Features Selection in Spatial Point Processes

Achmad Choiruddin

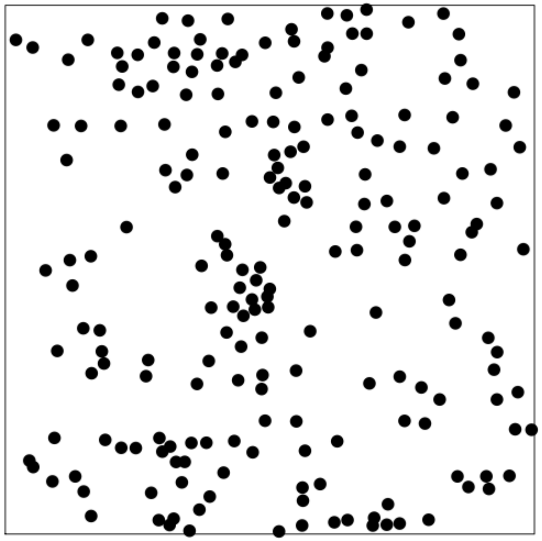
Supervisors : Jean-françois Coeurjolly and Frédérique Letué

- **Spatial Point Processes**
  - Few Examples
  - Introduction to Spatial Point Processes
  - Basic Point Processes
- **Features Selection**
  - Why should be?
  - What we propose
- **Further Research**
  - Selection of Covariates Techniques
  - More Point Processes Models
  - Estimation of Intensity Function

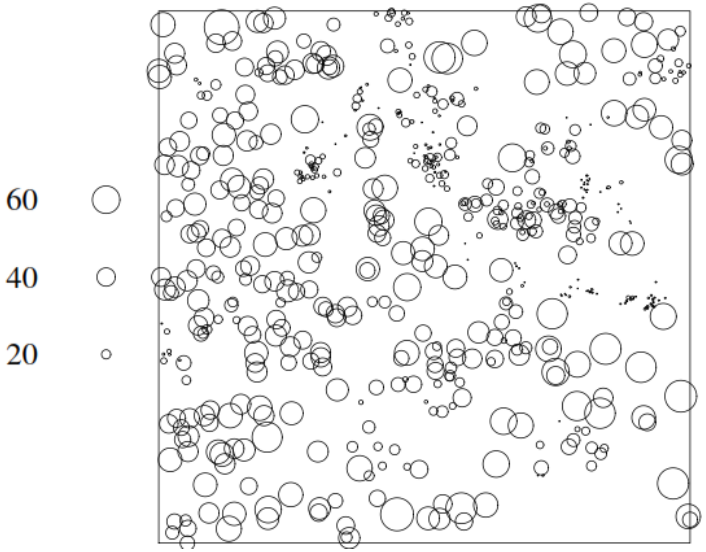
- **Spatial Point Processes**
  - Few Examples
  - Introduction to Spatial Point Processes
  - Basic Point Processes
- **Features Selection**
  - Why should be?
  - What we propose
- **Further Research**
  - Selection of Covariates Techniques
  - More Point Processes Models
  - Estimation of Intensity Function

Few Examples

- 204 seedlings and saplings of Japanese Black Pine
- 10 x 10 meters sampling region
- Objective : to understand ecological processes (competition for resources and spatial variation in the landscape)

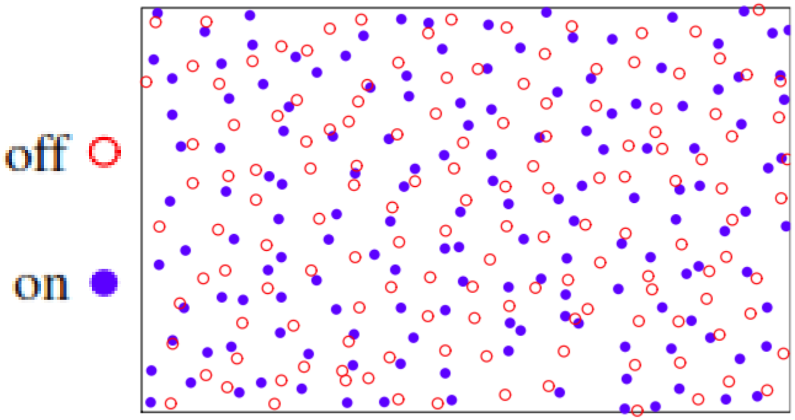


- Forest survey, size of each tree was measured by its diameter at breast height (dbh)
- 584 Longleaf Pine (Pinus palustris)
- 200 ×200 meters region in southern Georgia (USA)

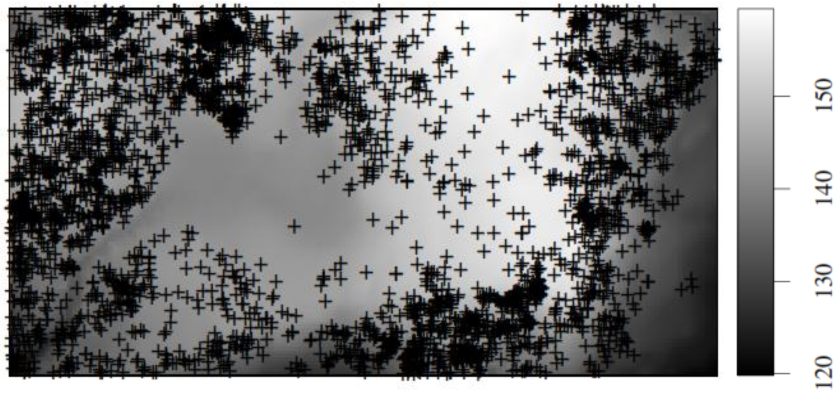


Few Examples

- Amacrine cells in the retina of a rabbit
- cells categorised as either "on" or "off" according to their response to stimuli



