Features selection for modelling the intensity of spatial point processes

Achmad Choiruddin
Advisors : Jean-françois Coeurjolly and Frédérique Letué

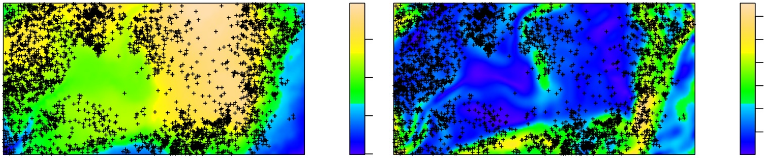
Laboratory Jean Kuntzmann, Univ. Grenoble Aples, France

4th Stochastic Geometry Days
Poitiers, August 24th-28th, 2015

- Spatial Point Processes
  - Brief introduction
  - Interests in spatial point processes
- Variable(s) Selection
  - Brief introduction
  - Penalty functions
- Variable(s) Selection for Spatial Point Processes
  - Recent studies
  - Interests of this study
  - Main results
- Simulation Study
- Conclusion

Spatial Point Processes

- Analysis of spatial point pattern data : modelling of random locations of point (or objects) observed on a continuous space
  - Example of tropical rainforest datasets : Location of *Beishmiedia pendula* trees (R Package spatstat), 1000×500 m in Barro Colorado island
  - 3604 locations of trees observed with two spatial covariates (the elevation and gradient of the elevation).



- One of interest : Modelling the intensity function.
- When the intensity function is assumed to be a parametric function of covariates : what is the appropriate procedure to select the "right covariates"?

Spatial Point Processes

- Let $\mathbf{X}$ be a spatial point process
- Suppose $\mathbf{x} = \{x_1, x_2, \ldots, x_m\}$ denotes a realization of a spatial point process $\mathbf{X}$ observed within a bounded region D
- If $\mathbf{X}$ is a Poisson point process, then the log-likehood function is $\ell(\beta) = \sum_{u \in X \cup D} log \rho(u;\beta) - \int_D \rho(u;\beta) du$
- $\rho(u;\beta)$ is the intensity function parameterized by a vector $\beta \in \mathbb{R}^p$
- Assume $\rho(u;\beta) = exp(\beta^\top z(u))$, where $z(u) = \{z_1(u), z_2(u), \cdots, z_p(u)\}$ are spatial covariates
- β can be estimated by maximizing the log-likelihood $\ell(\beta)$
- By Campbell Theorems, it is ensured that the score function $\ell^{(1)}(\beta)$ is unbiased estimating equation

Variable(s) Selection

- Variable(s) selection via maximizing the penalized likelihood :

$$Q(\beta) = \ell(\beta) - n\sum_{j=1}^{p} p_{\lambda_j}(|\beta_j|)$$

- Variable(s) selection for spatial point processes

$$Q(\beta) = \ell(\beta) - |D|\sum_{j=1}^{p} p_{\lambda_j}(|\beta_j|)$$

where :

- $\ell(\beta)$ is Poisson log-likelihood function
- $\ell(w, \beta) = \sum_{u \in X \cap D} w(u) log\rho(u; \beta) - \int_D w(u)\rho(u; \beta)du$ (Guan and Shen, 2010)
- for any $\lambda_j > 0$, we say that $p_{\lambda_j}(\theta)$ is a penalty function

Variable(s) Selection

Method	Penalty function
Ridge	$p_\lambda(\theta) = \frac{1}{2}\lambda\theta^2$
Lasso	$p_\lambda(\theta) = \lambda\|\theta\|$
Enet	$p_\lambda(\theta) = \lambda\{\frac{1}{2}(1-\alpha)\theta^2 + \alpha\|\theta\|\}$, for any $0 < \alpha < 1$
SCAD	$p'_\lambda(\theta) = \lambda\left(I(\theta \leq \lambda) + \frac{(a\theta - \beta_j)_+}{(a-1)\lambda} I(\theta > \lambda)\right)$
Ad Lasso	$p_{\lambda_j}(\theta) = \lambda_j\|\theta\|$
Ad Enet	$p_{\lambda_j}(\theta) = \{\frac{1}{2}\lambda(1-\alpha)\theta^2 + \alpha\lambda_j\|\theta\|\}$

