Statistical models and methods for spatial point processes

Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

May 30, 2013

Lectures:

1. Intro to point processes and moment measures
2. The Poisson process
3. Cox and cluster processes
4. The conditional intensity and Markov point processes
5. Estimating equations
6. Likelihood-based inference and MCMC (if time allows)

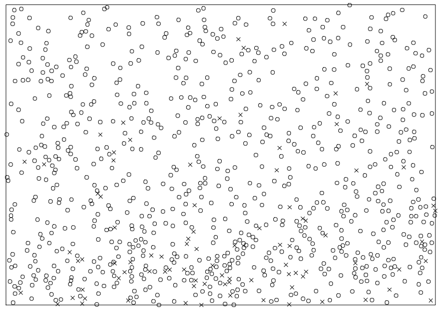
Aim: overview of

- spatial point process theory
- statistics for spatial point processes with emphasis on estimating equation inference
- not comprehensive: the most fundamental topics and my favorite things.
- all methods in Section 1-5 implemented in R package `spatstat`

1. Intro to point processes and moment measures
2. The Poisson process
3. Cox and cluster processes
4. The conditional intensity and Markov point processes
5. Estimating equations
6. Likelihood-based inference and MCMC (if time allows)

Mucous membrane cells

Centres of cells in mucous membrane:



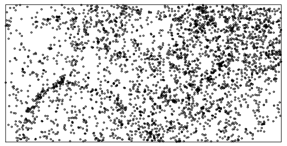
Repulsion due to physical extent of cells

Inhomogeneity - lower intensity in upper part

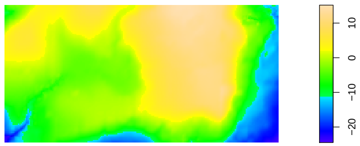
Bivariate - two types of cells

Same type of inhomogeneity for two types ?

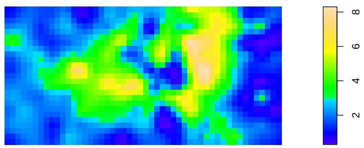
Data example: *Capparis Frondosa*



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



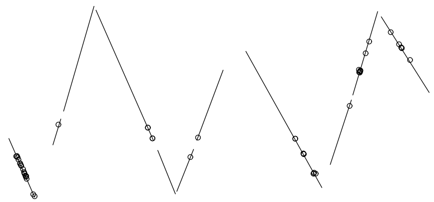
Elevation



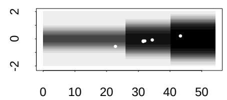
Potassium content in soil.

Objective: quantify dependence on environmental variables and clustering

Whale positions



Close up:



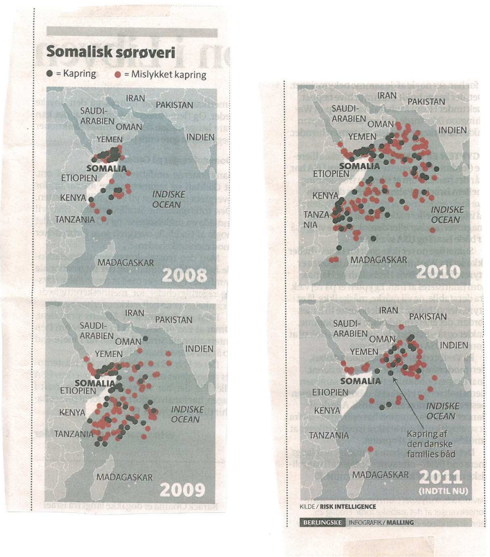
Aim: estimate whale intensity λ

Observation window W = narrow strips around transect lines

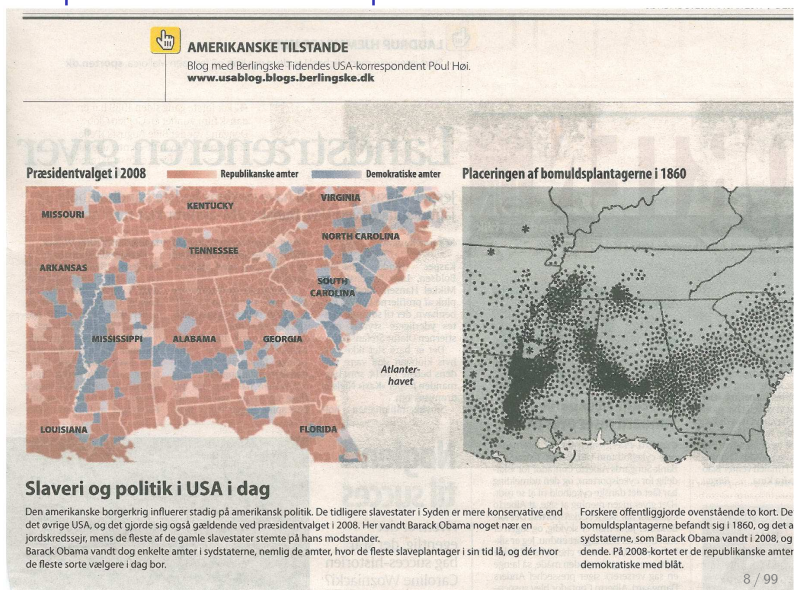
Varying detection probability: inhomogeneity (thinning)

Variation in prey intensity: clustering

Somalian pirates - two-type space-time



Cotton plantations in the Deep South



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. stochastic process of count variables $\{N(B)\}_{B \in \mathcal{B}_0}$ indexed by bounded Borel sets $\mathcal{B}_0$.
3. 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 (second and third lecture) or in terms of a probability density (fourth lecture).

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:

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

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

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

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

Second moment vs. second factorial moment measure

Second moment measure

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

Hence due to "diagonal terms" in sum not absolutely continuous.

Campbell formulae

By definition of intensity function and product density and the standard proof we obtain the useful Campbell formulae:

$$\mathbb{E}\sum_{u\in\mathbf{X}} h(u) = \int h(u)\rho(u)\mathrm{d}u$$

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

Pair correlation function

Pair correlation tendency to cluster/repel relative to case of independent points:

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

$= 1$ if independence (Poisson process, next section)

Covariance and pair correlation function

$$\mathbb{C}\mathrm{ov}[N(A),N(B)] = \int_{A\cap B}\rho(u)\mathrm{d}u + \int_A\int_B \rho(u)\rho(v)(g(u,v)-1)\mathrm{d}u\mathrm{d}v$$

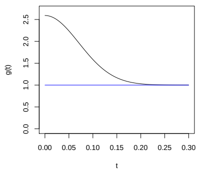
= Poisson variance + extra variance due to interaction

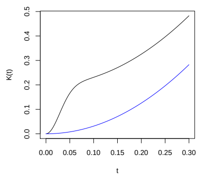
K-function

$$K(t) = \int_{\|h\|\leq t} g(h)\mathrm{d}h$$

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

Examples of pair correlation and K-functions:





Unbiased estimate of K-function (W observation window):

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

($e_{u,v}$ edge correction factor)

