Regularized Poisson and logistic methods for spatial point processes intensity estimation with a diverging number of covariates

Achmad Choiruddin[1]
*Work with: Jean-françois Coeurjolly[1,2] and Frédérique Letué[1]

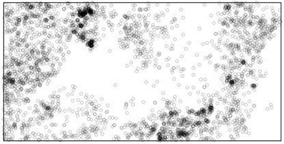
[1] Laboratory Jean Kuntzmann, Université Grenoble Alpes, France
[2] Department of Mathematics, Université du Québec à Montréal, Canada

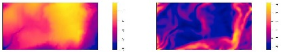
19th Workshop on Stochastic Geometry, Stereology and Image Analysis (SGSIA), Luminy, France
May 2017

Contents

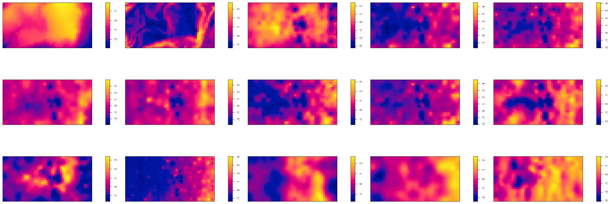
Motivation: Barro Colorado Island (BCI) study

- 3604 locations of **Beilschmiedia pendula Lauraceae (BPL)** tree
- Intensity: function of covariates
- $\rho(u;\boldsymbol{\beta}) = \exp(\boldsymbol{\beta}^\top \mathbf{z}(u))$, $\boldsymbol{\beta} \in \mathbb{R}^p$, $\mathbf{z}(u) = \{z_1(u), \cdots, z_p(u)\}^\top$ are spatial covariates at u
- $p =$ **moderate**!





In reality...



What if ...

- Parsimonius model
- Improve the prediction
- Computationally easy and fast to implement

- $\mathbf{X}$ is a spatial point process, D is the observed domain, $|D|$ is the volume of observed domain
- The Poisson log-likelihood

$$\ell(\boldsymbol{\beta}) = \sum_{u \in \mathbf{X} \cap D} \boldsymbol{\beta}^\top \mathbf{z}(u) - \int_D \exp(\boldsymbol{\beta}^\top \mathbf{z}(u)) \mathrm{d}u.$$

  Estimate $\boldsymbol{\beta}$ by solving $\ell^{(1)}(\boldsymbol{\beta}) = 0$

- Estimating equation-based methods for more general point processes

- $\mathbf{X}$ is a spatial point process, D is the observed domain, $|D|$ is the volume of observed domain
- The weighted Poisson log-likelihood (Guan and Shen, 2010)

$$\ell(w, \boldsymbol{\beta}) = \sum_{u \in \mathbf{X} \cap D} w(u) \boldsymbol{\beta}^\top \mathbf{z}(u) - \int_D w(u) \exp(\boldsymbol{\beta}^\top \mathbf{z}(u)) \mathrm{d}u.$$

Estimate $\boldsymbol{\beta}$ by solving $\ell^{(1)}(w, \boldsymbol{\beta}) = 0$

- Estimating equation-based methods for more general point processes

- $\mathbf{X}$ is a spatial point process, D is the observed domain, $|D|$ is the volume of observed domain

- The weighted logistic regression log-likelihood (Baddeley et al., 2014; Choiruddin et al., 2017)

$$\ell(w;\boldsymbol{\beta}) = \sum_{u\in\mathbf{X}\cap D} w(u)\log\left(\frac{\rho(\boldsymbol{\beta})}{g(u)+\rho(\boldsymbol{\beta})}\right) - \int_D w(u)g(u)\log\left(\frac{\rho(\boldsymbol{\beta})+g(u)}{g(u)}\right)\mathrm{d}u.$$

  Estimate $\boldsymbol{\beta}$ by solving $\ell^{(1)}(w,\boldsymbol{\beta}) = 0$

- Estimating equation-based methods for more general point processes

Regularized likelihood :

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

where :

