An introduction to statistics for spatial point processes

Jesper Møller and Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

June 12, 2006

Lectures:

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

2. Cox and cluster processes

3. The conditional intensity and Markov point processes

4. Likelihood-based inference and MCMC

Aim: overview of stats for spatial point processes - and spatial point process theory as needed.
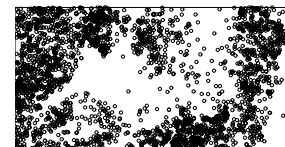
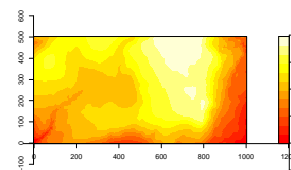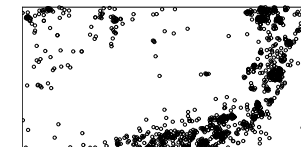Not comprehensive: the most fundamental topics and our favorite things.

---

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

2. Cox and cluster processes

3. The conditional intensity and Markov point processes

4. Likelihood-based inference and MCMC

Data example (Barro Colorado Island Plot)

Observation window $W = [0, 1000] \times [0.500]\text{m}^2$
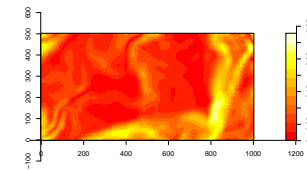
Beilschmiedia   Ocotea

Elevation   Gradient norm (steepness)

Sources of variation: elevation and gradient covariates *and* 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 (this and second lecture) or in terms of a probability density (third 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:

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

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

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

NB (exercise):

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

Campbell formula (by standard proof)

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

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):

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

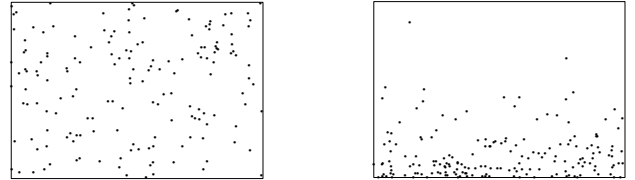($\Rightarrow$ non-parametric estimation if $\rho(u)\rho(v)$ known)

The Poisson process

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

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: $\rho(x, y) \propto e^{-10.6y}$

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 $g(u, v) = 1$

Exercises (30 minutes)

1. Show that the covariance between counts $N(A)$ and $N(B)$ is given by
$$\mathbb{C}\text{ov}[N(A), N(B)] = \mu^{(2)}(A \times B) + \mu(A \cap B) - \mu(A)\mu(B)$$

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

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

(Hint: use the Campbell formula)

3. (Practical `spatstat` exercise) Compute and interpret a non-parametric estimate of the K-function for the spruces data set.

(Hint: load `spatstat` using `library(spatstat)` and the spruces data using `data(spruces)`. Consider then the `Kest()` function.)

Distribution and moments of Poisson process

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

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

Similarly,

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

Proof of independent scattering (finite case)

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

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

$P(\mathbf{X} \cap A \in F, \mathbf{X} \cap B \in G) \quad (\mathbf{x} = \{x_1, \ldots, x_n\})$

$= \sum_{n=0}^{\infty} \frac{e^{-\mu(A \cup B)}}{n!} \int_{(A \cup B)^n} 1[\mathbf{x} \cap A \in F, \mathbf{x} \cap B \in G] \prod_{i=1}^{n} \rho(x_i) dx_1 \ldots dx_n$

$= \sum_{n=0}^{\infty} \frac{e^{-\mu(A \cup B)}}{n!} \frac{n!}{m!(n-m)!} \sum_{m=0}^{n} \int_{A^m} 1[\{x_1, x_2, \ldots, x_m\} \in F]$

$\int_{B^{n-m}} 1[\{x_{m+1}, \ldots, x_n\} \in G] \prod_{i=1}^{n} \rho(x_i) dx_1 \ldots dx_n$

$=$ (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 $\rho = \sum_{i=1}^{\infty} \rho(u)$ (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 $\int_S \rho_2(u) du < \infty$ and $\rho_2(u) = 0 \Rightarrow \rho_1(u) = 0$. Define $0/0 := 0$. Then

$P(\mathbf{X}_1 \in F)$

$= \sum_{n=0}^{\infty} \frac{e^{-\mu_1(S)}}{n!} \int_{S^n} 1[\mathbf{x} \in F] \prod_{i=1}^{n} \rho_1(x_i) dx_1 \ldots dx_n \quad (\mathbf{x} = \{x_1, \ldots, x_n\})$

