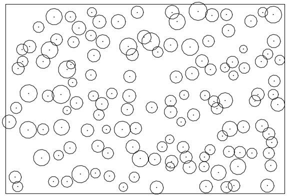
Aspects of spatial point process modelling and Bayesian inference

Jesper Møller,
Department of Mathematical Sciences, Aalborg University.

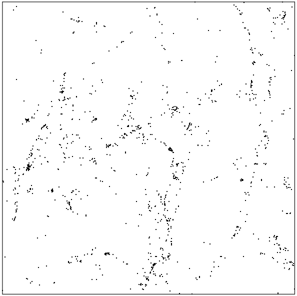
August 29, 2012

Data Examples: Norwegian spruces & Danish barrows

Spruces



Barrows



Note: specification of *observation window* very important – information about where points do not occur is just as important as information about where the points do occur.

Other examples of data

- One-dimensional point patterns:
  - Positions of car accidents on a highway during a month
  - Times of earthquakes in Japan
- Two-dimensional point patterns:
  - Positions of cities on a map
  - Positions of farms with mad cow disease in UK
  - Positions of broken wires in an electrical network
- Three-dimensional point patterns:
  - Positions of stars in the visible part of the universe
  - Positions of copper deposits underground
  - Times and positions and earthquakes in Japan

Note: Observation within a *bounded* window, i.e. a bounded subset of $\mathbb{R}^n$ - usually an interval/rectangle/box - and sometimes a more complicated shape. Need to account for boundary effects...

$n = 1$: 'The time axis' is directional; temporal point processes!

$n \geq 2$: We focus on spatial point processes (no time). No natural direction!

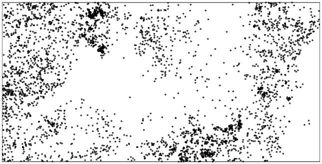
Spatial point pattern data

- agricultural research
- archeology
- computer science
- communication technology
- ecology
- forestry
- geography
- material science
- medical image analysis
- seismology
- spatial epidemiology
- statistical mechanics
- ...

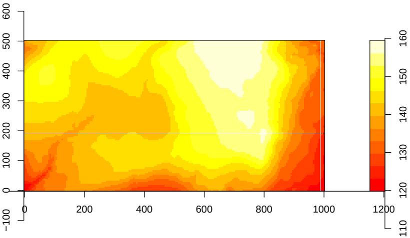
Two further examples of two-dimensional point patterns from plant and animal ecology illustrating important features of spatial point pattern data...

Tropical rain forests trees

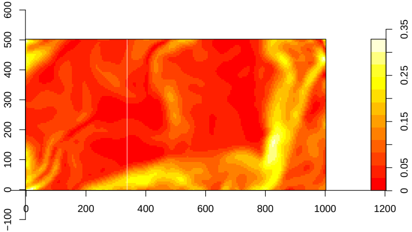
Beilschmiedia



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

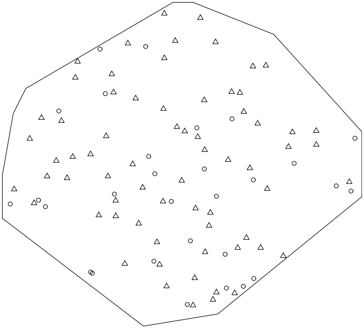


Altitude



Norm of altitude gradient (steepness)

Ants nests



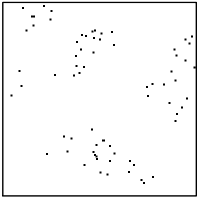
Observation window = polygon

Multitype point pattern: *Messor* (Δ) and *Cataglyphis* ($\circ$)

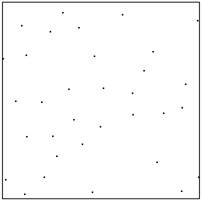
Cataglyphis ants feed on dead *Messors* $\Rightarrow$ *interaction* (hierarchical model)

Statistical inference for spatial point patterns

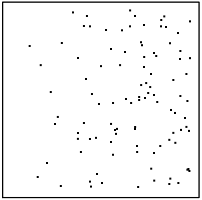
- Objective is to infer structure in spatial distribution of points:
  - interaction between points: inhibition/regularity or attraction/aggregation/clustering
  - inhomogeneity linked to covariates
- Spatial point processes are stochastic models for spatial point patterns.



Clustered



Regular



Inhomogeneous

What is a spatial point process?

- **Definitions:**
  1. a random counting measure N on $\mathbb{R}^d$
  2. a locally finite random subset $\mathbf{X}$ of $\mathbb{R}^d$
- Counting measure: $N(A)$ counts the number of points from $\mathbf{X}$ falling in any bounded Borel set $A \subset \mathbb{R}^d$.
- Locally finite: $\#(\mathbf{X} \cap A)$ finite for all bounded Borel sets $A \subset \mathbb{R}^d$.
- Equivalent if *simple point process* (i.e. no multiple points): $N(A) = \#(\mathbf{X} \cap A)$.

A bit of measure theory

We restrict attention to

locally finite simple point processes $\mathbf{X}$ defined on $\mathbb{R}^d$

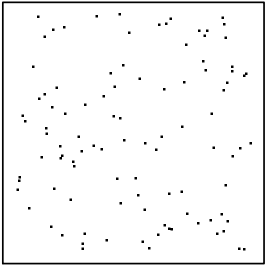
(extensions to other settings including non-simple point processes, marked point processes, multiple point processes, and lattice processes are rather straightforward).
Then

- *measurability* means that $N(A)$ is a random variable for any bounded Borel set $A \subset \mathbb{R}^d$;
- the distribution of $\mathbf{X}$ is uniquely determined by the *void probabilities*

$$v(A) = P(N(A) = 0), \quad A \subset \mathbb{R}^d \text{ compact.}$$

Simple example of point process: Binomial point process

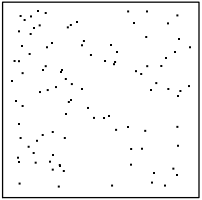
- Suppose f is a probability density on a Borel set $S \subseteq \mathbb{R}^d$ (usually bounded). Then $\mathbf{X}$ is a *binomial point process* with n points if $\mathbf{X} = \{x_1, \ldots, x_n\}$ consists of n *iid* points $x_i \sim f$.
- 'Binomial' since $N(A) \sim b(n, p)$ with $p = \int_A f(x) \mathrm{d}x$ and $A \subseteq S$.

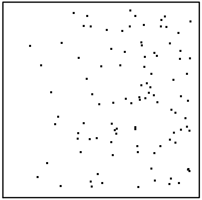


Example with $S = [0, 1] \times [0, 1]$, $n = 100$ and $f(x) = 1$.

Fundamental model: The Poisson process

- Assume μ locally finite measure on a Borel set $S \subseteq \mathbb{R}^d$ with $\mu(B) = \int_B \rho(u)\mathrm{d}u$ for all Borel sets $B \subseteq S$.
- $\mathbf{X}$ is a *Poisson process* on S with *intensity measure* μ and *intensity (function)* ρ if for any bounded Borel set $B \subseteq S$ with $\mu(B) > 0$:
  1. $N(B) \sim \mathrm{po}(\mu(B))$
  2. Given $N(B)$, points in $\mathbf{X} \cap B$ i.i.d. with density $\propto \rho(u)$, $u \in B$ (i.e. given $N(B)$, $\mathbf{X} \cap B$ is a binomial point process).
- Examples on $S = [0,1] \times [0,1]$:





Homogeneous: $\rho = 100$. Inhomogeneous: $\rho(x, y) \propto 200x$.

Classical assumptions: Stationarity and isotropy

- $\mathbf{X}$ on $\mathbb{R}^d$ is *stationary* if distribution invariant under translations:

$$\mathbf{X} \sim \mathbf{X} + s := \{s + u | u \in \mathbf{X}\}, \quad s \in \mathbb{R}^d.$$

- $\mathbf{X}$ on $\mathbb{R}^d$ is *isotropic* if distribution invariant under rotations:

$$\mathbf{X} \sim R\mathbf{X} := \{Ru | u \in \mathbf{X}\}, \quad R \text{ rotation around the origin.}$$

- Poisson process on $\mathbb{R}^d$ with constant intensity ρ: both stationary and isotropic.
- Many recent papers deal with non-stationary and anisotropic spatial point process models (see references at the end).

