Supplementary Materials for "Regularized Estimating Equations for Model Selection of Clustered Spatial Point Processes"

Andrew L. Thurman[a], Rao Fu[b], Yongtao Guan[c], Jun Zhu[d]

[a]*Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA 52242*
[b]*Department of Statistics, University of Wisconsin, Madison, WI 53706*
[c]*Department of Management Science, University of Miami, Miami, FL 33124*
[d]*Department of Statistics and Department of Entomology, University of Wisconsin, Madison, WI 53706*

---

1. Details of Cox Process

1.1. Second-order Properties

Second-order properties of a spatial point process indicate the spatial dependence of events in D. The factorial moment measure of order two is

$$\alpha_Y(B_1 \times B_2) = E\left[\sum_{s_1, s_2 \in Y}^{\neq} I\big\{(s_1, s_2) \in (B_1, B_2)\big\} \right],$$

where the sum $\sum\limits_{s_1,s_2\in Y}^{\neq}$ is taken over all distinct pairs of events s_1 and s_2. A second-order intensity density, $\rho_Y(s_1, s_2)$, is similarly defined such that

$$\alpha_Y(B_1 \times B_2) = \int_{B_1} \int_{B_2} \rho_Y(s_1, s_2) ds_1 ds_2.$$

With $\rho_Y(s_1, s_2)$ normalized by the first-order intensity values at s_1 and s_2, a reweighted pair correlation function of Y is defined as

$$g_Y(s_1, s_2) = \frac{\rho_Y(s_1, s_2)}{\lambda_Y(s_1)\lambda_Y(s_2)}.$$

Stricter conditions can be imposed on the spatial point process Y to facilitate estimation of g_Y. In particular, a spatial point process Y is said to be second-order intensity-reweighted

---

Email addresses: andrew-thurman@uiowa.edu (Andrew L. Thurman), rfu7@wisc.edu (Rao Fu), yguan@bus.miami.edu (Yongtao Guan), jzhu@stat.wisc.edu (Jun Zhu)

stationary (SIRWS) if the random process $N_Y^*(ds) = \sum_{s_1 \in Y} \lambda_Y(s_1)^{-1} I\big(s_1 \in ds\big)$ is second-order stationary (Baddeley, Møller, and Waagepetersen (2000)). If the pair correlation function g_Y satisfies $g_Y(s, s+h) = g_Y^{st}(h)$ for some function $g_Y^{st} : \mathbb{R}^2 \to \mathbb{R}^+$, then Y is SIRWS, and an inhomogeneous K-function of Y can be defined as

$$K_Y(t) = |B|^{-1} E\Big\{ \sum_{s_1 \in Y \bigcap B} \sum_{s_2 \in Y \setminus \{s_1\}} \frac{I(\|s_1 - s_2\| \leq t)}{\lambda_Y(s_1)\lambda_Y(s_2)} \Big\}, \tag{1}$$

where $t \geq 0$ and B is a bounded Borel set (Baddeley, Møller, and Waagepetersen (2000)). Essentially, $K_Y(t)$ gives the expected number of events to occur within distance t of a given event, normalized by values of the intensity function. Because Y is SIRWS, $K_Y(t)$ does not depend on the choice of B. Further, if g_Y is isotropic, in the sense that $g_Y^{st}(h) = g_Y^{iso}(\|h\|)$ for some function $g_Y^{iso} : \mathbb{R}^+ \to \mathbb{R}^+$, then

$$K_Y(t) = 2\pi \int_0^t u g_Y^{iso}(u) du \tag{2}$$

(Guan and Shen (2010)).

1.2. Poisson Process

Let $y_1, \ldots, y_n$ denote the observed spatial point pattern data comprising a set of n locations of events in D, and let $\mu(B) = \int_B \lambda(s) ds$ denote the intensity measure of the bounded Borel set $B \subset D$. A fundamental statistical model for spatial point pattern data is the spatial Poisson point process (henceforth, Poisson process) with intensity function $\lambda(\cdot)$ defined by the following conditions:

(1) Any bounded Borel set B in D has $\mu(B) \in (0, \infty)$ and $N(B) \sim$ Poisson$(\mu(B))$.

