Statistical inference for inhomogeneous Cox processes

Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

June 25, 2006

Outline

Inhomogeneous clustered data sets

Estimating functions for inhomogeneous Cox processes

Maximum likelihood inference for thinned Cox processes

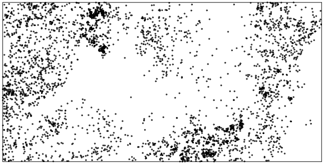
Inhomogeneous clustered data sets

Estimating functions for inhomogeneous Cox processes

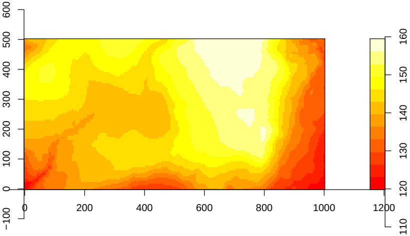
Maximum likelihood inference for thinned Cox processes

Tropical rain forests trees

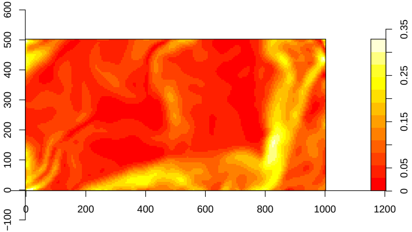
Beilschmiedia



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



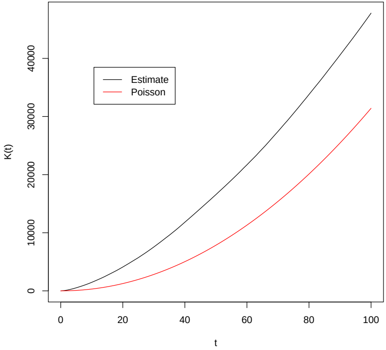
Altitude



Norm of altitude gradient (steepness)

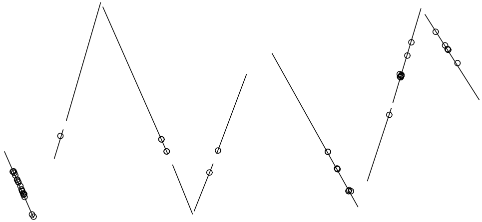
K-function (adjusted for inhomogeneity due to covariates)

Estimate of K adjusted for inhomogeneous intensity function $\rho(u)$:

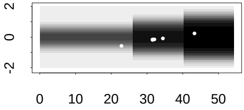


Poisson process not appropriate.

Whale positions



Close up:

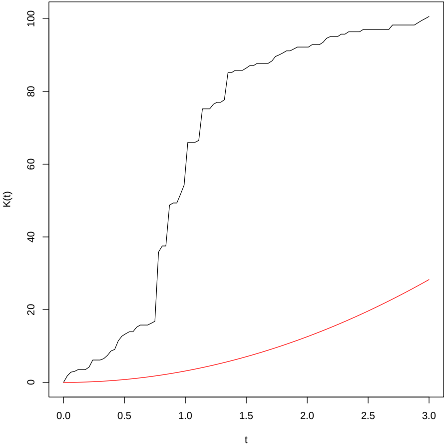


Observation window W = narrow strips around transect lines

Varying detection probability: inhomogeneity (thinning)

Variation in prey intensity: clustering

K-function (adjusted for inhomogeneity due to thinning)



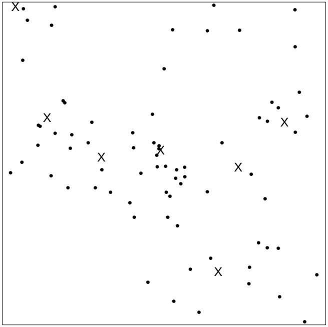
Poisson process not appropriate.

Inhomogeneous clustered data sets

Estimating functions for inhomogeneous Cox processes

Maximum likelihood inference for thinned Cox processes

Cluster process (Thomas process)



Mothers (crosses) Poisson point process $\mathbf{\Phi}$ with intensity $\kappa > 0$.

Offspring $\mathbf{X} = \cup_{c \in \mathbf{\Phi}} \mathbf{X}_c$ distributed around mothers c according to bivariate Gaussian density f.

ω: standard deviation of Gaussian density

α: mean of Poisson number of offspring for each mother.

