Convex and non-convex regularization methods for intensity estimation of inhomogeneous spatial point processes

Achmad Choiruddin
*Work with: Jean-françois Coeurjolly and Frédérique Letué

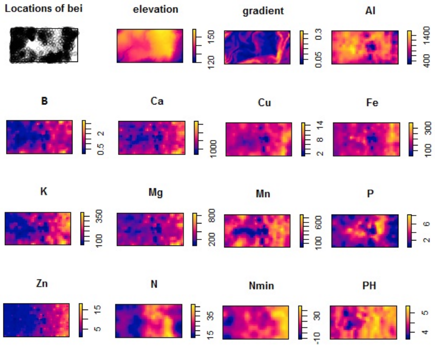
Laboratory Jean Kuntzmann, Univ. Grenoble Aples, France

Spatial Statistics and Image Analysis in Biology
Rennes, May 26th 2016

Content

- Inhomogeneous spatial point processes
  - Motivation
  - Intensity estimation without regularization
  - Intensity estimation with regularization
  - Convex and non-convex regularization methods
- Asymptotic Theory
  - Notations
  - Main results
- Simulation Study
  - Simulation set-up
  - Simulation results
- Application to forestry dataset
- Conclusion

- Tropical rainforest datasets : 3604 locations of *Beishmiedia pendula* trees (R Package spatstat), 1000 m $\times$ 500 m in BCI, with 15 covariates



- Intensity function: $\rho(u;\boldsymbol{\beta}) = \exp(\boldsymbol{\beta}^\top \mathbf{z}(u))$, $\boldsymbol{\beta} \in \mathbb{R}^p$, $\mathbf{z}(u) = \{z_1(u), z_2(u), \cdots, z_p(u)\}$ are spatial covariates
- when p is large, how to estimate $\boldsymbol{\beta}$?

Intensity estimation without regularization

- $\mathbf{X}$ is a spatial point process
- $\mathbf{x} = \{x_1, x_2, \dots, x_m\}$ denotes a realization of $\mathbf{X}$ observed within a bounded region D
- The Poisson log-likelihood

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

- By Campbell theorem, $\ell^{(1)}(\boldsymbol{\beta})$ appears to be an unbiased estimating equation, although $\mathbf{X} \neq$ Poisson point process.
- Study of intensity estimation for non-Poisson point processes (e.g., Waagepetersen, 2007; Møller and Waagepetersen, 2007, Guan and Shen, 2010; Guan et al., 2015)

Intensity estimation without regularization

- $\mathbf{X}$ is a spatial point process
- $\mathbf{x} = \{x_1, x_2, \dots, x_m\}$ denotes a realization of $\mathbf{X}$ observed within a bounded region D
- The weighted Poisson log-likelihood

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

- By Campbell theorem, $\ell^{(1)}(w, \boldsymbol{\beta})$ appears to be an unbiased estimating equation, although $\mathbf{X} \neq$ Poisson point process.
- Study of intensity estimation for non-Poisson point processes (e.g., Waagepetersen, 2007; Møller and Waagepetersen, 2007, Guan and Shen, 2010; Guan et al., 2015)

Penalized log-likelihood (e.g., Tibshirani, 1996; Fan and Li, 2001; Thurman et al., 2015):

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

where :

- $\ell(w, \boldsymbol{\beta})$ is the weighted Poisson log-likelihood
- $p_\lambda(\theta)$ is a penalty function parameterized by $\lambda \geq 0$ (non-negative, increasing function, $p_\lambda(0) = 0$)

Convex and non-convex regularization methods

Method	$\sum_{j=1}^{p} p_{\lambda_j}(\|\beta_j\|)$
Ridge	$\sum_{j=1}^{p} \frac{1}{2}\lambda\beta_j^2$
Lasso	$\sum_{j=1}^{p} \lambda\|\beta_j\|$
Enet	$\sum_{j=1}^{p} \lambda\{\alpha\|\beta_j\| + \frac{1}{2}(1-\alpha)\beta_j^2\}$
SCAD	$\sum_{j=1}^{p} p_\lambda(\|\beta_j\|)$, 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+	$\sum_{j=1}^{p} \left\{ \left(\lambda\|\beta_j\| - \frac{\beta_j^2}{2\gamma} \right) \mathbb{I}(\|\beta_j\| < \gamma\lambda) + \frac{1}{2}\gamma\lambda^2 \mathbb{I}(\|\beta_j\| \geq \gamma\lambda) \right\}$
Adaptive Lasso	$\sum_{j=1}^{p} \lambda_j\|\beta_j\|$
Adaptive enet	$\sum_{j=1}^{p} \lambda_j\{\alpha\|\beta_j\| + \frac{1}{2}(1-\alpha)\beta_j^2\}$