(2) Conditional on $N(B)$, the joint density f of the event locations $y_1, \ldots, y_n$ is proportional to the product of the intensity functions $\lambda(y_i)$: $f(y_1, \ldots, y_n | N(B) = n) \propto \prod_{i=1}^n \lambda(y_i)$.

Thus, conditional on the number of events, the locations of the events are independent and

$$\alpha(B_1 \times B_2) = \Big(\int_{B_1} \lambda(s_1) ds_1 \Big) \Big(\int_{B_2} \lambda(s_2) ds_2 \Big),$$

which implies $\rho(s_1, s_2) = \lambda(s_1)\lambda(s_2)$ and $g(s_1, s_2) = 1$. For more on spatial Poisson point processes, see Møller and Waagepetersen (2004).

1.3. Neyman-Scott Process

Suppose a Poisson process (or, parent process) is generated with constant intensity function $\kappa > 0$. Then, given the location of a generated event c of the parent process, a Poisson process Y_c (or, child process) is generated with inhomogeneous intensity function

$$\lambda_c(s; \beta, \omega) = h(s - c; \omega) \exp\left\{x(s)^T\beta\right\},$$

where $h(s-c;\omega)$ is a density function parameterized by ω, with the covariate vector $x(s)$ and regression coefficient vector β defined as before. The superposition of the child processes $Y = \bigcup_{c\in C} Y_c$ defines a Neyman-Scott process with the intensity function $\lambda(s;\beta) = \exp\left\{x(s)^T\beta^*\right\}$, where $\beta^* = (\log(\kappa) + \beta_0, \beta_1, \ldots, \beta_p)^T$ (Waagepetersen (2007)).

One example of a Neyman-Scott process is the Thomas process, with

$$h(s - c; \omega) = (2\pi\omega^2)^{-1} \exp\Big\{ - (2\omega^2)^{-1}(s - c)^T(s - c)\Big\}.$$

Conditional on a parent event at location c, children events are normally distributed around c. Smaller values of ω correspond to tighter clusters, and smaller values of κ correspond to fewer parents and, hence, more children events per parent for a fixed sample size of children events. The parameter vector $\theta = (\kappa, \omega)^T$ is referred to as the interaction parameter as it controls the spatial interactions (or, dependence) among events. The K-function of the Thomas process is, in closed-form,

$$K(t;\theta) = \pi t^2 + \kappa^{-1}\Big[1 - \exp\big\{ - t^2/(4\omega^2)\big\}\Big]$$

(Waagepetersen (2007)). From (2), the reweighted pair correlation function is

$$g(t) = 1 + (4\pi\kappa\omega^2)^{-1} \exp\big\{ - (4\omega^2)^{-1}t^2\big\}.$$

2. Computational Aspects

2.1. Estimation of Weight Function

Guan and Shen (2010) showed that the WEE can produce efficiency gains relative to the EE, and in particular, they proposed the weight function

$$w(s) = \{1 + \lambda(s) f(s)\}^{-1}, \quad (3)$$

where $f(s) = \int_D \{g(u-s) - 1\} \, du$ and $g(\cdot)$ is the reweighted pair correlation function. These weights were derived as minimizers of the trace of the asymptotic variance-covariance matrix of $\tilde{\beta}^{\text{WEE}}$. When clustering is present so that $g(u-s) > 1$, the weights are larger at locations where the intensity $\lambda(\cdot)$ is higher. The expression in (3) can be simplified if the process is assumed to be m-dependent, $g(u-s) = 1$ for $\|u-s\| > m$, in which case $f(s)$ can be approximated by $K(m) - \pi m^2$ for all s. To estimate $\lambda(s)$, the solution $\tilde{\beta}^{\text{EE}}$ is substituted to obtain the estimate $\hat{\lambda}(s) = \lambda(s; \tilde{\beta}^{\text{EE}})$. Alternatively, a kernel estimate $\hat{\lambda}(s)$ can be computed. Further, $K(m)$ can be estimated by

$$\hat{K}(m) = \sum_{y_1, y_2 \in Y \bigcap D} \frac{I\big(0 < \|y_1 - y_2\| < m\big)}{\exp\left[\{x(y_1) + x(y_2)\}^T \tilde{\beta}^{\text{EE}}\right]} e_{y_1, y_2}, \quad (4)$$

where e_{y_1,y_2} is an appropriate edge correction (Baddeley, Møller, and Waagepetersen (2000)). Essentially, $\hat{K}(m)$ is the number of events, $y_1, y_2 \in Y \bigcap D$, that are within distance m, adjusted for inhomogeneous intensity of events in space. Then, $\hat{w}(s) = \left\{1 + \hat{\lambda}(s)\hat{f}(s)\right\}^{-1}$.