Random intensity function:

$$\Lambda(u) = \alpha \sum_{c \in \mathbf{\Phi}} f(u - c; \omega)$$

Inhomogeneous Cox process

$z_{1:p}(u) = (z_1(u), \ldots, z_p(u))$ vector of p nonconstant covariates.

$\beta_{1:p} = (\beta_1, \ldots \beta_p)$ regression parameter.

Random intensity function:

$$\Lambda(u) = \alpha \exp(z(u)_{1:p}\beta_{1:p}^{\mathsf{T}}) \sum_{c \in \mathbf{\Phi}} f(u - c; \omega)$$

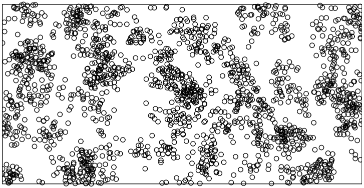
Rain forest example:

$$z_{1:2}(u) = (z_{\text{elev}}(u), z_{\text{grad}}(u))$$

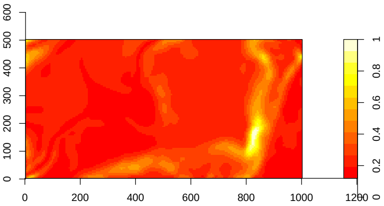
elevation/gradient covariate

Interpretation in terms of thinning

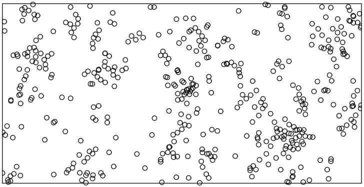
Homogeneous Cox process



Survival probabilities
$p(u) \propto \exp(z_{1:2}(u)\beta_{1:2}^{\mathsf{T}})$



After thinning (inhomogeneous Cox)



Parameter Estimation: regression parameters

Intensity function for inhomogeneous Cox:

$$\rho_\beta(u) = \kappa\alpha \exp(z(u)_{1:p}\beta_{1:p}^\mathsf{T}) = \exp(z(u)\beta^\mathsf{T})$$
$$z(u) = (1, z_{1:p}(u)) \quad \beta = (\log(\kappa\alpha), \beta_{1:p})$$

Parameter Estimation: regression parameters

Intensity function for inhomogeneous Cox:

$$\rho_\beta(u) = \kappa\alpha \exp(z(u)_{1:p}\beta_{1:p}^\mathsf{T}) = \exp(z(u)\beta^\mathsf{T})$$
$$z(u) = (1, z_{1:p}(u)) \quad \beta = (\log(\kappa\alpha), \beta_{1:p})$$

Consider indicators $N_i = \mathbf{1}[\mathbf{X} \cap C_i \neq \emptyset]$ of occurrence of points in disjoint C_i ($W = \cup C_i$) where $P(N_i = 1) \approx \rho_\beta(u_i)\mathrm{d}C_i$, $u_i \in C_i$.

Parameter Estimation: regression parameters

Intensity function for inhomogeneous Cox:

$$\rho_\beta(u) = \kappa\alpha \exp(z(u)_{1:p}\beta_{1:p}^\mathsf{T}) = \exp(z(u)\beta^\mathsf{T})$$
$$z(u) = (1, z_{1:p}(u)) \quad \beta = (\log(\kappa\alpha), \beta_{1:p})$$

Consider indicators $N_i = \mathbf{1}[\mathbf{X} \cap C_i \neq \emptyset]$ of occurrence of points in disjoint C_i ($W = \cup C_i$) where $P(N_i = 1) \approx \rho_\beta(u_i)\mathrm{d}C_i$, $u_i \in C_i$.

Composite likelihood

$$\prod_{i=1}^{n} (\rho_\beta(u_i)\mathrm{d}C_i)^{N_i}(1-\rho_\beta(u_i)\mathrm{d}C_i)^{1-N_i} \equiv \prod_{i=1}^{n} \rho_\beta(u_i)^{N_i}(1-\rho_\beta(u_i)\mathrm{d}C_i)^{1-N_i}$$

Limit ($\mathrm{d}C_i \to 0$) of log composite likelihood

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

Maximize using `spatstat` to obtain $\hat{\beta}$.

Asymptotic distribution of regression parameter estimates

Assume increasing mother intensity: $\kappa_n = n\tilde{\kappa} \to \infty$ and
$\mathbf{\Phi} = \cup_{i=1}^{n} \mathbf{\Phi}_i$, $\mathbf{\Phi}_i$ independent Poisson processes of intensity $\tilde{\kappa}$.

