An introduction to spatial Cox point processes with applications in plant and animal ecology

Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

September 19, 2006

Lectures:

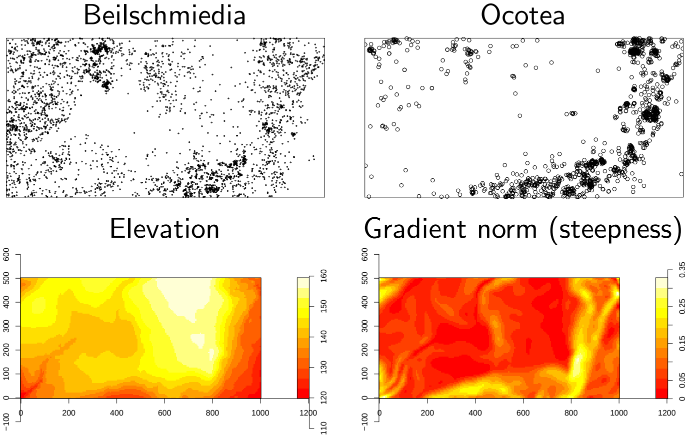
1. Intro to point processes, moment measures, and the Poisson process
2. Maximum likelihood estimation for a spatial Poisson process
3. Cox and cluster processes
4. Inference based on estimating equations
5. Bayesian inference
6. Maximum likelihood inference for whale data

'Recent' features:

- emphasis on inhomogeneous Cox processes depending on spatial covariates.
- likelihood-based inference using MCMC.

Tropical rain forest data (Barro Colorado Island)

Observation window: $S = [0, 1000] \times [0, 500]\text{m}^2$



Question: tree intensities related to elevation and gradient ?

Additional source of variation: clustering due to seed dispersal.

What is a spatial point process ?

Definitions:

1. a locally finite random subset $\mathbf{X}$ of $\mathbb{R}^2$ ($\#(\mathbf{X} \cap A)$ finite for all bounded subsets $A \subset \mathbb{R}^2$)
2. a random counting measure N on $\mathbb{R}^2$

Equivalent provided no multiple points: ($N(A) = \#(\mathbf{X} \cap A)$)

This course: appeal to 1. and skip measure-theoretic details.

In practice distribution specified by an explicit construction or in terms of a probability density.

Moments of a spatial point process

Fundamental characteristics of point process: mean and covariance of counts $N(A) = \#(\mathbf{X} \cap A)$.

Intensity measure μ:

$$\mu(A) = \mathbb{E}N(A), \quad A \subseteq \mathbb{R}^2$$

In practice often given in terms of *intensity function*

$$\mu(A) = \int_A \rho(u)\mathrm{d}u$$

Moments of a spatial point process

Fundamental characteristics of point process: mean and covariance of counts $N(A) = \#(\mathbf{X} \cap A)$.

Intensity measure μ:

$$\mu(A) = \mathbb{E}N(A), \quad A \subseteq \mathbb{R}^2$$

In practice often given in terms of *intensity function*

$$\mu(A) = \int_A \rho(u)\mathrm{d}u$$

Infinitesimal interpretation: $N(A)$ binary variable (presence or absence of point in A) when A very small. Hence

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

Second-order moments

Second order factorial moment measure:

$$\mu^{(2)}(A \times B) = \mathrm{E} \sum_{u,v \in \mathbf{X}}^{\neq} \mathbf{1}[u \in A,\ v \in B] \qquad A, B \subseteq \mathbb{R}^2$$

$$= \int_A \int_B \rho^{(2)}(u, v) \, \mathrm{d}u \, \mathrm{d}v$$

where $\rho^{(2)}(u, v)$ is the *second order product density*

Second-order moments

Second order factorial moment measure:

$$\mu^{(2)}(A \times B) = \mathrm{E} \sum_{u,v \in \mathbf{X}}^{\neq} \mathbf{1}[u \in A,\, v \in B] \qquad A, B \subseteq \mathbb{R}^2$$
$$= \int_A \int_B \rho^{(2)}(u,v)\, \mathrm{d}u\, \mathrm{d}v$$

where $\rho^{(2)}(u,v)$ is the *second order product density*

Campbell formula (by standard proof)

$$\mathbb{E} \sum_{u,v \in \mathbf{X}}^{\neq} h(u,v) = \iint h(u,v) \rho^{(2)}(u,v) \mathrm{d}u \mathrm{d}v$$

Covariance

$A, B \subseteq \mathbb{R}^2$:

$$\mathbb{E}[N(A)N(B)] = \mathbb{E} \sum_{u,v \in \mathbf{X}} \mathbf{1}[u \in A,\ v \in B] = \mu^{(2)}(A \times B) + \mu(A \cap B)$$

$$\mathbb{C}\mathrm{ov}(N(A), N(B)) = \mu^{(2)}(A \times B) + \mu(A \cap B) - \mu(A)\mu(B) \qquad (1)$$

Pair correlation function and K-function

Infinitesimal interpretation of $\rho^{(2)}$ ($u \in A$,$v \in B$):

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

Pair correlation function and K-function

Infinitesimal interpretation of $\rho^{(2)}$ ($u \in A$,$v \in B$):

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

Pair correlation: tendency to cluster or repel relative to case where points occur independently of each other

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

Pair correlation function and K-function

Infinitesimal interpretation of $\rho^{(2)}$ ($u \in A$,$v \in B$):

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

Pair correlation: tendency to cluster or repel relative to case where points occur independently of each other

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

Suppose $g(u,v) = g(u-v)$. K-function (cumulative quantity):

$$K(t) := \int_{\mathbb{R}^2} 1[\|u\| \leq t] g(u) \mathrm{d}u = \frac{1}{|B|} \mathbb{E} \sum_{\substack{u \in \mathbf{X} \cap B \\ v \in \mathbf{X}}}^{\neq} \frac{1[\|u-v\| \leq t]}{\rho(u)\rho(v)}$$

($\Rightarrow$ non-parametric estimation) ('$=$' by the Campbell formula)

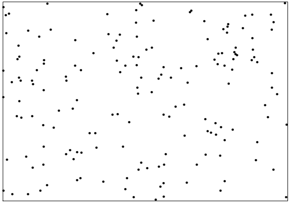
The Poisson process

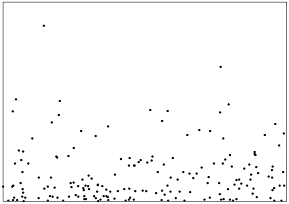
Assume μ locally finite measure on $\mathbb{R}^2$ with density ρ.

$\mathbf{X}$ is a Poisson process with intensity measure μ if for any bounded region B with $\mu(B) > 0$:

1. $N(B) \sim \text{Poisson}(\mu(B))$
2. Given $N(B)$, points in $\mathbf{X} \cap B$ i.i.d. with density $\propto \rho(u)$, $u \in B$

$B = [0, 1] \times [0, 0.7]$:





Homogeneous: $\rho = 150/0.7$ Inhomogeneous: $\rho(x, y) \propto \mathrm{e}^{-10.6y}$

Existence of Poisson process on $\mathbb{R}^2$: use definition on disjoint partitioning $\mathbb{R}^2 = \cup_{i=1}^{\infty} B_i$ of bounded sets B_i.

Existence of Poisson process on $\mathbb{R}^2$: use definition on disjoint partitioning $\mathbb{R}^2 = \cup_{i=1}^{\infty} B_i$ of bounded sets B_i.

Independent scattering:

- $A, B \subseteq \mathbb{R}^2$ disjoint $\Rightarrow$ $\mathbf{X} \cap A$ and $\mathbf{X} \cap B$ independent

Existence of Poisson process on $\mathbb{R}^2$: use definition on disjoint partitioning $\mathbb{R}^2 = \cup_{i=1}^{\infty} B_i$ of bounded sets B_i.