2.2. Standard Error Estimation

The asymptotic variance-covariance matrix of the penalized WEE estimates $\hat{\beta}^{\text{WEE}}$ is

$$\Sigma(w, \beta) = A(w,\beta)^{-1}\big\{B(w,\beta) + C(w,\beta)\big\}A(w,\beta)^{-1},$$

where

$$\begin{aligned} A(w,\beta) &= \int_D w(s)x(s)x(s)^T\lambda(s;\beta)ds, \\ B(w,\beta) &= \int_D w(s)^2 x(s)x(s)^T\lambda(s;\beta)ds, \\ C(w,\beta) &= \int_D w(s)x(s)\lambda(s;\beta)\left[\int_D w(u)x(u)^T\lambda(u;\beta)\{g(u-s)-1\}du\right]ds. \end{aligned}$$

A Riemann sum approximation of $A(w,\beta)$ and $B(w,\beta)$ can be used for computation. If the process is m-dependent, then

$$C(w,\beta) = \int_D w(s)x(s)\lambda(s;\beta)\left[\int_{D(s,m)} w(u)x(u)^T\lambda(u;\beta)\{g(u-s)-1\}du\right]ds,$$

where $D(s,m) = \{u \in \mathbb{R}^2 : \|u-s\| \leq m\}$. Further, if the function $w(s)x(s)\lambda(s;\beta)$ is assumed to be smooth, then when u is close to s, $w(s)x(s)\lambda(s;\beta) \approx w(u)x(u)\lambda(u;\beta)$. Thus, $C(w,\beta)$ can be approximated by

$$C(w,\beta) \approx \int_D w(s)x(s)\lambda(s;\beta)w(s)x(s)^T\lambda(s;\beta)\left[\int_{D(s,m)}\{g(u-s)-1\}du\right]ds$$

where $\int_{D(s,m)}\{g(u-s)-1\}du \approx K(m)-\pi m^2$ (Guan and Shen (2010)). This approximation relies on the assumptions that the dependence range is not too large (e.g., m-dependence), and the function $w(s)x(s)\lambda(s;\beta)$ is smooth in some sense. The advantage of this approximation is to avoid the computationally intensive evaluation of the double integral in $C(w,\beta)$. The final estimate $\hat{\beta}^{\text{WEE}}$ and the selected covariates are substituted to obtain the plug-in estimates of the standard errors.

The standard errors of the penalized EE estimates can be obtained similarly. Since the penalized EE method is a special case of the penalized WEE method with $w(s) = 1$ for all s, the terms $A(w,\beta)$, $B(w,\beta)$, and $C(w,\beta)$ can be computed analogously. Here, the standard errors have been estimated without assuming a specific clustered spatial point process model. In addition to the empirical estimation above, it is possible to estimate the standard errors parametrically (Waagepetersen (2007); Guan and Shen (2010)). For more on exploring the dependence structure of SIRWS point processes, see Baddeley, Møller, and Waagepetersen (2000).

3. Technical Details of Theorem 1

Let

$$\lambda_k(s_1,\ldots,s_k) = \lim_{\substack{|ds_i|\to 0\\ i=1,\ldots,k}}\left[\frac{E\{N(ds_1)\cdots N(ds_k)\}}{|ds_1|\cdots|ds_k|}\right]$$

denote the kth-order intensity and let

$$G_k(s_1, \ldots, s_k) = \lim_{\substack{|ds_i| \to 0 \\ i=1,\ldots,k}} \left[\frac{cum\{N(ds_1), \ldots, N(ds_k)\}}{|ds_1| \cdots |ds_k|} \right]$$