Score function asymptotically normal:

$$\frac{1}{\sqrt{n}} \frac{\mathrm{d}l(\beta)}{\mathrm{d}\log\alpha \mathrm{d}\beta_{1:p}} = \frac{1}{\sqrt{n}} \left(\sum_{u \in \mathbf{X} \cap W} z(u) - n\tilde{\kappa}\alpha \int_W z(u) \exp(z(u)_{1:p}\beta_{1:p}^{\mathsf{T}}) \mathrm{d}u \right)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[\sum_{c \in \mathbf{\Phi}_i} \sum_{u \in \mathbf{X}_c \cap W} z(u) - \tilde{\kappa}\alpha \int_W \exp(z_{1:p}(u)\beta_{1:p}^{\mathsf{T}}) \mathrm{d}u \Big] \approx N(0, V)$$

where $V = \mathbb{V}\mathrm{ar} \sum_{c \in \mathbf{\Phi}_i} \sum_{u \in \mathbf{X}_c \cap W} z(u)$

Asymptotic distribution of regression parameter estimates

Assume increasing mother intensity: $\kappa_n = n\tilde{\kappa} \to \infty$ and $\mathbf{\Phi} = \cup_{i=1}^{n} \mathbf{\Phi}_i$, $\mathbf{\Phi}_i$ independent Poisson processes of intensity $\tilde{\kappa}$.

Score function asymptotically normal:

$$\frac{1}{\sqrt{n}} \frac{\mathrm{d} l(\beta)}{\mathrm{d} \log \alpha \mathrm{d}\beta_{1:p}} = \frac{1}{\sqrt{n}} \left(\sum_{u \in \mathbf{X} \cap W} z(u) - n\tilde{\kappa}\alpha \int_W z(u) \exp(z(u)_{1:p} \beta_{1:p}^{\mathsf{T}}) \mathrm{d}u \right)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[\sum_{c \in \mathbf{\Phi}_i} \sum_{u \in \mathbf{X}_c \cap W} z(u) - \tilde{\kappa}\alpha \int_W \exp(z_{1:p}(u) \beta_{1:p}^{\mathsf{T}}) \mathrm{d}u \Big] \approx N(0, V)$$

where $V = \mathbb{V}\mathrm{ar} \sum_{c \in \mathbf{\Phi}_i} \sum_{u \in \mathbf{X}_c \cap W} z(u)$

By standard results for estimating functions (J observed information for Poisson likelihood):

$$\sqrt{\kappa_n} \big[(\log(\hat{\alpha}), \hat{\beta}_{1:p}) - (\log \alpha, \beta_{1:p}) \big] \approx N(0, J^{-1} V J^{-1})$$

Parameter Estimation: clustering parameters

Theoretical expression for (inhomogeneous) K-function:

$$K(t;\kappa,\omega) = \pi t^2 + \big(1 - \exp(-t^2/(2\omega)^2)\big)/\kappa.$$

Parameter Estimation: clustering parameters

Theoretical expression for (inhomogeneous) K-function:

$$K(t;\kappa,\omega) = \pi t^2 + \big(1 - \exp(-t^2/(2\omega)^2)\big)/\kappa.$$

Semi-parametric estimate

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

Parameter Estimation: clustering parameters

Theoretical expression for (inhomogeneous) K-function:

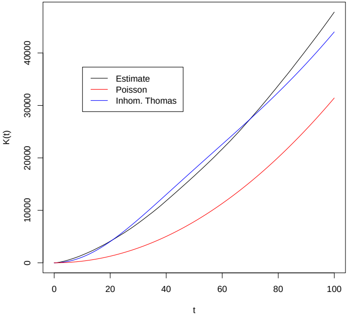
$$K(t;\kappa,\omega) = \pi t^2 + \left(1 - \exp(-t^2/(2\omega)^2)\right)/\kappa.$$