Variable(s) Selection for Spatial Point Processes

Recent studies :

Thurman, A. L and Zhu, J. (2014)
Employed an adaptive Lasso penalty to select variables for Poisson point processes model

Thurman, A. L; Fu, R; Guan, Y; and Zhu, J. (2015)
Extended to clustered spatial point processes

Interests of this study :

- Extend the theoritical part of Thurman et al., (2015) by considering more general penalty functions
- Propose a more accurate estimator of the asymptotic covariance matrix

Notations

- $\boldsymbol{\beta_0} = (\boldsymbol{\beta_{10}^T}, \boldsymbol{\beta_{20}^T})^\mathbf{T}$ denote a *p*-dimensional vector of true coefficient values
- $\boldsymbol{\beta_{10}}$ is the *s*-dimensional vector of nonzero coefficients and $\boldsymbol{\beta_{20}} = \mathbf{0}$ is (*p*-*s*)-dimensional
- z_1 and z_2 denote the respective *s* and (*p*-*s*) vectors of covariates

We then define

$$\begin{aligned} a_n &= \max_j \{p'_{\lambda_{n,j}}(|\beta_{j0}|) : \beta_{j0} \neq 0\} \\ b_n &= \min_j \{p'_{\lambda_{n,j}}(|\beta_{j0}|) : \beta_{j0} = 0\} \\ c_n &= \max_j \{p''_{\lambda_{n,j}}(|\beta_{j0}|) : \beta_{j0} \neq 0\} \\ \Pi &= |D| diag(p''_{\lambda_{n,j}}(|\beta_{10}|), \cdots, p''_{\lambda_{n,j}}(|\beta_{s0}|)) \end{aligned}$$

Assumptions

Consider the following regularity conditions (C.1)-(C.6), where o denotes the origin of $\mathbb{R}^d$:

- (C.1) For every $n \geq 1, D_n = nA = \{na : a \in A\}$, where $A \subset \mathbb{R}^d$ is convex, compact, and contains o in its interior.
- (C.2) For any $w(u)$ and $z(u)$, $\sup_{u \in \mathbb{R}^d} |w(u)| < \infty$ and $\sup_{u \in \mathbb{R}^d} ||z(u)|| < \infty$.
- (C.3) There exists an integer $\delta \geq 1$ such that for $k = 1, \ldots, 2 + \delta$, the product density $\rho^{(k)}$ exists and $\rho^{(k)} \leq K$, where $K \leq \infty$ is a constant.
- (C.4) For the strong mixing coefficients (see, e.g., Politis, Paparoditis, and Romano (1998)), we assume that there exists some $t > d(2 + \delta)/\delta$ such that $\alpha_{2,\infty}(m) = O(m^{-t})$.
- (C.5) The second order product density $\rho^{(2)}$ exists, there exists a p x p positive definite matrix I_0 such that for all sufficiently large $n, |D_n|^{-1}\{B_n(w, \beta) + C_n(w, \beta)\} \geq I_0$.
- (C.6) For every nonnegative penalty function $p_\lambda(\theta)$, its second derivative $p''_\lambda(\theta)$ exists.

Inspired by : Coeurjolly and Møller (2014), Waagepetersen and Guan (2007), Karacsony (2006).

Main Results

Theorem 1

Assume the conditions (C.1)-(C.6) hold. If $a_n = O(|D_n|^{-1/2})$ and $c_n \to 0$, then there exists a local maximizer $\hat{\boldsymbol{\beta}}$ of $Q(\beta)$ such that $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta_0}\| = O_p(|D_n|^{-1/2} + a_n)$.