$= \sum_{n=0}^{\infty} \frac{e^{-\mu_2(S)}}{n!} \int_{S^n} 1[\mathbf{x} \in F] e^{\mu_2(S) - \mu_1(S)} \prod_{i=1}^{n} \frac{\rho_1(x_i)}{\rho_2(x_i)} \prod_{i=1}^{n} \rho_2(x_i) dx_1 \ldots dx_n$

$= \mathbb{E}\left(1[\mathbf{X}_2 \in F] f(\mathbf{X}_2)\right)$

where

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

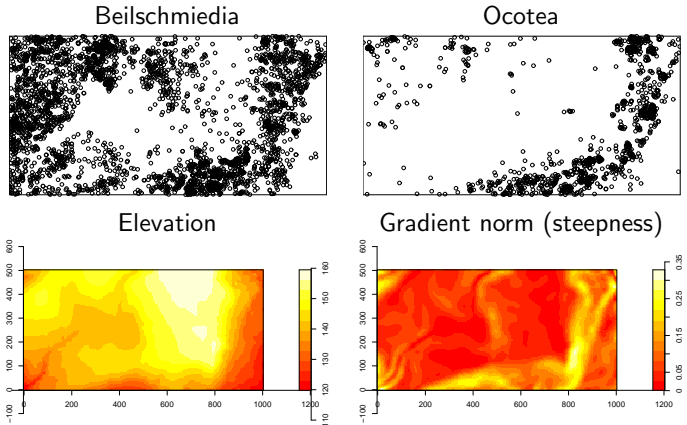Hence f is a density of $\mathbf{X}_1$ with respect to distribution of $\mathbf{X}_2$.

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

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

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)

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

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 (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

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

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

Implementation in spatstat

```
> 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
> fisherinf=vcov(beifit) #Fisher information matrix
> sqrt(diag(fisherinf)) #standard errors
(Intercept)        elev        grad
0.017500262 0.002287773 0.255860860
> rho=predict.ppm(beifit)
> Kbei=Kinhom(bei,rho) #warning: problem with large data sets.
> myKbei=myKest(cbind(bei$x,bei$y),rho,100,3,1000,500,F) #my own
                                          #procedure
```

K-functions

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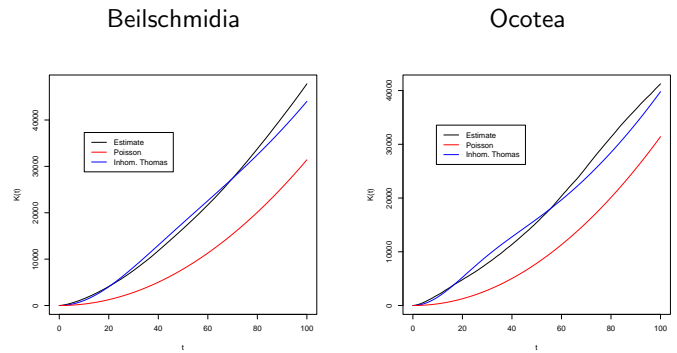
Poisson process: $K(t) = \pi t^2$ (since $g = 1$) less than K functions for data. Hence Poisson process models not appropriate.

Exercises (remaining time until 11:30)

1. Check that the Poisson expansion (1) indeed follows from the definition of a Poisson process.

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

3. (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 easy 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 $\mathbf{X}_{\text{thin}} = \{u \in \mathbf{X} | R(u) \leq p(u)\}$.)

Discussion of exercises: 11.30-12:00

Solution: second order product density for Poisson

$$\mathbb{E} \sum_{u,v \in \mathbf{X}}^{\neq} 1[u \in A, v \in B]$$

$$= \sum_{n=0}^{\infty} \frac{e^{-\mu(A \cup B)}}{n!} \int_{(A \cup B)^n} \sum_{u,v \in \mathbf{X}}^{\neq} 1[u \in A, v \in B] \prod_{i=1}^{n} \rho(x_i) dx_1 \ldots dx_n$$

$$= \sum_{n=2}^{\infty} \frac{e^{-\mu(A \cup B)}}{n!} 2\binom{n}{2} \int_{(A \cup B)^n} \int_{(A \cup B)^n} 1[x_1 \in A, x_2 \in B] \prod_{i=1}^{n} \rho(x_i) dx_1 \ldots dx_n$$

$$= \sum_{n=2}^{\infty} \frac{e^{-\mu(A \cup B)}}{(n-2)!} \mu(A)\mu(B)\mu(A \cup B)^{n-2}$$