Independent scattering:

- $A, B \subseteq \mathbb{R}^2$ disjoint $\Rightarrow \mathbf{X} \cap A$ and $\mathbf{X} \cap B$ independent
- $\rho^{(2)}(u, v) = \rho(u)\rho(v)$ and $g(u, v) = 1$

Superpositioning and thinning

If $\mathbf{X}_1, \mathbf{X}_2, \ldots$ are independent Poisson processes (ρ_i), then *superposition* $\mathbf{X} = \cup_{i=1}^{\infty} \mathbf{X}_i$ is a Poisson process with intensity function $\rho = \sum_{i=1}^{\infty} \rho(u)$ (provided ρ integrable on bounded sets).

Superpositioning and thinning

If $\mathbf{X}_1, \mathbf{X}_2, \ldots$ are independent Poisson processes (ρ_i), then *superposition* $\mathbf{X} = \cup_{i=1}^{\infty} \mathbf{X}_i$ is a Poisson process with intensity function $\rho = \sum_{i=1}^{\infty} \rho(u)$ (provided ρ integrable on bounded sets).

Conversely: *Independent π-thinning* of Poisson process $\mathbf{X}$: independent retain each point u in $\mathbf{X}$ with probability $\pi(u)$. Thinned process $\mathbf{X}_{\text{thin}}$ and $\mathbf{X} \setminus \mathbf{X}_{\text{thin}}$ are independent Poisson processes with intensity functions $\pi(u)\rho(u)$ and $(1 - \pi(u))\rho(u)$.

(Superpositioning and thinning results most easily verified using void probability characterization of Poisson process, see M & W, 2003)

Superpositioning and thinning

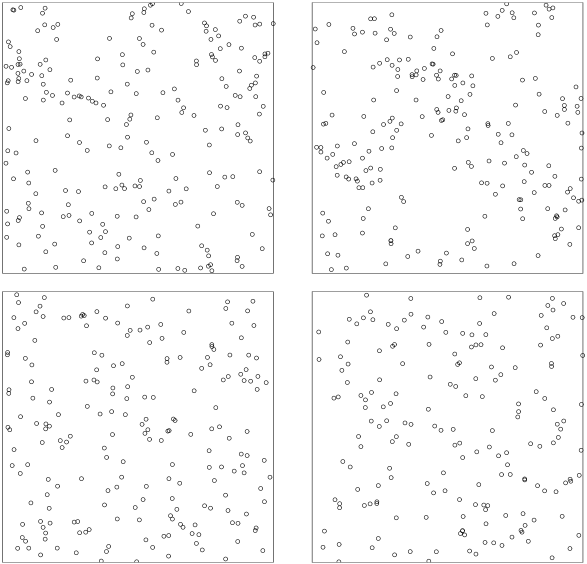
If $\mathbf{X}_1, \mathbf{X}_2, \ldots$ are independent Poisson processes (ρ_i), then *superposition* $\mathbf{X} = \cup_{i=1}^{\infty} \mathbf{X}_i$ is a Poisson process with intensity function $\rho = \sum_{i=1}^{\infty} \rho(u)$ (provided ρ integrable on bounded sets).

Conversely: *Independent π-thinning* of Poisson process $\mathbf{X}$: independent retain each point u in $\mathbf{X}$ with probability $\pi(u)$. Thinned process $\mathbf{X}_{\text{thin}}$ and $\mathbf{X} \setminus \mathbf{X}_{\text{thin}}$ are independent Poisson processes with intensity functions $\pi(u)\rho(u)$ and $(1 - \pi(u))\rho(u)$.

(Superpositioning and thinning results most easily verified using void probability characterization of Poisson process, see M & W, 2003)

For general point process $\mathbf{X}$: thinned process $\mathbf{X}_{\text{thin}}$ has product density $\pi(u)\pi(v)\rho^{(2)}(u, v)$ - hence g and K invariant under independent thinning.

Which one is not homogeneous Poisson?



Exercise

What is $K(t)$ for a Poisson process ?

For a stationary process with constant intensity function $\rho(u) = \kappa$, $\kappa K(t)$ may be interpreted as the expected number of points within distance less than t from a typical point - does this fit with the K-function for a Poisson process ?

Argue heuristically that a π-thinned process $\mathbf{X}_{\text{thin}}$ has intensity function $\pi(u)\rho(u)$ and second-order product density $\pi(u)\pi(v)\rho^{(2)}(u, v)$.

Check formula (1) for covariance of counts.

1. Intro to point processes, moment measures, and the Poisson process

2. Maximum likelihood estimation for a spatial Poisson process

3. Cox and cluster processes

4. Inference based on estimating equations

5. Bayesian inference

6. Maximum likelihood inference for whale data

Distribution and moments of Poisson process

$\mathbf{X}$ a Poisson process on S with $\mu(S) = \int_S \rho(u)\mathrm{d}u < \infty$ and F set of finite point configurations in S.

Distribution and moments of Poisson process

$\mathbf{X}$ a Poisson process on S with $\mu(S) = \int_S \rho(u)\mathrm{d}u < \infty$ and F set of finite point configurations in S.

By definition of a Poisson process

$$
\begin{aligned}
&P(\mathbf{X} \in F) \qquad (2)\\
&= \sum_{n=0}^{\infty} \frac{\mathrm{e}^{-\mu(S)}}{n!} \int_{S^n} 1[\{x_1, x_2, \ldots, x_n\} \in F] \prod_{i=1}^{n} \rho(x_i)\mathrm{d}x_1 \ldots \mathrm{d}x_n
\end{aligned}
$$

Distribution and moments of Poisson process

$\mathbf{X}$ a Poisson process on S with $\mu(S) = \int_S \rho(u)\mathrm{d}u < \infty$ and F set of finite point configurations in S.

By definition of a Poisson process

$$
\begin{aligned}
&P(\mathbf{X} \in F) \qquad (2) \\
&= \sum_{n=0}^{\infty} \frac{\mathrm{e}^{-\mu(S)}}{n!} \int_{S^n} 1[\{x_1, x_2, \dots, x_n\} \in F] \prod_{i=1}^{n} \rho(x_i)\mathrm{d}x_1 \dots \mathrm{d}x_n
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\mathbb{E}h(\mathbf{X}) \\
&= \sum_{n=0}^{\infty} \frac{\mathrm{e}^{-\mu(S)}}{n!} \int_{S^n} h(\{x_1, x_2, \dots, x_n\}) \prod_{i=1}^{n} \rho(x_i)\mathrm{d}x_1 \dots \mathrm{d}x_n
\end{aligned}
$$

Proof of independent scattering (finite case)

Consider bounded and disjoint $A, B \subseteq \mathbb{R}^2$.

$\mathbf{X} \cap (A \cup B)$ Poisson process.

Proof of independent scattering (finite case)

Consider bounded and disjoint $A, B \subseteq \mathbb{R}^2$.

$\mathbf{X} \cap (A \cup B)$ Poisson process. Hence

$$P(\mathbf{X} \cap A \in F, \mathbf{X} \cap B \in G) \quad (\mathbf{x} = \{x_1, \ldots, x_n\})$$
$$= \sum_{n=0}^{\infty} \frac{\mathrm{e}^{-\mu(A \cup B)}}{n!} \int_{(A \cup B)^n} 1[\mathbf{x} \cap A \in F, \mathbf{x} \cap B \in G] \prod_{i=1}^{n} \rho(x_i) \mathrm{d}x_1 \ldots \mathrm{d}x_n$$

Proof of independent scattering (finite case)

Consider bounded and disjoint $A, B \subseteq \mathbb{R}^2$.