Exercises

1. Show that the covariance between counts $N(A)$ and $N(B)$ is
$$\mathbb{C}\mathrm{ov}[N(A), N(B)] = \mu(A \cap B) + \alpha^{(2)}(A \times B) - \mu(A)\mu(B)$$
2. Verify covariance formula on slide 16 (covariance in terms of pair correlation function).
3. Show that in the isotropic case ($g(u,v) = g(\|u-v\|)$), $K'(r) = 2\pi r g(r)$.
4. Show that
$$K(t) := \int_{\mathbb{R}^2} 1[\|u\| \le t] g(u) \mathrm{d}u = \frac{1}{|B|} \mathbb{E} \sum_{\substack{u \in \mathbf{X} \cap B \\ v \in \mathbf{X}}}^{\neq} \frac{1[\|u-v\| \le t]}{\rho(u)\rho(v)}$$
(Hint: use the Campbell formula)
5. Show that the following estimate is unbiased:
$$\hat{K}(t) = \sum_{u,v \in \mathbf{X} \cap W}^{\neq} \frac{1[\|u-v\| \le t]}{\rho(u)\rho(v)|W \cap W_{u-v}|}$$
where W_{u-v} translated version of W.

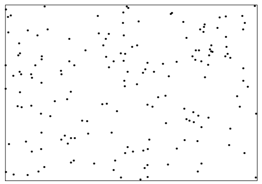
The Poisson process

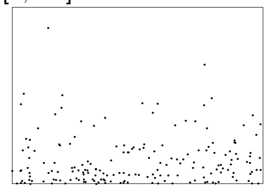
Assume μ locally finite measure on $\mathbb{R}^2$ with density ρ.

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

1. $N(B) \sim \mathrm{Poisson}(\mu(B))$
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.

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 $g(u,v) = 1$
- $\mathbb{C}\mathrm{ov}[N(A), N(B)] = \int_{A \cap B} \rho(u) \mathrm{d}u$

Characterization in terms of void probabilities

The distribution of $\mathbf{X}$ is uniquely determined by the void probabilities $P(\mathbf{X} \cap B = \emptyset)$, for bounded subsets $B \subseteq \mathbb{R}^2$.

Intuition: consider very fine subdivision of observation window – then at most one point in each cell and (joint) probabilities of absence/presence determined by void probabilities.

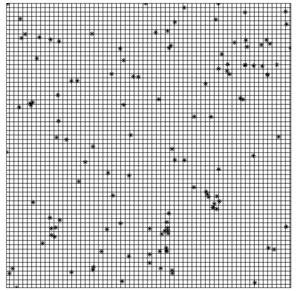
Hence, a point process $\mathbf{X}$ with intensity measure μ is a Poisson process if and only if

$$P(\mathbf{X} \cap B = \emptyset) = \exp(-\mu(B))$$

for any bounded subset B.

Homogeneous Poisson process as limit of Bernouilli trials

Consider disjoint subdivision $W = \cup_{i=1}^{n} C_i$ where $|C_i| = |W|/n$. With probability $\rho|C_i|$ a uniform point is placed in C_i.



Number of points in subset A is $b(n|A|/|W|, \rho|W|/n)$ which converges to a Poisson distribution with mean $\rho|A|$.

Hence, Poisson process default model when points occur independently of each other.

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.

Examples of F: all point configurations with total number of points in a given interval, point configurations where all pairs of points separated by distance δ,...

By definition of a Poisson process and law of total probability

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

Similarly,

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

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

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

$=$ (interchange order of summation and sum over m and $k = n - m$)
$P(\mathbf{X} \cap A \in F)P(\mathbf{X} \cap B \in G)$

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 $\rho = \sum_{i=1}^{\infty} \rho_i(u)$ (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.

Density (likelihood) of a finite Poisson process

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

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

where

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

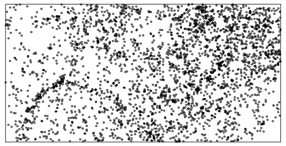
Hence f is a density of $\mathbf{X}_1$ with respect to distribution of $\mathbf{X}_2$.

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

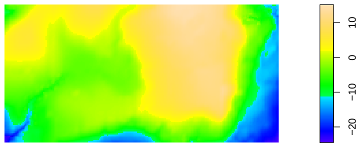
$$f(\mathbf{x}) = \mathrm{e}^{\int_S (1-\rho_1(u))\mathrm{d}u} \prod_{i=1}^{n} \rho_1(x_i)$$

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

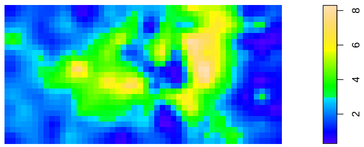
Back to the rain forest



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



Elevation



Potassium content in soil.

Objective: quantify dependence on environmental variables and clustering

Inhomogeneous Poisson process

Log linear intensity function

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

Estimate β from Poisson log likelihood (`spatstat`)

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

Model check using edge-corrected estimate of K-function

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

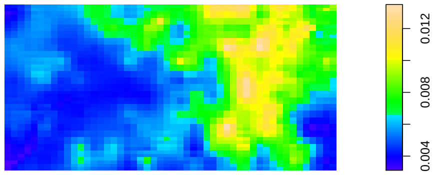
W_{u-v} translated version of W.

Capparis Frondosa and Poisson process ?

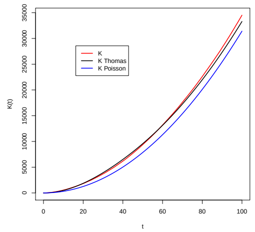
Fit model with covariates elevation, potassium,...

Fitted intensity function

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



Estimated K-function and $K(t) = \pi t^2$-function for Poisson process:



Not Poisson process - aggregation due to unobserved factors (e.g. seed dispersal)

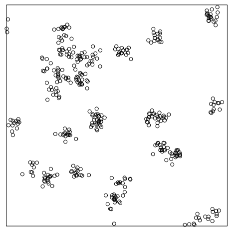
Exercises

1. What is $K(t)$ for a Poisson process ?
2. Check that the Poisson expansion (1) indeed follows from the definition of a Poisson process.
3. Compute the second order product density for a Poisson process $\mathbf{X}$.

   (Hint: compute second order factorial measure using the Poisson expansion for $\mathbf{X} \cap (A \cup B)$ for bounded $A, B \subseteq \mathbb{R}^2$.)
4. (if time) Assume that $\mathbf{X}$ has second order product density $\rho^{(2)}$ and show that g (and hence K) is invariant under independent thinning (note that a heuristic argument follows easily from the infinitesimal interpretation of $\rho^{(2)}$).

   (Hint: introduce random field $\mathbf{R} = \{R(u) : u \in \mathbb{R}^2\}$, of independent uniform random variables on $[0, 1]$, and independent of $\mathbf{X}$, and compute second order factorial measure for thinned process $\mathbf{X}_{\text{thin}} = \{u \in \mathbf{X} | R(u) \leq p(u)\}$.)