- $\ell(w, \boldsymbol{\beta})$ is either the Poisson or logistic log-likelihoods
- $p_\lambda(\theta)$ is a penalty function parameterized by $\lambda \geq 0$
- Computation: `spatstat` + `glmnet` (and `ncvreg`)

Method	$\sum_{j=1}^{p} p_{\lambda_j}(\vert\beta_j\vert)$
Ridge	$\sum_{j=1}^{p} \frac{1}{2}\lambda\beta_j^2$
Lasso	$\sum_{j=1}^{p} \lambda\vert\beta_j\vert$
Enet	$\sum_{j=1}^{p} \lambda\{\alpha\vert\beta_j\vert + \frac{1}{2}(1-\alpha)\beta_j^2\}$
SCAD	$\sum_{j=1}^{p} p_\lambda(\vert\beta_j\vert)$, with $p_\lambda(\theta) = \begin{cases} \lambda\theta & \text{if } (\theta \leq \lambda) \\ \frac{\gamma\lambda\theta - \frac{1}{2}(\theta^2+\lambda^2)}{\gamma-1} & \text{if } (\lambda < \theta < \gamma\lambda) \\ \frac{\lambda^2(\gamma^2-1)}{2(\gamma-1)} & \text{if } (\theta \geq \gamma\lambda) \end{cases}$
MC+	$\sum_{j=1}^{p} \left\{ \left(\lambda\vert\beta_j\vert - \frac{\beta_j^2}{2\gamma}\right)\mathbb{I}(\vert\beta_j\vert < \gamma\lambda) + \frac{1}{2}\gamma\lambda^2\mathbb{I}(\vert\beta_j\vert \geq \gamma\lambda) \right\}$
Adaptive Lasso	$\sum_{j=1}^{p} \lambda_j\vert\beta_j\vert$
Adaptive enet	$\sum_{j=1}^{p} \lambda_j\{\alpha\vert\beta_j\vert + \frac{1}{2}(1-\alpha)\beta_j^2\}$

Asymptotic theory

Conditions

- The p_n-dimensional vector of true coefficient values:

$$\boldsymbol{\beta}_0 = \{\beta_{01}, \dots, \beta_{0p_n}\}^\top = \{\beta_{01}, \dots, \beta_{0s}, \beta_{0(s+1)}, \dots, \beta_{0p_n}\}^\top = \{\boldsymbol{\beta}_{01}^\top, \boldsymbol{\beta}_{02}^\top\}^\top = \{\boldsymbol{\beta}_{01}^\top, \mathbf{0}^\top\}^\top$$

- $\mathbf{X}$ observed over $D = D_n, n = 1, 2, \dots$ which expands to $\mathbb{R}^d$ as $n \to \infty$.
- We allow $p_n \to \infty$ as $n \to \infty$ and we have a fixed s.
- We assume that $p_n^3/|D_n| \to 0$ as $n \to \infty$.
- We define

$$a_n = \max_{j=1,\dots,s} \{p'_{\lambda_{n,j}}(|\beta_{0j}|)\},$$

$$b_n = \inf_{j=s+1,\dots,p_n} \inf_{\substack{|\theta| \le \epsilon_n \\ \theta \neq 0}} p'_{\lambda_{n,j}}(|\theta|), \text{ for } \epsilon_n = K\sqrt{p_n/|D_n|}, \text{ and}$$

$$c_n = \max_{j=1,\dots,s} \{p''_{\lambda_{n,j}}(|\beta_{0j}|)\}.$$

Theorem 1

Under some regularity conditions, if $a_n = O(|D_n|^{-1/2})$ and $c_n \to 0$, there exists a local maximizer $\hat{\boldsymbol{\beta}}$ of $Q(w, \boldsymbol{\beta})$ such that $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\mathbf{0}}\| = O_{\mathrm{P}}\big(\sqrt{p_n}(|D_n|^{-1/2} + a_n)\big)$.