Lemma 1. *Assume the conditions (C.1)-(C.6) hold. If $a_n = O(|D_n|^{-1/2})$ and $|D_n|^{1/2} b_n \to \infty$ as $n \to \infty$, then with probability tending to 1, for any β_1 satisfying $\|\boldsymbol{\beta_1} - \boldsymbol{\beta_{10}}\| = O_p(|D_n|^{-1/2})$, for some constants C,*

$$Q\begin{pmatrix} \boldsymbol{\beta_1} \\ \mathbf{0} \end{pmatrix} = \max_{\|\beta_2\| \le C|D_n|^{-1/2}} Q\begin{pmatrix} \boldsymbol{\beta_1} \\ \boldsymbol{\beta_2} \end{pmatrix}.$$

Main Results

Theorem 2

Assume the conditions (C.1)-(C.6) hold. If $a_n = o(|D_n|^{-1/2})$ and $|D_n|^{1/2} b_n \to \infty$ as $n \to \infty$, then with probability tending to 1, the root-$|D_n|$ consistent local maximizers $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_{\mathbf{1}}^{\mathbf{T}}, \hat{\boldsymbol{\beta}}_{\mathbf{2}}^{\mathbf{T}})^T$ in Theorem 1 satisfies :

(i) Sparsity: $\hat{\boldsymbol{\beta}}_{\mathbf{2}} = \mathbf{0}$

(ii) Asymptotic Normality:

$|D_n|^{1/2} \Sigma_n(w, \beta_{10})^{-1/2} (\hat{\boldsymbol{\beta}}_{\mathbf{1}} - \boldsymbol{\beta}_{\mathbf{10}}) \xrightarrow{d} N(0, I_{sxs})$

Main Results

where $\Sigma_n(w, \beta_{10}) =$
$|D_n|\{A_n(w, \beta_{10}) + \Pi\}^{-1}\{B_n(w, \beta_{10}) + C_n(w, \beta_{10})\}\{A_n(w, \beta_{10}) + \Pi\}^{-1}$, and

$$A_n(w, \beta_{10}) = \int_{D_n} w(u) z_1(u) z_1(u)^T \rho(u; \beta_{10}) du$$

$$B_n(w, \beta_{10}) = \int_{D_n} w(u)^2 z_1(u) z_1(u)^T \rho(u; \beta_{10}) du$$

$$C_n(w, \beta_{10}) = \int_{D_n} w(u) z_1(u) \rho(u) \left[\int_{D_n} w(v) z_1(v)^T \rho(v) \{g(v-u) - 1\} dv \right] du$$

Remark. The estimator of covariance matrix is also influenced by the second derivative of the selected penalty function. For family of ℓ_1 penalty, $\Pi = 0$. In contrary, Π exists when considering Ridge, SCAD and elastic net, and adaptive elastic net penalty.

Discussions

- Requiring $a_n = O(|D_n|^{-1/2})$ and $|D_n|^{1/2} b_n \to \infty$ as $n \to \infty$ simultaneously
- Requiring $a_n = O(|D_n|^{-1/2})$: to ensure that penalized log-likelihood estimation is root-$|D_n|$ consistent
- Requiring $|D_n|^{1/2} b_n \to \infty$ as $n \to \infty$: to ensure that all zero coefficients are exclude from model
- However, if $C_1 a_n = C_2 b_n$, those two assumptions cannot be satisfied simultaneously

Discussions

Method	$a_n = \max\limits_{j=1,\dots s} \{p'_{\lambda_{n,j}}(\lvert\beta_{j0}\rvert)\}$	$b_n = \min\limits_{j=s+1,\dots,p} \{p''_{\lambda_{n,j}}(\lvert\beta_{j0}\rvert)\}$
Ridge	$\lambda_n \max\limits_{j=1,\dots s} \{\lvert\beta_{j0}\rvert\}$	$\lambda_n \min\limits_{j=s+1,\dots p} \{\lvert\beta_{j0}\rvert\}$
Lasso	λ_n	λ_n
Enet	$\lambda_n \left[(1-\alpha) \max\limits_{j=1,\dots s} \{\lvert\beta_{j0}\rvert\} + \alpha\right]$	$\lambda_n \left[(1-\alpha) \min\limits_{j=s+1,\dots p} \{\lvert\beta_{j0}\rvert\} + \alpha\right]$
SCAD	$(a\lambda_n - \min\limits_{j=s+1,\dots p} \{\lvert\beta_{j0}\rvert\})_+/(a-1)$	λ_n
Ad Lasso	$\max\limits_{j=1,\dots s} \{\lambda_{n,j}\}$	$\min\limits_{j=s+1,\dots p} \{\lambda_{n,j}\}$
Ad Enet	$\max\limits_{j=1,\dots s} \{\lambda_n(1-\alpha)\lvert\beta_{j0}\rvert + \lambda_{n,j}\alpha\}$	$\min\limits_{j=s+1,\dots p} \{\lambda_n(1-\alpha)\lvert\beta_{j0}\rvert + \lambda_{n,j}\alpha\}$