denote the cumulant functions of X, where $cum(Y_1, \ldots, Y_k)$ is the coefficient of $i^k t_1 \ldots t_k$ in the Taylor series expansion of $\log\left[E\{\exp(i \sum_{j=1}^{k} Y_j t_j)\}\right]$ about the origin. Let

$$\begin{aligned} \alpha(p;k) &= \sup\Big\{ |P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_1 \in \mathcal{F}(E_1), A_2 \in \mathcal{F}(E_2), E_2 = E_1 + s, \\ &\qquad |E_1| = |E_2| \leq p, d(E_1, E_2) \geq k \Big\}, \end{aligned}$$

where the supremum is over all compact and convex subsets $E_1 \subset \mathbb{R}^2$ and $s \in \mathbb{R}^2$; here, $d(E_1, E_2)$ is the maximal distance between E_1 and E_2 and $\mathcal{F}(E)$ is the σ-algebra generated by the random events of N that are in E. Let $f^{(i)}(\beta)$ denote the ith derivative of $f(\beta)$.

Consider the following regularity conditions.

(A.1) There exists $0 < C_1 \leq C_2 < \infty$ such that $C_1 n^2 \leq |D_n| \leq C_2 n^2$ and $C_1 n \leq |\partial D_n| \leq C_2 n$.

(A.2) $\sup_p p^{-1} \alpha(p;k) = O(k^{-\epsilon})$ for some $\epsilon > 2$.

(A.3) $\lambda(s;\beta)$ is bounded below from 0.

(A.4) $\lambda^{(2)}(s;\beta)$ is bounded and continuous with respect to β.

(A.5) $\sup_{s_1} \int_{D_n} \cdots \int_{D_n} |G_k(s_1, \ldots, s_k)| ds_2 \cdots ds_k < C$ for $k = 2, 3, 4$.

Briefly, (A.1) ensures that the spatial domain expands in all directions and the boundary is not too irregular. (A.2) is a mixing condition to control the rate of decay of dependence between sub-domains. (A.3)–(A.5) control the higher order terms in the Taylor series expansion of the score function.

Proof. First, note that $\hat{\beta} \xrightarrow{p} \beta^0$ (Guan and Shen (2010)); consistency of $\hat{\beta}$ was assumed in the proof of Guan and Loh (2007) but proven in Guan and Shen (2010). Let

$$
\begin{aligned}
u_n(w,\beta^0) &= \sum_{y\in Y\bigcap D_n} w(y)x(y) - \int_{D_n} w(s)x(s)\lambda(s;\beta^0)ds,\\
A_n(w,\beta^0) &= \int_{D_n} w(s)x(s)x(s)^T\lambda(s;\beta^0)ds,\\
B_n(w,\beta^0) &= \int_{D_n} w(s)^2x(s)x(s)^T\lambda(s;\beta^0)ds,\\
C_n(w,\beta^0) &= \\
&\int_{D_n} w(s)x(s)\lambda(s;\beta^0)\left[\int_{D_n} w(u)x(u)^T\lambda(u;\beta^0)\{g(u-s)-1\}du\right]ds.
\end{aligned}
$$

A Taylor expansion gives

$$
|D_n|^{1/2}(\hat\beta-\beta^0) = \{-u_n^{(1)}(w,\tilde\beta)\}^{-1}|D_n|^{-1/2}u_n(w,\beta^0), \tag{5}
$$

where $\tilde\beta$ is between $\hat\beta$ and β^0. By arguments similar to Guan and Loh (2007),

$$
\begin{aligned}
E\{-u_n^{(1)}(w,\beta^0)\} &= A_n(w,\beta^0),\\
E\{|D_n|^{-1/2}u_n(w,\beta^0)\} &= 0,\\
Var\{|D_n|^{-1/2}u_n(w,\beta^0)\} &= B_n(w,\beta^0)+C_n(w,\beta^0),
\end{aligned}
$$

and $|D_n|^{-1/2}u_n(w,\beta^0)$ converges to a multivariate normal distribution. Also,

$$
|D_n|^{-1}A_n(w,\beta^0) \to A(w,\beta^0),
$$

$$
|D_n|^{-1}B_n(w,\beta^0) \to B(w,\beta^0),
$$

$$
|D_n|^{-1}C_n(w,\beta^0) \to C(w,\beta^0),
$$

where $A(w,\beta^0)$, $B(w,\beta^0)$, and $C(w,\beta^0)$ are finite limits. Thus,

$$
|D_n|^{-1/2}u_n(w,\beta^0) \quad \xrightarrow{p} \quad N\big(0, B(w,\beta^0)+C(w,\beta^0)\big) \tag{6}
$$

The remainder of the proof bears similarity to Zhu et al. (2010), who proved a result for normally-distributed lattice data. See details in Thurman (2013).