$\mathbf{X} \cap (A \cup B)$ Poisson process. Hence

$$
\begin{aligned}
&P(\mathbf{X} \cap A \in F, \mathbf{X} \cap B \in G) \quad (\mathbf{x} = \{x_1, \ldots, x_n\}) \\
&= \sum_{n=0}^{\infty} \frac{\mathrm{e}^{-\mu(A \cup B)}}{n!} \int_{(A \cup B)^n} 1[\mathbf{x} \cap A \in F, \mathbf{x} \cap B \in G] \prod_{i=1}^{n} \rho(x_i) \mathrm{d}x_1 \ldots \mathrm{d}x_n \\
&= \sum_{n=0}^{\infty} \frac{\mathrm{e}^{-\mu(A \cup B)}}{n!} \frac{n!}{m!(n-m)!} \sum_{m=0}^{n} \int_{A^m} 1[\{x_1, x_2, \ldots, x_m\} \in F] \\
&\quad \int_{B^{n-m}} 1[\{x_{m+1}, \ldots, x_n\} \in G] \prod_{i=1}^{n} \rho(x_i) \mathrm{d}x_1 \ldots \mathrm{d}x_n
\end{aligned}
$$

Proof of independent scattering (finite case)

Consider bounded and disjoint $A, B \subseteq \mathbb{R}^2$.

$\mathbf{X} \cap (A \cup B)$ Poisson process. Hence

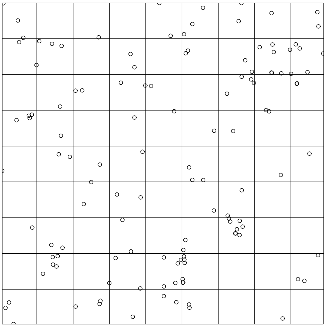
$$
\begin{aligned}
&P(\mathbf{X} \cap A \in F, \mathbf{X} \cap B \in G) \quad (\mathbf{x} = \{x_1, \dots, x_n\}) \\
=& \sum_{n=0}^{\infty} \frac{\mathrm{e}^{-\mu(A \cup B)}}{n!} \int_{(A \cup B)^n} 1[\mathbf{x} \cap A \in F, \mathbf{x} \cap B \in G] \prod_{i=1}^{n} \rho(x_i) \mathrm{d}x_1 \dots \mathrm{d}x_n \\
=& \sum_{n=0}^{\infty} \frac{\mathrm{e}^{-\mu(A \cup B)}}{n!} \frac{n!}{m!(n-m)!} \sum_{m=0}^{n} \int_{A^m} 1[\{x_1, x_2, \dots, x_m\} \in F] \\
& \int_{B^{n-m}} 1[\{x_{m+1}, \dots, x_n\} \in G] \prod_{i=1}^{n} \rho(x_i) \mathrm{d}x_1 \dots \mathrm{d}x_n \\
=& \text{ (interchange order of summation and sum over } m \text{ and } k = n - m) \\
& P(\mathbf{X} \cap A \in F) P(\mathbf{X} \cap B \in G)
\end{aligned}
$$

Density (likelihood) of a finite Poisson process (heuristic)

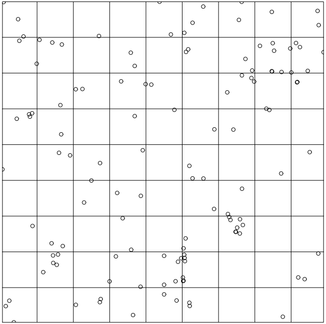
Consider indicators $N_i = \mathbf{1}[\mathbf{X} \cap C_i \neq \emptyset]$ of occurrence of points in disjoint C_i ($W = \cup C_i$) where

$$P(N_i = 1) \approx \rho(u_i)\mathrm{d}C_i, \quad u_i \in C_i$$

(quadrat count approach)



Density (likelihood) of a finite Poisson process (heuristic)



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

$$P(N_i = 1) \approx \rho(u_i)\mathrm{d}C_i, \quad u_i \in C_i$$

(quadrat count approach)

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

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

is

$$\prod_{u \in \mathbf{X} \cap W} \rho(u) \exp(-\int_W \rho(u)\,\mathrm{d}u)$$

Density (likelihood) of a finite Poisson process (formal)

$\mathbf{X}_1$ and $\mathbf{X}_2$ Poisson processes on S with intensity functions ρ_1 and ρ_2 where $\mu_1(S) = \int_S \rho_1(u)\mathrm{d}u < \infty$, $\mu_2(S) < \infty$, and $\rho_2(u) = 0 \Rightarrow \rho_1(u) = 0$, $u \in S$.

$$\begin{aligned}
&P(\mathbf{X}_1 \in F) \\
=&\sum_{n=0}^{\infty} \frac{\mathrm{e}^{-\mu_1(S)}}{n!} \int_{S^n} 1[\mathbf{x} \in F] \prod_{i=1}^{n} \rho_1(x_i)\mathrm{d}x_1 \dots \mathrm{d}x_n \quad (\mathbf{x} = \{x_1, \dots, x_n\}) \\
=&\sum_{n=0}^{\infty} \frac{\mathrm{e}^{-\mu_2(S)}}{n!} \int_{S^n} 1[\mathbf{x} \in F] \mathrm{e}^{\mu_2(S)-\mu_1(S)} \prod_{i=1}^{n} \frac{\rho_1(x_i)}{\rho_2(x_i)} \prod_{i=1}^{n} \rho_2(x_i)\mathrm{d}x_1 \dots \mathrm{d}x_n \\
=&\mathbb{E}\big(1[\mathbf{X}_2 \in F] f(\mathbf{X}_2)\big)
\end{aligned}$$

where

$$f(\mathbf{x}) = \mathrm{e}^{\mu_2(S)-\mu_1(S)} \prod_{i=1}^{n} \frac{\rho_1(x_i)}{\rho_2(x_i)}$$

Hence f is a density of $\mathbf{X}_1$ with respect to distribution of $\mathbf{X}_2$.

In particular (if S bounded): $\mathbf{X}_1$ has density

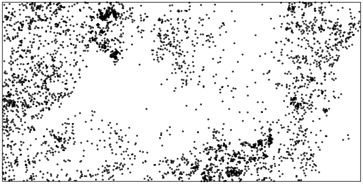
$$f(\mathbf{x}) = \exp\left(\int_S (1-\rho_1(u))\mathrm{d}u\right)\prod_{i=1}^{n}\rho_1(x_i)$$

with respect to unit rate Poisson process ($\rho_2 = 1$).

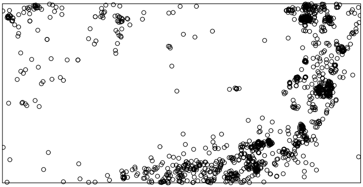
Data example: tropical rain forest trees

Observation window $W = [0, 1000] \times [0, 500]$

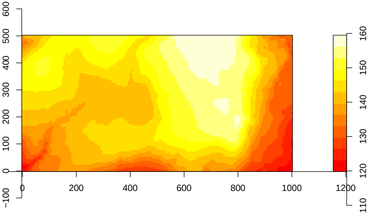
Beilschmiedia



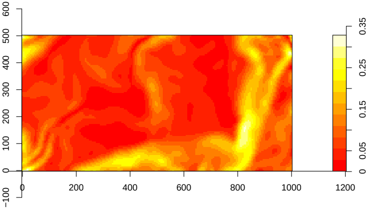
Ocotea



Elevation



Gradient norm (steepness)



Inhomogeneous Poisson process

Log linear intensity function

$$\rho(u;\beta) = \exp(z(u)\beta^{\mathsf{T}}), \quad z(u) = (1, z_{\text{elev}}(u), z_{\text{grad}}(u))$$

Inhomogeneous Poisson process

Log linear intensity function

$$\rho(u;\beta) = \exp(z(u)\beta^{\mathsf{T}}), \quad z(u) = (1, z_{\mathsf{elev}}(u), z_{\mathsf{grad}}(u))$$

Estimate β from Poisson log likelihood

$$\sum_{u \in \mathbf{X} \cap W} z(u)\beta^{\mathsf{T}} - \int_W \exp(z(u)\beta^{\mathsf{T}})\mathrm{d}u \quad (W = \text{ observation window})$$

Inhomogeneous Poisson process

Log linear intensity function

$$\rho(u;\beta) = \exp(z(u)\beta^{\mathsf{T}}), \quad z(u) = (1, z_{\text{elev}}(u), z_{\text{grad}}(u))$$