Semi-parametric estimate

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

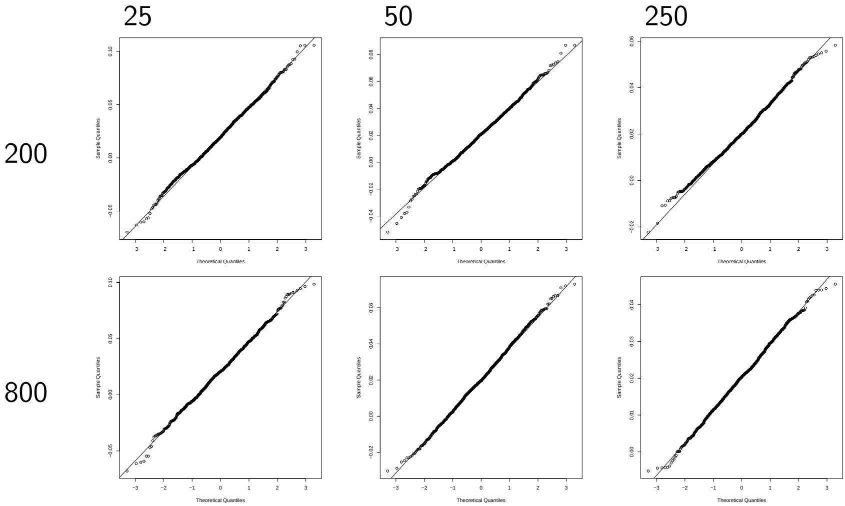
Estimate κ and ω by minimizing contrast

$$\int_0^{100} \left(K(t;\kappa,\omega)^{1/4} - \hat{K}(t)^{1/4}\right)^2 \mathrm{d}t$$



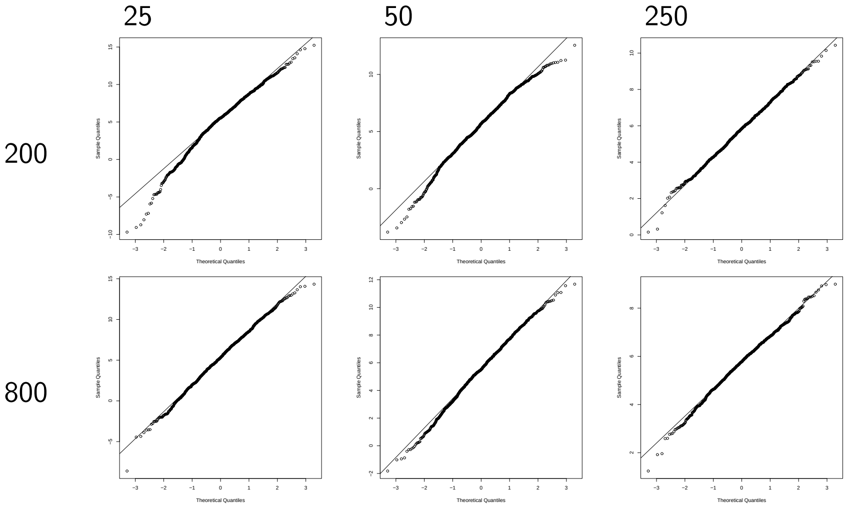
Simulation study

Quantile plots of $\hat{\beta}_{\text{elev}}$ (varying expected numbers 25, 50 and 250 of mothers and offspring, 200 or 800)



Simulation study II

Quantile plots of $\hat{\beta}_{\text{grad}}$ (varying expected numbers 25, 50 and 250 of mothers and offspring, 200 or 800)



Results for Beilschmiedia

Parameter estimates and confidence intervals (Poisson in red).

Elevation	Gradient	κ	α	ω
0.021 [-0.018,0.061]	5.842 [0.885,10.797]	8e-05	85.9	20.0
[0.017,0.026]	[5.340,6.342]			

Clustering: less information in data and wider confidence intervals than for Poisson process (independence).

Evidence of positive association between gradient and Beilschmiedia intensity.

Inhomogeneous clustered data sets

Estimating functions for inhomogeneous Cox processes

Maximum likelihood inference for thinned Cox processes

Shot-noise Cox process model for whales

Whales: stationary Cox process $\mathbf{Y}$ with random intensity function

$$\Lambda(u) = \sum_{(c,\gamma)\in\mathbf{\Phi}} \gamma k(u - c)$$

$\mathbf{\Phi}$ homogeneous marked Poisson process of marked cluster centres (c, γ) where $\gamma \sim \Gamma(\alpha, 1)$.

Shot-noise Cox process model for whales

Whales: stationary Cox process $\mathbf{Y}$ with random intensity function

$$\Lambda(u) = \sum_{(c,\gamma)\in\mathbf{\Phi}} \gamma k(u-c)$$

$\mathbf{\Phi}$ homogeneous marked Poisson process of marked cluster centres (c,γ) where $\gamma \sim \Gamma(\alpha, 1)$.