Notations

- $\mathbf{X}$ as a point process observed over $D = D_n, n = 1, 2, \ldots$ which expands to $\mathbb{R}^d$ as $n \to \infty$.
- Let $\boldsymbol{\beta}_0 = \{\beta_{01}, \ldots, \beta_{0p}\}^\top = \{\boldsymbol{\beta}_{01}^\top, \boldsymbol{\beta}_{02}^\top\}^\top = \{\boldsymbol{\beta}_{01}^\top, \mathbf{0}^\top\}^\top$ denote the p-dimensional vector of true coefficient values
- $p_\lambda(\theta)$ is twice continuously differentiable, we define the sequences a_n, b_n, and c_n by

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

$$b_n = \inf_j \inf_{\substack{|\theta| \le \epsilon_n \\ \theta \neq 0}} p'_{\lambda_{n,j}}(|\theta|), \text{ for } \epsilon_n = K|D_n|^{-1/2}, \text{ and}$$

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

Theorem 1

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

Theorem 2

Under some regularity conditions, if $|D_n|^{1/2}a_n \to 0$ and $|D_n|^{1/2}b_n \to \infty$ as $n \to \infty$, the root-$|D_n|$ 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 $\boldsymbol{\Sigma}_n(w,\boldsymbol{\beta}_0) = |D_n|\{\mathbf{A}_{n,11}(w,\boldsymbol{\beta}_0) + \boldsymbol{\Pi}_n\}^{-1}\{\mathbf{B}_{n,11}(w,\boldsymbol{\beta}_0) + \mathbf{C}_{n,11}(w,\boldsymbol{\beta}_0)\}\{\mathbf{A}_{n,11}(w,\boldsymbol{\beta}_0) + \boldsymbol{\Pi}_n\}^{-1}$

- $\boldsymbol{\Sigma}_n(w,\boldsymbol{\beta}_0)$ is the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}_1$
- For Lasso and adaptive Lasso, $\boldsymbol{\Pi}_n = \mathbf{0}$. For other penalties, $\boldsymbol{\Pi}_n = o(|D_n|)$ since $c_n \to 0$

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

Method	a_n	b_n	Satisfy?
Ridge	$\lambda_n \max\limits_{j=1,\dots s} \{|\beta_{0j}|\}$	0	No
Lasso	λ_n	λ_n	No
Enet	$\lambda_n \left[(1-\alpha) \max\limits_{j=1,\dots s} \{|\beta_{0j}|\} + \alpha\right]$	$\lambda_n \alpha$	No

Note! Need to obey $|D_n|^{1/2}a_n \to 0$ and $|D_n|^{1/2}b_n \to \infty$ 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	$\max_{j=1,\dots s}\{\lambda_{n,j}\big((1-\alpha)\lvert\beta_{0j}\rvert+\alpha\big)\}$	$\alpha \inf_{j=s+1,\dots p}\{\lambda_{n,j}\}$	Yes
SCAD	0^*	λ_n	Yes
MC+	0^*	λ_n	Yes

* if $\lambda_n \to 0$ for n sufficient large

Simulation set-up

- $D = [0, 1000] \times [0, 500]$
- $z_1 + z_2 + z_3 + \cdots + z_{20}$, z_1=the elevation and z_2=slope
- The true intensity function: $\rho(u; \boldsymbol{\beta}_0) = \{\beta_0 + \beta_1 z_1(u) + \beta_2 z_2(u)\}$, where $\beta_1 = 2$ and $\beta_2 = 0.75$
- The mean number of points over domain D, μ, is chosen to be 1600. Furthermore, we erode regularly the domain D such that the μ over $D \ominus R$ becomes 400
- 2000 replications from a Thomas point process, with κ ($\kappa = 5 \times 10^{-4}, \kappa = 5 \times 10^{-5}$) and $\omega = 20$
- Fitted using modified internal function in `spatstat` (Baddeley et al., 2015) `glmnet` (Friedman et al., 2010), and `ncvreg` (Breheny and Huang, 2011)

