INTRODUCTION TO SPATIAL POINT PROCESSES AND SIMULATION-BASED INFERENCE

Jesper Møller
Department of Mathematics
Aalborg University
Denmark
http://people.math.aau.dk/∼jm/

August 30, 2018

1 FUNDAMENTAL CONCEPTS AND TOOLS

1.1 INTRODUCTION

Some literature

These lectures are mainly based on

- 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. Møller and R.P. Waagepetersen (2017). Some recent developments in statistics for spatial point patterns. *Annual Review of Statistics and Its Applications*, **4**, 317-342.

Some other literature

See also the monographs

- P.J. Diggle (2013). Statistical Analysis of Spatial and Spatio-Temporal Point Patterns. Chapman and Hall/CRC, London. (Third edition.)
- J. Illian, A. Penttinen, H. Stoyan and D. Stoyan (2008). Statistical Analysis and Modelling of Spatial Point Patterns. John Wiley and Sons, Chichester.
- A. Baddeley, E. Rubak and R. Turner (2015). Spatial Point Patterns: Methodology and Applications with R. Chapman and Hall/CRC Press, London.

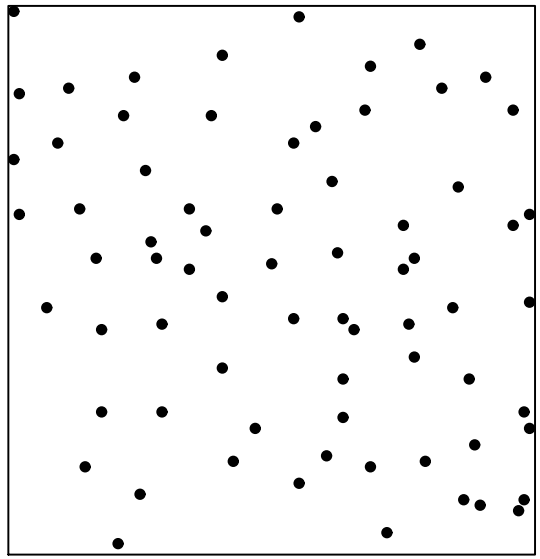
For references to others, see above or look at my homepage: http://people.math.aau.dk/∼jm/

Point processes in R

- `R`-package for dealing with spatial point processes: `spatstat`
- Homepage: `www.spatstat.org`
  Here you can find manuals, news, etc.
- Many algorithms implemented for
  - parameter estimation
  - simulation
  - model checking
- R codes related to these lectures can be downloaded:
  //people.math.aau.dk/∼jm/name
  with name = Rcode1Poisson.R, Rcode2SummaryStat.R,
  Rcode3Cox.R, Rcode4Markov.R, or Rcode5Estimation.R

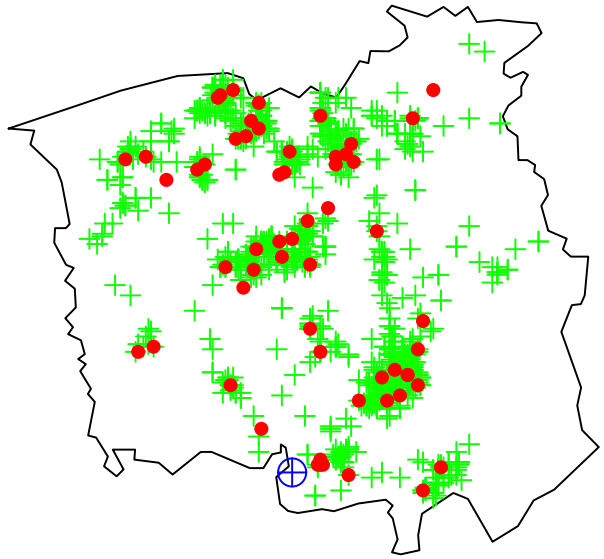
1.2 DATA EXAMPLES AND STATISTICAL INFERENCE

Data example 1: Swedish pines data



- Black rectangle: The observation window, 9.6×10 meters.
- Black dots: positions of pine saplings (young trees) in a Swedish forest (a point pattern).
- Note that a point pattern dataset consists of both the observation window and the point pattern.

Data example 2: Chorley-Ribble data

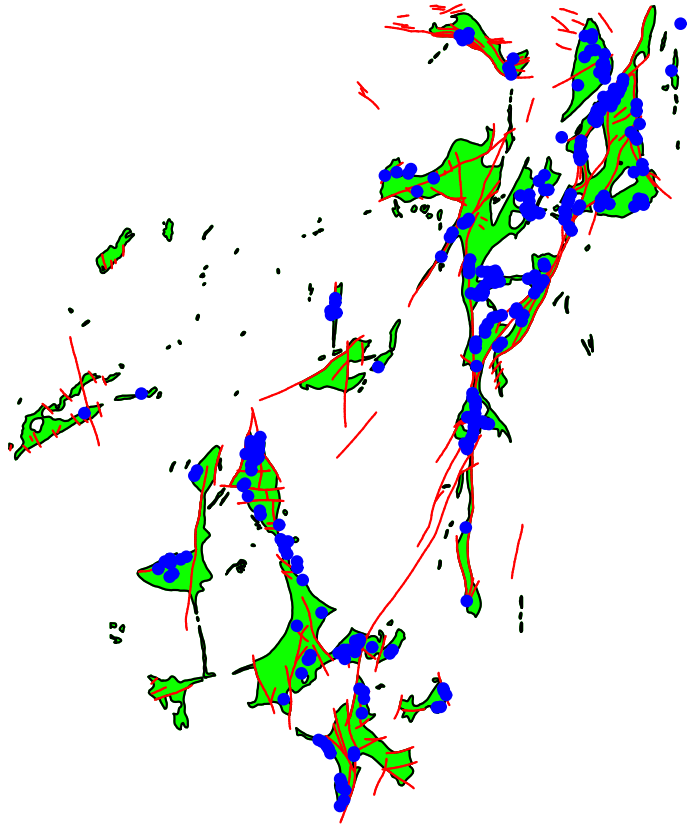


- Black boundary: The Chorley-Ribble region in Britain (observation window)
- Red dots: Cases of larynx cancer (a point pattern)
- Green plusses: Cases of lung cancer (another point pattern)
- Blue plus/circle: An old incinerator

Data example 2: Chorley-Ribble data

- Note that this dataset consists of
  - the observation window
  - two point patterns
  - covariate information (any other potentially relevant information, in this case the position of the incinerator)
- Question: Does the incinerator cause larynx cancer, i.e. is there a higher risk of cancer close to the incinerator?
- Statistical analysis is complicated by the spatial aspects of the data: There is a different amount of people that can get cancer, due to the (unknown!) non-constant population density. The irregular shape of the observation window may complicate "things".