Theorem 2

Under some regularity conditions, If $a_n|D_n|^{1/2} \to 0$, $b_n\sqrt{|D_n|/p_n^2} \to \infty$ and $\sqrt{p_n}c_n \to 0$ as $n \to \infty$, the root-$(|D_n|/p_n)$ consistent local maximizers $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1^\top, \hat{\boldsymbol{\beta}}_2^\top)^\top$ in Theorem 1 satisfy:

(i) Sparsity: $\mathrm{P}(\hat{\boldsymbol{\beta}}_2 = 0) \to 1$ as $n \to \infty$,

(ii) Asymptotic Normality: $|D_n|^{1/2}\boldsymbol{\Sigma}_n(w, \boldsymbol{\beta}_0)^{-1/2}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{01}) \xrightarrow{d} \mathcal{N}(0, \mathbb{I}_s)$,

where $\boldsymbol{\Sigma}_n(w, \boldsymbol{\beta}_0) = |D_n|\{\mathbf{A}_{n,11}(w, \boldsymbol{\beta}_0) + \boldsymbol{\Pi}_n\}^{-1}\{\mathbf{B}_{n,11}(w, \boldsymbol{\beta}_0) + \mathbf{C}_{n,11}(w, \boldsymbol{\beta}_0)\}\{\mathbf{A}_{n,11}(w, \boldsymbol{\beta}_0) + \boldsymbol{\Pi}_n\}^{-1}$

Note! Need to obey $|D_n|^{1/2}a_n \to 0$ and $\sqrt{|D_n|/p_n^2}b_n \to \infty$ as $n \to \infty$ to satisfy the Theorem 2

Method	a_n	b_n	Satisfy?
Ridge	$\lambda_n \max_{j=1,\dots s}\{\lvert\beta_{0j}\rvert\}$	0	No
Lasso	λ_n	λ_n	No
Enet	$\lambda_n \left[(1-\alpha)\max_{j=1,\dots s}\{\lvert\beta_{0j}\rvert\}+\alpha\right]$	$\lambda_n\alpha$	No
ALasso	$\max_{j=1,\dots s}\{\lambda_{n,j}\}$	$\inf_{j=s+1,\dots p}\{\lambda_{n,j}\}$	Yes
Aenet	$\max_{j=1,\dots s}\{\lambda_{n,j}\left((1-\alpha)\lvert\beta_{0j}\rvert+\alpha\right)\}$	$\alpha \inf_{j=s+1,\dots p}\{\lambda_{n,j}\}$	Yes
SCAD	0^*	$\lambda_n{}^{**}$	Yes
MC+	0^*	$\lambda_n - \frac{K\sqrt{p_n}}{\gamma\sqrt{\lvert D_n\rvert}}{}^{**}$	Yes

* if $\lambda_n \to 0$ for n sufficient large
** if $\sqrt{|D_n|/p_n^2}\lambda_n \to \infty$ for n sufficient large

Application

Barro Colorado Island (BCI) study

- Estimate BPL intensity: $\log \rho(u; \boldsymbol{\beta}) = \beta_1 z_1(u) + \cdots + \beta_{93} z_{93}(u)$
- 93 covariates: 2 topology, 13 soil nutrients, 78 interactions
- Regularized (un)weighted PL with LASSO, AL and SCAD

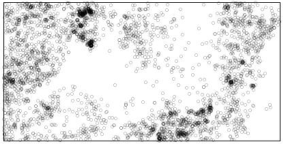
Method	Unweighted		Weighted	
	#Selected	#No	#Selected	#No
LASSO	77	16	45	48
AL	50	43	10	83
SCAD	58	35	3	90