Some notation and conventions

- Whenever we consider sets $S, B, \ldots \subseteq \mathbb{R}^d$, they are assumed to be Borel sets.
- $|S|$ denotes Lebesgue measure (length/area/volume/...).
- For a point process $\mathbf{X}$ on $S \subseteq \mathbb{R}^d$ and a subset $B \subseteq S$,
  - $\mathbf{X}_B = \mathbf{X} \cap B$ is the restriction of $\mathbf{X}$ to B
  - $n(\mathbf{X}_B)$ is the number of points in $\mathbf{X}_B$
  - $N(B)$ is generic notation for $n(\mathbf{X}_B)$

Summary statistics

- Summary statistics are numbers or functions describing characteristics of point processes, for example:
  - The mean number of points in a set B.
  - The covariance of the number of points in sets A and B.
  - The mean number of points within distance $r > 0$ of an 'arbitrary point of the process'.
  - The probability that there are no points within distance R of an 'arbitrary point of the process'.
- They are useful for:
  - Preliminary analysis
  - Model fitting (minimum contrast estimation, maximum composite likelihood estimation, ... (non-Bayesian!)).
  - Model checking (incl. Bayesian inference!)

Another useful tool: residuals. See
Baddeley, Turner, Møller and Hazelton (2005), *JRSS B*;
Baddeley, Møller and Pakes (2008), *AISM*;
Baddeley, Rubak and Møller (2011), *Statistical Science*.

First order moments

- *Intensity measure* μ:

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

- *Intensity function* ρ:

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

- Infinitesimal interpretation: when A very small, $N(A) \approx$ binary variable (presence or absence of point in A). Hence if A has area/volume/... $|A| = \mathrm{d}u$,

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

- Note: if $\rho(u)$ is constant, we say $\mathbf{X}$ is *homogeneous*; otherwise it is *inhomogeneous*.

Second order moments

- *Second order factorial moment measure* $\alpha^{(2)}$:

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

- *Second order product density* $\rho^{(2)}$:

$$\alpha^{(2)}(A \times B) = \int_A \int_B \rho^{(2)}(u,v)\,\mathrm{d}u\,\mathrm{d}v.$$

- Infinitesimal interpretation of $\rho^{(2)}$:

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

$(u \in A,\ |A| = \mathrm{d}u,\ v \in B,\ |B| = dv,\ A \cap B = \emptyset)$

- Note that covariances can be expressed using these:

$$\mathrm{Cov}[N(A), N(B)] = \alpha^{(2)}(A \times B) + \mu(A \cap B) - \mu(A)\mu(B).$$

Second order product density for a Poisson process

- If $\mathbf{X}$ is a Poisson process with intensity function $\rho(u)$, then its second order product density is given by

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

Pair correlation function

- Pair correlation:

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

  (here $a/0 = 0$ for all a).
- Interpretation of pair correlation function:
  - Poisson process: $g(u,v) = 1$.
  - If $g(u,v) > 1$, then attraction/aggregation/clustering.
  - If $g(u,v) < 1$, then repulsion/regularity.
- If X is stationary, then $g(u,v) = g(u-v)$;
  if X is also isotropic, then $g(u,v) = g(\|u-v\|) = g(r)$.

Non-parametric estimation of ρ (homogeneous case)

- Suppose that $\mathbf{X}_W$ is observed, where $W \subset \mathbb{R}^d$ is a bounded observation window.
- Estimate of ρ in the homogeneous case:

$$\hat{\rho} = n(\mathbf{X}_W)/|W|.$$

- $\mathbb{E}\hat{\rho} = \rho$.
- Poisson process: $\hat{\rho}$ =MLE.

Non-parametric estimation of ρ (inhomogeneous case)

- Estimate of $\rho(u)$ in the inhomogeneous case (Diggle, 1985):

$$\hat{\rho}(u) = \sum_{v \in \mathbf{X}_W} k(u-v)/c_W(v), \quad u \in W.$$

- *Kernel*: $k(u)$ is a probability density function.
- *Edge-correction factor*: $c_W(v) = \int_W k(u-v)\mathrm{d}u$.
- $\int_W \hat{\rho}(u)\mathrm{d}u$ is an unbiased estimate of $\mu(W)$.
- Sensitive to the choice of 'bandwidth'...
  (if covariate information is available, a parametric model for ρ may be preferred).

K (Ripley, 1977) and L-function (Besag, 1977)

- Assume $\mathbf{X}$ stationary with intensity $\rho > 0$ and pair correlation function $g(u,v) = g(u-v)$.
- Ripley's K-function: $K(r) = \int_{\|u\| \le r} g(u) \mathrm{d}u$, or

$$\rho K(r) = \mathbb{E} \frac{1}{\rho |A|} \sum_{u \in \mathbf{X}_A} \sum_{v \in \mathbf{X} \setminus \{u\}} 1[\|u-v\| \le r], \quad r > 0.$$

- Interpretation: $\rho K(r)$ is the expected number of points within distance r of an arbitrary point of X.
- (Besag's) L-function (variance stabilizing transformation):

$$L(r) = (K(r)/\omega_d)^{1/d}$$

  ($\omega_d = |\text{unit ball in } \mathbb{R}^d| = \pi^{d/2}/\Gamma(1+d/2)$).
- $L(r) - r$ is often plotted instead of $K(r)$:
  - Poisson process: $L(r) - r = 0$.
  - If $L(r) - r > 0$ ($L(r) - r < 0$), then attraction/aggregation/clustering (repulsion/regularity).

Inhomogeneous K and L-functions (Baddeley, Møller & Waagepetersen, 2000)

Def.: $\mathbf{X}$ is second-order intensity reweighted stationary (s.o.i.r.s.) if $g(u,v) = g(u-v)$. Then we still define $K(r) = \int_{\|u\| \leq r} g(u)\mathrm{d}u$ and $L(r) = (K(r)/\omega_d)^{1/d}$.

- s.o.i.r.s. is satisfied for any Poisson process, for many Cox process models (see later), and for an independent thinning of any stationary point process.
- Poisson case: $L(r) - r = 0$.
- If $\mathbf{X}$ is s.o.i.r.s. and $W_u = \{u + v : v \in W\}$, then

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

  is unbiased, but in practice an estimate for $\rho(u)$ is plugged in.

Model check using summary statistics: envelopes

- Compare (theoretical) summary statistic $T(r)$ from model with (non-parametric) estimate $\hat{T}_0(r)$ obtained from data.
- If $T(r)$ is intractable, it may be approximated using simulations, i.e. simulate n new point patterns and calculate estimates $\hat{T}_1(r), \ldots, \hat{T}_n(r)$.
- If $\hat{T}_{(1)}(r), \ldots, \hat{T}_{(n)}(r)$ are the ordered simulated statistics, then e.g.

$$P(\hat{T}_0(r) \leq \hat{T}_{(1)}(r) \text{ or } \hat{T}_0(r) \geq \hat{T}_{(n)}(r)) = 2/(n+1)$$

(if no ties). For $n = 39$, we have $2/(n+1) = 0.05$.

Point processes in R

- R-packages for dealing with point processes:
  - Spatial point processes: `spatstat`
  - (Temporal point processes: `PtProcess`)
- Manuals:
  - `www.spatstat.org/spatstat/doc/spatstatJSSpaper.pdf`
  - (`cran.at.r-project.org/web/packages/PtProcess/PtProcess.pdf`)
- Many algorithms implemented for
  - Parameter estimation
  - Simulation
  - Model checking

Bayesian inference for the Poisson process

- Aim: estimate the intensity function ρ of a Poisson process, imposing a parametric or 'non-paramteric' prior model; investigate the dependence of covariates...
- References to various contributions can be found at the end.
- We focus on some examples... and start with a short summary on Poisson processes (further reading: Møller & Waagepetersen (2004); Kingman (1993)!!)