Simulation Study

- Use real dataset *bei* (R package Spatstat)
- Use two covariates (elevation and gradient) - centered and scaled
- Generate 18 artificial covariates, where each covariate was normally-distributed with mean zero and variance one
- Employ : ridge, lasso, scad, elastic net, adaptive lasso, and adaptive elastic net penalty
- 300 times simulation

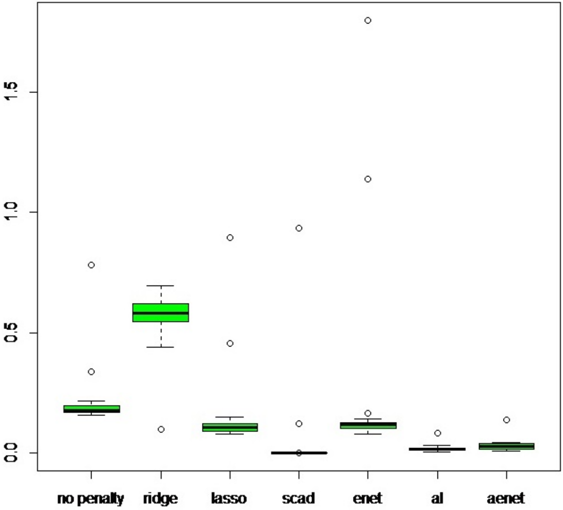
Table: Selection percentage

Method	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
Ridge	100	100	100	100	100	100	100	100	100	100
Lasso	100	100	29,3	29,7	29,3	29,3	37,7	29	34	30,3
Enet	100	100	27,7	31,7	31,3	27,7	29,7	32	29,3	29,7
SCAD	100	100	0	0,3	0	0,3	0	0,3	0,3	0,3
Ad Lasso	100	100	2	1,3	2,7	1	2	2,3	1	2,3
Ad Enet	100	100	1,7	3	3	4	2	2	3,3	3,3

Method	β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	β_{16}	β_{17}	β_{18}	β_{19}	β_{20}
Ridge	100	100	100	100	100	100	100	100	100	100
Lasso	26,7	31,7	32	25	32,3	30	31,3	29	28,3	33,3
Enet	34,7	34,3	31,7	28,3	31	31,7	29,7	31,7	30,7	33
SCAD	0	0,3	0	0	0	0,3	0	0	0	0
Ad Lasso	2,7	2,3	2,3	2,3	2,3	2	1	1,3	1	1,67
Ad Enet	1,3	3	1,67	3,67	3,3	2,3	4,3	2,7	3,7	4,7

Simulation Study

$SSE_j = \sum_{k=1}^{300} (\hat{\beta}_{jk} - \beta_{j0k})^2$, for $j = 1, \cdots, 20$



Conclusion

Conclusion :

- Employing ridge penalty, select all variables
- Employing lasso and elastic net, they perform variable selection and parameter estimation simultaneously but they seems to not have oracle property
- It looks that SCAD, adaptive lasso, and adaptive elastic net confirm the oracle property

Perspective :

- Design more complex simulation studies (exp. : consider high multicolinearity among the covariates)
- Apply to Gibbs point processes
- Consider when the number of parameters increase

References I

- Guan, Y and Shen Y. (2010). A weighted estimating equation approach for inhomogeneous spatial point process, *Biometrika*, **97**, 867-880.
- Møller, J. and Waagepetersen, R. P. (2004). Statistical inference and simulaion for spatial point processes, *Chapman and Hall/crc*.
- Thurman, A. L and Zhu, J. (2014). Variable selection for spatial Poisson point processes via a regularization method, *Statistical Methodology*, **17**, 113-125.
- Thurman, A. L; Fu, R; Guan, Y; and Zhu, J. (2015). Regularized estimating equations for model selection of clustered spatial point process, *forthcoming*.
- Yue, Y and Loh J. M. (2015). Variable selection for inhomogeneous spatial point process models, *The Canadian Journal of Statistics*, **43**, 288-305.