparametric is consistent when nuisance estimator is consistent.

Related Work

- **Related Work**:

  - Spatial Causal Inference: [Papadogeorgo et al., 2022], Parametric model for propsensity score, kernel estimation for treatment effect

  - Forestry, multivariate: [Hessellund et al,. 2022]. Assume common latent effect among all processes

  - Ecology, thinnning Process: [Fithian, et al, 2015], parametric for nuisance and target, thinned Poisson processess

Formulas

$$E\left[\sum_{\mathbf{u}_1,\ldots\mathbf{u}_m,\in X}\mathbf{1}(\mathbf{u_1}\in A_1,\ldots,\mathbf{u}_m\in A_m)\right]=\int_{\prod_{i=1}^m A_i}\lambda^{(m)}(\mathbf{u}_1,\ldots,\mathbf{u}_m)\mathrm{d}\mathbf{u}_1\ldots\mathrm{d}\mathbf{u}_m$$

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

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}^*,\eta^*,\nu_n^*,g)=Var(\frac{\partial}{\partial\boldsymbol{\theta}}\ell(\boldsymbol{\theta},\eta_{\boldsymbol{\theta}})\mid_{\boldsymbol{\theta}=\boldsymbol{\theta}^*})$$

$$\boldsymbol{S}(\boldsymbol{\theta}^*,\eta^*,\nu)=\mathbb{E}\left[-\frac{\partial^2}{\partial\boldsymbol{\theta}^2}\ell(\boldsymbol{\theta},\eta_{\boldsymbol{\theta}})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right]$$

$\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}$ is asymptotic variance

Spatial Point Process Formulas

$$\sup_{f\in\mathscr{F}_n^1}\left|\mathbb{G}_n^s[V_n^{(1)}f]\right| = o_p(1), \sup_{f\in\mathscr{F}_n^1}\left|\mathbb{G}_n^s[V_n^{(1)}(\hat{\eta}_{\theta^*,n}-\eta^*_{\theta^*,n})\circ\mathbf{z}]\right| = o_p(1)$$

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

$$\sup_{f\in\mathscr{F}_n^2}\left|\mathbb{G}_n^s[V_n^{(2)}\partial/\partial\theta(\hat{\eta}_{\theta^*,n}-\eta^*_{\theta^*,n})\circ\mathbf{z}]\right| = o_p(1)$$

$$\mathscr{F}_n^1 := \left\{(\hat{\eta}_{\theta^*,n}-\eta^*_{\theta^*,n})\circ\mathbf{z}\right\}$$

$$\mathscr{F}_n^2 := \left\{\partial/\partial\theta(\hat{\eta}_{\theta^*,n}-\eta^*_{\theta^*,n})\circ\mathbf{z}\right\}$$

$$V_n^{(1)}(f)(\,\cdot\,) := \partial^2/\partial\theta\partial\eta\log\lambda(\,\cdot\,;\theta,\eta^*_{\theta,n})\mid_{\theta=\theta^*}$$

$$V_n^{(2)}(f)(\,\cdot\,) := \partial^2/\partial\eta^2\log\lambda(\,\cdot\,;\theta,\eta^*_{\theta,n})\mid_{\theta=\theta^*}$$

Spatial Point Process

- **Notation**: $X \subset \mathbb{R}^2$: spatial point process, $A \subset \mathbb{R}^2$: observational window, $\mathbf{u} \in \mathbb{R}^2$: discrete points

- $\mathbf{u} \in \mathbb{R}^2$,$\mathbf{u} \mapsto \boldsymbol{y}(\mathbf{u}) \in \mathbb{R}^p$: target cov, $\mathbf{u} \mapsto \mathbf{z}(\mathbf{u}) \in \mathbb{R}^q$: nuisance cov

- **Recent work**: Fithian, et al. (2015), Hessellund et al. (2022), Papadogeorgo et al. (2022)

- **Our model**: intensity function $\lambda(\mathbf{u})$, chance of observe event at $\mathbf{u}$

$$\lambda(\mathbf{u}; \boldsymbol{\theta}, \eta) = \rho(\tau_{\boldsymbol{\theta}}(\boldsymbol{y}(\mathbf{u})), \eta(\boldsymbol{z}(\mathbf{u})))$$

  - Partial linear model [Robinson (1988)]

  - Cross-fitting [Chernozhukov et al. (2018)]

Cross-Fitting

Algorithm 1 V-Fold Cross-Fitting For Generalized Partial Linear Model

Require: $W_i = (X_i, Y_i, Z_i)$ be i.i.d. sample from $(X, Y, Z)$and number of folds V.

Step 1: Take a V-fold random partition$(I_v)_{v=1}^{V}$ of observation indices $[N] = \{1., \ldots, N\}$ such that the size of each fold $|I_v| = N/V$. Denote $I_v^c = \{1, \ldots, N\} \backslash I_v$.

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

Step 2a: Fix $\boldsymbol{\theta}$, estimate the nuisance parameter η using sample not in I_v.

$$\hat{\eta}_{\boldsymbol{\theta}}^{(v)} = \arg\max_{\eta \in H} \ell(\boldsymbol{\theta}, \eta; \{W_i\}_{i \in I_v^c})$$

Step 2b: Plug-in $\hat{\eta}_{\boldsymbol{\theta}}^{(v)}$, estimate the target parameter $\boldsymbol{\theta}$.

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

end for

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