Cluster process: Inhomogeneous Thomas process



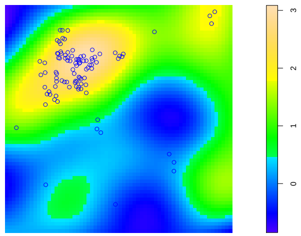
Parents stationary Poisson point process intensity κ

Poisson(α) number of offspring distributed around parents according to bivariate Gaussian density

Inhomogeneity: offspring survive according to probability

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

depending on covariates (independent thinning).



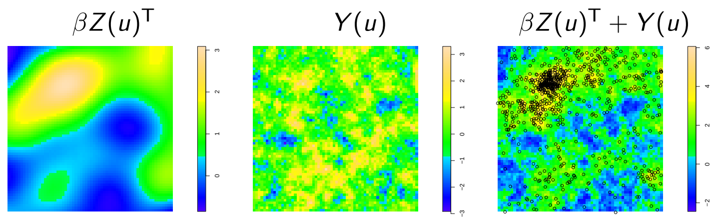
Cox processes

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 ("point process GLMM")

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

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



Z: systematic varition Y: random clustering around peaks in Y

Wide range of covariance models available for Y: exponential, Gaussian, Matérn,...(Tilmann's course)

Cox processes "bridge" between point processes and geostatistics.

Shot-noise Cox process

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

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:

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

Inhomogeneous Thomas: inhomogeneous shot-noise with Gaussian $k(\cdot)$ and $\gamma_v = \alpha > 0$.

Moments for Cox processes

Intensity function

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

Second-order product density

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

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

(overdispersion relative to a Poisson process)

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.

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

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

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

Specific models for $c_0(u-v) = \mathbb{C}\mathrm{ov}[\Lambda_0(u),\Lambda_0(v)]$

Log-Gaussian:

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

where Y Gaussian field.

Covariance (Laplace transform of normal distribution):

$$c_0(h) = \exp[\mathbb{C}\mathrm{ov}(Y(u),Y(u+h))] - 1$$

Shot-noise:

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

Covariance (convolution):

$$c_0(u-v) = \kappa\alpha^2\int_{\mathbb{R}^2}k(u)k(u+h)\mathrm{d}u$$

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

normal-variance mixture Cox/cluster processes

Suppose kernel $k(\cdot)$ given by variance-gamma density.

Y variance-gamma if $Y = \sqrt{W}U$ where $W \sim \Gamma$ and $U \sim N_p(0,I)$
$\Rightarrow$ closed under convolution.

Then Matérn covariance function:

$$c_0(h) = \sigma_0^2\frac{(\|h\|/\eta)^\nu K_\nu(\|h\|/\eta)}{2^{\nu-1}\Gamma(\nu)}$$

Suppose $k(\cdot)$ Cauchy density

$$k(u) = \frac{1}{2\pi\omega^2}[1+(\|u\|/\omega)^2)]^{-3/2}$$

(normal with inverse-gamma variance) then

$$c_0(r) = \sigma_0^2[1+(\|r\|/\eta)^2]^{-3/2}$$

Cauchy too ($\sigma_0^2 = \kappa\xi^2/(2\pi\eta)^2$ $\eta = 2\omega$)

Density of a Cox process

- Restricted to a bounded region W, the density is

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

- Not on closed form
- likelihood-based inference: MCMC or Laplace approximation (INLA for log Gaussian Cox processes)
- estimating equations based on closed form expressions for intensity and pair correlation

Exercises

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

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

2. Show that a cluster process with Poisson(α) number of iid offspring is a Cox process with random intensity function

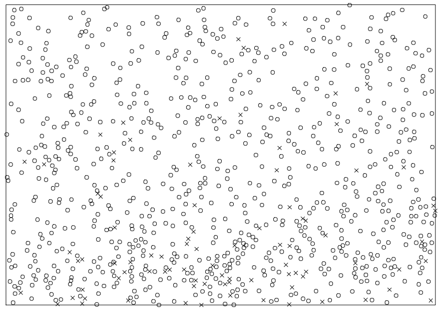
$$\Lambda(u) = \alpha \sum_{v \in \mathbf{C}} k(u - v)$$

(using notation from previous slide on cluster processes. Hint: use void probability characterisation and superposition result for Poisson process)

3. Compute the intensity and second-order product density for an inhomogeneous Thomas process. (Hint: interpret the Thomas process as a Cox process and use the Campbell formula)
4. Show that pair correlation for LCGP is $g(u,v) = \exp[\mathbb{C}\mathrm{ov}(Y(u), Y(v))]$

Mucous membrane cells

Centres of cells in mucous membrane:



Repulsion due to physical extent of cells

Inhomogeneity - lower intensity in upper part

Bivariate - two types of cells

Same type of inhomogeneity for two types ?

Density with respect to a Poisson process

$\mathbf{X}$ on bounded S has density f with respect to unit rate Poisson $\mathbf{Y}$ if

$$P(\mathbf{X} \in F) = \mathbb{E}(1[\mathbf{Y} \in F] f(\mathbf{Y}))$$
$$= \sum_{n=0}^{\infty} \frac{\mathrm{e}^{-|S|}}{n!} \int_{S^n} 1[\mathbf{x} \in F] f(\mathbf{x}) \mathrm{d}x_1 \ldots \mathrm{d}x_n \quad (\mathbf{x} = \{x_1, \ldots, x_n\})$$

Example: Strauss process

For a point configuration $\mathbf{x}$ on a bounded region S, let $n(\mathbf{x})$ and $s(\mathbf{x})$ denote the number of points and number of (unordered) pairs of R-close points ($R \geq 0$).

A *Strauss process* $\mathbf{X}$ on S has density

$$f(\mathbf{x}) = \frac{1}{c} \exp(\beta n(\mathbf{x}) + \psi s(\mathbf{x}))$$

with respect to a unit rate Poisson process $\mathbf{Y}$ on S and

$$c = \mathbb{E} \exp(\beta n(\mathbf{Y}) + \psi s(\mathbf{Y})) \qquad (2)$$

is the normalizing constant (unknown).

Note: only well-defined ($c < \infty$) if $\psi \leq 0$.

Intensity and conditional intensity

Suppose $\mathbf{X}$ has *hereditary* density f with respect to Y: $f(\mathbf{x}) > 0 \Rightarrow f(\mathbf{y}) > 0, \mathbf{y} \subset \mathbf{x}$.

Intensity function $\rho(u) = \mathbb{E} f(\mathbf{Y} \cup \{u\})$ usually unknown (except for Poisson and Cox/Cluster).

Instead consider *conditional intensity*

$$\lambda(u, \mathbf{x}) = \frac{f(\mathbf{x} \cup \{u\})}{f(\mathbf{x})}$$

(does not depend on normalizing constant !)

Note

$$\rho(u) = \mathbb{E} f(\mathbf{Y} \cup \{u\}) = \mathbb{E}\big[\lambda(u, \mathbf{Y}) f(\mathbf{Y})\big] = \mathbb{E}\lambda(u, \mathbf{X})$$

and

$\rho(u)\mathrm{d}A \approx P(\mathbf{X}$ has a point in $A) = \mathbb{E}P(\mathbf{X}$ has a point in $A | \mathbf{X} \setminus A), u \in A$

Hence, $\lambda(u, \mathbf{X})\mathrm{d}A$ probability that $\mathbf{X}$ has point in very small region A given $\mathbf{X}$ outside A.