Definition of the Poisson process

- Assume μ locally finite measure on a (Borel) set $S \subseteq \mathbb{R}^d$ with $\mu(B) = \int_B \rho(u)\mathrm{d}u$ for all (Borel) sets $B \subseteq S$.
- $\mathbf{X}$ is a *Poisson process* on S with *intensity measure* μ and *intensity (function)* ρ if for any bounded region B with $\mu(B) > 0$:
  1. $N(B) \sim \mathrm{po}(\mu(B))$
  2. Given $N(B)$, points in $\mathbf{X}_B$ are i.i.d. with density $\propto \rho(u)$, $u \in B$.

[[Verifying the existence: consider a subdivision $\mathbb{R}^d = \cup_i B_i$ (disjoint);
construct $\mathbf{X}_{B_i}$ and thereby $X = \cup_i \mathbf{X}_{B_i}$;
easy to show that $P(N(A) = 0) = \exp(-\mu(A))$, the void probability for the Poisson process.]]

Some properties: Independent scattering

- Suppose $\mathbf{X}$ is a Poisson process on S and $B_1, B_2, \ldots$ are disjoint subsets of S.
- Then $\mathbf{X}_{B_1}, \mathbf{X}_{B_2}, \ldots$ are independent Poisson processes.

Proof: Calculate void probabilities!

Some properties: Superpositioning

- Suppose $\mathbf{X}_i \sim \text{Poisson}(S, \rho_i)$, $i = 1, 2, \ldots$ are independent Poisson processes and that $\rho = \sum_i \rho_i$ is locally integrable.
- Then $\mathbf{X} = \cup_{i=1}^{\infty} \mathbf{X}_i$ is a disjoint union with probability one, and $\mathbf{X} = \cup_{i=1}^{\infty} \mathbf{X}_i$ is $\text{Poisson}(S, \rho)$.

Proof: Calculate void probabilities!

Some properties: Independent thinning

- Suppose we obtain $\mathbf{X}_{\text{thin}}$ by independently either keeping or deleting points $u \in \mathbf{X}$ according to probabilities $p(u)$:

$$\mathbf{X}_{\text{thin}} = \{u \in \mathbf{X} | R_u \leq p(u)\}$$

  where the R_u are independent uniform variables on $[0, 1]$ independent of $\mathbf{X}$.
- Then $\mathbf{X}_{\text{thin}}$ is an *independent thinning* of $\mathbf{X}$.
- Result: $\mathbf{X}_{\text{thin}}$ and $\mathbf{X} \setminus \mathbf{X}_{\text{thin}}$ are independent Poisson processes with intensity functions $p(u)\rho(u)$ and $(1 - p(u))\rho(u)$.

Proof: Calculate void probabilities!

Simulation of Poisson processes on a bounded set $S \subset \mathbb{R}^d$

Homogeneous case, intensity $\rho > 0$:

- Generate $n \sim \text{po}(\rho|S|)$.
- For $i = 1, \ldots, n$, generate $u_i \sim \text{unif}(S)$.

Inhomogeneous case, intensity $\rho(u) \leq \rho_{\text{max}}$, for some $\rho_{\text{max}} > 0$:

- Generate $\mathbf{X}$ as a homogeneous Poisson process with intensity ρ_{max}.
- For $i = 1, \ldots, n$, keep u_i with probability $\rho(u_i)/\rho_{\text{max}}$.

Densities for Poisson processes

- Recall that $\mathbf{X}_1$ is absolutely continuous wrt. $\mathbf{X}_2$ if $P(X_2 \in F) = 0 \Rightarrow P(X_1 \in F) = 0$.
- 1. For any numbers $\rho_1 > 0$ and $\rho_2 > 0$, Poisson$(\mathbb{R}^d, \rho_1)$ is absolutely continuous wrt. Poisson$(\mathbb{R}^d, \rho_2)$ if and only if $\rho_1 = \rho_2$.
  2. Suppose $\rho_1(\cdot)$ and $\rho_2(\cdot)$ are intensity functions so that $\mu_1(S)$ and $\mu_2(S)$ are finite and that $\rho_1(u) > 0 \Rightarrow \rho_2(u) > 0$. Then Poisson$(S, \rho_2)$ has density

$$f(\mathbf{x}) = \exp(\mu_1(S) - \mu_2(S)) \prod_{u \in \mathbf{x}} \frac{\rho_2(u)}{\rho_1(u)}$$

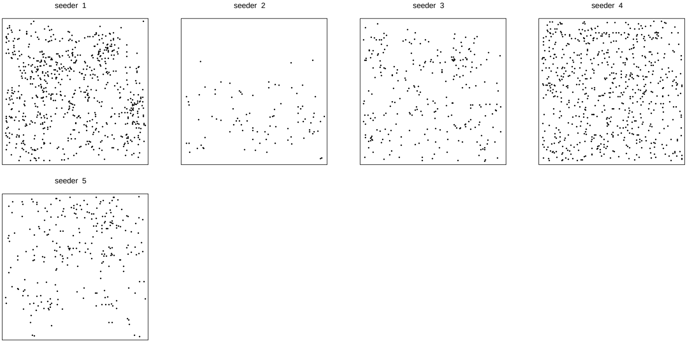
     wrt. Poisson(S, ρ_1).
- Example: for bounded S, Poisson(S, ρ) has density
$$f(\mathbf{x}) = \exp(|S| - \mu(S)) \prod_{u \in \mathbf{x}} \rho(u)$$
wrt. standard (unit-rate) Poisson process Poisson$(S, 1)$.

Example of a Bayesian analysis of a parametric model

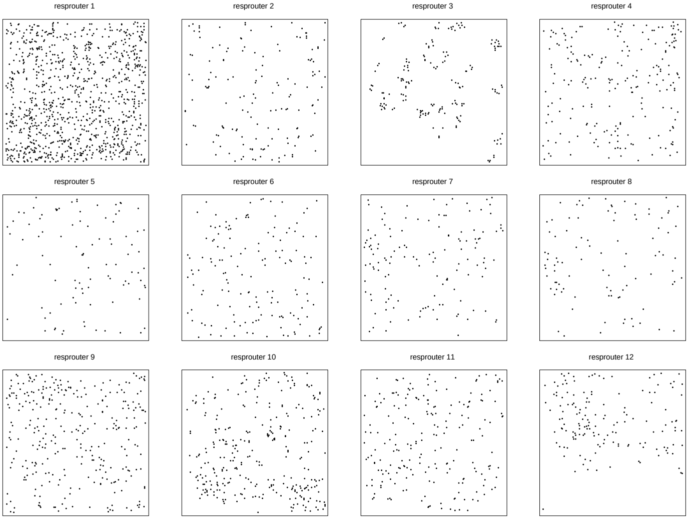
J.B. Illian, J. Møller and R.P. Waagepetersen (2009). Hierarchical spatial point process analysis for a plant community with high biodiversity. *Environmental and Ecological Statistics*, 16, 389-405.

- Discusses a multivariate Poisson point process model for spatial point patterns formed by a natural plant community with a high degree of biodiversity (22×22 m plot at Cataby in the Mediterranean type shrub- and heathland of the South-Western area of Western Australia).
- Next figures: point patterns of the 5 most abundant species of 'seeders' and the 19 most dominant (influential) species of 'resprouters' (have been at the exactly same location for a very long time).

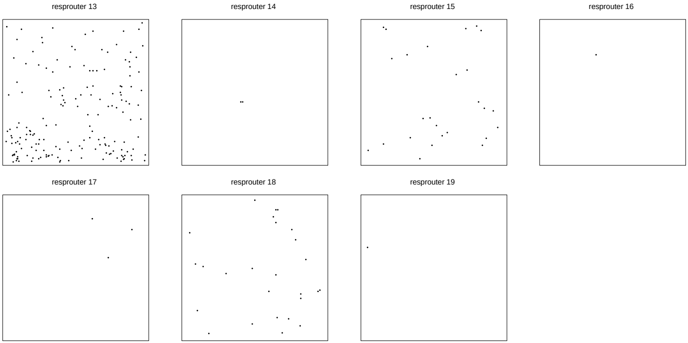
5 most abundant species of seeders:



19 most influential species of resprouters (the 12 first):



19 most influential species of resprouters (the next 7):



Likelihood for the seeders conditional on the resprouters

Our likelihood resembles approaches derived from ecological field theory...: We assume that the 5 seeders $\mathbf{Y}_1, \ldots, \mathbf{Y}_5$ conditional on the 19 resprouters $\mathbf{X}_1, \ldots, \mathbf{X}_{19}$ are independent Poisson processes with intensity functions

$$\lambda(\xi|\mathbf{x}, \boldsymbol{\theta}_i) = \exp\left(\boldsymbol{\theta}_i s(\xi|\mathbf{x})^\top\right), \qquad \xi \in W, \qquad i = 1, \ldots, 5,$$

where
$\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_{19})$ is the collection of all 19 resprouter patterns;
$\boldsymbol{\theta}_i = (\theta_{i0}, \ldots, \theta_{i19})$ is a vector of parameters;
$s(\xi|\mathbf{x}) = (1, t(\xi|\mathbf{x}_1), \ldots, t(\xi|\mathbf{x}_{19}))$ with

$$t(\xi|\mathbf{x}_j) = \sum_{\eta \in \mathbf{x}_j} h_\eta(\|\xi - \eta\|), \qquad j = 1, \ldots, 19,$$

describing the dependence of $\mathbf{x}_j$.