Estimate β from Poisson log likelihood

$$\sum_{u \in \mathbf{X} \cap W} z(u)\beta^{\mathsf{T}} - \int_W \exp(z(u)\beta^{\mathsf{T}})\mathrm{d}u \quad (W = \text{ observation window})$$

Model check using edge-corrected estimate of K-function

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

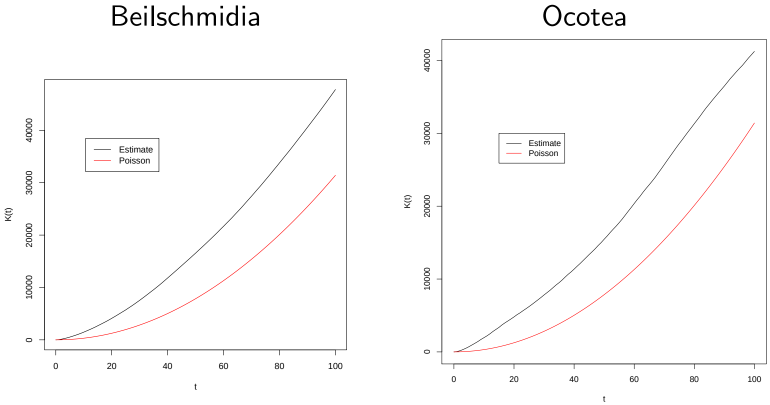
W_{u-v} translated version of W. $|A|$: area of $A \subset \mathbb{R}^2$.

Implementation in R

Using package spatstat (Baddeley & Turner):

```
> bei=ppp(beilpe$X,beilpe$Y,xrange=c(0,1000),yrange=c(0,500))
> beifit=ppm(bei,~elev+grad,covariates=list(elev=elevim,
  grad=gradim))
> coef(beifit) #parameter estimates
(Intercept)         elev         grad
-4.98958664  0.02139856  5.84202684
> asympcov=vcov(beifit) #Fisher information matrix
> sqrt(diag(asympcov) #standard errors
(Intercept)         elev         grad
0.017500262 0.002287773 0.255860860
> rho=predict.ppm(beifit)
> Kbei=Kinhom(bei,rho)
```

K-functions



Poisson process: $K(t) = \pi t^2$ (since $g = 1$) less than K functions for data. Hence Poisson process models not appropriate.

Exercise

Check that the Poisson expansion (2) indeed follows from the definition of a Poisson process.

Obtain by a heuristic argument the Poisson likelihood as a limit of likelihoods for independent Poisson counts.

Cox processes

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

Calculation of intensity and product density:

$$\rho(u) = \mathbb{E}\Lambda(u), \quad \rho^{(2)}(u, v) = \mathbb{E}[\Lambda(u)\Lambda(v)]$$

$$\mathbb{C}\mathrm{ov}(\Lambda(u), \Lambda(v)) > 0 \Leftrightarrow g(u, v) > 1 \quad \text{(clustering)}$$

Overdispersion for counts:

$$\mathbb{V}\mathrm{ar} N(A) = \mathbb{E}\mathbb{V}\mathrm{ar}[N(A) \mid \Lambda] + \mathbb{V}\mathrm{ar}\mathbb{E}[N(A) \mid \Lambda] = \mathbb{E}N(A) + \mathbb{V}\mathrm{ar}\mathbb{E}[N(A) \mid \Lambda]$$

Log Gaussian Cox process (LGCP)

- Poisson log linear model: $\log \rho(u) = z(u)\beta^{\mathsf{T}}$
- LGCP: in analogy with random effect models, take

$$\log \Lambda(u) = z(u)\beta^{\mathsf{T}} + U(u)$$

where $\mathbf{U} = (U(u))_{u \in \mathbb{R}^2}$ is a zero-mean Gaussian process

Log Gaussian Cox process (LGCP)

- Poisson log linear model: $\log \rho(u) = z(u)\beta^{\mathsf{T}}$
- LGCP: in analogy with random effect models, take

$$\log \Lambda(u) = z(u)\beta^{\mathsf{T}} + U(u)$$

where $\mathbf{U} = (U(u))_{u\in\mathbb{R}^2}$ is a zero-mean Gaussian process

- E.g. use power exponential covariance functions:

$$c(u, v) \equiv \mathbb{C}\mathrm{ov}[U(u), U(v)] = \sigma^2 \exp\left(-\|u - v\|^{\delta}/\alpha\right),$$

$\sigma, \alpha > 0,\ 0 \leq \delta \leq 2$

Log Gaussian Cox process (LGCP)

- Poisson log linear model: $\log \rho(u) = z(u)\beta^{\mathsf{T}}$
- LGCP: in analogy with random effect models, take

$$\log \Lambda(u) = z(u)\beta^{\mathsf{T}} + U(u)$$

where $\mathbf{U} = (U(u))_{u \in \mathbb{R}^2}$ is a zero-mean Gaussian process

- E.g. use power exponential covariance functions:

$$c(u, v) \equiv \mathbb{C}\mathrm{ov}[U(u), U(v)] = \sigma^2 \exp\left(-\|u - v\|^\delta / \alpha\right),$$

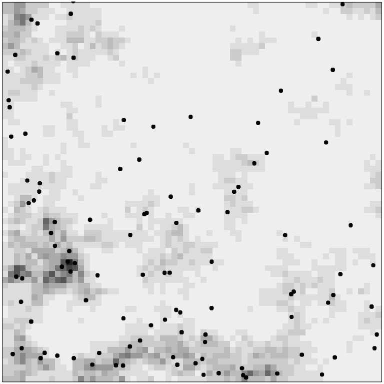
$\sigma, \alpha > 0,\ 0 \leq \delta \leq 2$

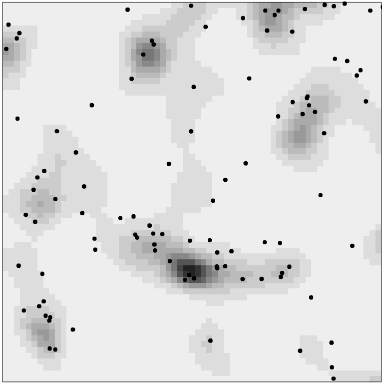
- Tractable product densities

$$\rho(u) = \mathbb{E}\Lambda(u) = \mathrm{e}^{z(u)\beta^{\mathsf{T}}} \mathbb{E}\mathrm{e}^{U(u)} = \exp\left(z(u)\beta^{\mathsf{T}} + c(u, u)/2\right)$$

$$g(u, v) = \frac{\mathbb{E}\left[\Lambda(u)\Lambda(v)\right]}{\rho(u)\rho(v)} = \ldots = \exp(c(u, v))$$

Two simulated homogeneous LGCP's


Exponential covariance function


Gaussian covariance function

Cluster processes

$\mathbf{M}$ 'mother' point process of cluster centres. Given $\mathbf{M}$, $\mathbf{X}_m$, $m \in \mathbf{M}$ are 'offspring' point processes (clusters) centered at m.

Intensity function for $\mathbf{X}_m$: $\alpha k(m,u)$ where k probability density and α expected size of cluster.