Aims:

- Regularized Poisson log-likelihood estimate (PL) Vs. Regularized weighted Poisson log-likelihood estimate (WPL)
- Behaviour of the estimate using different penalties

We look at:

- Selection properties: $\boldsymbol{\beta}_0 = \{\beta_{01}, \beta_{02}, 0, \ldots, 0\}^\top$
  - TPR=True positive rate
  - FPR= False positive rate
  - PPV= Positive prediction value
- Prediction properties: $\boldsymbol{\beta}_0 = \{2, 0.75, 0, \ldots, 0\}^\top$
  - Weighted bias (WB)
  - Weighted standard deviation (WSD)
  - Weighted root mean squared errors (WE)

Simulation Study

Simulation results: selection properties

Method	$\kappa = 5 \times 10^{-4}$						$\kappa = 5 \times 10^{-5}$					
	Regularized PL			Regularized WPL			Regularized PL			Regularized WPL		
	TPR	FPR	PPV	TPR	FPR	PPV	TPR	FPR	PPV	TPR	FPR	PPV
$\mu = 400$												
Ridge	100	100	10	100	100	10	100	100	10	100	100	10
Lasso	100*	23	40	97	11	74	87	38	29	46	1	71
Enet	100*	64	17	96	24	62	90	62	20	43	1	68
AL	100*	1	96	97	1	97	86	6	75	46	0*	74
Aenet	100*	2	87	97	2	94	86	12	61	45	0*	73
SCAD	100*	11	78	99	6	88	85	17	57	47	1	72
MC+	100*	11	76	98	7	86	84	16	56	46	1	71

Simulation Study

Simulation results: selection properties

Method	$\kappa = 5 \times 10^{-4}$						$\kappa = 5 \times 10^{-5}$					
	Regularized PL			Regularized WPL			Regularized PL			Regularized WPL		
	TPR	FPR	PPV	TPR	FPR	PPV	TPR	FPR	PPV	TPR	FPR	PPV
$\mu = 400$												
Ridge	100	100	10	100	100	10	100	100	10	100	100	10
Lasso	100*	23	40	97	11	74	87	38	29	46	1	71
Enet	100*	64	17	96	24	62	90	62	20	43	1	68
AL	100*	1	96	97	1	97	86	6	75	46	0*	74
Aenet	100*	2	87	97	2	94	86	12	61	45	0*	73
SCAD	100*	11	78	99	6	88	85	17	57	47	1	72
MC+	100*	11	76	98	7	86	84	16	56	46	1	71
$\mu = 1600$												
Ridge	100	100	10	100	100	10	100	100	10	100	100	10
Lasso	100	26	36	99	3	91	98	48	24	81	0*	99
Enet	100	70	15	99	11	74	99	78	14	83	0*	99
AL	100	0*	98	99	0*	100*	95	4	83	77	0*	99
Aenet	100	1	95	99	0*	99	97	8	69	80	0*	99
SCAD	100	6	91	99	3	94	97	12	61	82	0*	99
MC+	100	7	88	99	3	93	97	15	58	77	0*	99

* Approximation value

Simulation Study

Simulation results: prediction properties

Method	$\kappa = 5 \times 10^{-4}$						$\kappa = 5 \times 10^{-5}$					
	Regularized PL			Regularized WPL			Regularized PL			Regularized WPL		
	WB	WSD	WE	WB	WSD	WE	WB	WSD	WE	WB	WSD	WE
	$\mu = 400$											
Oracle	0.09	0.21	0.23	0.08	0.22	0.24	0.13	0.39	0.41	0.08	0.27	0.29
Ridge	0.09	0.21	0.23	0.07	0.22	0.23	0.10	0.39	0.40	0.09	0.28	0.30
Lasso	0.23	0.23	0.33	0.35	0.24	0.42	0.23	0.40	0.46	0.46	0.26	0.52
Enet	0.19	0.24	0.31	0.46	0.28	0.54	0.18	0.41	0.45	0.63	0.27	0.68
AL	0.12	0.22	0.25	0.13	0.23	0.26	0.14	0.37	0.40	0.17	0.25	0.30
Aenet	0.14	0.22	0.27	0.18	0.23	0.29	0.16	0.38	0.41	0.23	0.27	0.35
SCAD	0.28	0.18	0.34	0.30	0.19	0.36	0.26	0.34	0.43	0.42	0.31	0.52
MC+	0.29	0.18	0.35	0.31	0.19	0.36	0.27	0.34	0.43	0.42	0.31	0.52

