Optimal Estimation of the Intensity of a Spatial Cox Process

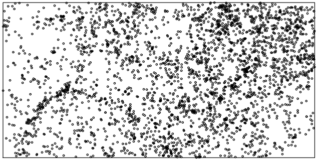
Rasmus Waagepetersen
Department of Mathematical Sciences
Aalborg University

August 22, 2012

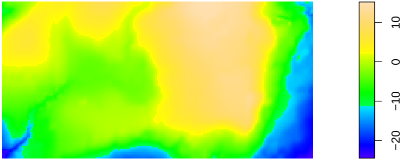
joint work

with Yongtao Guan and Abdollah Jalilian

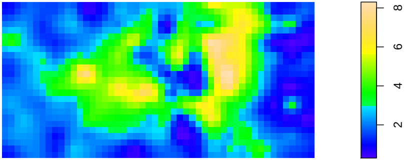
Data example: *Capparis Frondosa*



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



Elevation



Potassium content in soil.

Quantify dependence on environmental variables taking into account clustering due to e.g. seed dispersal.

Framework: spatial Cox point processes.

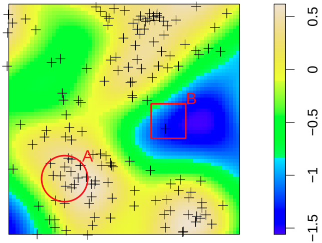
Poisson and Cox processes

X random set of points. $N(B)$ random number of points in $B \subseteq \mathbb{R}^2$.

X Poisson process with intensity function $\rho(\cdot)$:

counts $N(B)$ independent and Poisson with mean

$$\mathbb{E}N(B) = \int_B \rho(u)\mathrm{d}u$$



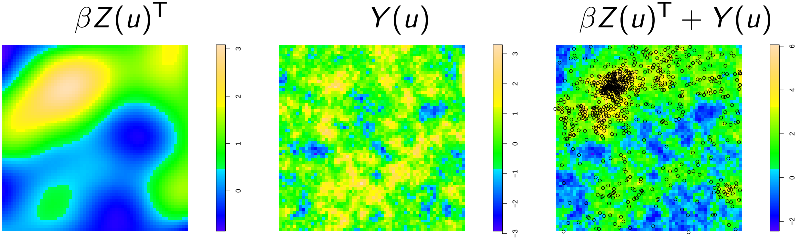
X is a *Cox process* driven by the *random* intensity function Λ if, conditional on $\Lambda = \lambda$, **X** is a Poisson process with intensity function λ.

Example: log Gaussian Cox process

log Gaussian Cox process ("point process GLMM")

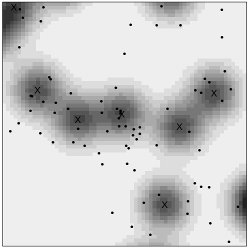
$$\Lambda(u) = \exp[\beta Z(u)^{\mathsf{T}} + Y(u)]$$

where $\{Y(u)\}$ Gaussian random field:



Shot-noise Cox process

$$\Lambda(u) = \sum_{v \in C} \gamma_v k(u - v)$$

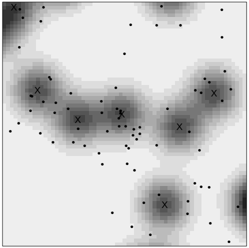


where

- C homogeneous Poisson with intensity κ
- $k(\cdot)$ probability density.
- γ_v *iid* positive random variables independent of C

Shot-noise Cox process

$$\Lambda(u) = \sum_{v \in C} \gamma_v k(u - v)$$



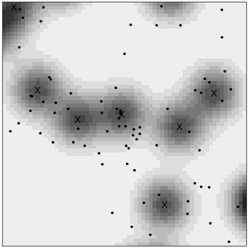
where

- C homogeneous Poisson with intensity κ
- $k(\cdot)$ probability density.
- γ_v *iid* positive random variables independent of C

NB: equivalent to cluster process with parents C, random cluster size γ_v and dispersal density k.

Shot-noise Cox process

$$\Lambda(u) = \sum_{v \in C} \gamma_v k(u - v)$$



where

- C homogeneous Poisson with intensity κ
- $k(\cdot)$ probability density.
- γ_v *iid* positive random variables independent of C

NB: equivalent to cluster process with parents C, random cluster size γ_v and dispersal density k.