$$= \mu(A)\mu(B) = \int_{A \times B} \rho(u)\rho(v) du dv$$

Solution: invariance of g (and K) under thinning

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

$$\mathbb{E} \sum_{u,v \in \mathbf{X}_{\text{thin}}}^{\neq} 1[u \in A, v \in B]$$

$$= \mathbb{E} \sum_{u,v \in \mathbf{X}}^{\neq} 1[R(u) \leq p(u), R(v) \leq p(v), u \in A, v \in B]$$

$$= \mathbb{E}\,\mathbb{E}\Big[\sum_{u,v \in \mathbf{X}}^{\neq} 1[R(u) \leq p(u), R(v) \leq p(v), u \in A, v \in B] \,\big|\, \mathbf{X}\Big]$$

$$= \mathbb{E} \sum_{u,v \in \mathbf{X}}^{\neq} p(u)p(v)1[u \in A, v \in B]$$

$$= \int_A \int_B p(u)p(v)\rho^{(2)}(u, v) du dv$$

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.

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:

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

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

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

  where $\boldsymbol{\Psi} = (\Psi(u))_{u \in \mathbb{R}^2}$ is a zero-mean Gaussian process
- Often sufficient to use power exponential covariance functions:

$$c(u,v) \equiv \mathbb{C}\mathrm{ov}[\Psi(u), \Psi(v)] = \sigma^2 \exp\left(-\|u-v\|^{\delta}/\alpha\right),$$
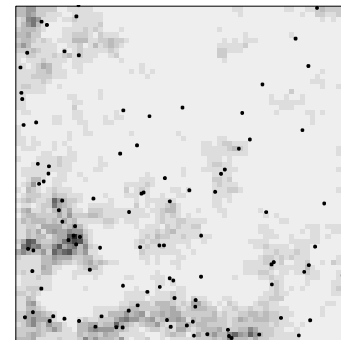
  $\sigma, \alpha > 0, \ 0 \le \delta \le 2$ (or linear combinations)
- Tractable product densities

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

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

Two simulated homogeneous LGCP's
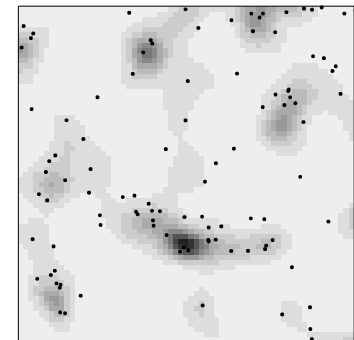
Exponential covariance function   Gaussian covariance function

Cluster processes

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

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

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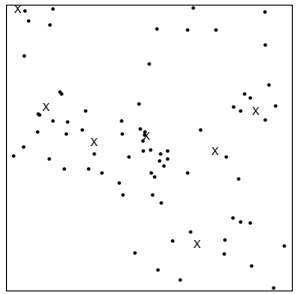$$\mathbf{X} = \cup_{m \in \mathbf{M}} \mathbf{X}_m$$

By superpositioning: if cond. on **M**, the $\mathbf{X}_m$ are independent Poisson processes, then **X** Cox process with random intensity function
$$\Lambda(u) = \alpha \sum_{m \in \mathbf{M}} f(m, u)$$

Nice expressions for intensity and product density if **M** Poisson on $\mathbb{R}^2$ with intensity function $\rho(\cdot)$ (Campbell):

$$\mathbb{E}\Lambda(u) = \mathbb{E}\alpha \sum_{m \in \mathbf{M}} f(m, u) = \alpha \int f(m, u)\rho(m)\mathrm{d}m \quad (= \kappa\alpha \text{ if } \rho(\cdot) = \kappa$$
$$\text{and } f(m, u) = f(u - m))$$

Example: modified Thomas process

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

Offspring $\mathbf{X} = \cup_m \mathbf{X}_m$ distributed around mothers according to bivariate isotropic Gaussian density f.

ω: standard deviation of Gaussian density

α: Expected number of offspring for each mother.

Cox process with random intensity function:

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

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:

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

Rain forest example:

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

elevation/gradient covariate.

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

► Fourth lecture: likelihood-based inference (missing data MCMC approach)

► Now: simulation free estimation

Parameter Estimation: regression parameters

Intensity function for inhomogeneous Thomas ($\rho(\cdot) = \kappa$):

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

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:

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

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):

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

Parameter Estimation: clustering parameters

Theoretical expression for (inhomogeneous) K-function:

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

Estimate κ and ω by matching theoretical K with semi-parametric estimate (minimum contrast)

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

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.