Data example 3: Murchison data



- Green area: An area of greenstone (the observation window).
- Red lines: Faults in the ground (covariate information).
- Blue dots: Gold deposits (the point pattern).
- Question: Are the faults (easily observable) a good indicator of where to dig for gold?

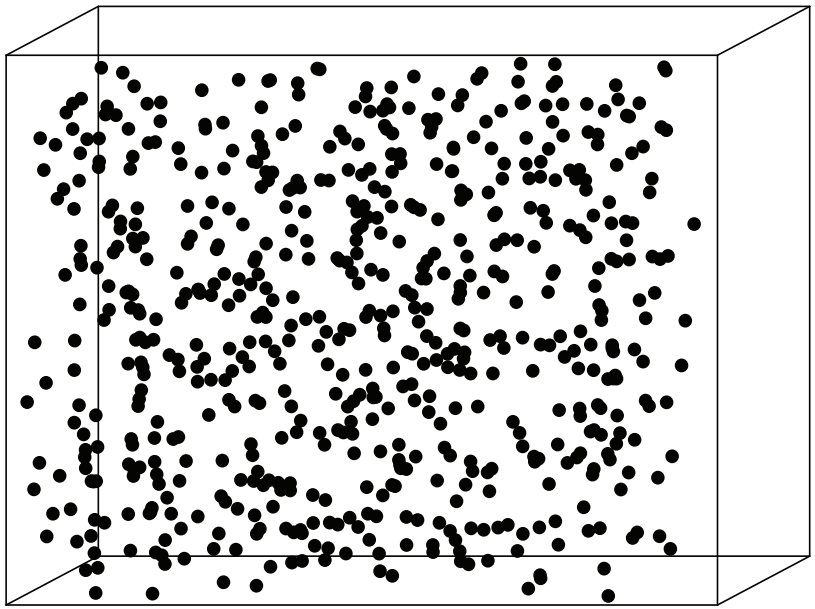
Other examples of data

- One-dimensional point patterns:
  - Times of earthquakes in some region
  - Times of market crashes
  - Times of production failures in a factory
  - Positions of car accidents on a highway during a month
- Two-dimensional point patterns:
  - Positions of restaurants in a city
  - Positions of farms with mad cow disease in some region
  - Positions of broken wires in an electrical network
- Three-dimensional point patterns:
  - Positions of stars or galaxies in the visible part of the universe
  - Positions of copper deposits underground
  - Times and (2D-)positions of earthquakes

Note: In practice, the observation windows are typically bounded subsets of $\mathbb{R}^d$. Some techniques are available specifically for point patterns on the time line (directional), but in these lectures we focus on models and methods useful for any dimension (not directional if $d \geq 2$).

Yet another example of a 3D point pattern dataset

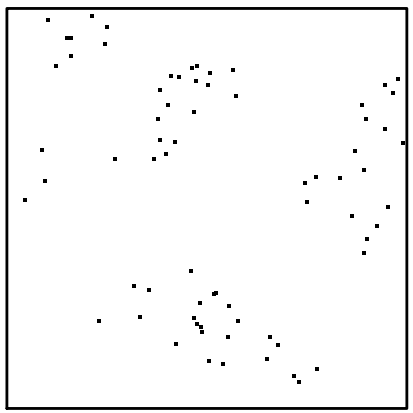
Locations of 623 pyramidal cells in a $508 \times 138 \times 320$ μm^3 large region (Brodmann area 4 of the grey matter of a human brain) :



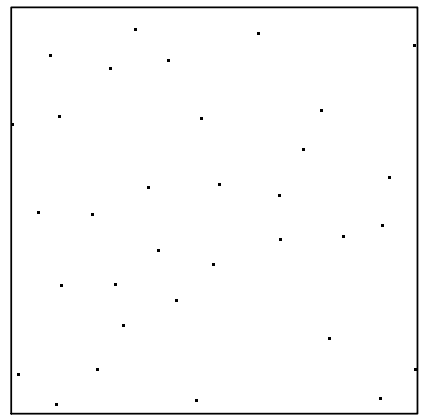
Minicolumn hypothesis (Mountcastle, 1957): cells developed in the centre of the brain and traversed radial paths to the pial surface $\rightarrow$ a columnar arrangement perpendicular to the pial surface of the brain, i.e., parallel to the z-axis.

Statistical inference for spatial point patterns

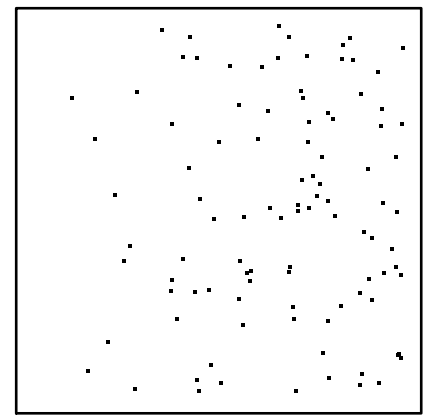
- Objective is to infer structure in spatial distribution of points:
  - interaction between points: regularity or clustering ('random')
  - inhomogeneity linked to covariates ('systematic')
  - to investigate an hypothesis (e.g. the minicolumn hypothesis)



Clustered



Regular



Inhomogeneous

- Briefly, spatial point processes are stochastic models for spatial point patterns. Such models and various statistical tolls have been developed depending on the problem and the type of spatial point pattern dataset.

1.3 SPATIAL POINT PROCESSES AND SIMPLE MODELS

What is a spatial point process?

We consider only point patterns $\subset \mathbb{R}^d$ which are locally finite.