Density and conditional intensity - factorization

One-to-one correspondence between density and conditional intensity (up to normalizing constant)

$$f(\{x_1, \ldots, x_n) = f(\emptyset) \prod_{i=1}^{n} \lambda(x_i, \{x_1, \ldots, x_{i-1})$$

Markov point processes

Def: suppose that f hereditary and $\lambda(u, \mathbf{x})$ only depends on $\mathbf{x}$ through $\mathbf{x} \cap b(u, R)$ for some $R > 0$ (*local Markov property*). Then f is *Markov* with respect to the R-close neighbourhood relation.

Thm (Hammersley-Clifford) The following are equivalent.

1. f is Markov.
2. $$f(\mathbf{x}) = \prod_{\mathbf{y} \subseteq \mathbf{x}} \phi(\mathbf{y})$$
   where $\phi(\mathbf{y}) = 1$ whenever $\|u - v\| \geq R$ for some $u, v \in \mathbf{y}$.

Pairwise interaction process: $\phi(\mathbf{y}) = 1$ whenever $n(\mathbf{y}) > 2$.

NB: in H-C, R-close neighbourhood relation can be replaced by an arbitrary symmetric relation between pairs of points.

Modelling the conditional intensity function

Suppose we specify a model for the conditional intensity. Two questions:

1. does there exist a density f with the specified conditional intensity ?
2. is f well-defined (integrable) ?

Solution:

1. find f by identifying interaction potentials (Hammersley-Clifford) or guess f.
2. sufficient condition (local stability): $\lambda(u, \mathbf{x}) \leq K$

NB some Markov point processes have interactions of any order in which case H-C theorem is less useful (e.g. area-interaction process).

Some examples

Strauss (pairwise interaction):

$$\lambda(u, \mathbf{x}) = \exp\left(\beta + \psi \sum_{v \in \mathbf{x}} 1[\|u - v\| \leq R]\right), \quad f(\mathbf{x}) = \frac{1}{c} \exp\left(\beta n(\mathbf{x}) + \psi s(\mathbf{x})\right)$$

Overlap process (pairwise interaction marked point process):

$$\lambda((u, m), \mathbf{x}) = \frac{1}{c} \exp\left(\beta + \psi \sum_{(u', m') \in \mathbf{x}} |b(u, m) \cap b(u', m')|\right) \quad (\psi \leq 0)$$

where $\mathbf{x} = \{(u_1, m_1), \ldots, (u_n, m_n)\}$ and $(u_i, m_i) \in \mathbb{R}^2 \times [a, b]$.

Area-interaction process:

$$f(\mathbf{x}) = \frac{1}{c} \exp\left(\beta n(\mathbf{x}) + \psi V(\mathbf{x})\right), \quad \lambda(u, \mathbf{x}) = \exp\left(\beta + \psi (V(\{u\} \cup \mathbf{x}) - V(\mathbf{x})\right)$$

$V(\mathbf{x}) = |\cup_{u \in \mathbf{x}} b(u, R/2)|$ is area of union of balls $b(u, R/2)$, $u \in \mathbf{x}$.

NB: $\phi(\cdot)$ complicated for area-interaction process.

The Georgii-Nguyen-Zessin formula ('Law of total probability')

$$\mathbb{E} \sum_{u \in \mathbf{X}} k(u, \mathbf{X} \setminus \{u\}) = \int_S \mathbb{E}[\lambda(u, \mathbf{X}) k(u, \mathbf{X})] \, \mathrm{d}u = \int_S \mathbb{E}^![k(u, \mathbf{X}) \mid u] \rho(u) \, \mathrm{d}u$$

$\mathbb{E}^![\cdot \mid u]$: expectation with respect to the conditional distribution of $\mathbf{X} \setminus \{u\}$ given $u \in \mathbf{X}$ (*reduced Palm distribution*)

Density of reduced Palm distribution:

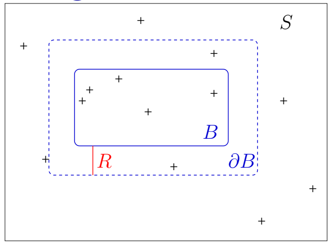
$$f(\mathbf{x} \mid u) = f(\mathbf{x} \cup \{u\}) / \rho(u)$$

NB: GNZ formula holds in general setting for point process on $\mathbb{R}^d$.

The spatial Markov property and edge correction

Let $B \subset S$ and assume $\mathbf{X}$ Markov with interaction radius R.

Define: ∂B points in $S \setminus B$ of distance less than R



Factorization (Hammersley-Clifford):

$$f(\mathbf{x}) = \prod_{\mathbf{y} \subseteq \mathbf{x} \cap (B \cup \partial B)} \phi(\mathbf{y}) \prod_{\substack{\mathbf{y} \subseteq x \setminus B: \\ \mathbf{y} \cap S \setminus (B \cup \partial B) \neq \emptyset}} \phi(\mathbf{y})$$

Hence, conditional density of $\mathbf{X} \cap B$ given $\mathbf{X} \setminus B$

$$f_B(\mathbf{z}|\mathbf{y}) \propto f(\mathbf{z} \cup \mathbf{y})$$

depends on $\mathbf{y}$ only through $\partial B \cap \mathbf{y}$.

Edge correction using the border method

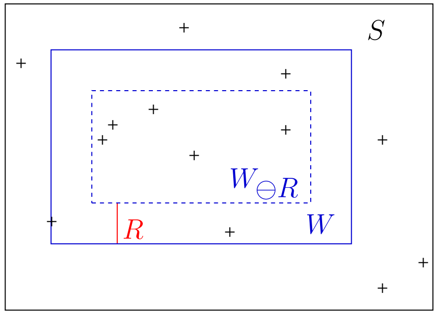
Suppose we observe $\mathbf{x}$ realization of $\mathbf{X} \cap W$ where $W \subset S$.

Problem: density (likelihood) $f_W(\mathbf{x}) = \mathbb{E} f(\mathbf{x} \cup Y_{S \setminus W})$ unknown.

Border method: base inference on

$$f_{W_{\ominus R}}(\mathbf{x} \cap W_{\ominus R} | \mathbf{x} \cap (W \setminus W_{\ominus R}))$$

i.e. conditional density of $\mathbf{X} \cap W_{\ominus R}$ given $\mathbf{X}$ outside $W_{\ominus R}$.



Exercises

1. Suppose that S contains a disc of radius $\epsilon \leq R/2$. Show that (2) is not finite, and hence the Strauss process not well-defined, when ψ is positive.

   (Hint: $\sum_{n=0}^{\infty} \frac{(\pi \epsilon^2)^n}{n!} \exp(n\beta + \psi n(n-1)/2) = \infty$ if $\psi > 0$.)
2. Show that local stability for a spatial point process density ensures integrability. Verify that the area-interaction process is locally stable.
3. what is the unnormalized density of a Strauss (β, ψ) with respect to a Poisson process of intensity $\exp(\beta)$?
4. Starting with the conditional intensity for a Strauss process, identify the potential function ϕ
5. (if time) Verify the Georgii-Nguyen-Zessin formula for a finite point process.