Generalisations

- *Shot noise Cox processes* driven by $\Lambda(u) = \sum_{(c,\gamma)\in\Phi} \gamma k(c,u)$
  where $c \in \mathbb{R}^2$, $\gamma > 0$ (Φ = marked Poisson process)
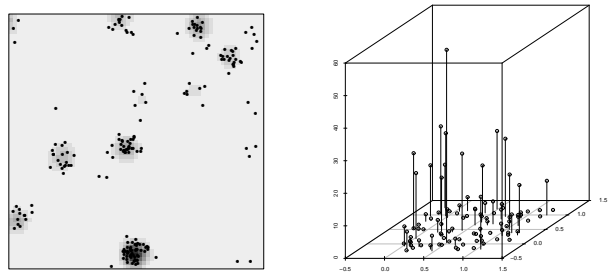
- Generalized SNCP's... (Møller & Torrisi, 2005)

Exercises (remaining time until 13:00)

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 number of iid offspring is a Cox process with random intensity function

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

(using notation from previous slide on cluster processes. Hint: if $\mathcal{D}(\mathbf{X}|\mathbf{M})$ only depends on $\mathbf{M}$ through Λ then $\mathcal{D}(\mathbf{X}|\mathbf{M}) = \mathcal{D}(\mathbf{X}|\Lambda)$)

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)

Lunch: we return at 14:00

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

2. Cox and cluster processes

3. The conditional intensity and Markov point processes

4. Likelihood-based inference and MCMC

Density with respect to a Poisson process

X on bounded S has density f with respect to unit rate Poisson **Y**
if

$$P(\mathbf{X} \in F) = \mathbb{E}\big(1[\mathbf{Y} \in F]f(\mathbf{Y})\big)$$

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

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* **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 **Y** on S and

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

is the normalizing constant (unknown).

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

Intensity and conditional intensity

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

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

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

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.
$$f(\mathbf{x}) = \exp(\sum_{\mathbf{y} \subseteq \mathbf{x}} U(\mathbf{y}))$$

   where $U(\mathbf{y}) = 0$ whenever $\|u - v\| \geq R$ for some $u, v \in \mathbf{y}$.

Pairwise interaction process: $U(\mathbf{y}) = 0$ 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):

$$\lambda(u, \mathbf{x}) = \exp\left(\beta + \psi \sum_{v \in \mathbf{x}} 1[\|u - v\| \leq R]\right), \quad f(\mathbf{x}) = \frac{1}{c} \exp\left(\beta n(\mathbf{x}) + \psi s(\mathbf{x})\right) \quad (\psi \leq 0)$$

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: $U(\cdot)$ 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 (easily shown):

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

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

Useful e.g. for residual analysis (paper).

Statistical inference based on pseudo-likelihood

$\mathbf{x}$ observed within bounded S. Parametric model $\lambda_\theta(u, \mathbf{x})$.

Let $N_i = 1[\mathbf{x} \cap C_i \neq \emptyset]$ where C_i disjoint partitioning of $S = \cup_i C_i$.

$P(N_i = 1 \mid \mathbf{X} \cap S \setminus C_i) \approx \lambda_\theta(u_i, \mathbf{X}) \mathrm{d}C_i$ where $u_i \in C_i$. Hence composite likelihood based on the N_i:

$$\prod_{i=1}^n (\lambda_\theta(u_i, \mathbf{x}) \mathrm{d}C_i)^{N_i} (1 - \lambda_\theta(u_i, \mathbf{x}) \mathrm{d}C_i)^{1-N_i} \equiv \prod_{i=1}^n \lambda_\theta(u_i, \mathbf{x})^{N_i} (1 - \lambda_\theta(u_i, \mathbf{x}) \mathrm{d}C_i)^{1-N_i}$$

which tends to *pseudo likelihood* function

$$\prod_{u \in \mathbf{x}} \lambda_\theta(u, \mathbf{x}) \exp\left(-\int_S \lambda_\theta(u, \mathbf{x}) \mathrm{d}u\right)$$

Score of pseudo-likelihood: unbiased estimating function by GNZ.

Pseudo-likelihood estimates asymptotically normal but asymptotic variance must be found by parametric bootstrap.
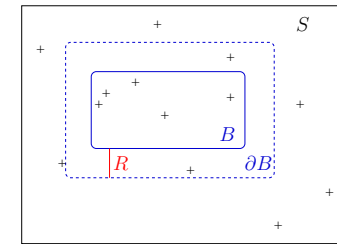
Flexible implementation for log linear conditional intensity (fixed R) in `spatstat`

Estimation of interaction range R: profile likelihood (?)