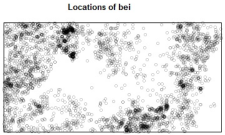
Simulation results: prediction properties

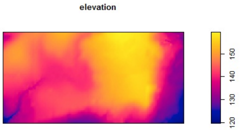
Method	$\kappa = 5 \times 10^{-4}$						$\kappa = 5 \times 10^{-5}$					
	Regularized PL			Regularized WPL			Regularized PL			Regularized WPL		
	WB	WSD	WE	WB	WSD	WE	WB	WSD	WE	WB	WSD	WE
$\mu = 400$												
Oracle	0.09	0.21	0.23	0.08	0.22	0.24	0.13	0.39	0.41	0.08	0.27	0.29
Ridge	0.09	0.21	0.23	0.07	0.22	0.23	0.10	0.39	0.40	0.09	0.28	0.30
Lasso	0.23	0.23	0.33	0.35	0.24	0.42	0.23	0.40	0.46	0.46	0.26	0.52
Enet	0.19	0.24	0.31	0.46	0.28	0.54	0.18	0.41	0.45	0.63	0.27	0.68
AL	0.12	0.22	0.25	0.13	0.23	0.26	0.14	0.37	0.40	0.17	0.25	0.30
Aenet	0.14	0.22	0.27	0.18	0.23	0.29	0.16	0.38	0.41	0.23	0.27	0.35
SCAD	0.28	0.18	0.34	0.30	0.19	0.36	0.26	0.34	0.43	0.42	0.31	0.52
MC+	0.29	0.18	0.35	0.31	0.19	0.36	0.27	0.34	0.43	0.42	0.31	0.52
$\mu = 1600$												
Oracle	0.04	0.15	0.16	0.03	0.13	0.13	0.13	0.50	0.52	0.10	0.35	0.37
Ridge	0.03	0.15	0.15	0.03	0.13	0.13	0.15	0.49	0.52	0.11	0.35	0.37
Lasso	0.11	0.16	0.19	0.23	0.16	0.28	0.20	0.49	0.53	0.30	0.28	0.42
Enet	0.09	0.16	0.18	0.34	0.19	0.39	0.19	0.50	0.54	0.52	0.26	0.58
AL	0.05	0.15	0.16	0.06	0.14	0.15	0.12	0.46	0.48	0.14	0.28	0.31
Aenet	0.05	0.15	0.16	0.07	0.14	0.16	0.15	0.48	0.50	0.15	0.29	0.33
SCAD	0.19	0.14	0.24	0.23	0.15	0.27	0.17	0.36	0.40	0.36	0.25	0.44
MC+	0.19	0.14	0.24	0.23	0.15	0.28	0.17	0.36	0.40	0.33	0.25	0.42

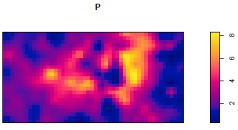
Application to forestry datasets

using adaptive Lasso method

Method	Elev	Slope	Al	B	Ca	Cu	Fe	K	Mg	Mn	P	Zn	N	Nmin	pH
PL	0.39	0.28	0	0.33	0.14	0.09	0.07	-0.04	-0.20	0.12	-0.60	-0.43	0	-0.12	-0.14
WPL	0.41	0.53	0	0	0	0	0	0	0	0.24	-0.49	-0.36	0	0	0







Conclusion:

- Satisfy our Theorem 2: SCAD, MC+, Adaptive Lasso, Adaptive Enet
- Regularized PL may be the appropriate method
- Adaptive Lasso may perform best

Perspective:

- Non asymptotic results
- $"p >> n"$
- Extension to the conditional intensity function of fractional point processes

Some References

- Baddeley, A; Rubak, E; and Turner, R .(2015). Spatial point patterns: methodology and application with R, *CRC press*.
- Coeurjolly, J-F and Møller, J. (2014). Variational approach to estimate the intensity of spatial point processes, *Bernoulli*, **20**(3), 1097-1125.
- Guan, Y and Shen Y. (2010). A weighted estimating equation approach for inhomogeneous spatial point process, *Biometrika*, **97**(4), 867-880.
- Møller, J and Waagepetersen, R. P. (2004). Statistical inference and simulaion for spatial point processes,*Chapman and Hall/crc*.
- Møller, J and Waagepetersen, R. P. (2007). Modern statistics for spatial point processes*, *Scandinavian Journal of Statistics*, **34**(4), 643-684.
- 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.