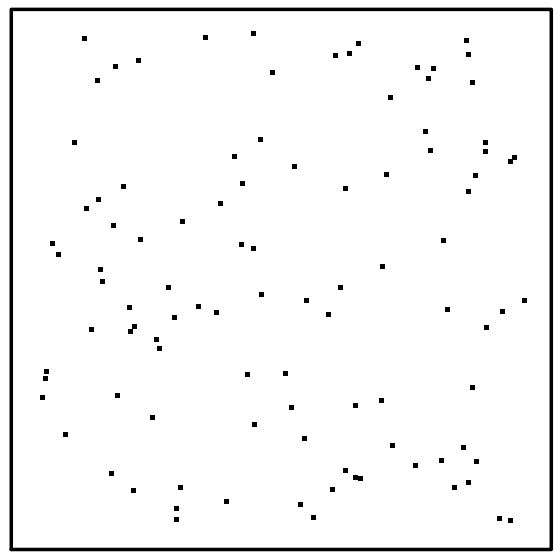
- **Definition:** *A spatial point process* $\mathbf{X}$ *is a locally finite random subset of* $\mathbb{R}^d$.
- Equivalently we have a counting measure $N(A) := \#\mathbf{X} \cap A$ for (Borel) sets $A \subset \mathbb{R}^d$.
- *Measurability* of X means that $N(A)$ is a random variable for any bounded (Borel) set $A \subset \mathbb{R}^d$.
- Fact: the distribution of $\mathbf{X}$ is uniquely determined by the *void probabilities*

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

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

Simple example of point process: Binomial point process

- Let f be a probability density function (pdf) on a (Borel) set $S \subseteq \mathbb{R}^d$. Then $\mathbf{X}$ is a *binomial point process* with n points in S and pdf f if $\mathbf{X}$ consists of n *iid* points, each with pdf f.

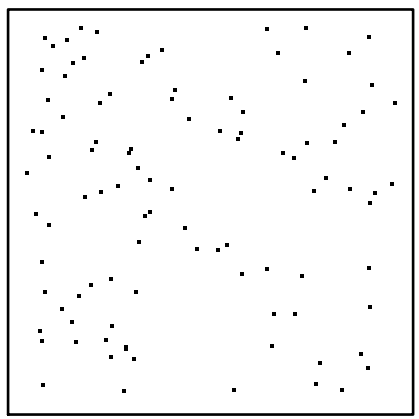


(Example with $S = [0,1] \times [0,1]$, $n = 100$ and $f(x) = 1$.)

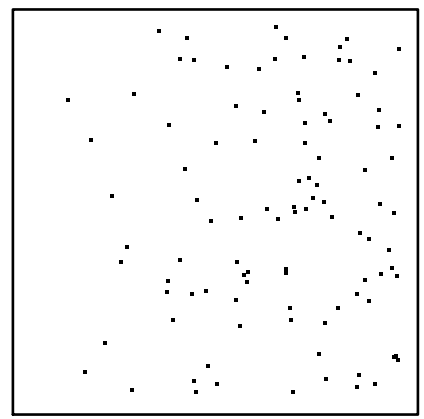
- 'Binomial' since the number of points falling in a (Borel) subset $A \subseteq S$ is binomially distributed $b(n,p)$ with $p = \int_A f(x)\,\mathrm{d}x$.

Fundamental example: The Poisson process

- Assume μ locally finite measure on a given (Borel) set $S \subseteq \mathbb{R}^d$ s.t. $\mu(B) = \int_B \rho(u)\,\mathrm{d}u$ for all (Borel) sets $B \subseteq S$.
- **Def: 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 $\mathbf{X} \cap B$ i.i.d. with density $\propto \rho(u)$, $u \in B$ (i.e. a binomial point process).
- Examples on $S = [0,1] \times [0,1]$:



Homogeneous: $\rho = 100$



Inhomogeneous: $\rho(x, y) = 200x$

Properties of the Poisson process

Using the definition and void probabilities it is fairly easy to show

- the existence of the Poisson process;
- that the class of Poisson processes is closed under
  - independent thinning,
  - independent super-positioning,
  - independent random shifts.

It is easy to simulate a Poisson process X within a bounded set $B \subseteq S$ if $\sup_B \rho \leq \rho_{\max}$:

- Straightforward to generate homogeneous Poisson process $\mathbf{X}_{\max}$ on B with intensity $\rho_{\max}$.
- Obtain $\mathbf{X} \cap B \subseteq \mathbf{X}_{\max}$ by independent thinning, with retention probability $\rho(x)/\rho_{\max}$ for each $x \in \mathbf{X}_{\max}$.

Poisson processes in R

Have a look at some of the things we can do with Poisson processes in R using the `spatstat` package... (Rcode1Poisson.R)

Two fundamental properties: Stationarity and isotropy

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

$$\mathbf{X} \sim \{u + s | u \in \mathbf{X}\} \qquad \text{for any } s \in \mathbb{R}^d.$$

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

$$\mathbf{X} \sim \{Ru | u \in \mathbf{X}\}$$

  for any $d \times d$ rotation matrix R.
- A Poisson process on $\mathbb{R}^d$ with constant intensity ρ is both stationary and isotropic.

Overview of the rest of the lectures

- Summary statistics – useful tools for preliminary analysis and model checking.
- Cox processes – models for aggregated/clustered point patterns.
- Markov processes – mostly models for regular/inhibitive point patterns.
- Simulation of point processes, including MCMC based simulation.
- Inference – mainly estimation of parameters.

1.3 (FUNCTIONAL) SUMMARY STATISTICS

Summary statistics

- Summary statistics are numbers or functions describing characteristics of point processes, for example:
  - The mean number of points in a set $B \subseteq \mathbb{R}^d$.
  - The covariance of the number of points in sets A and B.
  - The mean number of further points within distance $r > 0$ of an 'arbitrary point of the process'.
  - The probability that there is no further point within distance R of an 'arbitrary point of the process'.
- They are useful for:
  - Preliminary analysis.
  - Model fitting (minimum contrast estimation).
  - Model checking.

Another useful tool is 'residuals', see
Baddeley, Turner, M. and Hazelton (2005),
Baddeley, M. and Pakes (2008),
Baddeley, Rubak and M. (2011).

Notation and conventions

- Whenever we consider sets $S, B, \ldots \subseteq \mathbb{R}^d$, they are assumed to be Borel sets. (In practice: don't worry!)
- 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;
  - for a finite set $\mathbf{x} \subset \mathbb{R}^d$, $n(\mathbf{x})$ is the number of points in $\mathbf{x}$;
  - so $N(B) = n(\mathbf{X}_B)$.

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 $\mathbf{X}$ is *inhomogeneous*.

Non-parametric estimation of ρ (homogeneous case)

Let $W \subset \mathbb{R}^d$ with $0 < |W| < \infty$. Suppose $\mathbf{X}_W$ is homogeneous and a realization is observed.
Natural estimate of ρ:

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

because

- $\mathbb{E}\hat{\rho} = \rho$;
- in case of a homogeneous Poisson process, $\hat{\rho}$ is the MLE.