Cluster process:

$$\mathbf{X} = \cup_{m \in \mathbf{M}} \mathbf{X}_m$$

Intersection of cluster and Cox processes

By superpositioning: if cond. on $\mathbf{M}$, the $\mathbf{X}_m$ are independent Poisson processes, then $\mathbf{X}$ Cox process with random intensity function

$$\Lambda(u) = \alpha \sum_{m \in \mathbf{M}} k(m, u)$$

Intersection of cluster and Cox processes

By superpositioning: if cond. on $\mathbf{M}$, the $\mathbf{X}_m$ are independent Poisson processes, then $\mathbf{X}$ Cox process with random intensity function

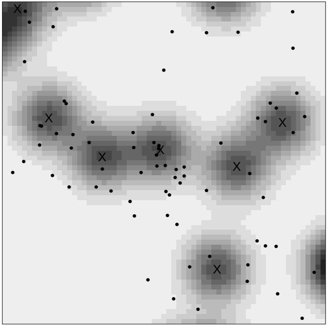
$$\Lambda(u) = \alpha \sum_{m \in \mathbf{M}} k(m, u)$$

Nice expressions for intensity and product density if $\mathbf{M}$ stationary Poisson on $\mathbb{R}^2$ with intensity function $\rho(\cdot) = \kappa$ and $k(m, u) = k(u - m)$(Campbell):

$$\mathbb{E}\Lambda(u) = \mathbb{E}\alpha \sum_{m \in \mathbf{M}} k(u - m) = \alpha \int k(u - m)\rho(m)\kappa \mathrm{d}m = \kappa\alpha$$

$$\begin{aligned}\mathbb{E}\Lambda(u)\Lambda(v) &= (\kappa\alpha)^2 + \mathbb{E}\alpha \sum_{m \in \mathbf{M}} k(u - m)k(v - m) \\ &= (\kappa\alpha)^2 + \kappa\alpha \int k(u - m)k(v - m)\mathrm{d}m\end{aligned}$$

Example: modified Thomas process



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

Clusters $\mathbf{X}_m, m \in \mathbf{M}$ Poisson processes with k = bivariate isotropic Gaussian density.

ω: standard deviation of Gaussian density

α: Expected number of offspring for each mother.

Cox process with random intensity function:

$$\Lambda(u) = \alpha \sum_{m \in \mathbf{M}} k(u - m; \omega)$$

Inhomogeneous Thomas 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:

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

$\mathbf{M}$: 'mother' Poisson point process of intensity κ.

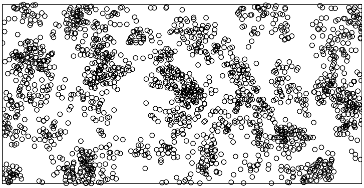
Rain forest example:

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

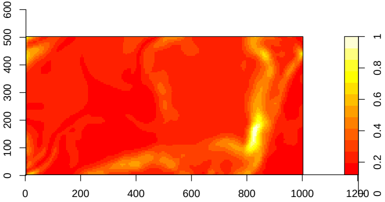
elevation/gradient covariate.

Interpretation in terms of thinning

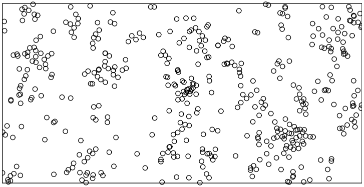
Homogeneous Cox process



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



After thinning (inhomogeneous Cox)



Density of a Cox process

- Restricted to a bounded region W, the density is

$$f(\mathbf{x}) = \mathbb{E}\left[\exp\left(|W| - \int_W \Lambda(u)\,\mathrm{d}u\right) \prod_{u \in \mathbf{X} \cap W} \Lambda(u)\right]$$

- Not on closed form
- Lecture 4 (tomorrow): inference based on estimating equations.
- Lecture 5 and 6: likelihood-based inference using MCMC.

1. Intro to point processes, moment measures, and the Poisson process

2. Maximum likelihood estimation for a spatial Poisson process

3. Cox and cluster processes

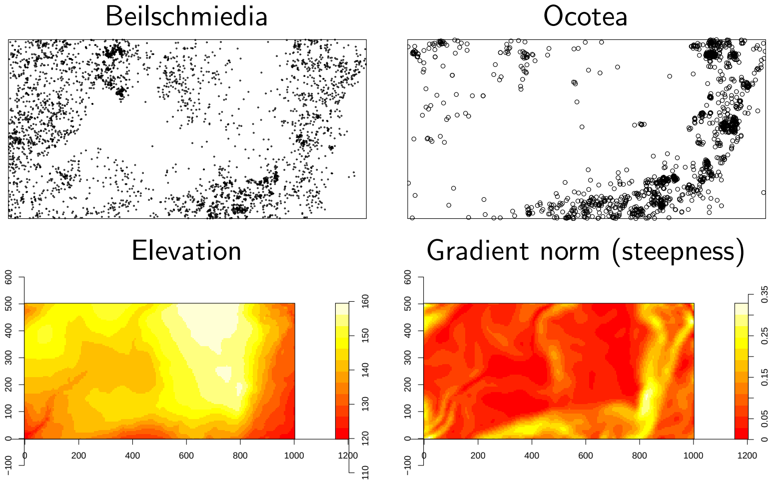
4. Inference based on estimating equations

5. Bayesian inference

6. Maximum likelihood inference for whale data

Data example: tropical rain forest trees

Observation window $W = [0, 1000] \times [0, 500]$



Sources of variation: elevation and gradient covariates *and* clustering/aggregation due to unobserved covariates and/or seed dispersal.

Summary of basic concepts

Intensity function

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

Summary of basic concepts

Intensity function

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

Second order product density

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

Summary of basic concepts

Intensity function

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

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 $g(u,v) = g(u-v)$)

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

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

Summary of basic concepts

Intensity function

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

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: $g(u-v) > 1$.

Density/likelihood for Poisson process

$$\exp(|W| - \int_W \rho(u)\,\mathrm{d}u) \prod_{u \in \mathbf{X} \cap W} \rho(u) \equiv \exp(-\int_W \rho(u)\,\mathrm{d}u) \prod_{u \in \mathbf{X} \cap W} \rho(u)$$

Inhomogeneous Thomas process

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

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

Random intensity function:

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

M Poisson mother point process of intensity κ.

Rain forest example:

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

elevation/gradient covariate.

Parameter Estimation: regression parameters

Intensity function for inhomogeneous Thomas:

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

$$z(u) = (1, z_{1:p}(u)) \quad \beta = (\log(\kappa\alpha), \beta_{1:p})$$

Parameter Estimation: regression parameters

Intensity function for inhomogeneous Thomas:

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

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 Thomas:

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

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 $P(N_i = 1) \approx \rho_\beta(u_i)\mathrm{d}C_i$, $u_i \in C_i$

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

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

is

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

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

Asymptotic distribution of regression parameter estimates

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

Score function asymptotically normal:

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

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

where $V = \mathbb{V}\mathrm{ar} \sum_{m \in \mathbf{M}_i} \sum_{u \in \mathbf{X}_m \cap W} z(u)$ ($\mathbf{X}_m$ offspring for mother m).

Asymptotic distribution of regression parameter estimates

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

Score function asymptotically normal:

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

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

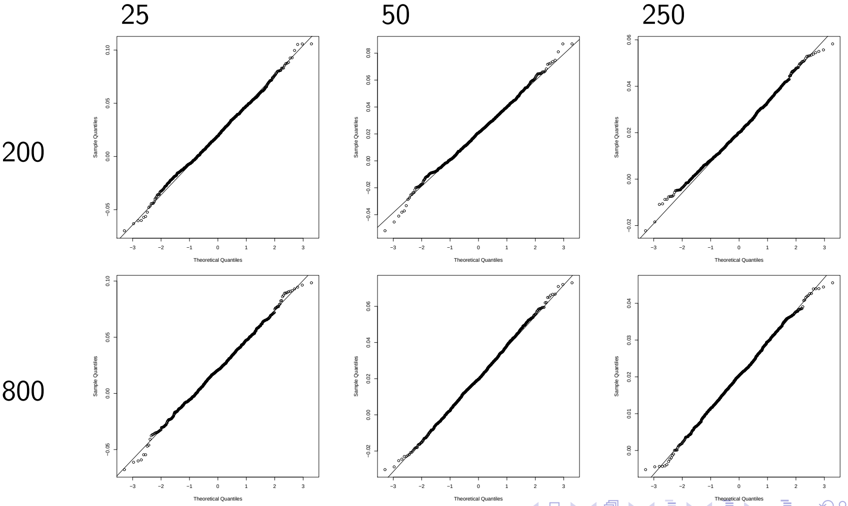
where $V = \mathbb{V}\mathrm{ar} \sum_{m \in \mathbf{M}_i} \sum_{u \in \mathbf{X}_m \cap W} z(u)$ ($\mathbf{X}_m$ offspring for mother m).

