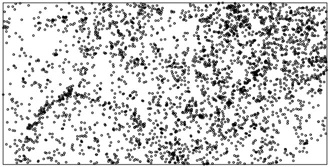
Estimating functions for inhomogeneous spatial point processes

Rasmus Waagepetersen
Spar Nord Bank and Aalborg University
Denmark

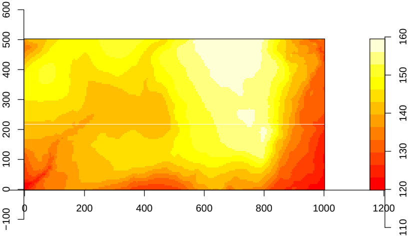
March 10, 2008

Tropical rain forests trees

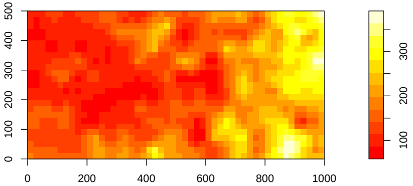
Capparis Frondosa



- *observation window* $= 1000 \text{ m} \times 500 \text{ m}$
- seed dispersal $\Rightarrow$ *clustering*
- *covariates* $\Rightarrow$ *inhomogeneity*



Elevation



Potassium content in soil.

Intensity function and product density

Intensity function of point process **X** on $\mathbb{R}^2$:

$$\rho(u)\mathrm{d}A \approx P(\mathbf{X} \text{ has a point in A})$$

Intensity function and product density

Intensity function of point process $\mathbf{X}$ on $\mathbb{R}^2$:

$$\rho(u)\mathrm{d}A \approx P(\mathbf{X} \text{ has a point in A})$$

Second order product density

$$\rho^{(2)}(u, v)dAdB \approx P(\mathbf{X} \text{ has a point in each of } A \text{ and } B)$$

Intensity function and product density

Intensity function of point process $\mathbf{X}$ on $\mathbb{R}^2$:

$$\rho(u)\mathrm{d}A \approx P(\mathbf{X} \text{ has a point in A})$$

Second order product density

$$\rho^{(2)}(u,v)dAdB \approx P(\mathbf{X} \text{ has a point in each of } A \text{ and } B)$$

Pair correlation and K-function (provided $g(u,v) = g(u-v)$)

$$g(u,v) = \frac{\rho^{(2)}(u,v)}{\rho(u)\rho(v)} \quad \text{and } K(t) = \int_{\mathbb{R}^2} 1[\|u\| \le t]g(u)\mathrm{d}u$$

NB: for Poisson process, $g(u-v) = 1$, clustering: $g(u-v) > 1$.

Parametric models

Study influence of covariates using log-linear model for intensity function:

$$\rho(u;\beta) = \exp(z(u)\beta^{\mathsf{T}})$$

and quantify clustering using parameter ψ in parametric model

$$K(t;\psi) = \int_{\|v\|\leq t} g(v;\psi)\mathrm{d}v$$

for K/g-function.

Estimating function for β

Maximum likelihood estimation only easy in case of a Poisson process $\mathbf{X}$ in which case log likelihood is

$$l(\beta) = \sum_{u \in \mathbf{X} \cap W} z(u)\beta^{\mathsf{T}} - \int_W \rho(u;\beta)\mathrm{d}u$$

Poisson score estimating function based on point process $\mathbf{X}$ observed in W:

$$u_1(\beta) = \sum_{u \in \mathbf{X} \cap W} z(u) - \int_W z(u)\rho(u;\beta)\mathrm{d}u$$

also applicable for *non-Poisson* point processes with intensity function $\rho(\cdot;\beta)$ (Schoenberg, 2005, Waagepetersen, 2007)

Estimating function for ψ

Estimate of K-function:

$$\hat{K}_\beta(t) = \sum_{u,v \in \mathbf{X} \cap W} \frac{1[0 < \|u - v\| \le t]}{\rho(u;\beta)\rho(v;\beta)} e_{u,v}$$

Unbiased if $\beta = \beta^*$ ‘true’ regression parameter.