Covariates	Unweighted			Weighted		
	LASSO	ALASSO	SCAD	LASSO	ALASSO	SCAD
Elev	0.32	0.40	0.33	0.40	0.32	0
Slope	0.39	0.40	0.36	0.42	0.44	0
Cu	0.56	0.31	0.61	0.39	0.33	0
Mn	0.14	0.14	0.09	0.15	0.22	0
P	-0.48	-0.43	-0.54	-0.33	-0.57	-1.07
Zn	-0.75	-0.66	-0.83	-0.58	-0.40	0
Al:P	-0.30	-0.29	-0.31	-0.28	-0.16	0
Mg:P	0.62	0.29	0.45	0.48	0.42	0
Zn:N	0.21	0.35	0.30	0	0	0.62
N.Min:pH	0.44	0.44	0.49	0.25	0.27	0

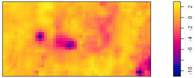
10 common selected covariates

Covariates	Unweighted			Weighted		
	LASSO	ALASSO	SCAD	LASSO	ALASSO	SCAD
Elev	0.32	0.40	0.33	0.40	0.32	0
Slope	0.39	0.40	0.36	0.42	0.44	0
Cu	0.56	0.31	0.61	0.39	0.33	0
Mn	0.14	0.14	0.09	0.15	0.22	0
P	-0.48	-0.43	-0.54	-0.33	-0.57	-1.07
Zn	-0.75	-0.66	-0.83	-0.58	-0.40	0
Al:P	-0.30	-0.29	-0.31	-0.28	-0.16	0
Mg:P	0.62	0.29	0.45	0.48	0.42	0
Zn:N	0.21	0.35	0.30	0	0	0.62
N.Min:pH	0.44	0.44	0.49	0.25	0.27	0

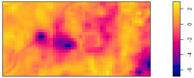
Estimates of BPL intensity (log scale)



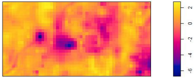
3604 locations of BPL



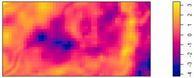
LASSO



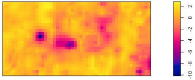
WLASSO



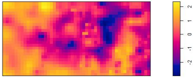
AL



WAL



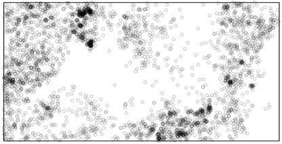
SCAD

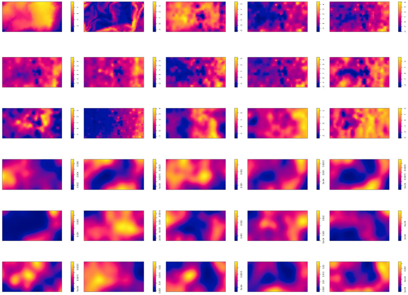


WSCAD

- Regularized versions of EE derived from Poisson and logistic regression likelihoods
- Large classes of spatial point processes and many penalty functions
- Nice theoretical properties and computationally preferable
- Satisfy our Theorems: SCAD, MC+, ALasso, Aenet
- Regularized WPL may be more appropriate for a clustered process
- Adaptive Lasso may perform best

Possible extensions





- Deal with very large datasets → some troubles
- The Dantzig selector vs. penalization methods
- Multivariate, spatio-temporal point processes

Some References

- Baddeley, A; Rubak, E; and Turner, R. (2015). Spatial point patterns: methodology and application with R, *CRC press*.
- Baddeley, A; Coeurjolly, J.-F; Rubak, E; and Waagepetersen , R. (2014). Logistic regression for spatial Gibbs point processes, *Biometrika*, **101**(2):377–392.
- Choiruddin, A; Coeurjolly, J-F; and Letué, F. (2017). Convex and non-convex regularization methods for spatial point processes intensity estimation, *ArXiv preprint arXiv:1703.02462*.
- Friedman, J; Hastie, T; Hofling H; Tibshirani, R; et al. (2007). Pathwise coordinate optimization. The Annals of Applied Statistics, 1(2):302–332.
- Guan, Y and Shen Y. (2010). A weighted estimating equation approach for inhomogeneous spatial point process, *Biometrika*, **97**(4), 867-880.
- Thurman, A. L; Fu, R; Guan, Y; and Zhu, J. (2015). Regularized estimating equations for model selection of clustered spatial point process, *Statistica Sinica*,**25**(1), 173-188.