   (Hint: consider first the case of a finite Poisson-process $\mathbf{Y}$ in which case the identity is known as the Slivnyak-Mecke theorem, next apply $\mathbb{E} g(\mathbf{X}) = \mathbb{E}\big[g(\mathbf{Y}) f(\mathbf{Y})\big]$.)

Summary: Cox/cluster vs. Markov

	$\lambda(u\|\mathbf{X})$	$\rho(u)$	GNZ	Campbell	interaction
Markov	yes	no	yes	no	repulsive
Cox	no	yes	no	yes	clustering

Estimating function

Estimating function: $e(\theta)$ $[e(\theta, \mathbf{X})]$ function of θ and data $\mathbf{X}$.

Parameter estimate $\hat{\theta}$ solution of

$$e(\theta) = 0$$

$\hat{\theta}$ unbiased $\mathbb{E}\hat{\theta} = \theta^*$ if $e(\theta)$ unbiased $\mathbb{E}e(\theta^*) = 0$ (θ^* true value).

$$\mathbb{V}\mathrm{ar}\hat{\theta} = S^{-1}\Sigma S^{-1} \quad \Sigma = \mathbb{V}\mathrm{ar}e(\theta^*)$$

where sensitivity:

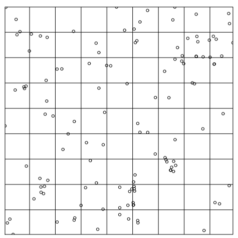
$$S = -\mathbb{E}[\frac{\mathrm{d}}{\mathrm{d}\theta}e(\theta)]$$

minus expected derivative of $e(\theta)$

How do we construct unbiased estimating functions involving $\mathbf{X}$ and θ ?

Composite and pseudo-likelihood

Disjoint subdivision $W = \cup_{i=1}^m C_i$ in 'cells' C_i.



$u_i \in C_i$ 'center' point.

Random indicator variables:

$$Y_i = 1[\mathbf{X} \text{ has a point in } C_i]$$

(presence/absence of points in C_i).

$$P(Y_i = 1) = |C_i|\rho_\theta(u_i) \text{ and } P(Y_i = 1|\mathbf{X} \setminus C_i) = |C_i|\lambda_\theta(u_i, \mathbf{X})$$

Idea: form composite likelihoods based on Y_i, e.g.

$$\prod_i P(Y_i = 1)^{Y_i}(1 - P(Y_i = 1))^{1-Y_i}$$

Consider limit when $|C_i| \to 0$.

Log composite likelihood (in fact log likelihood for Poisson):

$$\sum_{u \in \mathbf{X}} \log \rho_\theta(u) - \int_W \rho_\theta(u)\mathrm{d}u$$

Log pseudo-likelihood (Besag, 1977)

$$\sum_{u \in \mathbf{X}} \log \lambda_\theta(u, \mathbf{X} \setminus u) - \int_W \lambda_\theta(u, \mathbf{X})\mathrm{d}u$$

Scores:

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

and

$$\sum_{u \in \mathbf{X}} \frac{\lambda'_\theta(u, \mathbf{X} \setminus u)}{\lambda_\theta(u, \mathbf{X} \setminus u)} - \int_W \lambda'_\theta(u, \mathbf{X})\mathrm{d}u$$

unbiased estimating functions by Campbell/GNZ.

Issue:

- integrals

$$\int_W \rho'_\theta(u)\mathrm{d}u \text{ and } \int_W \lambda'_\theta(u,\mathbf{X})\mathrm{d}u$$

often not explicitly computable.

Numerical quadrature may introduce bias.

Monte Carlo approximation

Let **D** 'quadrature/dummy' point process of intensity κ and independent of **X**.

By GNZ

$$\mathbb{E}\int_W \lambda'(u,\mathbf{X})\mathrm{d}u = \mathbb{E}\sum_{u\in\mathbf{X}\cup\mathbf{D}} \frac{\lambda'(u,\mathbf{X}\setminus u)}{\lambda(u,\mathbf{X}\setminus u)+\kappa}$$

By Campbell

$$\int_W \rho'(u)\mathrm{d}u = \mathbb{E}\sum_{u\in\mathbf{X}\cup\mathbf{D}} \frac{\rho'(u)}{\rho(u)+\kappa}$$

Idea: replace integrals in pseudo- or composite likelihood with unbiased estimates using **D**.

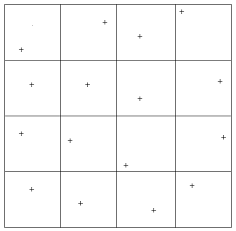
Dummy point process

Should be easy to simulate and mathematically tractable.

Possibilities:

1. Poisson process
2. binomial point process (fixed number of independent points)
3. stratified binomial point process

Stratified:



Approximate pseudo- and composite likelihood scores:

$$s(\theta) = \sum_{u\in\mathbf{X}} \frac{\lambda'_\theta(u,\mathbf{X}\setminus u)}{\lambda_\theta(u,\mathbf{X}\setminus u)} - \sum_{u\in(\mathbf{X}\cup\mathbf{D})} \frac{\lambda'_\theta(u,\mathbf{X}\setminus u)}{\lambda_\theta(u,\mathbf{X}\setminus u)+\kappa} \tag{3}$$

$$s(\theta) = \sum_{u\in\mathbf{X}} \frac{\rho'_\theta(u)}{\rho_\theta(u)} - \sum_{u\in(\mathbf{X}\cup\mathbf{D})} \frac{\rho'_\theta(u)}{\rho_\theta(u)+\kappa} \tag{4}$$

Note: of *logistic regression/case control* form with 'probabilities'

$$p(u|\mathbf{X}) = \frac{\lambda_\theta(u,\mathbf{X}\setminus u)}{\lambda_\theta(u,\mathbf{X}\setminus u)+\kappa}$$

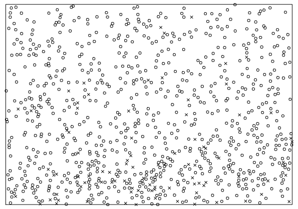
and

$$p(u) = \frac{\rho_\theta(u)}{\rho_\theta(u)+\kappa}$$

I.e. probabilities that $u\in\mathbf{X}$ given $u\in\mathbf{X}\cup\mathbf{D}$.

Hence computations straightforward with `glm()` software !

Example: mucous membrane



86 (type 1) + 807 (type 2) points.

1×0.7 observation window.

Marked point $u = (x, y, m)$ where $m = 1$ or 2 (two types of points).

Bivariate Strauss point process with

$$\lambda_\theta(u, \mathbf{X}) = \exp[q_{m,\theta}(y) + \psi n_R(u, \mathbf{X})]$$

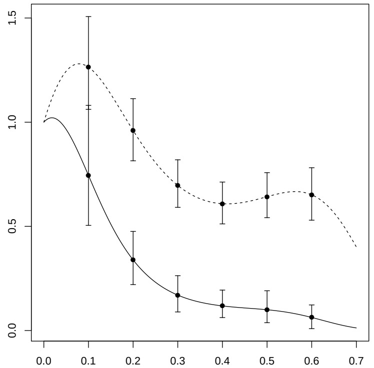
$q_{m,\theta}(y)$: polynomial in spatial y-coordinate.