Minimum contrast estimation: minimize

$$\int_0^r \left(\hat{K}_\beta(t) - K(t;\psi)\right)^2 \mathrm{d}t$$

or solve estimating equation

$$u_{2,\beta}(\psi) = |W| \int_0^r \left(\hat{K}_\beta(t) - K(t;\psi)\right) \frac{\mathrm{d}K(t;\psi)}{\mathrm{d}\psi} \mathrm{d}t = 0$$

Two-step estimation

Estimate $(\hat{\beta}, \hat{\psi})$ by solving

1. $u_1(\beta) = 0$
2. $u_{2,\hat{\beta}}(\psi) = 0$

or, equivalently, solve

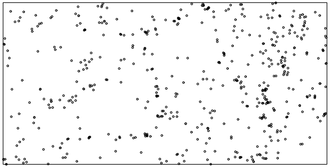
$$u(\beta, \psi) = \big(u_1(\beta), u_{2,\beta}(\psi)\big) = 0.$$

Waagepetersen and Guan (2007): asymptotic normality of $(\hat{\beta}, \hat{\psi})$ for *mixing* point processes (e.g. Poisson cluster processes).

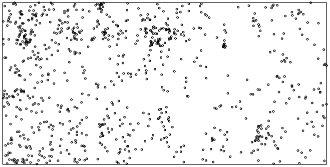
Modes of seed dispersal and clustering

Consider three species with different modes of seed dispersal:

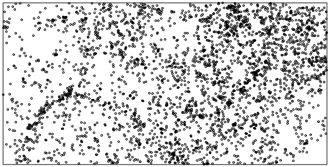
explosive capsules



wind

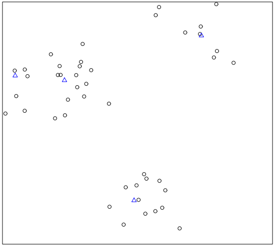


bird/mammal



Is degree of clustering related to mode of seed dispersal ?

Modified Thomas process



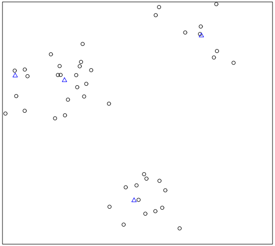
Mothers (triangles) stationary Poisson point process intensity κ

Offspring distributed around mothers according to bivariate isotropic Gaussian density with standard deviation ω

K-function:

$$K_{(\omega,\kappa)}(t) = \pi t^2 + [1 - \exp(-t^2/(2\omega)^2)]/\kappa$$

Modified Thomas process



Mothers (triangles) stationary Poisson point process intensity κ

Offspring distributed around mothers according to bivariate isotropic Gaussian density with standard deviation ω

K-function:

$$K_{(\omega,\kappa)}(t) = \pi t^2 + [1 - \exp(-t^2/(2\omega)^2)]/\kappa$$

Inhomogeneity: offspring survive according to probability

$$p(u) \propto \exp(z(u)\beta^{\mathsf{T}})$$

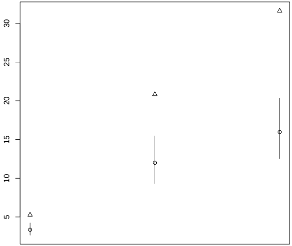
depending on covariates (independent thinning).

Fit Thomas cluster process with log linear model for intensity function.

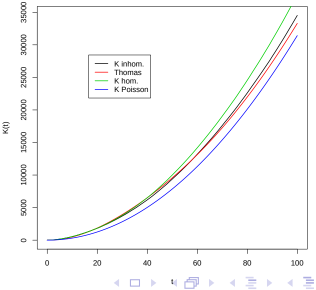
Tree intensities depend on elevation, potassium, nitrogen, and phosphorous.

Consider ω = 'width' of clusters.

Estimates of ω for explosive, wind and bird/mammal:

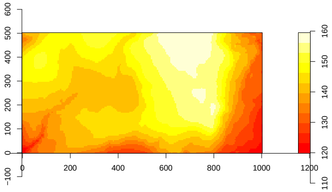


Estimates of K-functions for bird/mammal dispersed species



Missing covariate data

Elevation covariate



interpolated from elevation observations on grid.

However, evaluating

$$u(\beta) = \sum_{u \in \mathbf{X}} z(u) - \int_W z(u)\lambda(u;\beta)\mathrm{d}u$$

requires $z(u)$ *observed* for any $u \in W$!