Here $\|\cdot\|$ denotes Euclidean distance, and we have chosen a simple smooth interaction function

$$h_\eta(r) = \begin{cases} \left(1-(r/R_\eta)^2\right)^2 & \text{if } 0 < r \leq R_\eta \\ 0 & \text{else} \end{cases}$$

for $r \geq 0$, where $R_\eta \geq 0$ defines the radii of interaction of a given resprouter at location η.
So

$$\log \lambda(\xi|\mathbf{x}) = \theta_{i0} + \sum_{j=1}^{19} \theta_{ij} \sum_{\eta \in \mathbf{x}_j} h_\eta(\|\xi - \eta\|)$$

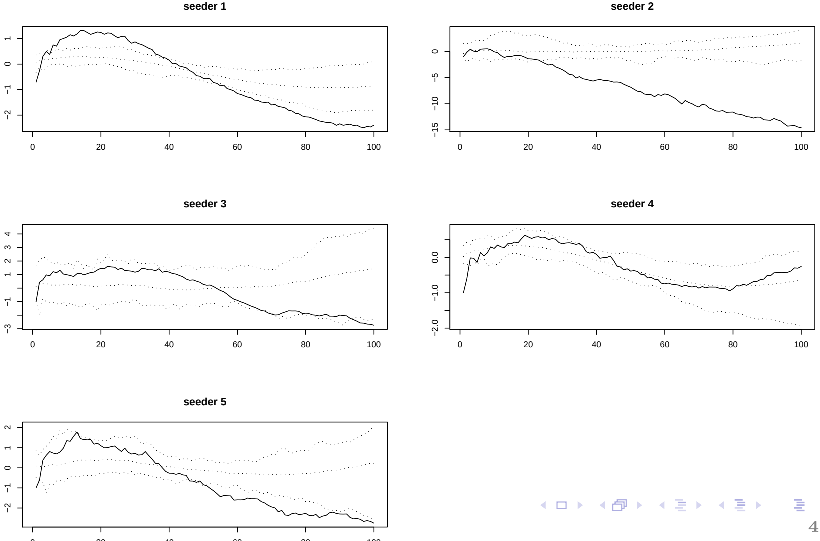
where $\theta_{i0} \in \mathbb{R}$ is an intercept and for $j = 1, \ldots, 19$, $\theta_{ij} \in \mathbb{R}$ controls the influence of the jth resprouter on the ith seeder:
$\theta_{ij} > 0$ means a positive/attractive association;
$\theta_{ij} < 0$ means a negative/repulsive association.

Hence the conditional log likelihood function based on the 5 seeder point patterns $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_5)$ is

$$l(\boldsymbol{\theta}, \mathbf{R}; \mathbf{y}|\mathbf{x}) = \sum_{i=1}^{5} \left[\boldsymbol{\theta}_i \sum_{\xi \in \mathbf{y}_i} s(\xi|\mathbf{x})^\top - \int_W \exp\left(\boldsymbol{\theta}_i s(\xi|\mathbf{x})^\top \right) \mathrm{d}\xi \right]$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_5)$ is the vector of all 100 parameters θ_{ij} and $\mathbf{R}$ is the vector of all 3168 radii $R_\eta,\ \eta \in \mathbf{x}_j,\ j = 1, \ldots, 19$.
In comparison, there are $N_1 + \cdots + N_5 = 1954$ seeders.
MLE: 'hopeless' unless we assume known and equal interaction radii for resprouters of the same type—but this assumption is highly unrealistic, since the plants vary in size. Also results based on summary statistics indicate that a more appropriate model would have to take intra-specific interaction into account.
A Bayesian setting seems needed...

Estimated inhomogeneous $(L(r) - r)$-functions for seeders 1-5 with 95% envelopes simulated from the model with known interaction radii for resprouters and fited by maximum likelihood. Distance $r > 0$ is in cm.



Prior information

After extensive and detailed discussions with the scientist who collected the data, we used his knowledge to elicit informative priors on the interaction radii.

Range of zone of influence (in cm) for resprouters:

1.	10-40	11.	20-30
2.	5-15	12.	25-75
3.	15-60	13.	30-50
4.	25-75	14.	50-130
5.	10-25	15.	150-400
6.	10-20	16.	50-200
7.	10-25	17.	50-200
8.	10-25	18.	50-250
9.	2-10	19.	10-250
10.	20-100		

Prior assumptions

- The interaction radii R_η are independent;
- for each $\eta \in \mathbf{x}_j$, $R_\eta \sim \mathrm{N}(\mu_j, \sigma_j^2)$ restricted to $[0, \infty)$, where (μ_j, σ_j^2) is chosen so that under the unrestricted $\mathrm{N}(\mu_j, \sigma_j^2)$, the range of the zone of influence in the table is a central 95% interval;
- given the R_η, the θ_{ij} are i.i.d., following a relatively non-informative $\mathrm{N}(0, \sigma^2)$-distribution (the specification of σ is discussed in the paper).

These a priori independence assumptions are essentially made, since we have no prior knowledge on how to specify a correlation structure for all the R_η and all the θ_{ij}. For the same reason, $(\boldsymbol{\theta}, \mathbf{R})$ and $\mathbf{X}$ are assumed to be independent.

Posterior

Hence the posterior density for $(\theta, \mathbf{R})$ is

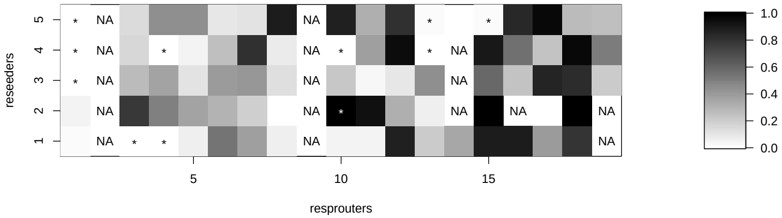
$$\pi(\boldsymbol{\theta}, \mathbf{R}|\mathbf{x}, \mathbf{y}) \propto$$
$$\exp\left(-\sum_{i=1}^{5}\left[\theta_{i0}^2/(2\sigma^2) - \sum_{j=1}^{19}\theta_{ij}^2/(2\sigma^2)\right] - \sum_{j=1}^{19}\sum_{\eta\in\mathbf{x}_j}(R_\eta - \mu_j)^2/(2\sigma_j^2)\right)$$
$$\times\exp\left(\sum_{i=1}^{5}\left[\boldsymbol{\theta}_i\sum_{\xi\in\mathbf{y}_i}s(\xi|\mathbf{x})^\top - \int_W \exp(\boldsymbol{\theta}_i s(\xi|\mathbf{x})^\top)\mathrm{d}\xi\right]\right), \quad \theta_{ij}\in\mathbb{R}, R_\eta \geq 0.$$

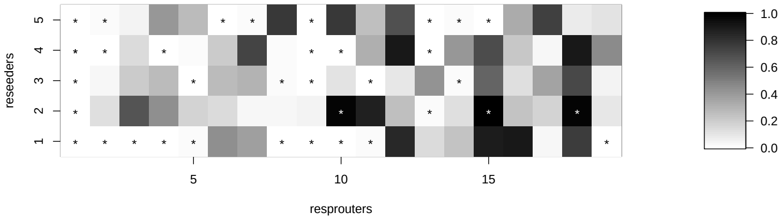
((Hybrid Markov chain Monte Carlo algorithm/Metropolis within Gibbs, using random walk Metropolis updates))

(((...The proposal distributions for these random walk updates are multivariate normal with diagonal covariance matrices. The vector of proposal standard deviations for $\boldsymbol{\theta}_i$ is given by $k\hat{\sigma}_{i|\mathbf{y}}$, where k is a user specified parameter and $\hat{\sigma}_{i|\mathbf{y}}$ is an estimate of the vector of posterior standard deviations for $\boldsymbol{\theta}_i$ obtained from a pilot run. The value of k was chosen to give acceptance rates around 25 %. The vector of proposal standard deviations for $\mathbf{R}$ is given by the vector of prior standard deviations divided by 2...)))