Non-parametric estimation of ρ (inhomogeneous case)

Let $W \subset \mathbb{R}^d$ with $0 < |W| < \infty$. Suppose $\mathbf{X}_W$ is inhomogeneous and a realization is observed.
Non-parametric estimate of $\rho(u)$ (Diggle, 1985):

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

where

- k is *kernel* (i.e. a pdf),
- $c_W(v) = \int_W k(u - v)\,\mathrm{d}u$ is an *edge-correction factor*.

Then $\int_W \hat{\rho}(u)\,\mathrm{d}u$ is an unbiased estimate of $\mu(W)$.

Practical issues in kernel estimation

- Typically a kernel depends on a band-width $b > 0$.
- For example, an isotropic mean-zero normal density function is a kernel with band-width $b = \sigma$:

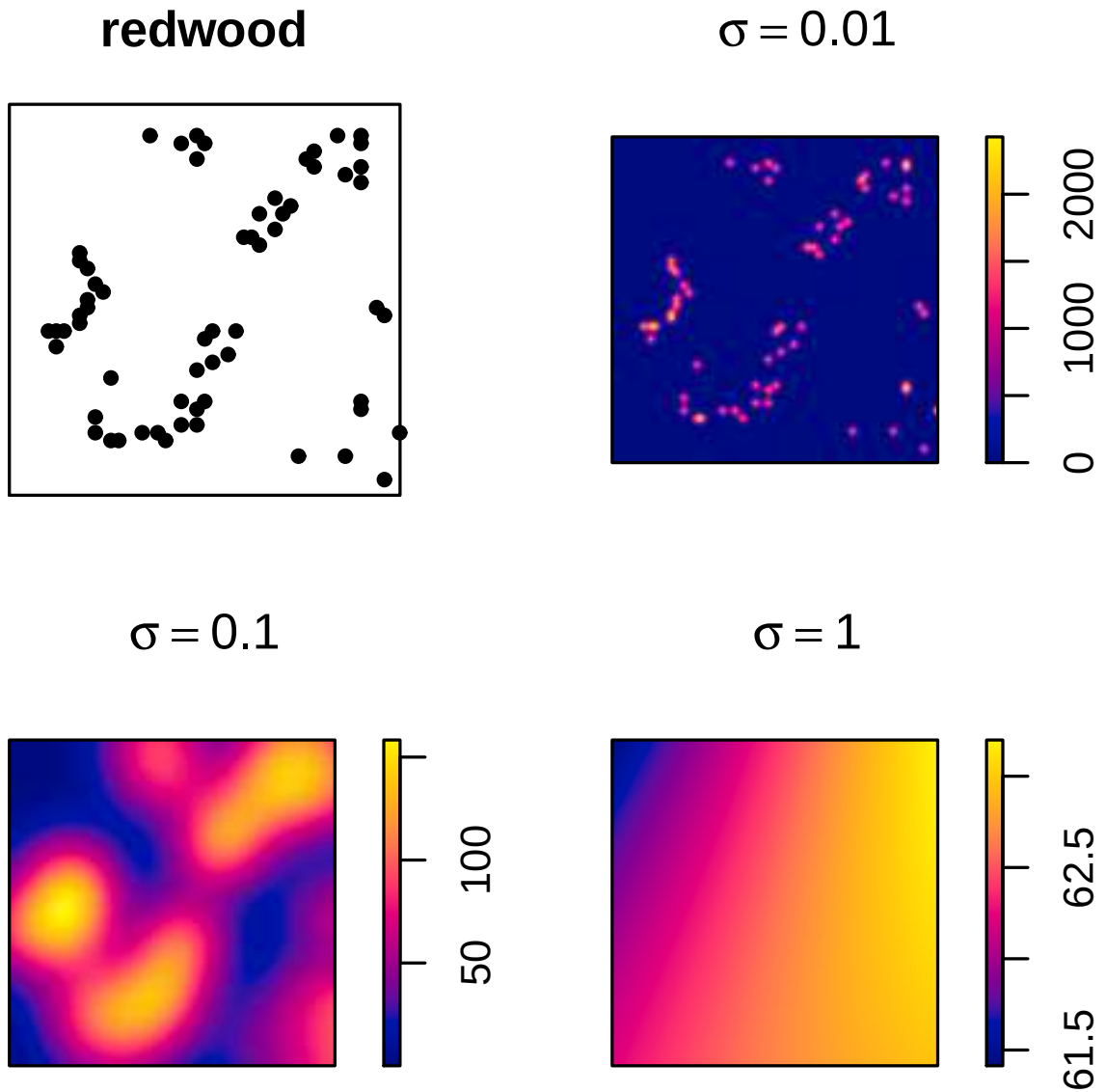
$$k(u) = \frac{1}{(\sqrt{2\pi}\sigma)^d} \exp\left(-\frac{\|u\|^2}{2\sigma^2}\right).$$

- In general,

$$k_b(x) = k_1(x/b)/b^d$$

  where k_1 is a given pdf.
- In practice, the shape of k is less important than b.
- Sensitive to the choice of 'band-width'...
  (if covariate information is available, a parametric model for ρ may be preferred).

Examples when using a normal kernel



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/second order intensity* $\rho^{(2)}$:

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

- Infinitesimal interpretation of $\rho^{(2)}$: For $u \in A$ and $v \in B$ with $|A| = \mathrm{d}u$, $|B| = dv$, and $A \cap B = \emptyset$,

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

- 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 Poisson process

- If $\mathbf{X}$ is a Poisson process with intensity function ρ, then

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

- Proof: The so-called extended Slivnyak-Mecke formula (omitted here).

Pair correlation function (pcf)

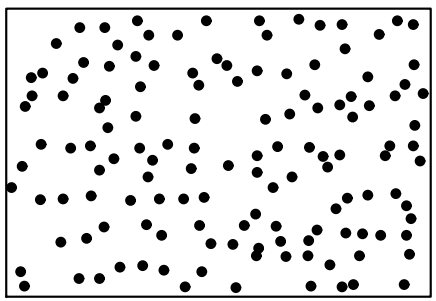
The pcf is defined by

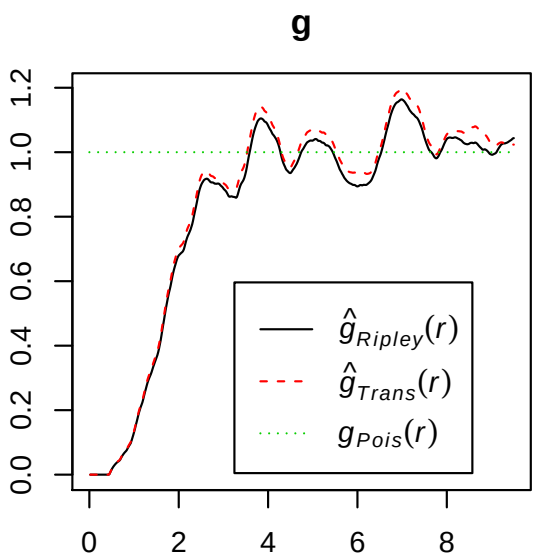
$$g(u,v) = \frac{\rho^{(2)}(u,v)}{\rho(u)\rho(v)} \approx \frac{P(\mathbf{X} \text{ has a point in each of } A \text{ and } B)}{P(\mathbf{X} \text{ has a point in } A)P(\mathbf{X} \text{ has a point in } B)}$$

(here $u \in A$ and $B \in v$ are infinitesimally small sets, and $a/0 := 0$ for $a \in \mathbb{R}$).