Approximations of log likelihood: `spatstat`

Approximation of integral used in `R` package `spatstat` (Baddeley and Turner)

$$\int_W z(u)\lambda(u;\beta)\mathrm{d}u \approx \sum_{u\in\mathbf{Q}} w(u)z(u)\lambda(u;\beta)$$

where quadrature points $\mathbf{Q} = \mathbf{X} \cup \mathbf{D}$ union of *observed points* $\mathbf{X}$ and *dummy* points $\mathbf{D}$.

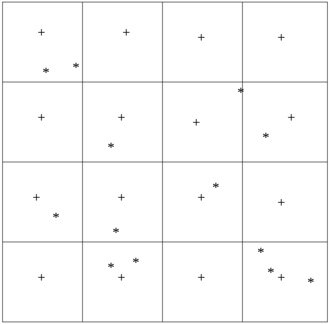
Note: $\mathbf{X}$ data points (e.g. locations of trees) supplied by 'nature'

$\mathbf{D}$ dummy points controlled by scientist.

Two types of weights $w(u)$: `grid` or `dirichlet`

Quadrature schemes in spatstat

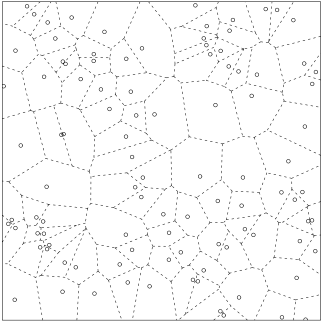
Grid



$w(u) = \frac{|C_v|}{\#(\mathbf{X} \cap C_v)+1}, \; u \in C_v$

where $W = \cup_{v \in \mathbf{D}} C_v$

Dirichlet



$w(u)$ area of *Dirichlet cell* for u in Dirichlet tesselation generated by $\mathbf{Q}$.

spatstat: relation to generalized linear models

Estimating function

$$u^{\text{spat}}(\beta) = \sum_{u \in \mathbf{X}} z(u) - \sum_{u \in \mathbf{X} \cup \mathbf{D}} w(u) z(u) \lambda(u; \beta)$$

formally equivalent to score function of weighted Poisson regression (log link):

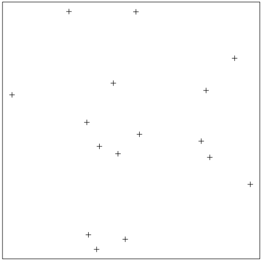
$$u^{\text{spat}}(\beta) = \sum_{u \in \mathbf{X} \cup \mathbf{D}} w(u) z(u) \left[y_u - \lambda(u; \beta) \right]$$

with weights $w(u)$ and 'observations' $y_u = 1[u \in \mathbf{X}]/w(u)$.

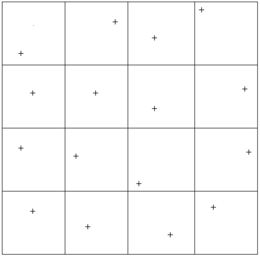
Hence implementation straightforward using e.g. `glm()` in `R` (Poisson family, log link).

Problem: hard to obtain distribution of parameter estimates from approximate score functions when deterministic quadrature points used.

Random dummy points



Simple: M *random* uniform dummy points.



Stratified: One uniformly sampled dummy point in each of M cells.

Monte Carlo approximation of integral

Rathbun et al. (2006): Monte Carlo approx. of integral:

$$\int_W z(u)\lambda(u)\mathrm{d}u \approx \frac{1}{M}\sum_{u\in\mathbf{D}} z(u)\lambda(u;\beta)$$

where **D** random dummy points.

Estimating function:

$$u^{rath}(\beta) = \sum_{u\in\mathbf{X}} z(u) - \frac{1}{M}\sum_{u\in\mathbf{D}} z(u)\lambda(u;\beta)$$

Distribution of parameter estimate (Poisson process case)

Poisson score asymptotically normal

$$u(\beta) \approx N\big(0, i(\beta)\big)$$

where $i(\beta)$ Fisher information.

Distribution of parameter estimate (Poisson process case)

Poisson score asymptotically normal

$$u(\beta) \approx N\big(0, i(\beta)\big)$$

where $i(\beta)$ Fisher information.