Posterior results

- Next figure: the lower plot shows a grey scale plot of posterior probabilities $P(\theta_{ij} > 0|\mathbf{y})$. The starred fields are those for which 0 is outside the central 95 % posterior interval for θ_{ij}. (The upper plot concerns MLE-results which are mostly ignored for this talk.)
- E.g. resprouter 1 seems to have a clear repulsive effect on seeders, while resprouters 15 and 18 have a distinct attractive effect on seeders.
- The Bayesian approach yields more clear-cut results than the maximum likelihood inference, since the intermediate grey scales are less frequent in the lower plot: more θ_{ij}'s have strong evidence for being different from zero (if 'strong evidence' is interpreted as being significant at the 95% level or being outside the 95% posterior interval, respectively).


reseeders
resprouters
NA
1.0
0.8
0.6
0.4
0.2
0.0


reseeders
resprouters
1.0
0.8
0.6
0.4
0.2
0.0

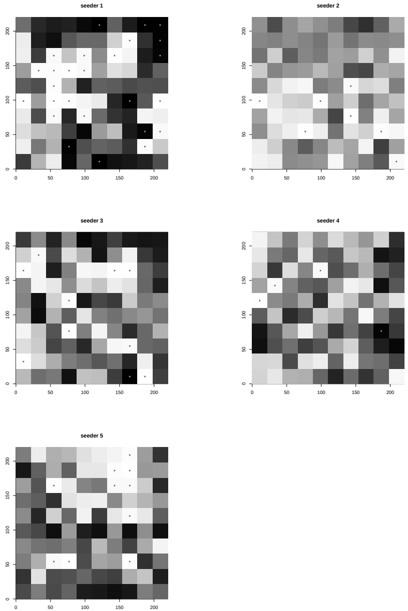
Model assessment

We follow the idea of posterior predictive model assessment (Gelman et al., 1996) and compare various summary statistics with their posterior predictive distributions, depending possibly both on the points $\mathbf{Y}$ and the parameters $\boldsymbol{\theta}$ and $\mathbf{R}$. Any posterior predictive distribution is obtained from simulations:

- we generate a posterior sample $(\boldsymbol{\theta}_{i,1}, \mathbf{R}_1), \ldots, (\boldsymbol{\theta}_{i,m}, \mathbf{R}_m)$, and for each $(\boldsymbol{\theta}_{i,k}, \mathbf{R}_k)$ new data $\mathbf{y}_{i,k}$ from the conditional distribution of $\mathbf{Y}_i$ given $(\boldsymbol{\theta}_{i,k}, \mathbf{R}_k)$;
- we use $m = 100$ (approximately) independent simulations obtained by subsampling a Markov chain of length $200,000$.

Next figure: 'residual' plots based on quadrat counts.

- W is divided into 100 equally sized quadrats. For each seeder we count the number of plants within each quadrat.
- The grey scales reflect the probabilities that counts drawn from the posterior predictive distribution are less or equal to the observed quadrat counts where dark means high probability.
- The posterior predictive distribution of the quadrat counts is obtained from posterior predictive samples $\mathbf{y}_{i,k}$, $k = 1, \ldots, 100$, as mentioned above.
- The stars mark quadrats where the observed counts are 'extreme' in the sense of being either below the 2.5% quantile or above the 97.5% quantile of the posterior predictive distribution.

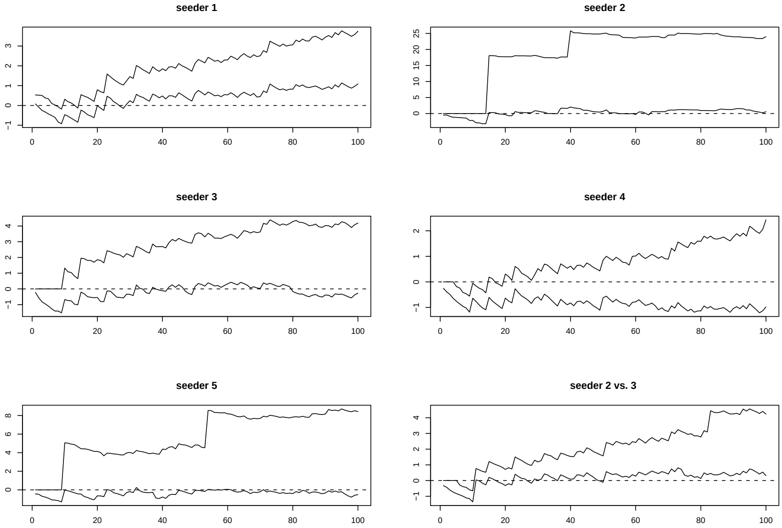

seeder 1
seeder 2
seeder 3
seeder 4
seeder 5
0
50
100
150
200

- The plot for seeder 1 indicate a lack of fit due to many ‘extreme’ counts. However, a ‘systematic’ discrepancy from the assumed model for the intensity is not obvious and the lack of fit could be caused by clustering due to seed dispersal around parent plants.
- The residual plots for the other seeders do not provide obvious evidence against our model except perhaps for a small group of adjacent ‘extreme’ counts for seeder 5.

Next figure:

- Denote by $\hat{L}(r; \mathbf{Y}_i, \boldsymbol{\theta}, \mathbf{R})$ the estimate of the L function obtained from the point process $\mathbf{Y}_i$ and the intensity function corresponding to the interaction parameter vector $\boldsymbol{\theta}$ and interaction radii $\mathbf{R}$.
- Consider the posterior predictive distribution of the differences $\Delta_i(r) = \hat{L}(r; \mathbf{y}_i, \boldsymbol{\theta}_i, \mathbf{R}) - \hat{L}(r; \mathbf{Y}_i, \boldsymbol{\theta}_i, \mathbf{R})$, $r > 0$, $i = 1, \ldots, 5$ (the 5 seeder species), i.e. the distribution obtained when we generate $(\mathbf{Y}_i, \boldsymbol{\theta}_i, \mathbf{R})$ under the posterior predictive distribution given the data $\mathbf{y}$.
- If zero is an extreme value in the posterior predictive distribution of $\Delta_i(r)$ for a range of distances r, we may question the fit of our model.
- As for the quadrat counts, the posterior predictive distribution is computed from a posterior predictive sample $\hat{L}(r; \mathbf{y}_i, \boldsymbol{\theta}_{i,k}, \mathbf{R}_k) - \hat{L}(r; \mathbf{y}_{i,k}, \boldsymbol{\theta}_{i,k}, \mathbf{R}_k)$, $k = 1, \ldots, 100$.

- The figure presents estimated upper and lower boundaries of the 95 % posterior envelopes for the posterior predictive distributions of $\Delta_i(r)$, $r > 0$, for the 5 seeder species.
- The wide envelopes probably arise because of the posterior uncertainty regarding the interaction radii; the intensity function at a seeder location may a posteriori be very variable if it is highly uncertain whether the seeder location falls within a resprouter influence zone or not.
- There is evidence of clustering for seeder 1 and perhaps also for seeders 2, 3, 5. This may be explained by offspring clustering around locations of parent plants.
- To look for interactions between the seeder species, we finally considered cross L functions for the 10 pairs of seeders. The lower right posterior predictive plot for seeder 2 vs. seeder 3 in the figure indicates repulsion at small distances and otherwise positive association between these seeders. The remaining plots (not shown) do not contradict the assumptions of independence between the seeders.


seeder 1
seeder 2
seeder 3
seeder 4
seeder 5
seeder 2 vs. 3

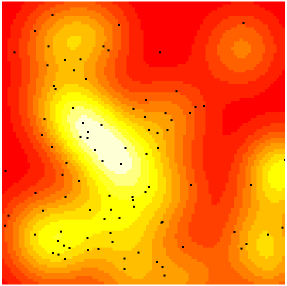
Concluding remarks

- Our analysis shows the difficulty of modelling spatial interactions in a plant community which requires very complex models with a large number of parameters.
- The Bayesian approach is more useful than the frequentist approach as it allowed a more flexible and realistic model.
- Taking biological background information into account in our analysis naturally lead to a hierarchical Bayesian model.
- The model does not sufficiently capture all interactions that may be present in the dataset: It does not consider an intra-species interaction for each seeder type and, similarly, assumes that the seeder species are independent given the resprouters.