Inhomogeneous shot-noise:

$$\Lambda(u) = \exp[\beta Z(u)^{\mathsf{T}}] \sum_{v \in C} \gamma_v k(u - v)$$

Moments for Cox processes

Intensity function

$$\rho(u) = \mathbb{E}\Lambda(u)$$

Second-order product density

$$\rho^{(2)}(u,v) = \mathbb{E}\Lambda(u)\Lambda(v) = \mathbb{C}\mathrm{ov}[\Lambda(u),\Lambda(v)] + \rho(u)\rho(v)$$

Moments for Cox processes

Intensity function

$$\rho(u) = \mathbb{E}\Lambda(u)$$

Second-order product density

$$\rho^{(2)}(u, v) = \mathbb{E}\Lambda(u)\Lambda(v) = \mathbb{C}\mathrm{ov}[\Lambda(u), \Lambda(v)] + \rho(u)\rho(v)$$

Pair correlation function

$$g(u, v) = \frac{\rho^{(2)}(u, v)}{\rho(u)\rho(v)} = 1 + \frac{\mathbb{C}\mathrm{ov}[\Lambda(u), \Lambda(v)]}{\rho(u)\rho(v)}$$

Moments for Cox processes

Intensity function

$$\rho(u) = \mathbb{E}\Lambda(u)$$

Second-order product density

$$\rho^{(2)}(u,v) = \mathbb{E}\Lambda(u)\Lambda(v) = \mathbb{C}\mathrm{ov}[\Lambda(u),\Lambda(v)] + \rho(u)\rho(v)$$

Pair correlation function

$$g(u,v) = \frac{\rho^{(2)}(u,v)}{\rho(u)\rho(v)} = 1 + \frac{\mathbb{C}\mathrm{ov}[\Lambda(u),\Lambda(v)]}{\rho(u)\rho(v)}$$

$$\begin{aligned}
\mathbb{C}\mathrm{ov}[N(A),N(B)] &= \int_{A\cap B} \mathbb{E}\Lambda(u)\mathrm{d}u + \int_A\int_B \mathbb{C}\mathrm{ov}[\Lambda(u),\Lambda(v)]\mathrm{d}u\mathrm{d}v \\
&= \int_{A\cap B} \rho(u)\mathrm{d}u + \int_A\int_B \rho(u)\rho(v)[g(u,v)-1]\mathrm{d}u\mathrm{d}v \\
&= \text{ Poisson variance } + \text{over dispersion due to } \Lambda
\end{aligned}$$

Log-linear model

Both log Gaussian and shot-noise Cox process of the form

$$\Lambda(u) = \Lambda_0(u) \exp[\beta Z(u)^\mathsf{T}]$$

where Λ_0 stationary non-negative reference process.

Log-linear intensity (assume $\mathbb{E}\Lambda_0(u) = 1$)

$$\rho(u) = \mathbb{E}\Lambda(u) = \exp[\beta Z(u)^\mathsf{T}]$$

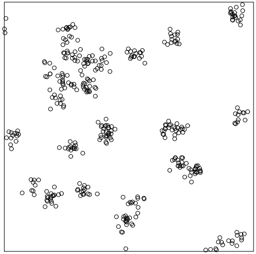
Pair correlation function ($\mathbb{E}\Lambda_0(u) = 1$):

$$g(h) = 1 + c_0(h) \quad c_0(h) = \mathbb{C}\mathrm{ov}[\Lambda_0(u), \Lambda_0(u+h)]$$

Interpretation: Cox process $\mathbf{X}$ independent inhomogeneous thinning of stationary $\mathbf{X}_0$ with random intensity function Λ_0.

Thinning interpretation: inhomogeneous Thomas process

$\mathbf{X}_0$

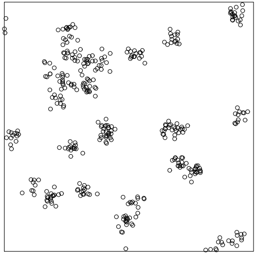


Λ_0 shot-noise process $\Rightarrow$ $\mathbf{X}_0$ cluster process:

Offspring distributed around Poisson parents according to Gaussian density

Thinning interpretation: inhomogeneous Thomas process

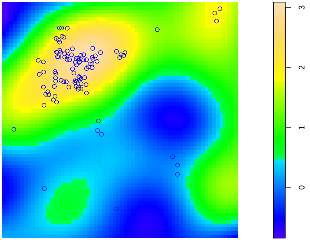
$\mathbf{X}_0$



Λ_0 shot-noise process $\Rightarrow$ $\mathbf{X}_0$ cluster process:

Offspring distributed around Poisson parents according to Gaussian density

$\mathbf{X}$



Inhomogeneity: offspring in $\mathbf{X}_0$ survive according to probability

$$p(u) \propto \exp[\beta Z(u)^{\mathsf{T}}]$$

depending on covariates (independent thinning).

Specific models for $g(u-v) = 1 + \mathbb{C}\mathrm{ov}[\Lambda_0(u), \Lambda_0(v)]$

Log-Gaussian:

$$\Lambda_0(u) = \exp[Y(u)]$$

where Y Gaussian field.