- Interpretation of pcf:
  - Poisson process: $g(u,v) = 1$.
  - If $g(u,v) > 1$ when u and v are not "too far", then realizations tend to be more aggregated/clustered than for a Poisson process.
  - If $g(u,v) < 1$ when u and v are not "too far", then realizations tend to be more regular than for a Poisson process.
- If X is stationary, then $g(u,v) = g(u-v)$, and if X is also isotropic, then $g(u,v) = g(r)$ where $r = \|u-v\|$.
- Non-parametric kernel estimation can be used (sensitive to choice of band width).

Example: Norwegian spruces in a 56×38 m region





ρ, $\rho^{(2)}$ and g under thinning

- If $\mathbf{X}_{\text{thin}}$ is a thinned process obtained from $\mathbf{X}$ with retention probability $p(u)$, then using an obvious notation,
  1. $\rho_{\text{thin}}(u) = p(u)\rho(u)$,
  2. $\rho^{(2)}_{\text{thin}}(u,v) = p(u)p(v)\rho^{(2)}(u,v)$,
  3. $g_{\text{thin}}(u,v) = g(u,v)$.
- Proof (sketch): We can let

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

where for all $u \in \mathbb{R}^d$ the $R(u)$ are independent uniform random variables on $[0,1]$ which are independent on $\mathbf{X}$.
Then consider μ_{thin} and $\alpha^{(2)}_{\text{thin}}$, where we first condition on the $R(u)$ and next use that for non-negative (Borel) functions h_1 and h_2,

$$\mathbb{E}\sum_{u\in\mathbf{X}} h_1(u) = \int h_1(u)\rho(u)\,\mathrm{d}u,$$

$$\mathbb{E}\sum_{u,v\in\mathbf{X}}^{\neq} h_2(u,v) = \int\int h_2(u,v)\rho^{(2)}(u,v)\,\mathrm{d}u\,\mathrm{d}v.$$

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 if $0 < |A| < \infty$,

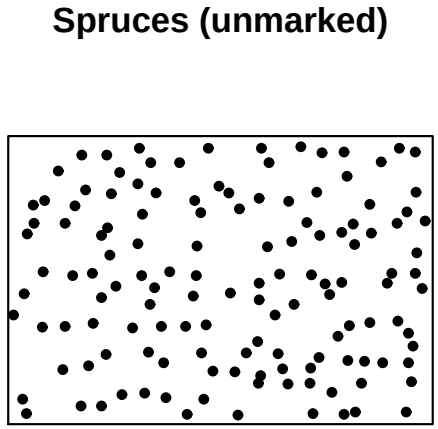
$$K(r) = \mathbb{E}\frac{1}{\rho^2|A|}\sum_{u\in\mathbf{X}_A}\sum_{v\in\mathbf{X}\setminus\{u\}} 1[\|u - v\| \le r], \quad r > 0.$$

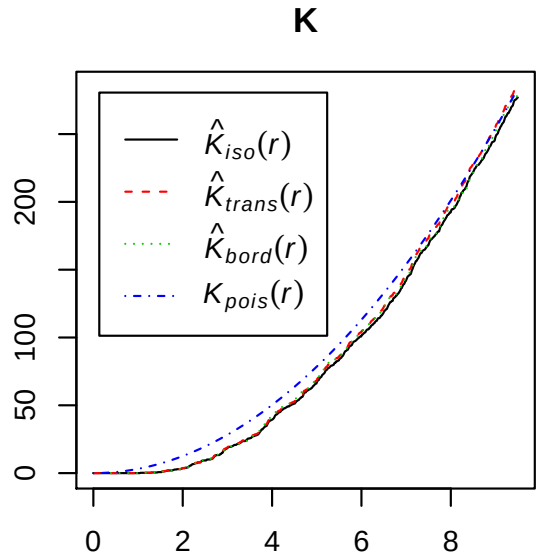
- Interpretation: $\rho K(r)$ is the expected number of further points within distance r of an arbitrary point of X.
- *Besag's L-function* (variance stabilizing transformation):

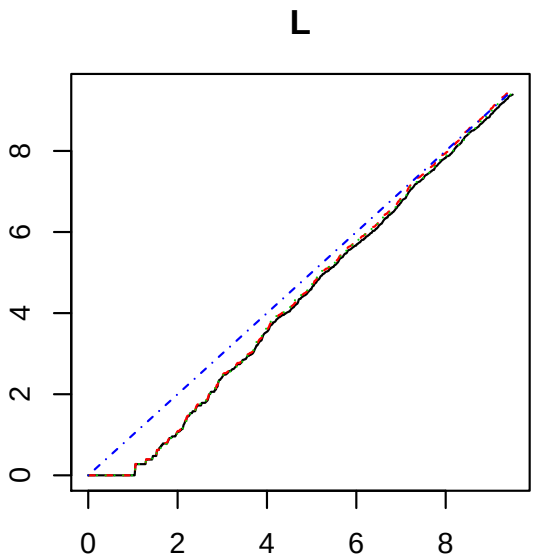
$$L(r) = (K(r)/\omega_d)^{1/d}$$

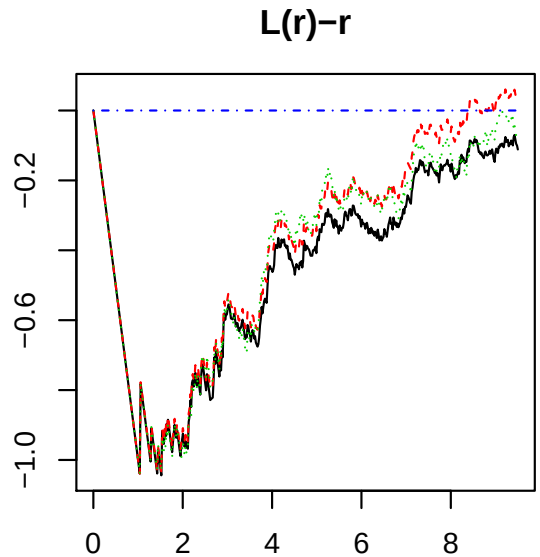
  where $\omega_d = \pi^{d/2}/\Gamma(1 + d/2)$ (volume of the unit ball in $\mathbb{R}^d$).
- More convenient to plot $L(r) - r$ than $K(r)$:
  - Poisson process: $L(r) - r = 0$
  - If $L(r) - r > 0$ (or if $L(r) - r < 0$) whenever $r > 0$ is not "too large": more clustering (or regularity) than for a Poisson process.