$p(u)$ detection probability of observing whale at location u.

Observed whales: $\mathbf{X}$ thinning of all whales $\mathbf{Y}$ i.e. inhomogeneous Cox process with random intensity function

$$p(u)\Lambda(u)$$

Note: $\mathbf{X}_{\neg\mathrm{obs}} = \mathbf{Y} \setminus \mathbf{X}$ and $\mathbf{X}$ independent Poisson processes given $\mathbf{\Phi}$.

Parameters

Assume $k(\cdot)$ bivariate Gaussian density truncated to have bounded support.

Parameters:

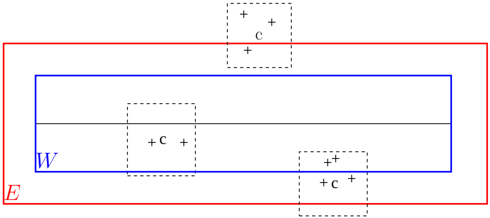
κ intensity of cluster centres c

$\alpha = \mathbb{E}\gamma$ (expected cluster size)

ω standard deviation of Gaussian density

Likelihood function for one transect

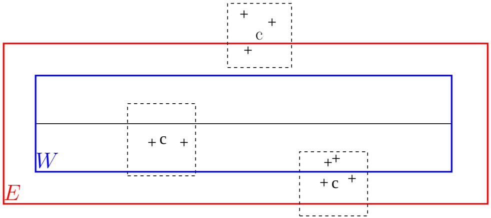
A transect:



W: support of $p(\cdot)$. E: $k(u-c)=0$ if $c \in \mathbb{R}^2 \setminus E$ and $u \in W$.

Likelihood function for one transect

A transect:



W: support of $p(\cdot)$. E: $k(u-c)=0$ if $c \in \mathbb{R}^2 \setminus E$ and $u \in W$.

Likelihood: $\theta = (\kappa, \alpha, \omega)$

1. $\mathbf{x}$ observed whales in W with conditional Poisson density

$$f(\mathbf{x}|\mathbf{\Phi};\omega) = \exp\Big(\int_W (1 - p(u)\Lambda(u))\mathrm{d}u\Big) \prod_{u \in \mathbf{x}} p(u)\Lambda(u)$$

2.

$$L(\theta) = \mathbb{E}_{(\kappa,\alpha)} f_\theta(\mathbf{x}|\mathbf{\Phi};\omega) = \mathbb{E}_{(\kappa,\alpha)} f(\mathbf{x}|\mathbf{\Phi} \cap E;\omega)$$

Derivatives of likelihood function

$\mathbf{\Phi}_E = \mathbf{\Phi} \cap E$ finite marked Poisson process with density

$$f(\phi; \kappa, \alpha) = \mathrm{e}^{|E|(1-\kappa)} \kappa^{n(\phi)} \prod_{(c,\gamma)\in\phi} \gamma^{\alpha-1} \exp(-\gamma)/\Gamma(\alpha)$$

Joint density of $\mathbf{X}$ and $\mathbf{\Phi}_E$:

$$f(\mathbf{x}, \phi; \kappa, \alpha, \omega) = f(\mathbf{x}|\phi; \omega) f(\phi; \kappa, \alpha)$$

Derivatives of likelihood function

$\mathbf{\Phi}_E = \mathbf{\Phi} \cap E$ finite marked Poisson process with density

$$f(\phi; \kappa, \alpha) = \mathrm{e}^{|E|(1-\kappa)} \kappa^{n(\phi)} \prod_{(c,\gamma)\in\phi} \gamma^{\alpha-1} \exp(-\gamma)/\Gamma(\alpha)$$

Joint density of $\mathbf{X}$ and $\mathbf{\Phi}_E$:

$$f(\mathbf{x}, \phi; \kappa, \alpha, \omega) = f(\mathbf{x}|\phi; \omega) f(\phi; \kappa, \alpha)$$

Let

$$V_\theta(\mathbf{X}, \mathbf{\Phi}_E) = \mathrm{d} \log f(\mathbf{X}, \mathbf{\Phi}_E; \theta)/\mathrm{d}\theta$$