- Incorporating all these aspects into a single model, though, is computationally very hard.
- Note that ignoring intra-species interaction does not necessarily invalidate estimates of intensity function parameters (cf. work by Schoenberg (2005) and Waagepetersen (2006)).
- However, it is clear that ignoring clustering leads to too narrow posterior credibility intervals. Hence the results regarding significant parameters should be taken with a pinch of salt.
- From an ecological perspective, we were able both to confirm existing knowledge on species' interactions and to generate new biological questions and hypotheses on species' interactions.

Cox processes

- $\mathbf{X}$ is a Cox process driven by a random intensity function ρ if $\mathbf{X}$ conditional on ρ is a Poisson process with intensity function ρ.



- Thus any Bayesian model for a Poisson process is a Cox process...
- Includes the previous analysis of the parametric model for the seeders conditional on the resprouters.
- Log Gaussian Cox processes (LGCP) and shot noise Cox process (SNCP) are the two most popular model classes. Used for spatial as well as space-time point process modelling of aggregated/clustered point patterns.
  References: See my homepage
  (http://people.math.aau.dk/∼jm/)
  and Peter Diggle's homepage
  (http://www.lancs.ac.uk/∼diggle/).

Log Gaussian Cox processes

- Definition: $\mathbf{X}$ is a LGCP if $\log \rho$ is a Gaussian process (Møller et al. (1998)).
- Moment expressions are very tractable. E.g. $g = \exp(c)$ where c is the covariance function of the Gaussian process.
- Earlier discretizations of the Gaussian process and the use of time-consuming MCMC algorithms (Langevin-Hastings) were used.
- Today software based on INLA (Rue, Martino & Chopin (2009)) provides a very fast way of calculating posterior results for $\log \rho$ and parameters of the Gaussian process (without MCMC!).

Shot noise Cox processes

- Definition: $\mathbf{X}$ is a SNCP if

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

where $k(c,\cdot)$ is a kernel and $\mathbf{\Phi} \sim \text{Poisson}(\mathbb{R}^d \times]0,\infty[, \zeta)$ (Møller (2003) and the references therein).

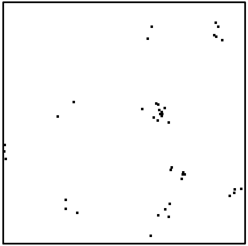
- Then $\mathbf{X}$ can be viewed as a *Poisson cluster process*:

$$\mathbf{X} \sim \cup_{(c,\gamma)\in\mathbf{\Phi}} \mathbf{X}_{(c,\gamma)}$$

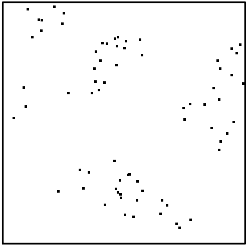
where conditional on $\mathbf{\Phi}$, the $\mathbf{X}_{(c,\gamma)}$ $\sim$Poisson$(\mathbb{R}^d, \gamma k(c,\cdot))$ are independent 'clusters'.

- Matérn cluster process: all $\gamma = \alpha$ (a single parameter), ζ is Lebesgue measure on $\mathbb{R}^d$ times $\kappa\delta(\gamma - \alpha)$ on $]0,\infty[$ where $\kappa > 0$, and $k(c,\cdot)$ is the uniform density on ball(c,r).

Matérn cluster process: cluster centres $\sim$ Poisson$(\mathbb{R}^2, \kappa)$; cluster associated to centre $c \sim$ Poisson(ball$(0, r)$,α).



$\kappa = 10, r = 0.05, \alpha = 5$



$\kappa = 10, r = 0.1, \alpha = 5$

- Conjugated prior if $k(c, \cdot) = \delta(c = \cdot)$ is degenerated and $\mathbf{\Phi}$ is a Poisson-gamma process, but usually we don't want the kernel to be degenerated (Wolpert and Ickstadt...)
- MCMC: If we don't aim at identifying the clusters, it is most convenient to include $\mathbf{\Phi}$ in the posterior and use a hybrid MCMC algorithm, where $\mathbf{\Phi}$ is updated by a Metropolis-Hastings birth-death algorithm (Geyer & Møller (1994)), and using e.g. random walk Metropolis updates of the parameters for the (hyper-)prior model of $\mathbf{\Phi}$.

Finite point processes specified by a density

- Assume $S \subset \mathbb{R}^d$ is bounded and f is a density for a point process $\mathbf{X}$ on S wrt. to the unit rate Poisson process on S, i.e.

$$P(\mathbf{X} \in F) = \sum_{n=0}^{\infty} \frac{\mathrm{e}^{-|S|}}{n!} \int_{S^n} 1[\{x_1, x_2, \dots, x_n\} \in F] \, f(\{x_1, \dots, x_n\}) \mathrm{d}x_1 \dots \mathrm{d}x_n.$$

- Often specified by an unnormalized density:

$$h(\mathbf{x}) = c\, f(\mathbf{x}), \quad \mathbf{x} \subset S \text{ finite}.$$

- Problem: calculation of the normalising constant

$$c = \sum_{n=0}^{\infty} \frac{\mathrm{e}^{-|S|}}{n!} \int_{S^n} h(\{x_1, \dots, x_n\}) \mathrm{d}x_1 \dots \mathrm{d}x_n.$$

Stability conditions and existence

- *Integrability (= existence)*: $c < \infty$ ($c > 0$ usually trivial)
- Suppose $c^* = \int_S K(u)\mathrm{d}u < \infty$ for some $K : S \to [0, \infty)$.
- *Local stability*: $h(x \cup u) \leq K(u)h(x)$
  (where $x \cup u = x \cup \{u\}$).
- *Ruelle stability*: $h(x) \leq \alpha \prod_{u \in X} K(u)$ for some $\alpha < \infty$.
- Proposition:
  Local stability $\Rightarrow$ Ruelle stability $\Rightarrow$ integrability.

Hereditary condition and Papangelou conditional intensity

- Hereditary density: f (or h) is *hereditary* if $f(\mathbf{y}) > 0 \Rightarrow f(\mathbf{x}) > 0$ whenever $\mathbf{x} \subset \mathbf{y}$.
- *Papangelou conditional intensity*:

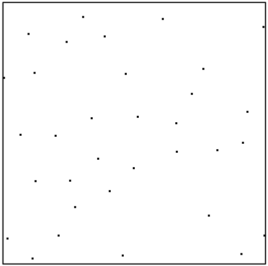
$$\lambda(\mathbf{x}, u) = \frac{f(\mathbf{x} \cup u)}{f(\mathbf{x})} = \frac{h(\mathbf{x} \cup u)}{h(\mathbf{x})}, \qquad u \notin \mathbf{x},$$

  where $a/0 = 0$ for all a. (NB: does not depend on c !!)
- Interpretation: $\lambda(\mathbf{x}, u)\mathrm{d}u$ is the probability of having a point in an infinitesimal region around u given the rest of $\mathbf{X}$ is $\mathbf{x}$.
- For the Poisson process with intensity $\rho(u)$,

$$\lambda(\mathbf{x}, u) = \rho(u).$$

Example: Strauss (1975) process

- Density: $f(\mathbf{x}) = \frac{1}{c}\beta^{n(\mathbf{x})}\gamma^{s(\mathbf{x})}$, where $\beta, \gamma \geq 0$, and $s(\mathbf{x})$ is the number of pairs of points within distance R.
- $\lambda(\mathbf{x}, u) = \beta\gamma^{s(\mathbf{x},u)}$ where $s(\mathbf{x}, u)$ is the number of R-close points in $\mathbf{x}$ to u. So $\mathbf{X}$ exists and is repulsive if $\gamma \leq 1$. (Non-existence if $\gamma \leq 1$, cf. Kelly & Ripley (1976)).