The spatial Markov property and edge correction

Let $B \subset S$ and assume **X** Markov with interaction radius R.

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

Factorization (Hammersley-Clifford):

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

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 **y** only through $\partial B \cap \mathbf{y}$.
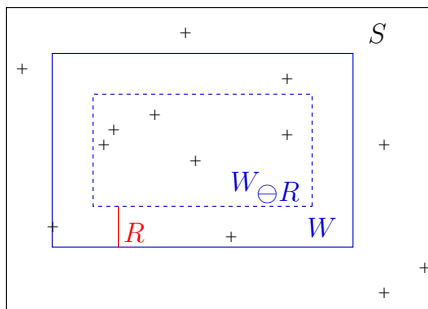
Edge correction using the border method

Suppose we observe **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 **X** outside $W_{\ominus R}$.

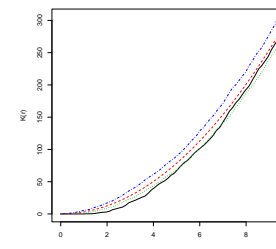## Example: spruces

Check fit of a homogeneous Poisson process using K-function and simulations:

```
> library(spatstat)
> data(spruces)
> plot(Kest(spruces)) #estimate K function
> Kenve=envelope(spruces,nrank=2)# envelopes "alpha"=4 %
 Generating 99 simulations of CSR ...
 1, 2, 3, 4, 5, 6, 7, 8, 9, 10,
 11, 12, 13, 14, 15, 16, 17, 18, 19, 20,
 ......
```

Strauss model for spruces

```
> fit=ppm(unmark(spruces),~1,Strauss(r=2),rbord=2)
> coef(fit)
 (Intercept) Interaction
-1.987940    -1.625994
> summary(fit)#details of model fitting
> simpoints=rmh(fit)#simulate point pattern from fitted model
> Kenvestrauss=envelope(fit,nrank=2)
```
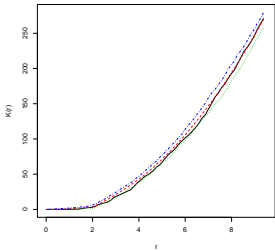
Exercises (remaining time until 15:30)

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. (spatstat) The multiscale process is an extension of the Strauss process where the density is given by

$$f(\mathbf{x}) \propto \exp(\beta n(\mathbf{x}) + \sum_{m=1}^{k} \psi_m s_m(\mathbf{x}))$$

   where $s_m(\mathbf{x})$ is the number of pairs of points u_i, u_j with $\|u_i - u_j\| \in]r_{m-1}, r_m]$ where $0 = r_0 < r_1 < r_2 < \cdots < r_k$. Fit a multiscale process with $k = 4$ and of interaction range $r_k = 5$ to the spruces data. Check the model using the K-function.

   (Hint: use the spatstat function ppm with the PairPiece potential. The function envelope can be used to compute envelopes for the K-function under the fitted model.)

Exercises (cntd).

4. (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 $\mathbb{E}g(\mathbf{X}) = \mathbb{E}[g(\mathbf{Y})f(\mathbf{Y})]$.)

5. (if time) Check using the GNZ formula, that the score of the pseudo-likelihood is an unbiased estimating function.

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 $\mathbf{X}^{\mathrm{prop}} = \mathbf{X}^i \cup \{u\}$ 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 $\mathbf{X}^{\mathrm{prop}} = \mathbf{X}^i \setminus \{u\}$ 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 $\mathbf{X}^{i+1} = \mathbf{X}^{\mathrm{prop}}$; 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 $\mathbf{X}$ with random intensity function

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

observed within W ($\mathbf{M}$ 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:

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

Likelihood

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

+ 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.

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

$$\frac{c(\theta)}{c(\theta_0)} = \mathbb{E}_{\theta_0} \frac{h_\theta(\mathbf{X})}{h_{\theta_0}(\mathbf{X})} \tag{3}$$

2. Show that the formula

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

follows from (3) by interpreting $L(\theta)$ as the normalizing constant of $f(\mathbf{m}|\mathbf{x}; \theta) \propto f(\mathbf{x}, \mathbf{m}; \theta)$.

3. (practical exercise) Compute MLEs for a multiscale process applied to the spruces data. Use the `newtonraphson.mpp()` procedure in the package `MppMLE`.