Appendix

Additional notations and conditions

- We define the matrices $\mathbf{A}_n(w,\boldsymbol{\beta}_0), \mathbf{B}_n(w,\boldsymbol{\beta}_0), \mathbf{C}_n(w,\boldsymbol{\beta}_0)$ by

$$\mathbf{A}_n(w,\boldsymbol{\beta}_0) = \int_{D_n} w(u)\mathbf{z}_1(u)\mathbf{z}(u)^\top \rho(u;\boldsymbol{\beta}_0)\mathrm{d}u$$

$$\mathbf{B}_n(w,\boldsymbol{\beta}_0) = \int_{D_n} w(u)^2\mathbf{z}(u)\mathbf{z}(u)^\top \rho(u;\boldsymbol{\beta}_0)\mathrm{d}u$$

$$\mathbf{C}_n(w,\boldsymbol{\beta}_0) = \int_{D_n} w(u)\mathbf{z}(u)\rho(u;\boldsymbol{\beta}_0)\left[\int_{D_n} w(v)\mathbf{z}(v)^\top \rho(v;\boldsymbol{\beta}_0)\{g(v-u)-1\}\mathrm{d}v\right]$$

- $\mathbf{A}_{n,11}(w,\boldsymbol{\beta}_0)$ (resp. $\mathbf{B}_{n,11}(w,\boldsymbol{\beta}_0), \mathbf{C}_{n,11}(w,\boldsymbol{\beta}_0)$) is the $s \times s$ top-left corner of $\mathbf{A}_n(w,\boldsymbol{\beta}_0)$ (resp. $\mathbf{B}_n(w,\boldsymbol{\beta}_0), \mathbf{C}_n(w,\boldsymbol{\beta}_0)$)
- $\mathbf{\Pi}_n = |D_n|\mathrm{diag}\{p''_{\lambda_{n,1}}(|\beta_{01}|), \dots, p''_{\lambda_{n,s}}(|\beta_{0s}|)\}$

Appendix

Additional notations and conditions

Consider the following conditions (C.1)-(C.7), where o denotes the origin of $\mathbb{R}^d$:

(C.1) For every $n \geq 1, D_n = nE = \{ne : e \in E\}$, where $E \subset \mathbb{R}^d$ is convex, compact, and contains o in its interior.

(C.2) $\sup_{u \in \mathbb{R}^d} |w(u)| < \infty$ and $\sup_{u \in \mathbb{R}^d} ||\mathbf{z}(u)|| < \infty$.

(C.3) There exists an integer $\delta \geq 1$ such that for $k = 1, \ldots, 2+\delta$, the product density $\rho^{(k)}$ exists and satisfies $\rho^{(k)} \leq K$, where $K < \infty$ is a constant.

(C.4) For the strong mixing coefficients, we assume that there exists some $t > d(2+\delta)/\delta$ such that $\alpha_{2,\infty}(m) = O(m^{-t})$.

(C.5) There exists a $p \times p$ positive definite matrix $\mathbf{I}_0$ such that for all sufficiently large n, we have $|D_n|^{-1}\mathbf{A}_n(w, \boldsymbol{\beta}_0) \geq \mathbf{I}_0$.

(C.6) There exists a $p \times p$ positive definite matrix $\mathbf{I}'_0$ such that for all sufficiently large n, $|D_n|^{-1}\{\mathbf{B}_n(w, \boldsymbol{\beta}_0) + \mathbf{C}_n(w, \boldsymbol{\beta}_0)\} \geq \mathbf{I}'_0$.

(C.7) The penalty function $p_\lambda(\theta)$ is twice continuously differentiable in θ for any $\theta \neq 0$.