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

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

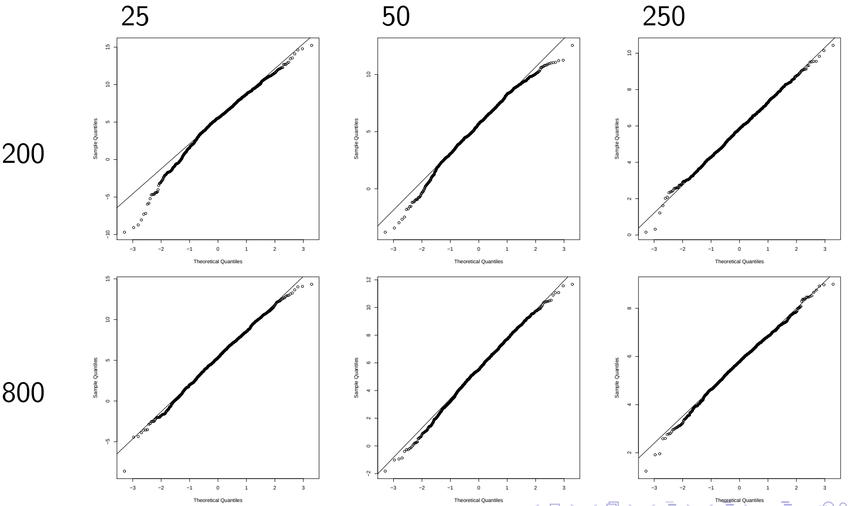
Simulation study

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)



Parameter Estimation: clustering parameters

Theoretical expression for (inhomogeneous) K-function:

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

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:

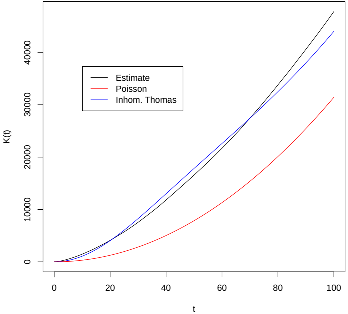
$$K(t;\kappa,\omega) = \pi t^2 + \big(1 - \exp(-t^2/(2\omega)^2)\big)/\kappa.$$

Semi-parametric estimate

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

Estimate κ and ω by minimizing contrast

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



Results for Beilschmiedia

Parameter estimates and confidence intervals (Poisson in red).

Elevation	Gradient	κ	α	ω
0.02 [-0.02,0.06]	5.84 [0.89,10.80]	8e-05	85.9	20.0
[0.02,0.03]	[5.34,6.34]			

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

Evidence of positive association between gradient and Beilschmiedia intensity.

Exercises

Check using the Campbell formula that the intensity function

$$\rho_\beta(u) = \mathbb{E}\Lambda(u) = \mathbb{E}\Big[\exp(z(u)_{1:p}\beta_{1:p}^{\mathsf{T}})\alpha \sum_{m \in \mathbf{M}} k(u - m; \omega)\Big]$$

for the inhomogeneous Thomas process is equal to
$\kappa\alpha \exp(z(u)_{1:p}\beta_{1:p}^{\mathsf{T}})$.

For a Cox process with random intensity function Λ, show that

$$\rho(u) = \mathbb{E}\Lambda(u), \quad \rho^{(2)}(u, v) = \mathbb{E}[\Lambda(u)\Lambda(v)]$$

Hint: compute first $\mathbb{E}N(A)$ and $\mathbb{E}\sum_{u,v \in \mathbf{X}}^{\neq} 1[u \in A, v \in B]$ using conditioning on Λ.

1. Intro to point processes, moment measures, and the Poisson process

2. Maximum likelihood estimation for a spatial Poisson process

3. Cox and cluster processes

4. Inference based on estimating equations

5. Bayesian inference

6. Maximum likelihood inference for whale data

Bayesian inference for LGCP

θ parameter vector with prior $p(\theta)$. $\mathbf{U}$ Gaussian field.

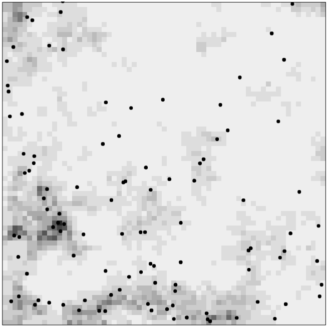
Likelihood $L(\theta) = \mathbb{E}_\theta f(\mathbf{x}|\mathbf{U}, \theta)$ not available in closed form.

Approximate Gaussian field $\mathbf{U}$ by finite Gaussian vector

$$\tilde{\mathbf{U}} = (U(u_1), U(u_2), \ldots, U(u_n))$$

Demarginalisation: consider joint posterior of $\tilde{\mathbf{U}}$ and θ:

$$p(\theta, \tilde{\mathbf{u}}|\mathbf{x}) \propto f(\mathbf{x}|\tilde{\mathbf{u}}, \theta) p(\tilde{\mathbf{u}}|\theta) p(\theta)$$



Conventional Markov chain Monte Carlo methods apply but specialized numerical methods (FFT) needed to cope with highdimensional Gaussian vector $\tilde{\mathbf{U}}$.

LGCP model for tropical rain forest data

$$\Lambda(u) = \exp\left(\beta_1 + \beta_2 z_2(u) + \beta_3 z_3(u) + U(u)\right)$$

where **U** is zero-mean Gaussian with

$$c(u, v) = \sigma^2 \exp(-\|u - v\|/\alpha), \quad \sigma > 0,\ \alpha > 0$$

Impose independent (improper) uniform priors for $\beta = (\beta_1, \beta_2, \beta_3)$, σ, α

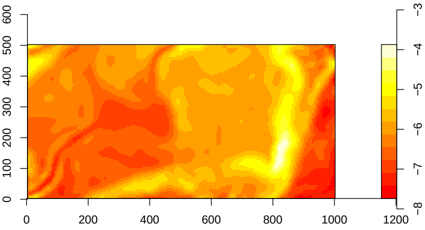
Posterior sim.: MCMC hybrid algorithm:

- Gaussian process is discretized to a 200×100 grid and updated by Langevin-Hastings
- $\beta = (\beta_1, \beta_2, \beta_3)$, σ, α updated by standard MCMC algorithms

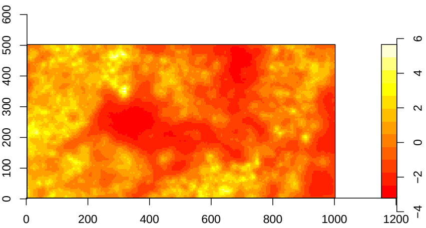
Posterior means, 95% central posterior intervals, etc.