□

References

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

Guan, Y. and Loh, J.M. (2007). A thinned block bootstrap variance estimation procedure for inhomogeneous spatial point patterns. *Journal of the American Statistical Association* **102**, 1377-1386.

Guan, Y. and Shen, Y. (2010). A weighted estimating equation approach to inhomogeneous spatial point processes. *Biometrika* **97**, 867-880.

Møller, J. and Waagepetersen, R. (2004). *Statistical inference and simulation for spatial point processes.* Chapman & Hall, Boca Raton.

Thurman, A.L. (2013). On Model Selection of Spatial Point Processes. Ph.D. dissertation, University of Wisconsin, Madison, Department of Statistics.

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

Zhu, J., Huang, H.-C., and Reyes, P. (2010). On selection of spatial linear models for lattice data. *Journal of the Royal Statistical Society B* **72**, 389-402.

Supplementary Tables and Figures

		AL		SCAD		Oracle	
μ	κ	Nominal	Median	Nominal	Median	Nominal	Median
				EE			
4	5	0.52	0.53	0.52	0.53	0.53	0.50
	10	0.39	0.39	0.39	0.39	0.40	0.38
	50	0.20	0.18	0.20	0.18	0.20	0.17
16	5	0.56	0.54	0.56	0.54	0.56	0.52
	10	0.37	0.38	0.37	0.38	0.37	0.37
	50	0.17	0.16	0.17	0.16	0.17	0.16
				WEE			
4	5	0.39	0.33	0.39	0.33	0.40	0.31
	10	0.28	0.25	0.28	0.25	0.28	0.24
	50	0.19	0.15	0.19	0.15	0.19	0.15
16	5	0.34	0.28	0.34	0.28	0.34	0.27
	10	0.25	0.20	0.25	0.20	0.25	0.20
	50	0.13	0.11	0.13	0.11	0.13	0.11

Table S1: *For different values of the model parameters μ ($\times 100$) and κ ($\times 10^{-5}$), nominal standard errors of estimates of β_1 and median of standard error estimates of β_1, under adaptive Lasso (AL) and smoothly clipped absolute deviation (SCAD) penalties and the true model (Oracle), using either estimating equations (EE) or weighted estimating equations (WEE).*

		AL		SCAD		Oracle	
μ	κ	Nominal	Median	Nominal	Median	Nominal	Median
EE							
4	5	0.40	0.42	0.40	0.42	0.40	0.39
	10	0.28	0.30	0.28	0.30	0.27	0.28
	50	0.12	0.13	0.12	0.13	0.12	0.12
16	5	0.39	0.43	0.39	0.45	0.38	0.40
	10	0.30	0.30	0.30	0.32	0.30	0.28
	50	0.12	0.12	0.12	0.13	0.12	0.12
WEE							
4	5	0.30	0.27	0.30	0.28	0.30	0.26
	10	0.21	0.20	0.21	0.20	0.20	0.19
	50	0.12	0.11	0.12	0.10	0.12	0.10
16	5	0.29	0.25	0.29	0.25	0.28	0.24
	10	0.21	0.17	0.21	0.17	0.20	0.17
	50	0.09	0.08	0.09	0.08	0.09	0.08

Table S2: *For different values of the model parameters μ ($\times 100$) and κ ($\times 10^{-5}$), nominal standard errors of estimates of β_2 and median of standard error estimates of β_2, under adaptive Lasso (AL) and smoothly clipped absolute deviation (SCAD) penalties and the true model (Oracle), using either estimating equations (EE) or weighted estimating equations (WEE).*