$n_R(u, \mathbf{X})$: number of neighbors within range $R = 0.008$.

3600 stratified dummy points (random marks 1 or 2).

Fitted polynomials

Fitted polynomials (with confidence intervals for selected y values):



Polynomials significantly different according to logistic likelihood ratio test (parametric bootstrap).

Issue: $\mathbf{X}$ inhomogeneous

$$\lambda_m(u) = \exp[q_m(y)]\mathbb{E}\exp[\psi n_R(u, \mathbf{X})]$$

so intensity function not proportional to log polynomial function.

Baddeley and Nair (2012): approximation of intensity functions for Gibbs point processes

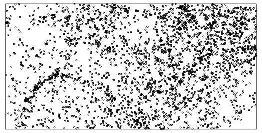
Decomposition of variance

	3600				**14400**		
	estm.	sd	sd_{pl}	inc. (%)	sd	sd_{pl}	inc. (%)
$q_1(0.1)$	6.004	0.195	0.189	3.608	0.191	0.189	0.812
$q_1(0.3)$	4.528	0.267	0.263	1.332	0.264	0.263	0.301
$q_1(0.5)$	3.994	0.406	0.404	0.555	0.404	0.404	0.146
$q_2(0.1)$	7.800	0.091	0.078	15.623	0.082	0.079	3.801
$q_2(0.3)$	7.204	0.083	0.075	10.923	0.076	0.075	2.589
$q_2(0.5)$	7.123	0.086	0.077	10.558	0.080	0.078	2.824
ψ	−2.594	0.344	0.341	0.971	0.342	0.341	0.197

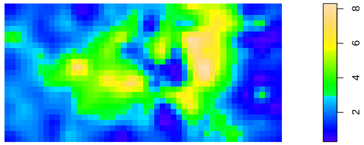
$sd_{pl} \approx$ standard deviation for pseudo-likelihood without approximation.

Example: rain forest trees

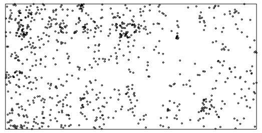
Capparis Frondosa



Potassium content in soil.



Loncocharpus Heptaphyllus



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

Clustered point patterns: Cox point process natural model.

Objective: infer regression model $\rho_\beta(u) = \exp[\beta Z(u)^{\mathsf{T}}]$

Composite likelihood targeted at estimating intensity function.

Problem: covariates sampled on (coarse) deterministic grid.

Plots shown: interpolated values of covariates.

Hence unbiased Monte Carlo approximation not applicable.

For now: integral in log composite likelihood

$$\sum_{u \in \mathbf{X}} \log \rho_\beta(u) - \int_W \rho_\beta(u) \mathrm{d}u$$

approximated using numerical quadrature based on interpolated values.

Need to convince biologists to use random sampling designs.

Another issue: optimality ?

Composite likelihood score

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

optimal for Poisson (likelihood).

Which f makes

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

optimal for Cox point process (positive dependence between points) ?

Optimal first-order estimating equation

Optimal choice of f: smallest variance

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

where

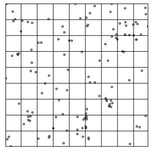
$$S_f = -\mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\beta^{\mathsf{T}}} e_f(\beta) \quad \Sigma_f = \mathbb{V}\mathrm{ar} e_f(\beta)$$

Possible to obtain optimal f as solution of certain Fredholm integral equation.

Numerical solution of integral equation leads to estimating function of quasi-likelihood type.

Quasi-likelihood

Integral equation approximated using Riemann sum dividing W into cells C_i with representative points u_i.



Resulting estimating function is *quasi-likelihood*

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

based on

$$Y=(Y_1,\ldots,Y_m), \quad Y_i=1[\mathbf{X} \text{ has point in } C_i].$$

μ mean of Y:

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

V covariance of Y (involves covariance of random intensity):

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

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.

Estimation of pair correlation function

Suppose parametric model $g(\cdot;\psi)$ for pair correlation.

Some options:

1. minimum contrast estimation based on K-function.
2. second-order composite likelihood: composite likelihood based on indicators for joint occurrence of points in pairs of cells:

$$X_{ij}=1[N_i>0 \text{ and } N_j>0]$$

$$\begin{aligned}P_{\beta,\psi}(X_{ij}=1)&=\rho^{(2)}(u,v;\beta,\psi)|C_i||C_j|\\&=\rho_\beta(u_i)\rho_\beta(v_j)g(u_i-u_j;\psi)|C_i||C_j|\end{aligned}$$

Minimum contrast estimation for ψ

Computationally easy alternative if $\mathbf{X}$ second-order reweighted stationary so that K-function well-defined.

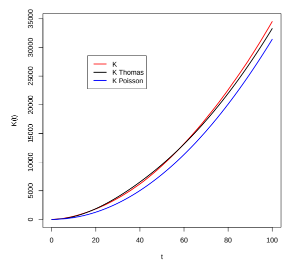
Estimate of K-function:

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

Unbiased if β 'true' regression parameter.

Minimum contrast estimation: minimize squared distance between theoretical K and $\hat{K}$:

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



Second-order composite likelihood

Second-order composite likelihood (given $\hat{\beta}$):

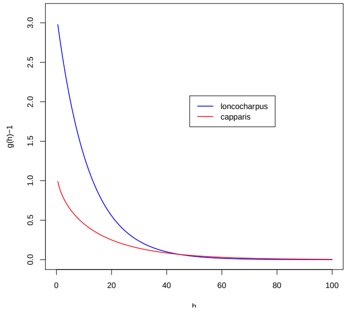
$$\mathrm{CL}_2(\psi|\hat{\beta}) = \prod^{\neq}_{\substack{u,v \in \mathbf{X} \cap W \\ \|u-v\| \leq R}} \rho^{(2)}(u,v;\hat{\beta},\psi) \times$$
$$\exp\left[-\iint_{\|u-v\|\leq R} \rho^{(2)}(u,v;\hat{\beta},\psi)\mathrm{d}u\mathrm{d}v\right]$$

NB: second-order reweighted stationarity (translation invariant pair correlation) not required.

Fitted pair correlation functions $g(\cdot)$ for Capparis and Loncocharpus

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 $\nu = 0.25$

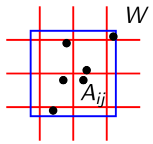
Two-step estimation

Obtain estimates $(\hat{\beta}, \hat{\psi})$ in two steps

1. obtain $\hat{\beta}$ using composite likelihood
2. obtain $\hat{\psi}$ using minimum contrast/second order composite likelihood

Asymptotic results - first order estimating function

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



Then

$$e_f(\beta) = \sum_{ij:A_{ij}\subseteq W} U_{ij}$$

where

$$U_{ij} = \sum_{u\in\mathbf{X}\cap A_{ij}} f_\beta(u) - \int_{A_{ij}} f_\beta(u)\rho_\beta(u)\mathrm{d}u$$

Assuming $\mathbf{X}$ is mixing, $\{U_{ij}\}_{ij}$ mixing random field and

$$|W|^{-1/2} e_f(\beta) \approx N(0, \Sigma_f)$$

(CLT for mixing random field $\{U_{ij}\}_{ij}$).