$S = [0, 1] \times [0, 1], \beta = 100, \gamma = 0, R = 0.1.$

Pairwise interaction process

- A pairwise interaction density is of the form

$$f(\mathbf{x}) \propto \prod_{u \in \mathbf{x}} \varphi(u) \prod_{\{u,v\} \subseteq \mathbf{x}} \varphi(\{u,v\}), \quad \varphi(\cdot) \geq 0.$$

- This is hereditary and

$$\lambda(\mathbf{x}, u) = \varphi(u) \prod_{v \in \mathbf{x}} \varphi(\{u,v\}).$$

- (((Markov w.r.t. $u \sim v$ iff $\varphi(\{u,v\}) \neq 1$)))
- If $\varphi(\{u,v\}) \leq 1$, then locally stable and $\mathbf{X}$ is repulsive.
- If $\varphi(\{u,v\}) \geq 1$, then usually $\mathbf{X}$ does not exist.
- Simple example of a *finite Gibbs point process*. By including higher order interaction terms, we obtain a general finite Gibbs point process. In turn this can be extended to *infinite Gibbs point processes*... In most cases these are also models for repulsive/regular point patterns.

Simulation of finite Gibbs point processes

Usually of birth-death types (add/delete one point) and based on $\lambda(\mathbf{x}, u)$ only.

- Geyer & Møller (1994): Metropolis-Hastings birth-death algorithm. (Special case of Green's reversible jump MCMC.)
- Kendall & Møller (2000): Spatial birth-death processes and dominating coupling from the past $\rightarrow$ perfect simulation algorithm.

Likelihoods with “unknown” normalizing constants

Consider a parametric model with likelihood

$$l(\theta|\mathbf{y}) = f_\theta(\mathbf{y}) = \frac{1}{Z_\theta} q_\theta(y)$$

where

- $q_\theta(y)$ is a known unnormalized density,
- Z_θ is an intractable normalizing constant.

Examples:

- Finite Gibbs point processes.
- (((Finite Gibbs/Markov random fields.)))

Posterior

- Impose a prior $\pi(\theta)$.

- Posterior
$$\pi(\theta|y) \propto \pi(\theta)q_\theta(y)/Z_\theta$$
depends on Z_θ; and conventional Metropolis-Hastings algorithms for simulation from the posterior depends on ratios of normalizing constants!?.

Auxiliary variable technique

Møller et al. (2004, 2006): First truly “exact”/”pure” MCMC algorithm for performing Bayesian inference for models with intractable normalising constants.

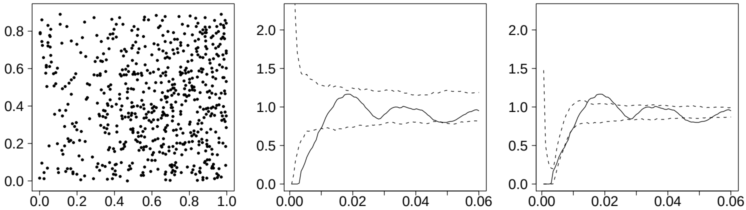
Murray et al. (2006): Exchange algorithm—sligthly simpler and more efficient.

- Both algorithms are based on perfect simulation of an auxiliary variable x generated from the observation model—or runnung an MCMC algorithm for long enough...
- They don't depend on the intractable normalizing constant.

Example of Bayesian inference for a pairwise interaction point process

K.K. Berthelsen and J. Møller (2008). Non-parametric Bayesian inference for inhomogeneous Markov point processes. *Australian and New Zealand Journal of Statistics*, 50, 627-649.

Data



Left: Locations of 617 cells in a 2D section of the mocous membrane of the stomach of a healthy rat (LHS: stomach cavity begins; RHS: muscle tissue begins).
Centre: Non-parametric estimate $\hat{g}$ and 95%-envelopes calculated from 200 simulations of a fitted inhomogeneous Poisson process.
Right: $\hat{g}$ and 95%-envelopes calculated from 200 simulations of the model fitted by Nielsen (2000) (non-Bayesian; to obtain inhomogeneity, she considered a transformation of a Strauss point process...).

Inhomogeneous pairwise interaction point process

Suppose the likelihood is given by the density

$$f_{\beta,\varphi}(\mathbf{y}) = \frac{1}{Z_{\beta,\varphi}} \prod_i \beta(y_i) \prod_{i<j} \varphi(\|y_i - y_j\|)$$

w.r.t. Poisson$(W, 1)$ where

- $W = [0, a] \times [0, b]$ is the observation window;
- $\beta(u_1, u_2) = \beta(u_1) \geq 0$ models the horizontal inhomogeneity;
- $0 \leq \varphi(\cdot) \leq 1$ is a non-decreasing pairwise interaction function.

Prior for $\beta(u_1, u_2) = \beta(u_1)$

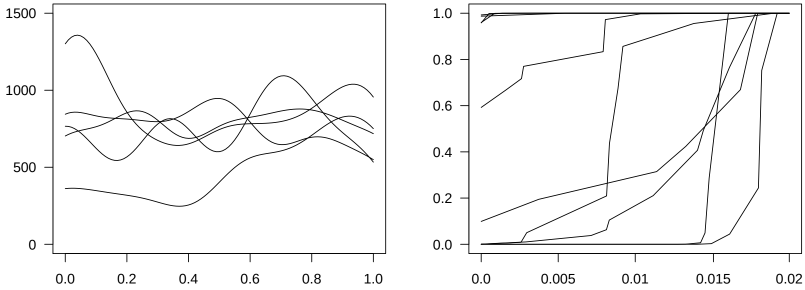
Shot noise process

$$\beta(u_1) = \gamma \sum_j \varphi\left(\frac{u_1 - c_j}{\sigma_1}\right) / \sigma_1$$

where φ is the $N(0,1)$-density;
$\psi = \{c_j\} \sim$Poisson$([-\Delta, a+\Delta], \kappa_1)$,
and independently of ψ, we impose a Gamma prior for $\gamma > 0$.

- The higher κ_1, the more kernels and more felexibilty.
- On the other hand, a high value of κ_1 leads to slow mixing in our MCMC algorithm for the posterior.
- Detailed discussion in the paper on the choice of $\Delta > 0$, $\kappa_1 > 0$, $\sigma_1 > 0$, and the Gamma prior for $\gamma > 0$.

Examples of prior realizations



Left panel: Five independent realisations of β under its prior distribution.
Right panel: Ten independent realisations of φ under its prior distribution.

Prior for φ

A first prior for φ:

$$\varphi(r) = \mathbf{1}[r > r_p] + \sum_{i=1}^{p} \mathbf{1}[r_{i-1} < r \leq r_i] \left(\frac{r - r_{i-1}}{r_i - r_{i-1}} (\gamma_{i+1} - \gamma_i) + \gamma_i \right)$$

where

- $r_1 < \ldots < r_p$ follow Poisson($[0, r_{\max}], \kappa_2$);
- $0 < \gamma_1 < \ldots < \gamma_p < \gamma_{p+1} = 1$ and setting $\gamma_0 = 0$, $\delta_i = \gamma_i - \gamma_{i-1}$,, $(\zeta_1, \ldots, \zeta_p) = (\ln(\delta_2/\delta_1), \ldots, \ln(\delta_{p+1}/\delta_p))$, then conditionally on $(r_1, \ldots, r_p)$, $\zeta_p, \ldots, \zeta_1$ is a Markov chain (random walk) with $\zeta_p \sim \mathrm{N}(0, \sigma_2^2)$ and $\zeta_i | \zeta_{i+1} \sim \mathrm{N}(\zeta_{i+1}, \sigma_2^2)$, $i = p-1, \ldots, 1$.
- E.g. $r_{\max} = 0.02$. See the paper for the choice of hyperparameters $\kappa_2 > 0$ and $\sigma_2 > 0$.

