Two-step estimation for inhomogeneous spatial point processes

Rasmus Waagepetersen
Spar Nord Bank and Aalborg University
Denmark

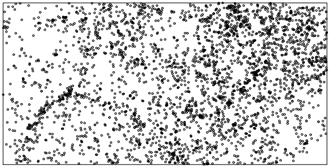
February 27, 2010

Background: Tropical rain forest ecology

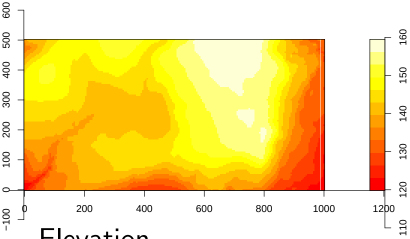
Fundamental questions: which factors govern the spatial distribution of rain forest trees and support the high biodiversity of rain forests ?

- environment: topography, soil composition,...
- seed dispersal: by wind, birds or mammals...
- competition between species

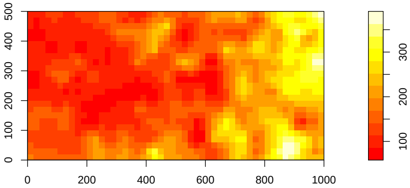
Example: *Capparis Frondosa*



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



Elevation



Potassium content in soil.

Quantify dependence on environmental variables and seed dispersal using statistics for spatial point processes.

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

$$g(u,v) = \frac{\rho^{(2)}(u,v)}{\rho(u)\rho(v)}$$

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 $g(u,v) = g(u-v)$ i.e. $\mathbf{X}$ second-order reweighted stationary)

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

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:

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

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

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

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:

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

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 (2009): asymptotic normality of $(\hat{\beta}, \hat{\psi})$ for *mixing* point processes (e.g. Poisson cluster processes).

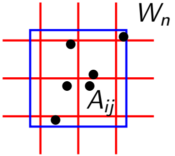
Essential requirement: $u(\beta^*, \psi^*)$ asymptotically normal - then asymptotic normality of $(\hat{\beta}, \hat{\psi})$ follows by Taylor expansion

$$(\hat{\beta}, \hat{\psi}) - (\beta^*, \psi^*) \approx u(\beta^*, \psi^*) \left[\frac{\mathrm{d}u(\beta, \psi)}{\mathrm{d}(\beta\psi)} \right]^{-1}$$

CLT for estimating function

Consider increasing observation windows W_n.

Divide $\mathbb{R}^2$ into quadratic cells
$A_{ij} = [i, i+1[\times[j, j+1[$



Express Poisson score in terms of lattice process X_{ij}, $i, j \in \mathbb{Z}$:

$$u_1^n(\beta) = \sum_{u \in \mathbf{X} \cap W_n} z(u) - \int_{W_n} z(u)\rho(u;\beta)\mathrm{d}u =$$

$$\sum_{i,j}\left[\sum_{u \in \mathbf{X} \cap W_n \cap A_{ij}} z(u) - \int_{W_n \cap A_{ij}} z(u)\rho(u;\beta)\mathrm{d}u\right] = \sum_{ij: A_{ij} \subseteq W_n} X_{ij} + o_P(1)$$

Similarly:

$$u^n_{2,\beta}(\psi) = |W_n| \int_0^r \left(\hat{K}_\beta(t) - K(t;\psi)\right) \frac{\mathrm{d}K(t;\psi)}{\mathrm{d}\psi} \mathrm{d}t = \sum_{ij:A_{ij}\subseteq W_n} Y_{ij} + o_P(1)$$

since

$$\hat{K}_\beta(t) = \sum_{u\in\mathbf{X}\cap W_n} \sum_{v\in\mathbf{X}\cap W_n} \frac{1[0<\|u-v\|\leq t]}{\rho(u;\beta)\rho(v;\beta)} e_{u,v} = \sum_{ij:A_{ij}\subseteq W_n} \hat{K}_{\beta,ij}(t) + o_P(1)$$

where

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

estimate of K-function based on $\mathbf{X}\cap A_{ij}$.

$\{X_{ij}\}$ and $\{Y_{ij}\}$ multivariate lattice processes.