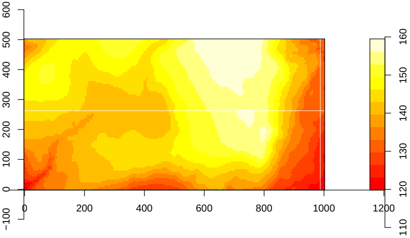
Posterior mean of systematic and random parts of Λ

$\mathbb{E}[\beta_1 + \beta_2 z_2(u) + \beta_3 z_3(u)|\mathbf{x}]$

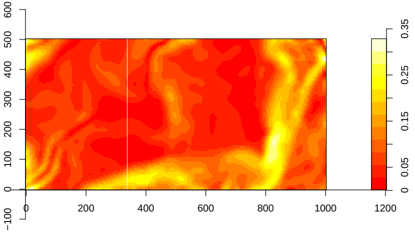


$\mathbb{E}[\tilde{\mathbf{U}}|\mathbf{x}]$





Altitude z_2



Norm of altitude gradient z_3

An example of posterior predictive model checking

Posterior predictive distribution: given the data $\mathbf{x}$ simulate

first $\theta = (\beta_1, \beta_2, \beta_3, \sigma, \alpha)$ and $\mathbf{U}$ from the posterior

and next $\mathbf{X}_{\text{sim}}$ from the Poisson process with intensity

$$\Lambda(u) = \exp\left(\beta_1 + \beta_2 z_2(u) + \beta_3 z_3(u) + U(u)\right)$$

An example of posterior predictive model checking

Posterior predictive distribution: given the data $\mathbf{x}$ simulate

first $\theta = (\beta_1, \beta_2, \beta_3, \sigma, \alpha)$ and $\mathbf{U}$ from the posterior

and next $\mathbf{X}_{\text{sim}}$ from the Poisson process with intensity

$$\Lambda(u) = \exp\left(\beta_1 + \beta_2 z_2(u) + \beta_3 z_3(u) + U(u)\right)$$

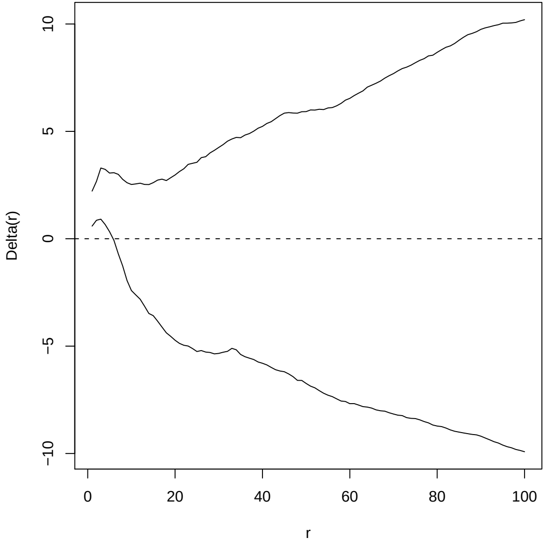
Consider

- $L(r; \mathbf{x}, \theta)$ = estimate of $L(r) = \sqrt{K(r)/\pi}$-function
- posterior predictive distribution of

$$\Delta(r) = L(r; \mathbf{x}, \theta) - L(r; \mathbf{X}_{\text{sim}}, \theta), \quad r > 0$$

(i.e. $\mathbf{X}_{\text{sim}}$ generated from posterior predictive distribution)

95% central posterior predictive envelopes of $\Delta(r)$



Bayesian inference for cluster process

Assume finite mother point process $\mathbf{M}$ with density $p(\mathbf{m}|\theta)$.

Consider joint posterior density of mothers $\mathbf{M}$ and θ

$$p(\theta, \mathbf{m}|\mathbf{x}) \propto f(\mathbf{x}|\mathbf{m}, \theta)p(\mathbf{m}|\theta)p(\theta)$$

Posterior of $\mathbf{M}$ finite point process with full conditional density

$$p(\mathbf{m}|\mathbf{x}, \theta) \propto f(\mathbf{x}|\mathbf{m}, \theta)p(\mathbf{m}|\theta)$$

Simulation using birth-death MCMC algorithm.

Use standard MCMC to update θ with full conditional

$$p(\theta|\mathbf{x}, \mathbf{m}) \propto f(\mathbf{x}|\mathbf{m}, \theta)p(\mathbf{m}|\theta)p(\theta)$$

MCMC birth-death simulation of a spatial point process

Birth-death Metropolis-Hastings algorithm for generating ergodic sample $\mathbf{X}^0, \mathbf{X}^1, \ldots$ from point process density p 'on S':

Suppose current state is $\mathbf{X}^i$, $i \geq 0$.

1. Either: with probability $1/2$
   - (birth) generate new point u uniformly on S and accept $\mathbf{X}^{\text{prop}} = \mathbf{X}^i \cup \{u\}$ with probability

$$\min\left\{1, \frac{p(\mathbf{X}^i \cup \{u\})|S|}{p(\mathbf{X}^i)(n+1)}\right\}$$

   or

MCMC birth-death simulation of a spatial point process

Birth-death Metropolis-Hastings algorithm for generating ergodic sample $\mathbf{X}^0, \mathbf{X}^1, \ldots$ from point process density p 'on S':

Suppose current state is $\mathbf{X}^i$, $i \geq 0$.

1. Either: with probability $1/2$
   - (birth) generate new point u uniformly on S and accept $\mathbf{X}^{\text{prop}} = \mathbf{X}^i \cup \{u\}$ with probability
   $$\min\left\{1, \frac{p(\mathbf{X}^i \cup \{u\})|S|}{p(\mathbf{X}^i)(n+1)}\right\}$$
   or
   - (death) select uniformly a point $u \in \mathbf{X}^i$ and accept $\mathbf{X}^{\text{prop}} = \mathbf{X}^i \setminus \{u\}$ with probability
   $$\min\left\{1, \frac{p(\mathbf{X}^i \setminus \{u\})n}{p(\mathbf{X}^i)|S|}\right\}$$
   (if $\mathbf{X}^i = \emptyset$ do nothing)

MCMC birth-death simulation of a spatial point process

Birth-death Metropolis-Hastings algorithm for generating ergodic sample $\mathbf{X}^0, \mathbf{X}^1, \ldots$ from point process density p 'on S':

Suppose current state is $\mathbf{X}^i$, $i \geq 0$.

1. Either: with probability $1/2$
   - (birth) generate new point u uniformly on S and accept $\mathbf{X}^{\text{prop}} = \mathbf{X}^i \cup \{u\}$ with probability
   $$\min\left\{1, \frac{p(\mathbf{X}^i \cup \{u\})|S|}{p(\mathbf{X}^i)(n+1)}\right\}$$
   or
   - (death) select uniformly a point $u \in \mathbf{X}^i$ and accept $\mathbf{X}^{\text{prop}} = \mathbf{X}^i \setminus \{u\}$ with probability
   $$\min\left\{1, \frac{p(\mathbf{X}^i \setminus \{u\})n}{p(\mathbf{X}^i)|S|}\right\}$$
   (if $\mathbf{X}^i = \emptyset$ do nothing)
2. if accept $\mathbf{X}^{i+1} = \mathbf{X}^{\text{prop}}$; otherwise $\mathbf{X}^{i+1} = \mathbf{X}^i$.

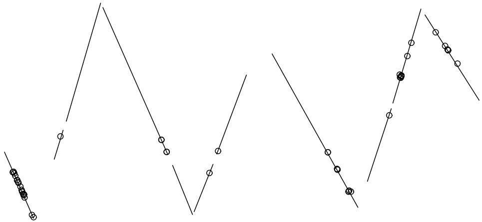
Initial state $\mathbf{X}_0$: arbitrary (e.g. empty or simulation from Poisson process).

Generated Markov chain $\mathbf{X}_0, \mathbf{X}_1, \ldots$ irreducible and aperiodic and hence ergodic: $\frac{1}{m}\sum_{i=0}^{m-1} k(\mathbf{X}^i) \to \mathbb{E}k(\mathbf{X}))$

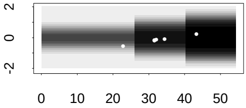
Moreover, usually geometrically ergodic and CLT:

$$\sqrt{m}\Big(\frac{1}{m}\sum_{i=0}^{m-1} k(\mathbf{X}^i) - \mathbb{E}k(\mathbf{X})\Big) \to N(0, \sigma_k^2)$$

Bayesian inference for whale positions



Close up:



Observation window W = narrow strips around transect lines

Varying detection probability: inhomogeneity (thinning)

Variation in prey intensity: clustering

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 $p(u)$ known.

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)

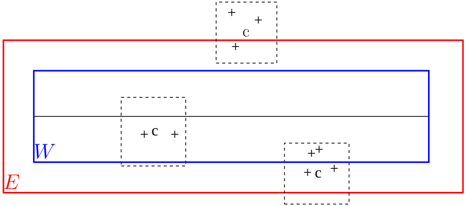
ω standard deviation of Gaussian density

Priors:

- Prior knowledge suggests that whales comes in groups of 1-3: $\alpha \sim N(2, 1)$
- uniform priors on bounded intervals for κ and ω.