Asymptotic results cntd.

Estimate $\hat{\beta}$ solves

$$e_f(\beta) = 0$$

And (Taylor)

$$e_f(\beta) \approx |W|(\hat{\beta} - \beta)S_f \Leftrightarrow (\hat{\beta} - \beta) = |W|^{-1} e_f(\beta) S_f^{-1}$$

where

$$S_f = -\mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\beta^\mathsf{T}} e_f(\beta)/|W|$$

It follows that

$$\hat{\beta} \approx N(\beta, V_f/|W|)$$

where

$$V_f = S_f^{-1} \Sigma_f S_f^{-1}$$

Alternative: "infill"/increasing intensity-asymptotics

If $\mathbf{X}$ infinitely divisible (e.g. Poisson or Poisson-cluster) then $\mathbf{X} = \cup_{i=1}^n \mathbf{X}_n$ where $\mathbf{X}_i$ *iid* and intensity of $\mathbf{X}$ is $\rho_\beta(u) = n\tilde{\rho}(u;\beta)$ where $\tilde{\rho}_\beta$ intensity of $\mathbf{X}_i$

$$e_f(\beta) = \sum_{i=1}^n \Big[\sum_{u \in \mathbf{X}_i} f_\beta(u) - \int_W f_\beta(u)\tilde{\rho}(u;\beta)\mathrm{d}u \Big]$$

Ordinary CLT applies.

Exercises

1. Check using the GNZ formula, that the score of the pseudo-likelihood is an unbiased estimating function.
2. show that the approximate pseudo- and composite likelihood scores (3) and (4) are of logistic regression score form when the intensity or conditional intensity is log linear
3. Check that the derivative of minimum contrast criterion and the score of the second order composite likelihood function are unbiased estimating functions when β is equal to the true value.
4. Derive the second-order product density of a stratified binomial point process with one point in each cell.
5. How can you partition a Poisson-cluster process $\mathbf{X}$ into a union $\cup_{i=1}^n \mathbf{X}_i$ of *iid* Poisson-cluster processes ?

Maximum likelihood inference for point processes

Concentrate on point processes specified by unnormalized density $h_\theta(\mathbf{x})$,

$$f_\theta(\mathbf{x}) = \frac{1}{c(\theta)} h_\theta(\mathbf{x})$$

Problem: $c(\theta)$ in general unknown $\Rightarrow$ unknown log likelihood

$$l(\theta) = \log h_\theta(\mathbf{x}) - \log c(\theta)$$

Importance sampling

Importance sampling: θ_0 fixed reference parameter:

$$l(\theta) \equiv \log h_\theta(\mathbf{x}) - \log \frac{c(\theta)}{c(\theta_0)}$$

and

$$\frac{c(\theta)}{c(\theta_0)} = \mathbb{E}_{\theta_0} \frac{h_\theta(\mathbf{X})}{h_{\theta_0}(\mathbf{X})}$$

Hence

$$\frac{c(\theta)}{c(\theta_0)} \approx \frac{1}{m} \sum_{i=0}^{m-1} \frac{h_\theta(\mathbf{X}^i)}{h_{\theta_0}(\mathbf{X}^i)}$$

where $\mathbf{X}^0, \mathbf{X}^1, \ldots,$ sample from f_{θ_0} (later).

Exponential family case

$$h_\theta(\mathbf{x}) = exp(t(\mathbf{x})\theta^\mathsf{T})$$

$$l(\theta) = t(\mathbf{x})\theta^\mathsf{T} - \log c(\theta)$$

$$\frac{c(\theta)}{c(\theta_0)} = \mathbb{E}_{\theta_0} \exp(t(\mathbf{X})(\theta - \theta_0)^\mathsf{T})$$

Caveat: unless $\theta - \theta_0$ 'small', $\exp(t(\mathbf{X})(\theta - \theta_0)^\mathsf{T})$ has very large variance in many cases (e.g. Strauss).

Path sampling (exp. family case)

Derivative of cumulant transform:

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \log \frac{c(\theta)}{c(\theta_0)} = \mathbb{E}_\theta t(\mathbf{X})$$

Hence, by integrating over differentiable path $\theta(t)$ (e.g. line) linking θ_0 and θ_1:

$$\log \frac{c(\theta_1)}{c(\theta_0)} = \int_0^1 \mathrm{E}_{\theta(s)}[t(\mathbf{X})] \frac{\mathrm{d}\theta(s)^\mathsf{T}}{\mathrm{d}s} \mathrm{d}s$$

Approximate $E_{\theta(s)} t(\mathbf{X})$ by Monte Carlo and $\int_0^1$ by numerical quadrature (e.g. trapezoidal rule).

NB Monte Carlo approximation on log scale more stable.

Maximisation of likelihood (exp. family case)

Score and observed information:

$$u(\theta) = t(\mathbf{x}) - \mathrm{E}_\theta t(\mathbf{X}), \quad j(\theta) = \mathrm{Var}_\theta t(\mathbf{X}),$$

Newton-Rahpson iterations:

$$\theta^{m+1} = \theta^m + u(\theta^m) j(\theta^m)^{-1}$$

Monte Carlo approximation of score and observed information: use importance sampling formula

$$\mathrm{E}_\theta k(\mathbf{X}) = \mathrm{E}_{\theta_0}\left[k(\mathbf{X}) \exp\left(t(\mathbf{X})(\theta - \theta_0)^\mathsf{T}\right)\right] / (c_\theta / c_{\theta_0})$$

with $k(\mathbf{X})$ given by $t(\mathbf{X})$ or $t(\mathbf{X})^\mathsf{T} t(\mathbf{X})$.

MCMC simulation of spatial point processes

Birth-death Metropolis-Hastings algorithm for generating ergodic sample $\mathbf{X}^0, \mathbf{X}^1, \ldots$ from locally stable density f 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
     $$\min\left\{1, \frac{f(\mathbf{X}^i \cup \{u\})|S|}{f(\mathbf{X}^i)(n+1)}\right\}$$
     or
   - (death) select uniformly a point $u \in \mathbf{X}^i$ and accept $\mathbf{X}^{\text{prop}} = \mathbf{X}^i \setminus \{u\}$ with probability
     $$\min\left\{1, \frac{f(\mathbf{X}^i \setminus \{u\})n}{f(\mathbf{X}^i)|S|}\right\}$$
     (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).