Score function and observed information

$$u(\kappa, \alpha) = \frac{\mathrm{d} \log L(\theta)}{\mathrm{d}\theta} = \mathbb{E}_\theta[V_\theta(\mathbf{X}, \mathbf{\Phi}_E)|\mathbf{X} = \mathbf{x}] \quad \text{and}$$

$$j(\kappa, \alpha) = -\mathbb{E}_\theta[\frac{\mathrm{d} V_\theta(\mathbf{X}, \mathbf{\Phi}_E)}{\mathrm{d}\theta^\mathsf{T}}|\mathbf{X} = \mathbf{x}] - \mathbb{V}\mathrm{ar}_\theta[V_\theta(\mathbf{X}, \mathbf{\Phi}_E)|\mathbf{X} = \mathbf{x}]$$

Importance sampling

$\theta = (\kappa, \alpha, \omega)$

$\mathbf{\Phi}_0, \mathbf{\Phi}_1, \dots, \mathbf{\Phi}_{n-1}$ sample from $f(\phi|\mathbf{x};\theta_0) = f(\mathbf{x},\phi;\theta_0)/f(\mathbf{x};\theta_0)$ for fixed $\theta_0 = (\kappa_0, \alpha_0, \omega_0)$

$$\begin{aligned}
\mathbb{E}_\theta[k(\mathbf{\Phi})|\mathbf{X}=\mathbf{x}] &= \frac{f(\mathbf{x};\theta_0)}{f(\mathbf{x},\theta)}\mathbb{E}_{\theta_0}\left[k(\mathbf{\Phi})\frac{f(\mathbf{x},\mathbf{\Phi};\theta)}{f(\mathbf{x};\mathbf{\Phi};\theta_0)}|\mathbf{X}=\mathbf{x}\right] \\
&\approx \frac{f(\mathbf{x};\theta_0)}{f(\mathbf{x},\theta)}\frac{1}{n}\sum_{m=0}^{n-1} k(\mathbf{\Phi}_m)\frac{f(\mathbf{x},\mathbf{\Phi}_m;\theta)}{f(\mathbf{x};\mathbf{\Phi}_m;\theta_0)}
\end{aligned}$$

$$\frac{f(\mathbf{x};\theta)}{f(\mathbf{x},\theta_0)} = \frac{L(\theta)}{L(\theta_0)} \approx \frac{1}{n}\sum_{m=0}^{n-1}\frac{f(\mathbf{x},\mathbf{\Phi}_m;\theta)}{f(\mathbf{x};\mathbf{\Phi}_m;\theta_0)}$$

Hence Monte Carlo approximations of likelihood ratios, score, and observed information.

Markov chain Monte Carlo

Conditional density of $\boldsymbol{\Phi}_E$ given $\mathbf{X} = \mathbf{x}$:

$$f(\phi|\mathbf{x}) \propto f(\phi)f(\mathbf{x}|\phi) = f(\phi)\,\mathrm{e}^{-\int_W p(u)\Lambda(u|\phi)\mathrm{d}u} \prod_{u\in\mathbf{x}} p(u)\Lambda(u|\phi)$$

Computation of $\int_W p(u)\Lambda(u|\phi)\mathrm{d}u$ not straightforward.

Markov chain Monte Carlo

Conditional density of $\mathbf{\Phi}_E$ given $\mathbf{X} = \mathbf{x}$:

$$f(\phi|\mathbf{x}) \propto f(\phi)f(\mathbf{x}|\phi) = f(\phi)\,\mathrm{e}^{-\int_W p(u)\Lambda(u|\phi)\mathrm{d}u} \prod_{u\in\mathbf{x}} p(u)\Lambda(u|\phi)$$

Computation of $\int_W p(u)\Lambda(u|\phi)\mathrm{d}u$ not straightforward.

Demarginalisation impute $\mathbf{X}_{\neg\text{obs}} = (\mathbf{Y} \cap W) \setminus \mathbf{X}$:

Full conditional distributions for $(\mathbf{\Phi}, X_{\neg\text{obs}})$:

$$\mathbf{X}_{\neg\text{obs}}|\mathbf{\Phi}_E, \mathbf{X} : \quad \text{Poisson}\big((1 - p(\cdot))\Lambda(\cdot|\phi)\big)$$

$$\mathbf{\Phi}_E|\mathbf{X}_{\neg\text{obs}}, \mathbf{X} : \quad f(\phi|\mathbf{x}, \mathbf{x}_{\neg\text{obs}}) \propto f(\phi)\,\mathrm{e}^{-\int_W \Lambda(u|\phi)\mathrm{d}u} \prod_{u\in\mathbf{x}\cup\mathbf{x}_{\neg\text{obs}}} \Lambda(u|\phi)$$