Posterior

Recall that β is specified by the Poisson process ψ and the Gamma variate γ, and φ by the marked Poisson process $\chi = \{(r_1, \gamma_1), \ldots, (r_p, \gamma_p)\}$.
The posterior density for $\theta = (\psi, \gamma, \chi)$

$$\pi(\theta|y) \propto \kappa_1^{n(\psi)} \gamma^{\alpha_1 - 1} \mathrm{e}^{-\gamma/\alpha_2} \kappa_2^p \mathbf{1}[0 < \gamma_1 < \ldots < \gamma_p < 1]/(\delta_1 \times \cdots \times \delta_{p+1})$$
$$\times \left(2\pi\sigma_2^2\right)^{-p/2} \exp\left(- \sum_{i=1}^{p} \left(\zeta_i - \zeta_{i+1}\right)^2 / \left(2\sigma_2^2\right) \right)$$
$$\times \frac{1}{Z_\theta} \prod_i \beta_{\psi,\gamma}(y_i) \prod_{i<j} \varphi_\chi(\|y_i - y_j\|)$$

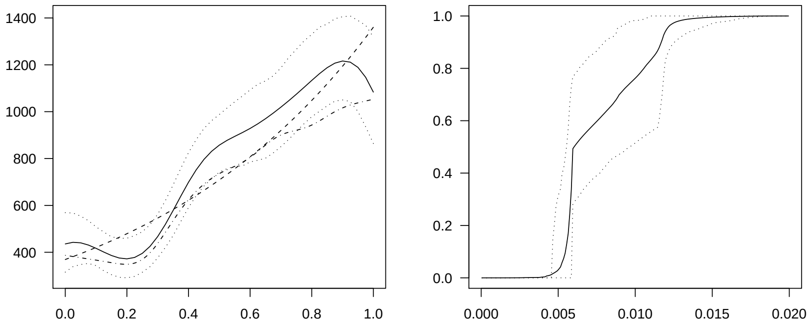
depends on Z_θ. So we apply the auxiliary variable algorithm...

Second prior for φ

A first Bayesian analysis indicated the need for including a hard core parameter $h \sim$Uniform$[0, r_{\max}]$:

$$\varphi_{\text{new}}(r; h, \chi) = \begin{cases} 0 & \text{if } r < h \\ \varphi_{\text{old}}\left(\frac{(r-h)r_{\max}}{r_{\max}-h}; \chi\right) & \text{if } h \leq r \leq r_{\max} \\ 1 & \text{if } h > r_{\max} \end{cases}$$

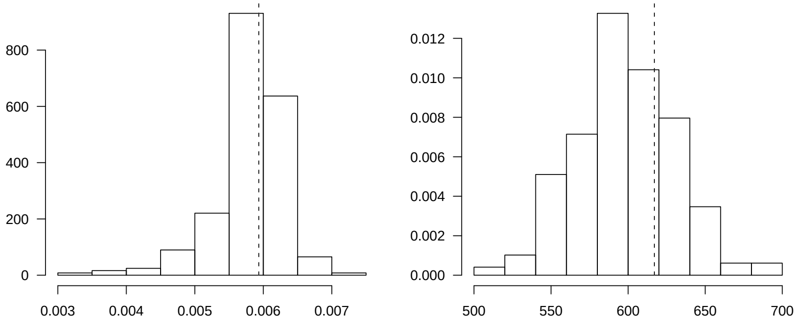
Some final posterior results



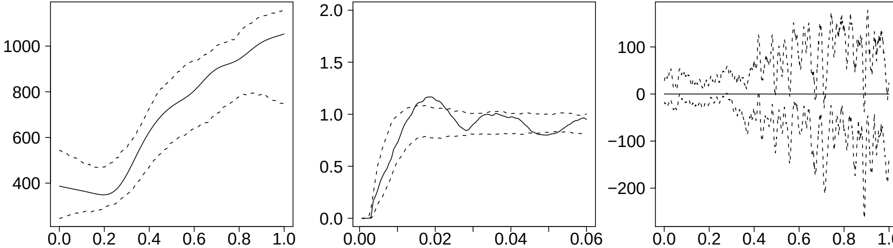
Solid line: Posterior mean for β (left) and φ (right).
Dotted lines: Pointwise 95% central posterior intervals.
Dashed line (left): β estimated by Nielsen (2000).
Dot-dashed line (left): Non-parametric estimate of β.

Some results for model checking

Consider the posterior predictive distribution.



Observed value (dashed line) and posterior predictive distribution of minimum inter-point distance (left panel) and number of points (right panel).



Left and centre panels: Observed (solid lines) non-parametric estimates $\hat{\rho}(u_1)$ (left panel) and $\hat{g}(r)$ (middle panel) together with pointwise 95% central posterior predictive intervals (dashed lines).
Right panel: Ignored in this talk.

Summary on Bayesian statistics for spatial point processes

- Poisson point processes: likelihood term is tractable, so rather straightforward (using MCMC or possibly even simpler methods).
- Cox processes: Include the unobserved random intensity into the posterior...
  - For a LGCP, as the Gaussian process on the observation window is not observed include this (approximated on a grid) into the posterior and use INLA in a hybrid MCMC algorithm.
  - For a SNCP, as the centre process is not observed, include this into the posterior and use for this the Metropolis-Hastings birth-death algorithm in a hybrid MCMC algorithm.
- Gibbs point processes: Here the problem is the intractable normalizing constant of the likelihood which also enters in the posterior. Use the auxiliary variable method.

Some literature (most material is non-Bayesian!)

- P.J. Diggle (2003). Statistical Analysis of Spatial Point Patterns. Arnold, London. (Second edition.)
- J. Møller and R.P. Waagepetersen (2004). Statistical Inference and Simulation for Spatial Point Processes. Chapman and Hall/CRC, Boca Raton.
- J. Møller and R.P. Waagepetersen (2007). Modern statistics for spatial point processes (with discussion). *Scandinavian Journal of Statistics*, **34**, 643-711.
- J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan (2008). Statistical Analysis and Modelling of Spatial Point Patterns. John Wiley and Sons, Chichester.
- A.E. Gelfand, P. Diggle, M. Fuentes, and P. Guttorp (2010). A Handbook of Spatial Statistics. Chapman and Hall/CRC. (Chapter 4)
- W.S. Kendall and I. Molchanov (eds.) (2010). New Perspectives in Stochastic Geometry. Oxford University Press, Oxford.

Some literature on Bayesian statistics (own work plus paper by Guttorp and Thorarinsdottir)

- P.G. Blackwell and J. Møller (2003). Bayesian analysis of deformed tessellation models. *Advances in Applied Probability*, 35, 4-26.
- Ø. Skare, J. Møller and E.B.V. Jensen (2007). Bayesian analysis of spatial point processes in the neighbourhood of Voronoi networks. *Statistics and Computing*, 17, 369-379.
- V. Benes, K. Bodlak, J. Møller and R.P. Waagepetersen (2005). A case study on point process modelling in disease mapping. *Image Analysis and Stereology, 24*, 159 - 168.
- J. Møller and R.P. Waagepetersen (2007). Modern statistics for spatial point processes (with discussion). *Scandinavian Journal of Statistics*, **34**, 643-711.

- K.K. Berthelsen and J. Møller (2008). Non-parametric Bayesian inference for inhomogeneous Markov point processes. *Australian and New Zealand Journal of Statistics*, 50, 627-649.
- J.B. Illian, J. Møller and R.P. Waagepetersen (2009). Hierarchical spatial point process analysis for a plant community with high biodiversity. *Environmental and Ecological Statistics*, 16, 389-405.
- P. Guttorp and T.L. Thorarinsdottir (2012) Bayesian inference for non-Markovian point processes (Chapter 4 in E. Porcu et al. (eds.) (2012) Advances and Challenges in Space-time Modelling and Natural Events).
- J. Møller and H. Toftager (2012). Geometric anisotropic spatial point pattern analysis and Cox processes. Research Report R-2012-01, Department of Mathematical Sciences, Aalborg University. Submitted for journal publication.

- J. Møller and J.G. Rasmussen (2012). A sequential point process model and Bayesian inference for spatial point patterns with linear structures. To appear in *Scandinavian Journal of Statistics*, 39.

Other talk...