CLT for Monte Carlo approximation (simple dummy points):

$$M^{1/2}\big[\frac{1}{M}\sum_{u \in \mathbf{D}} z(u)\lambda(u;\beta) - \int_W z(u)\lambda(u;\beta)\mathrm{d}u\big] \xrightarrow{d} N(0, G)$$

$$u^{rath}(\beta) = u(\beta) + \big[\frac{1}{M}\sum_{u \in \mathbf{D}} z(u)\lambda(u;\beta) - \int_W z(u)\lambda(u;\beta)\mathrm{d}u\big]$$

$$\approx N(0, i(\beta) + G/M)$$

Distribution of parameter estimate (Poisson process case)

Poisson score asymptotically normal

$$u(\beta) \approx N\big(0, i(\beta)\big)$$

where $i(\beta)$ Fisher information.

CLT for Monte Carlo approximation (simple dummy points):

$$M^{1/2}\left[\frac{1}{M}\sum_{u\in\mathbf{D}} z(u)\lambda(u;\beta) - \int_W z(u)\lambda(u;\beta)\mathrm{d}u\right] \xrightarrow{d} N(0,G)$$

$$u^{rath}(\beta) = u(\beta) + \left[\frac{1}{M}\sum_{u\in\mathbf{D}} z(u)\lambda(u;\beta) - \int_W z(u)\lambda(u;\beta)\mathrm{d}u\right]$$
$$\approx N(0, i(\beta) + G/M)$$

Suppose $\hat\beta$ solution of $u^{rath}(\beta) = 0$. Then

$$\hat\beta - \beta \approx u^{rath}(\beta) i(\beta)^{-1} \Rightarrow \hat\beta \approx N\big(\beta, i(\beta)^{-1} + i(\beta)^{-1}\ G\ i(\beta)^{-1}/M\big)$$

Monte Carlo versions of `spatstat` (Waagepetersen, 2007)

Monte Carlo version of `dirichlet` (simple or stratified dummy points)

$$u^{\mathrm{dir}}(\beta) = \sum_{u \in \mathbf{X}} z(u) - \sum_{u \in \mathbf{X} \cup \mathbf{D}} z(u)\lambda(u;\beta)\frac{\lambda(u;\beta)}{\lambda(u;\beta)+M}$$

(logistic regression/case-control)

Monte Carlo versions of spatstat (Waagepetersen, 2007)

Monte Carlo version of `dirichlet` (simple or stratified dummy points)

$$u^{\text{dir}}(\beta) = \sum_{u \in \mathbf{X}} z(u) - \sum_{u \in \mathbf{X} \cup \mathbf{D}} z(u)\lambda(u;\beta)\frac{\lambda(u;\beta)}{\lambda(u;\beta)+M}$$

(logistic regression/case-control)

Monte Carlo version of `grid`: (stratified dummy points)

$$u^{\text{grid}}(\beta) = \sum_{u \in \mathbf{X}} z(u) - \sum_{u \in \mathbf{X} \cup \mathbf{D}} z(u)\lambda(u;\beta)\frac{|C_u|}{\#(\mathbf{X} \cap C_u)+1}$$

(Poisson GLM with weights $w(u)$)

Both cases easy to implement using `glm()`

Distribution of parameter estimates

Grid: $u^{\text{grid}}(\beta)$ and $u^{\text{rath}}(\beta)$ with stratified dummy points same asymptotic distribution - hence same asymptotic distribution for parameter estimates.

Dirichlet: more subtle behaviour

- stratified dummy points: $u^{\text{dir}}(\beta)$ less efficient than $u^{\text{grid}}(\beta)/u^{\text{rath}}(\beta)$
- simple dummy points: asymptotic covariance matrix tends to that of MLE when $M \to \infty$.

NB: 'infill' asymptotics - both intensity of $\mathbf{X}$ and M tends to infinity. **NB:** asymptotic results also available for cluster processes.

References

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

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

Waagepetersen, R. (2007) Estimating functions for inhomogeneous spatial point processes with incomplete covariate data, *Biometrika*, to appear.

Waagepetersen, R. and Guan, Y. (2007). Two-step estimation for inhomogeneous spatial point processes, submitted.