Markov chain Monte Carlo

Conditional density of $\mathbf{\Phi}_E$ given $\mathbf{X} = \mathbf{x}$:

$$f(\phi|\mathbf{x}) \propto f(\phi)f(\mathbf{x}|\phi) = f(\phi)\,\mathrm{e}^{-\int_W p(u)\Lambda(u|\phi)\mathrm{d}u} \prod_{u\in\mathbf{x}} p(u)\Lambda(u|\phi)$$

Computation of $\int_W p(u)\Lambda(u|\phi)\mathrm{d}u$ not straightforward.

Demarginalisation impute $\mathbf{X}_{\neg\mathrm{obs}} = (\mathbf{Y} \cap W) \setminus \mathbf{X}$:

Full conditional distributions for $(\mathbf{\Phi}, X_{\neg\mathrm{obs}})$:

$$\mathbf{X}_{\neg\mathrm{obs}}|\mathbf{\Phi}_E, \mathbf{X} : \quad \mathrm{Poisson}\big((1 - p(\cdot))\Lambda(\cdot|\phi)\big)$$

$$\mathbf{\Phi}_E|\mathbf{X}_{\neg\mathrm{obs}}, \mathbf{X} : \quad f(\phi|\mathbf{x}, \mathbf{x}_{\neg\mathrm{obs}}) \propto f(\phi)\,\mathrm{e}^{-\int_W \Lambda(u|\phi)\mathrm{d}u} \prod_{u\in\mathbf{x}\cup\mathbf{x}_{\neg\mathrm{obs}}} \Lambda(u|\phi)$$

MCMC (Metropolis-within-Gibbs):

- $\mathbf{X}_{\neg\mathrm{obs}}|\mathbf{\Phi}_E, \mathbf{X}$: straightforward.
- $\mathbf{\Phi}_E|\mathbf{X}_{\neg\mathrm{obs}}, \mathbf{X}$: birth/death MCMC updates (Geyer & Møller 1994).

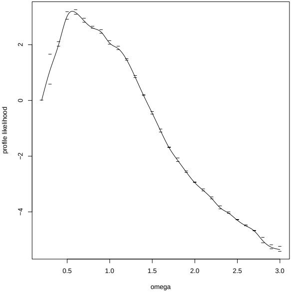
Maximization of likelihood

Likelihood based on all transects: multiply likelihoods for the different transects (approximately independent)

Maximize with respect to (κ, α) for finite set of ω values (Newton-Raphson)

Profile log likelihood function

$l_p(\omega) = \max_{\kappa,\alpha} \log L(\kappa, \alpha, \omega)$:

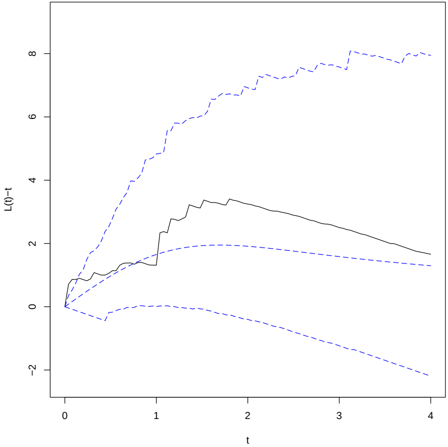


MLE: $\hat{\kappa} = 0.025$ $\hat{\alpha} = 2.4$ $\hat{\omega} = 0.6$.

95 % Confidence interval for whale intensity $\lambda = \kappa\alpha$: $[0.03; 0.08]$ (parametric bootstrap)

Model check using K-function

Plot based on $L(t) - t = \sqrt{K(t)/\pi} - t$



Summary

Estimating functions:

- computationally fast
- `R` packages available: `spatstat` and `InhomCluster`

Maximum likelihood estimation

- statistically more efficient
- long computations
- no standard software

References

Waagepetersen, R. (2006) An estimating function approach to inference for inhomogeneous Neyman-Scott processes, *Biometrics*, to appear.

Waagepetersen, R. and Schweder, T. (2006) Likelihood-based inference for clustered line transect data, *Journal of Agricultural, Biological, and Environmental Statistics*, to appear.

Møller, J. and Waagepetersen, R. (2003) *Statistical inference and simulation for spatial point processes*, Chapman & Hall/CRC Press.