Example: Norwegian spruces in a 56×38 m region









Inhomogeneous K and L-functions (Baddeley, M & 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 $K(r) = \int_{\|u\| \le r} g(u)\,\mathrm{d}u$ and $L(r) = (K(r)/\omega_d)^{1/d}$.

- s.o.i.r.s. is satisfied for any Poisson process, many Cox process models (see later), any stationary point process, and any 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

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

  is an unbiased estimate of $K(r)$.
- In practice an estimate for $\rho(u)$ is plugged in.

Summary statistics based on interpoint distances

Assume $\mathbf{X}$ is stationary with constant intensity ρ.

- *Empty space function*:

$$F(r) = P(d(0, \mathbf{X}) \leq r) = P(\mathbf{X} \cap b(0, r) \neq \emptyset), \quad r > 0,$$

  i.e. the distribution function for the length between a fixed location to its nearest point in $\mathbf{X}$.

- *Nearest-neighbour distribution function*:

$$G(r) = \mathbb{E} \frac{1}{\rho |A|} \sum_{u \in \mathbf{X} \cap A} 1[d(u, \mathbf{X} \setminus \{u\}) \leq r], \quad r > 0,$$

  which can be regarded as the distribution function for the length from a 'typical' point in $\mathbf{X}$ to its nearest neighbour in $\mathbf{X}$.

- Poisson process: $F(r) = G(r) = 1 - \exp(-\rho \omega_d r^d)$.

J-function

- Often $J(r) = (1 - G(r))/(1 - F(r))$ is considered.
- Interpretation of $J(r)$:
  - Poisson process: $J(r) = 1$.
  - If $J(r) < 1$ (or if $J(r) > 1$), then realizations tend to be more clustered (or regular) than for a Poisson process.
- Extension to the case of s.o.i.r.s.: see Van Lieshout (2011).

Non-parametric estimation of F and G

- Minus sampling: use only $W_{\ominus r} = \{u \in W : b(u, r) \subseteq W\}$.
- Estimation of F (I is a systematic grid of fixed locations):

$$\hat{F}(r) = \frac{1}{n(I \cap W_{\ominus r})} \sum_{u \in I \cap W_{\ominus r}} 1[d(u, \mathbf{X}) \leq r].$$

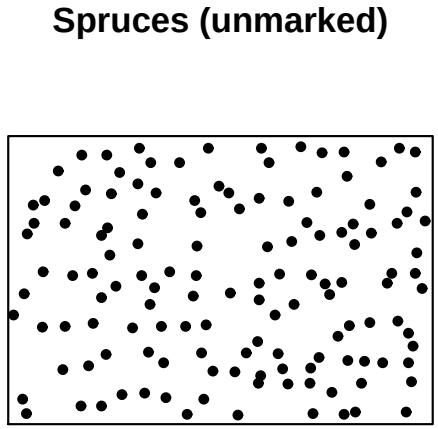
- Estimation of G:

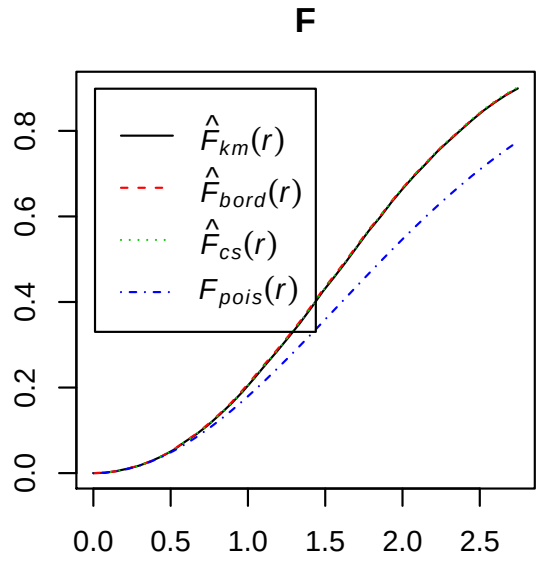
$$\hat{G}(r) = \frac{1}{\rho |W_{\ominus r}|} \sum_{u \in \mathbf{X} \cap W_{\ominus r}} 1[d(u, \mathbf{X} \setminus \{u\}) \leq r].$$

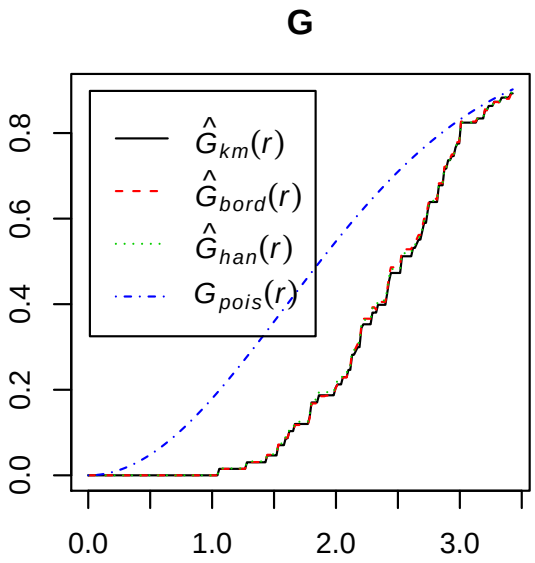
- Estimation of J:

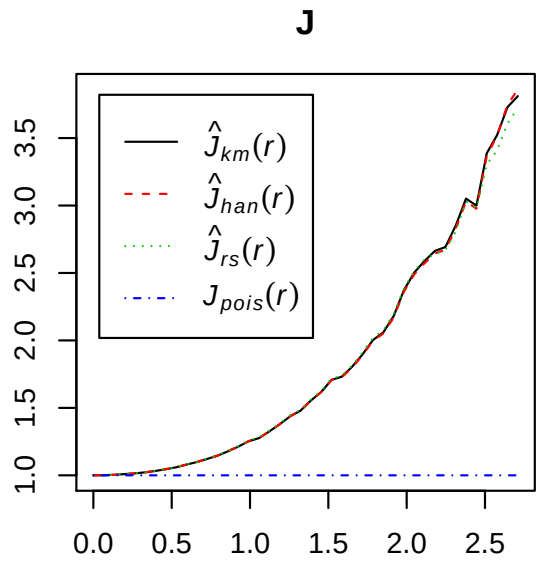
$$\hat{J}(r) = (1 - \hat{G}(r))/(1 - \hat{F}(r)).$$

Example: Norwegian spruces in a 56×38 m region









Model check using summary statistics: envelopes

- Model check: compare summary statistic $T(r)$ from model with estimate $\hat{T}(r)$ obtained from data.
- If $T(r)$ is intractable, it may be approximated using simulations: simulate n new point patterns, calculate estimates $\hat{T}_1(r), \ldots, \hat{T}_n(r)$, and consider

$$T(r) \approx \bar{T}(r) = \frac{1}{n} \sum_{i=1} T_i(r).$$

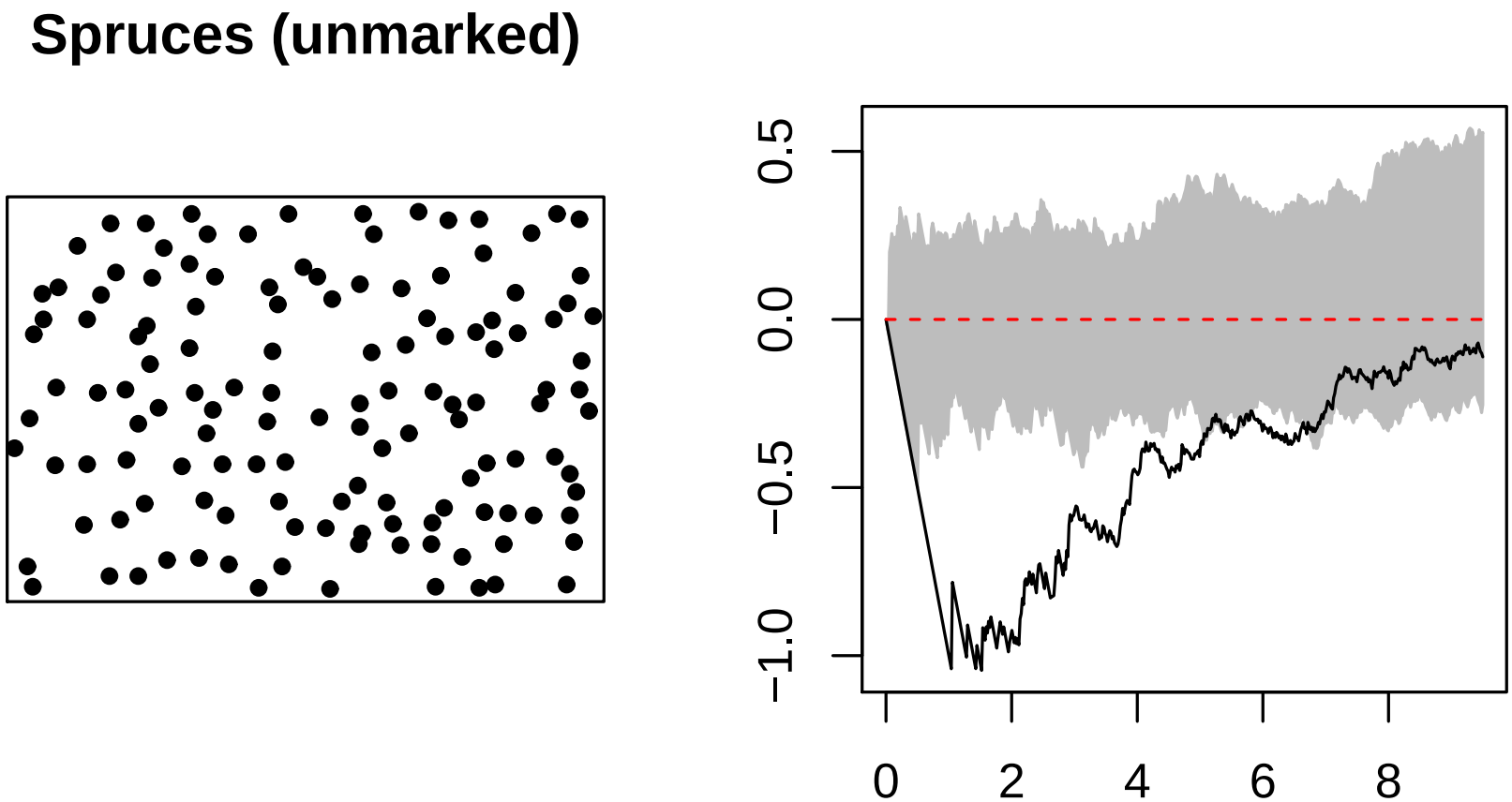
- Example of a *pointwise envelope*: if $\hat{T}_{(1)}(r), \ldots, \hat{T}_{(n)}(r)$ are the ordered statistics, then e.g.

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

(if no ties), i.e. for $n = 39$, we have $2/(n+1) = 0.05$, and so for each $r > 0$ an approximate 95% envelope is given by $[\hat{T}_{(1)}(r), \hat{T}_{(n)}(r))]$.