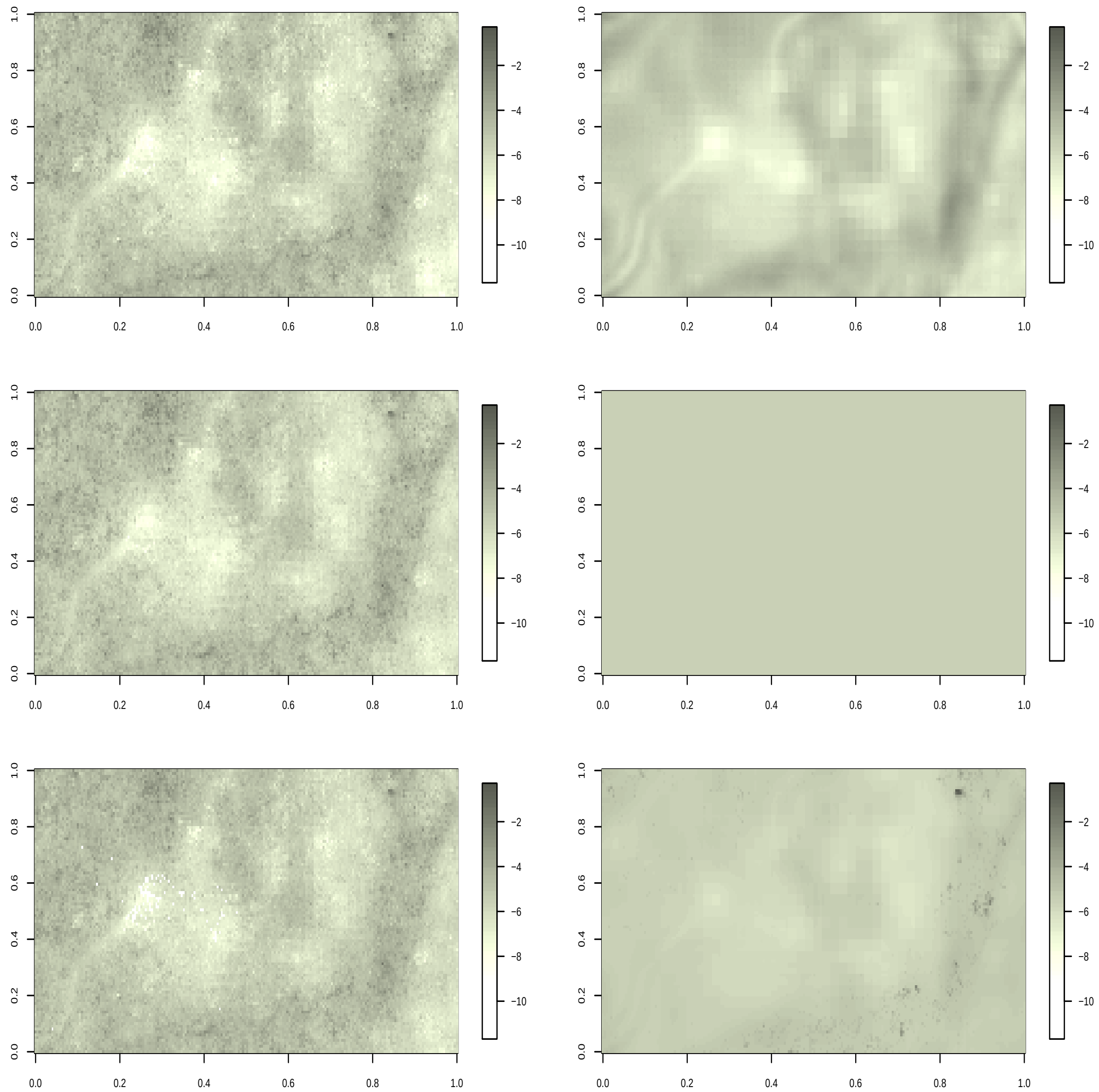

Figure S1: *Estimates of the log intensity functions for the B. pendula data under adaptive Lasso (AL) (row 1), smoothly clipped absolute deviation (SCAD) (row 2), and adaptive elastic net (AENET) (row 3) penalties, using either estimating equations (EE) (column 1) or weighted estimating equations (WEE) (column 2).*