Pair corrrelation (Laplace transform)

$$g(h) = \exp[\mathbb{C}\mathrm{ov}(Y(u), Y(u+h))]$$

Specific models for $g(u - v) = 1 + \mathbb{C}\mathrm{ov}[\Lambda_0(u), \Lambda_0(v)]$

Log-Gaussian:

$$\Lambda_0(u) = \exp[Y(u)]$$

where Y Gaussian field.

Pair corrrelation (Laplace transform)

$$g(h) = \exp[\mathbb{C}\mathrm{ov}(Y(u), Y(u+h))]$$

Shot-noise:

$$\Lambda_0(u) = \sum_{v \in C} \gamma_v k(u - v)$$

Pair correlation (convolution):

$$g(u - v) = \kappa\alpha^2 \int_{\mathbb{R}^2} k(u)k(u+h)\mathrm{d}u$$

$(\alpha = \mathbb{E}\gamma_v)$

Estimation of intensity function

Want to estimate β in regression model for intensity function:

$$\rho_\beta(u) = \mathbb{E}\Lambda(u) = \exp[\beta Z(u)^\mathsf{T}]$$

MLE is possible via numerical approximations (Laplace, MCMC) but time-consuming.

Here: estimating functions based on first and second order properties of Cox processes (like quasi-likelihood or GEE for generalized linear mixed models).

Campbell formula and first-order estimating functions

Campbell:

$$\mathbb{E} \sum_{u \in \mathbf{X}} f(u) = \int_{\mathbb{R}^2} f(u) \rho_\beta(u) \mathrm{d}u$$

Then

$$e_f(\beta) = \sum_{u \in \mathbf{X} \cap W} f_\beta(u) - \int_W f_\beta(u) \rho_\beta(u) \mathrm{d}u$$

unbiased estimating function:

$$\mathbb{E} e_f(\beta) = 0$$

Parameter estimate $\hat{\beta}$ solution

$$e_f(\hat{\beta}) = 0$$

Poisson score/composite likelihood estimating function

Choice

$$f_\beta(u) = \frac{\mathrm{d}}{\mathrm{d}\beta} \log \rho_\beta(u) = \frac{\rho'_\beta(u)}{\rho_\beta(u)}$$

leads to *composite likelihood/Poisson likelihood* score

$$\sum_{u \in \mathbf{X} \cap W} \frac{\rho'_\beta(u)}{\rho_\beta(u)} - \int_W \rho'_\beta(u)\mathrm{d}u$$

This is optimal choice for Poisson process (MLE) but what is optimal f_β in the clustered case ?

Optimal first-order estimating equation

Optimal choice of f_β: smallest variance

$$\mathbb{V}\mathrm{ar}\hat{\beta} = V_f = S_f^{-1} \Sigma_f S_f^{-1}$$

where

$$S_f = -\mathbb{E} \frac{\mathrm{d}}{\mathrm{d}\beta^\mathsf{T}} e_f(\beta) \quad \Sigma_f = \mathbb{V}\mathrm{ar} e_f(\beta)$$

Optimal first-order estimating equation

Optimal choice of f_β: smallest variance

$$\mathbb{V}\mathrm{ar}\hat{\beta} = V_f = S_f^{-1}\Sigma_f S_f^{-1}$$

where

$$S_f = -\mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\beta^{\mathsf{T}}} e_f(\beta) \quad \Sigma_f = \mathbb{V}\mathrm{ar} e_f(\beta)$$

Optimal f_β solution of Fredholm equation

$$f_\beta(u) + \int_W t(u,v) f_\beta(v)\mathrm{d}u = \frac{\mathrm{d}}{\mathrm{d}\beta}\log\rho_\beta(u), \quad u \in W,$$

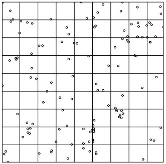
where integral equation kernel is

$$t(u,v) = \rho_\beta(v)[g(u,v) - 1]$$

Note: optimal f_β depends on pair correlation (second order/covariance property)!

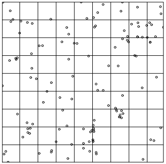
Numerical approximation and quasi-likelihood

Approximate solution of Fredholm equation using numerical quadrature: Riemann sum dividing W into cells C_i with representative points u_i.



Numerical approximation and quasi-likelihood

Approximate solution of Fredholm equation using numerical quadrature: Riemann sum dividing W into cells C_i with representative points u_i.



Resulting estimating function is *quasi-likelihood*

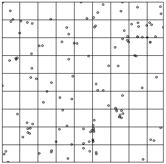
$$(N - \mu)V^{-1}D$$

based on

$$N = (N_1, \ldots, N_m), \quad N_i \text{ count of points in } C_i.$$

Numerical approximation and quasi-likelihood

Approximate solution of Fredholm equation using numerical quadrature: Riemann sum dividing W into cells C_i with representative points u_i.