Note: Metropolis-Hastings ratio does not depend on normalizing constant:

$$\frac{f(\mathbf{X}^i \cup \{u\})|S|}{f(\mathbf{X}^i)(n+1)} = \lambda(u, \mathbf{X}^i)\frac{|S|}{(n+1)}$$

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, geometrically ergodic and CLT:

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

Missing data

Suppose we observe $\mathbf{x}$ realization of $\mathbf{X} \cap W$ where $W \subset S$.
Problem: likelihood (density of $\mathbf{X} \cap W$)

$$f_{W,\theta}(\mathbf{x}) = \mathbb{E} f_\theta(\mathbf{x} \cap \mathbf{Y}_{S \setminus W})$$

not known - not even up to proportionality ! ($\mathbf{Y}$ unit rate Poisson on S)

Possibilities:

- Monte Carlo methods for missing data.
- Conditional likelihood

$$f_{W_{\ominus R},\theta}(\mathbf{x} \cap W_{\ominus R} | \mathbf{x} \cap (W \setminus W_{\ominus R})) \propto \exp(t(\mathbf{x})\theta^\mathsf{T})$$

(note: $\mathbf{x} \cap (W \setminus W_{\ominus R})$ fixed in $t(\mathbf{x})$)

Likelihood-based inference for Cox/Cluster processes

Consider Cox/cluster process **X** with random intensity function

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

observed within W (**M** Poisson with intensity κ).

Assume $f(m, \cdot)$ of bounded support and choose bounded $\tilde{W}$ so that

$$\Lambda(u) = \alpha \sum_{m \in \mathbf{M} \cap \tilde{W}} f(m, u) \quad \text{for } u \in W$$

$(\mathbf{X} \cap W, \mathbf{M} \cap \tilde{W})$ finite point process with density:

$$f(\mathbf{x}, \mathbf{m}; \theta) = f(\mathbf{m}; \theta) f(\mathbf{x} | \mathbf{m}; \theta) = \mathrm{e}^{|\tilde{W}|(1-\kappa)} \kappa^{n(\mathbf{m})} \mathrm{e}^{|W| - \int_W \Lambda(u) \mathrm{d}u} \prod_{u \in \mathbf{x}} \Lambda(u)$$

Likelihood

$$L(\theta) = \mathbb{E}_\theta f(\mathbf{x}|\mathbf{M}) = L(\theta_0) \mathbb{E}_{\theta_0} \Big[\frac{f(\mathbf{x}, \mathbf{M} \cap \tilde{W}; \theta)}{f(\mathbf{x}, \mathbf{M} \cap \tilde{W}; \theta_0)} \Big| \mathbf{X} \cap W = \mathbf{x} \Big]$$

\+ derivatives can be estimated using importance sampling/MCMC
\- however more difficult than for Markov point processes.

Bayesian inference: introduce prior $p(\theta)$ and sample posterior

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

(data augmentation) using birth-death MCMC.

Maximum likelihood estimation for log Gaussian Cox processes

Likelihood (probability density) for Cox process given observed point pattern **x**:

$$f_\theta(\mathbf{x}) = \mathbb{E}_\theta[\exp(-\int_W \Lambda(u) \mathrm{d}u) \prod_{u \in \mathbf{x}} \Lambda(u)]$$

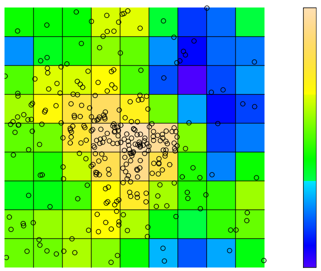
Problem for Monte Carlo approximation: $\Lambda = \{\Lambda(u)\}_{u \in W}$ infinitely dimensional quantity.

LCGP: approximate inference by discretizing random field $\Lambda(u) = \exp(\beta Z(u)^\mathsf{T} + Y(u))$

Counts N_i Poisson with mean

$$\exp(\beta Z(u_i)^\mathsf{T} + Y(u_i))|C_i|$$

(Poisson GLMM)



Computations: MCMC+FFT or INLA (Laplace approximations using Markov random fields for Gaussian field).

Exercises

1. Check the importance sampling formulas

$$\mathrm{E}_\theta k(\mathbf{X}) = \mathrm{E}_{\theta_0}\left[k(\mathbf{X})\frac{h_\theta(\mathbf{X})}{h_{\theta_0}(\mathbf{X})}\right]/(c_\theta/c_{\theta_0})$$

   and

$$\frac{c(\theta)}{c(\theta_0)} = \mathbb{E}_{\theta_0}\frac{h_\theta(\mathbf{X})}{h_{\theta_0}(\mathbf{X})} \tag{5}$$

2. Show that the formula

$$L(\theta)/L(\theta_0) = \mathbb{E}_{\theta_0}\Big[\frac{f(\mathbf{x},\mathbf{M}\cap\tilde{W};\theta)}{f(\mathbf{x},\mathbf{M}\cap\tilde{W};\theta_0)}\,\Big|\,\mathbf{X}\cap W = \mathbf{x}\Big]$$

   follows from (5) by interpreting $L(\theta)$ as the normalizing constant of $f(\mathbf{m}|\mathbf{x};\theta) \propto f(\mathbf{x},\mathbf{m};\theta)$.

Solution: second order product density for Poisson

$$\begin{aligned}
&\mathbb{E}\sum_{u,v\in\mathbf{X}}^{\neq} 1[u\in A, v\in B]\\
=&\sum_{n=0}^{\infty}\frac{\mathrm{e}^{-\mu(A\cup B)}}{n!}\int_{(A\cup B)^n}\sum_{u,v\in\mathbf{X}}^{\neq}1[u\in A,v\in B]\prod_{i=1}^{n}\rho(x_i)\mathrm{d}x_1\ldots\mathrm{d}x_n\\
=&\sum_{n=2}^{\infty}\frac{\mathrm{e}^{-\mu(A\cup B)}}{n!}2\binom{n}{2}\int_{(A\cup B)^2}\int_{(A\cup B)^{n-2}}1[x_1\in A,x_2\in B]\prod_{i=1}^{n}\rho(x_i)\mathrm{d}x_1\ldots\mathrm{d}
\\
=&\sum_{n=2}^{\infty}\frac{\mathrm{e}^{-\mu(A\cup B)}}{(n-2)!}\mu(A)\mu(B)\mu(A\cup B)^{n-2}\\
=&\mu(A)\mu(B)=\int_{A\times B}\rho(u)\rho(v)\mathrm{d}u\mathrm{d}v
\end{aligned}$$

Solution: invariance of g (and K) under thinning

Since $\mathbf{X}_{\text{thin}} = \{u\in\mathbf{X} : R(u)\le p(u)\}$,

$$\begin{aligned}
&\mathbb{E}\sum_{u,v\in\mathbf{X}_{\text{thin}}}^{\neq} 1[u\in A,v\in B]\\
=&\mathbb{E}\sum_{u,v\in\mathbf{X}}^{\neq}1[R(u)\le p(u),R(v)\le p(v),u\in A,v\in B]\\
=&\mathbb{E}\,\mathbb{E}\Big[\sum_{u,v\in\mathbf{X}}^{\neq}1[R(u)\le p(u),R(v)\le p(v),u\in A,v\in B]\,\Big|\,\mathbf{X}\Big]\\
=&\mathbb{E}\sum_{u,v\in\mathbf{X}}^{\neq}p(u)p(v)1[u\in A,v\in B]\\
=&\int_A\int_B p(u)p(v)\rho^{(2)}(u,v)\mathrm{d}u\mathrm{d}v
\end{aligned}$$