- 3605 trees in a tropical rainforest, with a supplementary grid map of the terrain elevation (altitude)
- The covariate *Z(u)* is the altitude at location *u*
- Research questions for investigation include whether the forest density depends on the terrain



Introduction to Spatial Point Processes

- **Point Processes on R^d**
  - A spatial point processes X is a random countable subset of a space S
  - N_{lf} : space of locally finite configurations, i.e.,

$$N_{lf} = \{X \subseteq S: n(X_B) < \infty \ for \ all \ bounded \ B \subseteq S\}$$

  - x, y, … denote points in B while ξ, φ, … denote points in S

- **Intensity ($\rho(u)$)**
  - Dividing the total number of points by the area
  - Basic descriptive characteristic of a point process (First Moment)

- **Marked Point Processes**
  - Let Y be a point process on $T \subseteq R^d$, given some space M, if a random mark process $m_\xi \in M$ is attached to each point $\xi \in Y$, then

$$X = \{(\xi, m_\xi): \xi \in Y\}$$

is called a marked point process with points in T and mark space M

Basic Point Processes

- **Poisson Point Process**
  - $n(X_B)$ is a Poisson random variable
  - If $B1$, $B2$, ... are disjoint regions of space then $n(X_{B1})$, $n(X_{B2})$, ... are independent random variables
  - Given that $n(X_B) = n$, the n points are independent and uniformly distributed in B
  - Described by it's intensity function $\rho(u)$
  - Likelihood for θ

$$\mathcal{L}(\theta; x) = \prod_{u \epsilon X_W} \rho(u;\theta) \exp\left(\int_W (1 - \rho(u;\theta))\, du\right.$$

$$Log\ \mathcal{L}(\theta; x) = \sum_{u \epsilon X_W} \log \rho(u;\theta) + \left(\int_W (1 - \rho(u;\theta))\, du\right.$$

- **Spatial Point Processes**
  - Few Examples
  - Introduction to Spatial Point Processes
  - Basic Point Processes

- **Features Selection**
  - Why should be?
  - What we propose

- **Further Research**
  - Selection of Covariates Techniques
  - More Point Processes Models
  - Estimation of Intensity Function

Why should be?

- **Real Case**
  - Study of 1251 trees in tropical forest of Barro Colorado Island (Panama)
  - 15 covariates include 13 types of soil-borne resources, terrain elevation, and slope
- **Multicollinearity**
  - Can make us select wrong model
  - Choose wrong variables, or even, spurious variables
  - Give rise to issue of over fitting
  - Under fit which leads to bias
- **Need Selection of Covariates**
  - Mitigate over fitting which results in variance inflation
  - Mitigate under fitting which leads to bias
  - Find 'true' covariates and make it more interpretable

What we propose

- **Poisson Point Process**
  - Log likelihood for θ

$$Log\ \mathcal{L}(\theta; x) = \sum_{u \epsilon X_W} \log \rho(u;\theta) + (\int_W (1 - \rho(u;\theta))\, du$$

$$= \sum_{u \epsilon X_W} \log \rho(u;\theta) - (\int_W \rho(u;\theta)\, du$$

- **Coarse quadratum approximation**

$$\int_W \rho(u;\theta) \approx \sum_{j=1}^{n+m} \rho(u_j;\theta)\, w_j$$

What we propose

$$Log\ \mathcal{L}(\theta; x) \approx \sum_{i=1}^{n} \log \rho(u_j; \theta) - \sum_{j=1}^{n+m} \rho(u_j; \theta)\, w_j$$

$$Log\ \mathcal{L}(\theta; x) \approx \sum_{j=1}^{n+m} (I_j \log \rho(u_j; \theta) - \rho(u_j; \theta) w_j)$$

$$Log\ \mathcal{L}(\theta; x) \approx \sum_{j=1}^{n+m} (y_j \log \rho(u_j; \theta) - \rho(u_j; \theta)) w_j \qquad for\ y_j = {}^{I_j}/_{w_j}$$

Equivalent to fitting a weighted Poisson generalized linear model

What we propose

- **Penalized Maximum Likelihood**

$$-\mathcal{L}(\theta; x) + np_{\gamma}(\theta)$$

Minimizing that penalized likelihood :
Simultaneously select variables and estimate the parameter

- Ridge (L2) $$p_{\gamma}(\theta) = \gamma \sum_{j=1}^{p} \theta_j^2$$

- LASSO (L1) $$p_{\gamma}(\theta) = \gamma \sum_{j=1}^{p} |\theta_j|$$

- **Spatial Point Processes**
  - Few Examples
  - Introduction to Spatial Point Processes
  - Basic Point Processes
- **Features Selection**
  - Why should be?
  - What we propose
- **Further Research**
  - Selection of Covariates Techniques
  - More Point Processes Models
  - Estimation of Intensity Function

Further Research

- **Selection of Covariates Techniques**
  - Least Absolute Shrinkage and Selection Operator (LASSO)
  - Adaptive LASSO
  - Elastic net
  - Dantzig Selector
  - Smoothly Clipped Absolute Deviation (SCAD)

- **More Point Processes Models**
  - Cox Point Process
  - Gibbs Point Process

- **Estimation of Intensity Function**
  - Parametric
  - Nonparametric

Features Selection in Spatial Point Processes

Achmad Choiruddin

Supervisors : Jean-françois Coeurjolly and Frédérique Letué

Thank you for your attention!!