Posterior based on one transect

A transect:

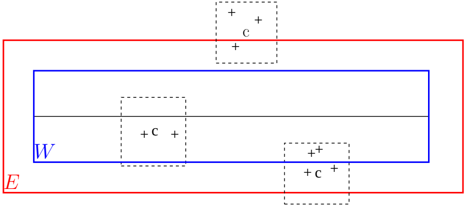


W: support of $p(\cdot)$.

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

Posterior based on one transect

A transect:



W: support of $p(\cdot)$.

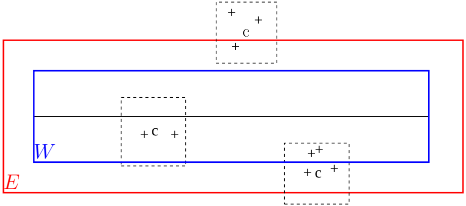
E: $k(u-c)=0$ if $c \in \mathbb{R}^2 \setminus E$ and $u \in W$.

Conditional Poisson density for $\mathbf{x}$ observed whales in W:

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

Posterior based on one transect

A transect:



W: support of $p(\cdot)$.

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

Conditional Poisson density for $\mathbf{x}$ observed whales in W:

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

NB: $f(\mathbf{x}|\mathbf{\Phi};\omega)$ depends only on finite point process $\mathbf{\Phi}_E = \mathbf{\Phi}\cap E$ with density

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

Posterior

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

Markov chain Monte Carlo for cluster centres

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

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

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

Markov chain Monte Carlo for cluster centres

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

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

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 $(\mathbf{\Phi}, X_{\neg\mathrm{obs}})$:

$$\begin{aligned}
\mathbf{X}_{\neg\mathrm{obs}}|\mathbf{\Phi}_E, \mathbf{X} : &\quad \mathrm{Poisson}\big((1 - p(\cdot))\Lambda(\cdot|\phi)\big) \\
\mathbf{\Phi}_E|\mathbf{X}_{\neg\mathrm{obs}}, \mathbf{X} : &\quad f(\phi|\mathbf{x}, \mathbf{x}_{\neg\mathrm{obs}}) \propto f(\phi)\,\mathrm{e}^{-\int_W \Lambda(u|\phi)\mathrm{d}u} \prod_{u \in \mathbf{x} \cup \mathbf{x}_{\neg\mathrm{obs}}} \Lambda(u|\phi)
\end{aligned}$$

Markov chain Monte Carlo for cluster centres

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

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

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

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

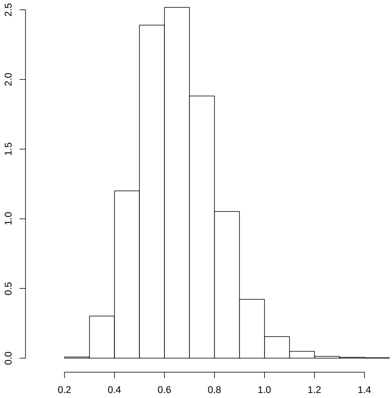
Full conditional distributions for $(\mathbf{\Phi}, X_{\neg\text{obs}})$:

$$\begin{aligned} \mathbf{X}_{\neg\text{obs}}|\mathbf{\Phi}_E, \mathbf{X}: &\quad \text{Poisson}\big((1 - p(\cdot))\Lambda(\cdot|\phi)\big) \\ \mathbf{\Phi}_E|\mathbf{X}_{\neg\text{obs}}, \mathbf{X}: &\quad f(\phi|\mathbf{x}, \mathbf{x}_{\neg\text{obs}}) \propto f(\phi)\,\mathrm{e}^{-\int_W \Lambda(u|\phi)\mathrm{d}u} \prod_{u \in \mathbf{x} \cup \mathbf{x}_{\neg\text{obs}}} \Lambda(u|\phi) \end{aligned}$$

MCMC (Metropolis-within-Gibbs):

- $\mathbf{X}_{\neg\text{obs}}|\mathbf{\Phi}_E, \mathbf{X}$: straightforward.
- $\mathbf{\Phi}_E|\mathbf{X}_{\neg\text{obs}}, \mathbf{X}$: birth/death MCMC updates.

Posterior distribution of ω



Posterior mean of κ and α: 0.0273 and 2.2

Posterior 95 % interval for whale intensity $\lambda = \kappa\alpha$: $[0.04; 0.08]$

Exercise

Check that the density of the finite mother point process $\mathbf{\Phi}_E = \mathbf{\Phi} \cap E$ is indeed

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

(hint: the density of $\Gamma(\alpha,\beta)$ is $\beta^{-\alpha}\gamma^{\alpha-1}\exp(-\gamma/\beta)/\Gamma(\alpha)$)

Why is it advantageous in the birth-death MCMC computations to have a kernel $k(\cdot)$ of bounded support ? (hint: write out explicitly the Metropolis-Hastings ratio for a birth or a death of a mother point).

1. Intro to point processes, moment measures, and the Poisson process

2. Maximum likelihood estimation for a spatial Poisson process

3. Cox and cluster processes

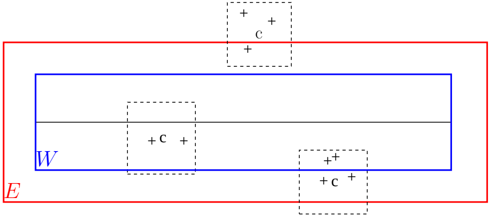
4. Inference based on estimating equations

5. Bayesian inference

6. Maximum likelihood inference for whale data

Likelihood function for one transect

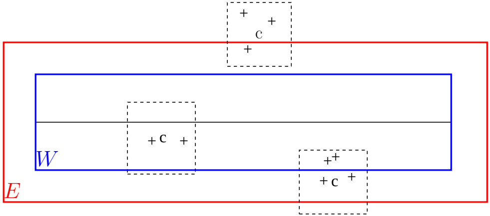
A transect:



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

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

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

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

Derivatives of likelihood function

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

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

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

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

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

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

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

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

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

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

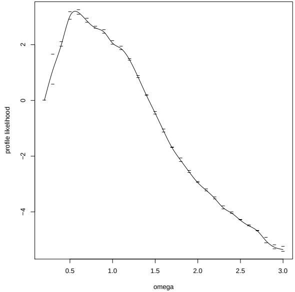
Maximization of likelihood

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

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

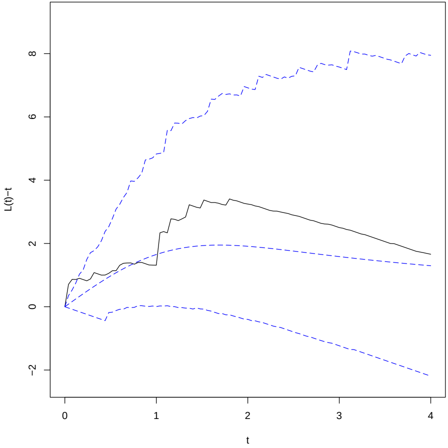


MLE: $\hat{\kappa} = 0.025$ $\hat{\alpha} = 2.4$ $\hat{\omega} = 0.6$.

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: inference for spatial Cox processes

Estimating functions:

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

Likelihood-based estimation

- statistically more efficient
- long computations
- Bayes computationally easier than MLE
- no standard software

Point patterns with repulsion

Markov point processes useful models for repulsion.

Example: Strauss process is specified by a density

$$f(\mathbf{x}) \propto \beta^{n(\mathbf{x})}\gamma^{s(\mathbf{x})}$$

with respect to unit rate Poisson process.

- $n(\mathbf{x})$: number of points
- $s(\mathbf{x})$: number of pairs of points $\{u, v\}$ with $\|u - v\| \leq R$
- only well-defined for $\gamma \leq 1$