Resulting estimating function is *quasi-likelihood*

$$(N - \mu) V^{-1} D$$

based on

$$N = (N_1, \ldots, N_m), \quad N_i \text{ count of points in } C_i.$$

μ mean of N:

$$\mu_i = \mathbb{E} N_i = \rho_\beta(u_i)|C_i| \text{ and } D = \left[\mathrm{d}\mu(u_i)/\mathrm{d}\beta_l \right]_{il}$$

V covariance of N:

$$V_{ij} = \mathbb{C}\mathrm{ov}[N_i, N_j] = \mu_i 1[i = j] + \mu_i \mu_j [g(u_i, u_j) - 1]$$

Practical implementation: IGLS

Solve

$$(N - \mu(\beta)V(\beta)^{-1}D(\beta) = 0$$

using iterative generalized least squares:

$$(\beta^{(l+1)} - \beta^{(l)})D(\beta^{(l)})^{\mathsf{T}}V(\beta^{(l)})^{-1}D(\beta^{(l)}) = (N - \mu(\beta^{(l)}))V(\beta^{(l)})^{-1}D(\beta^{(l)})$$

One issue: use fine discretization (large m) $\Rightarrow$ V highdimensional matrix - e.g. V 10000×10000.

Use tapering and sparse matrix Cholesky from `Matrix` library in R.

Simulation study

Consider variance of $\hat{\beta}$ obtained from either composite likelihood or quasi-likelihood.

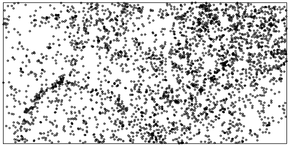
Simulations of inhomogeneous modified Thomas process depending on spatial covariates.

Reduction in variance for quasi-likelihood relative to composite likelihood: 10% to 65%.

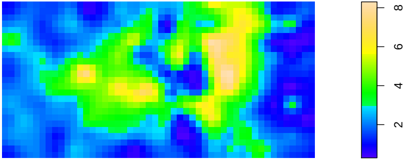
Large reductions when strong clustering and strong inhomogeneity.

Example: tree species Capparis Frondosa and Loncocharpus Heptaphyllus

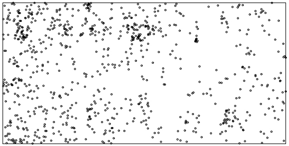
Capparis Frondosa



Potassium content in soil.



Loncocharpus Heptaphyllus

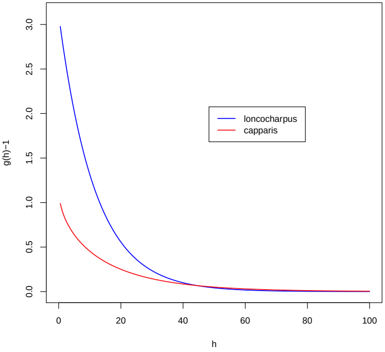


Covariates pH, elevation, gradient, potassium,...

Fitted pair correlation functions $g(\cdot)$

Use shot-noise Cox process with dispersal kernel given by variance-gamma density.

Then $g(h) - 1$ *Matérn covariance* function depending on smoothness/shape parameter ν.



Loncocharpus:
Matérn $\nu = 0.5$

Capparis:
Matérn $\nu = 0.25$

Results with composite likelihood and quasi-likelihood

species		$\widehat{\beta}$
Loncocharpus	CL	$-6.49 - 0.021\texttt{Nmin} - 0.11\texttt{P} - 0.59\texttt{pH} - 0.11\texttt{twi}$ $(81.06^*, 7.45^*, 58.78, 282.89^*, 53.19^*) \times 10^{-3}$
	QL	$-6.49 - 0.023\texttt{Nmin} - 0.12\texttt{P} - 0.55\texttt{pH} - 0.084\texttt{twi}$ $(80.15^*, 6.95^*, 55.23^*, 266.10^*, 45.47) \times 10^{-3}$
Capparis	CL	$-5.07 + 0.028\texttt{ele} - 1.10\texttt{grad} + 0.0043\texttt{K}$ $(79.54^*, 9.98^*, 1200.36, 1.16^*) \times 10^{-3}$
	QL	$-5.10 + 0.019\texttt{ele} - 2.50\texttt{grad} + 0.0039\texttt{K}$ $(77.77^*, 8.86^*, 935.02^*, 1.02^*) \times 10^{-3}$

Estimated standard errors always smallest for QL. Covariate `grad` significant according to QL but not for CL.

References

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

Jalilian, Guan and Waagepetersen (2012). Decomposition of variance for spatial Cox processes, *Scandinavian Journal of Statistics*, to appear.

Guan, Jalilian and Waagepetersen (2012). Optimal first order estimating functions for spatial point processes, submitted.

Thanks for your attention !