Apply Bolthausen/Guyoun CLT for mixing lattice processes to random field $\{Z_{ij}\}_{ij}$ of linear combinations

$$Z_{ij} = X_{ij}x^\mathsf{T} + Y_{ij}y^\mathsf{T}.$$

Mixing

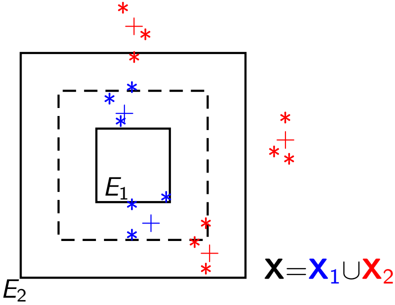
Consider $E_1, E_2 \subseteq \mathbb{R}^2$ and point configurations F_1 and F_2.

Need polynomial decay of dependence

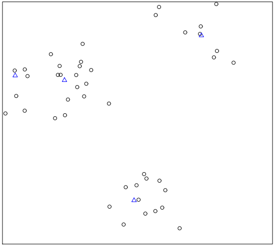
$$|P(\mathbf{X} \cap E_1 \in F_1, \mathbf{X} \cap E_2 \in F_2) - P(\mathbf{X} \cap E_1 \in F_1)P(\mathbf{X} \cap E_2 \in F_2)|$$

as function of distance between E_1 and E_2.

This can easily be verified for a Poisson cluster process where cluster density decays fast enough.



Modified Thomas process



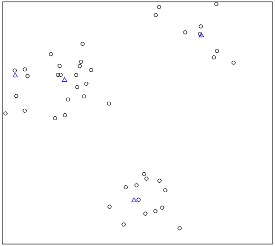
Mothers (triangles) stationary Poisson point process intensity κ

Offspring distributed around mothers according to 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 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

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

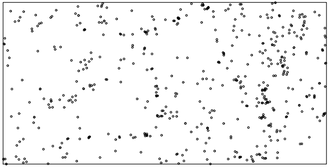
depending on covariates (independent thinning).

Inhomogenous Thomas process independent thinning of Thomas $\Rightarrow$ mixing

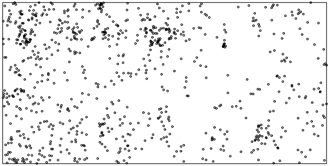
Modes of seed dispersal and clustering

Three species with different modes of seed dispersal:

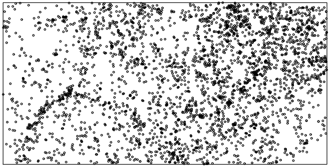
Acalypha Diversifolia explosive capsules



Loncocharpus Heptaphyllus wind



Capparis Frondosa bird/mammal



Is degree of clustering related to mode of seed dispersal ?

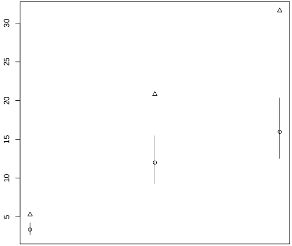
Fit Thomas cluster process with log linear model for intensity function.

Acalypha and *Capparis*: positive dependence on elevation and potassium.

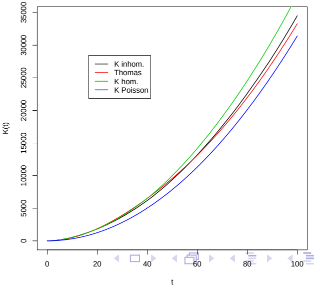
Loncocharpus: negative dependence on nitrogen and phosphorous.

Recall ω = 'width' of clusters.

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



Estimates of K-functions for bird/mammal dispersed species



Further statistical and biological issues:

- choice of integration limit r for minimum contrast estimation

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

- variance of $\hat{K}_{\hat{\beta}}(t)$ smaller than variance of $\hat{K}_{\beta*}(t)$ hence better to use $\hat{\beta}$ than β^* when estimating ψ.
- joint modelling of several species (competition)
- how to quantify the relative importance of different sources of spatial variation ?

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. and Guan, Y. (2009). Two-step estimation for inhomogeneous spatial point processes, *Journal of the Royal Statistical Society, Series B*, **71**, 685-702.