Example: 134 Norwegian spruces in a 56 × 38 m region

- $L(r) - r$ with (approximate) 95% envelope (i.e. $n = 39$) under a stationary Poisson process with intensity $134/(56 \times 38)$:



Pointwise and global envelopes

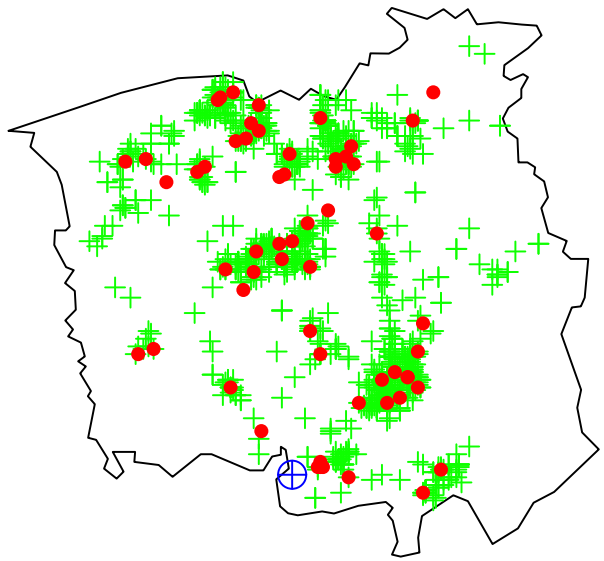
- It is only for each fixed value of r that we have constructed an (approximate) 95% envelope. Hence a *pointwise envelope*.
- If "rejection" means "being outside the envelope" and we consider several values of r, then we need a much wider 95% envelope in order to obtain a test of level 5%.
- Solution (more technical): Myllymäki et al. (2016) and Mrkvička et al. (2017) consider a so-called *global rank envelope test* providing a correct p-value for the fitted model being "true" and with a graphical interpretation in form of a *global (rank) envelope*. This means that for a 95% global envelope, with (approximate probability) 0.95 we expect $\hat{T}(r)$ to be within the envelope for all of several r values (typically a discretization of an interval).
- Details: see their paper.
- An R library for Global Envelope Tests: https://github.com/myllym/GET

Summary statistics in R

- The `spatstat` package contains many functions for estimating summary statistics for point patterns.
- Have a look at some of them... (Rcode2SummaryStat.R)

Example: Analysing the Chorley-Ribble data

Recall this dataset (Slide 9) (red dots: larynx cancer cases; green pluses: lung cancer cases; blue plus/circle: incinerator)



How do the larynx cancer cases depend on the distance $\lambda_{incin}(u)$ (u in the Chorley-Ripley region) to the incinerator, when accounting for the population density?
Problem: Do not know the population density.

Population density

Idea: Use instead that a non-parametric estimate of the intensity function of lung cancer cases (denoted λ_{pop}) may be in good agreement with the population density – see the following map.

Statistical analysis

Summary of the results obtained by a group of my third year students (for the Master's degree), see //people.math.aau.dk/∼jm/SS1.html:

- A homogeneous Poisson process was obviously not fitting well.
- An appropriate band-width $\sigma = 1, 1.5, 2$ for the non-parametric estimate λ_{pop} was chosen
  ($\sigma = 1.5$ was first preferred, but at the very end $\sigma = 1$ was chosen as it provided a better fit of the model).
- Fitted 3 models for the log-intensity of larynx cancer cases:
  - $\alpha + \beta\lambda_{pop} + \gamma\lambda_{incin}$
  - $\alpha + \beta \log \lambda_{pop} + \gamma\lambda_{incin}$
  - $\alpha + \beta \log \lambda_{pop} + \log \gamma\lambda_{incin}$
- Considering functional summary statistics together with envelopes, the last model with $\sigma = 1$ was providing the best fit; then $\hat{\beta} = 1.3$ and $\hat{\gamma} = -0.3$.
- So, when accounting for the population density, the intensity of larynx cancer cases is estimated to be a decreasing function of the distance to the incinerator.

2 TWO MAIN CLASSES OF SPATIAL POINT PROCESS MODELS: COX AND MARKOV POINT PROCESSES

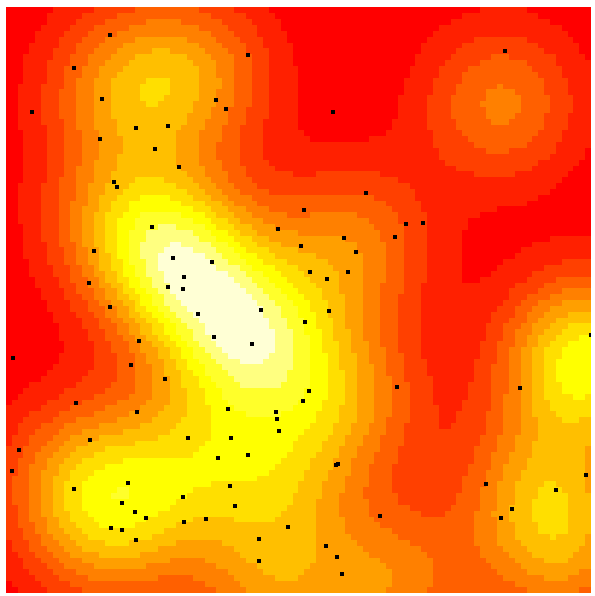
2.1 COX PROCESSES

Motivation for Cox processes

- Suppose we want to model a point pattern of trees given that we know the soil quality. May try with an inhomogeneous Poisson process.
- For example, if the soil quality is given by $h(x, y) \in \mathbb{R}$, then we could try with a Poisson process with intensity $\rho(x, y) = \exp(\alpha + \beta h(x, y))$ (where α, β are unknown parameters).
- But what if we don't know the soil type – or other covariates of importance?
- For the Chorley-Ribble dataset, if also we want to model the lung cancer cases, we miss covariate information.
- In analogy with random effect models, we may introduce a Poisson process with a random intensity – a Cox process (Cox, 1955).

Definition of a Cox process

- *Random intensity function*: $Z = \{Z(u) : u \in S\}$ is a locally integrable non-negative random field on $S \subseteq \mathbb{R}^d$.
- *Cox process*: $\mathbf{X}$ is a Cox process driven by the random intensity function Z (on S) if, conditional on $Z = z$, $\mathbf{X}$ is a Poisson process with intensity function z.



Summary statistics for Cox processes

- Intensity, product density, and pair correlation function:

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

$$g(u, v) = \frac{\mathbb{E}[Z(u)Z(v)]}{\mathbb{E}Z(u)\mathbb{E}Z(v)}$$

- In general for Cox processes (apart from a few exceptions), $g(u, v) > 1$ for all $u, v \in S$, i.e. models for aggregation/clustering.
- Overdispersion for counts in all Cox processes:

$$\mathbb{V}[N(A)] \geq \mathbb{E}[N(A)]$$

  This also indicates aggregation/clustering.
- F, G, J hard to obtain for general Cox processes, but expressions are available for specific Cox process models.

Simulation of Cox processes

- Two-step procedure:
  - Simulate Z.
  - Simulate $\mathbf{X}$ given $Z = z$ as an inhomogeneous Poisson process with intensity z.
- Need to find some way of simulating Z – this depends on the choice of stochastic model for Z.

Density and void probability of Cox processes

- Restricted to a bounded region W, the density is

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

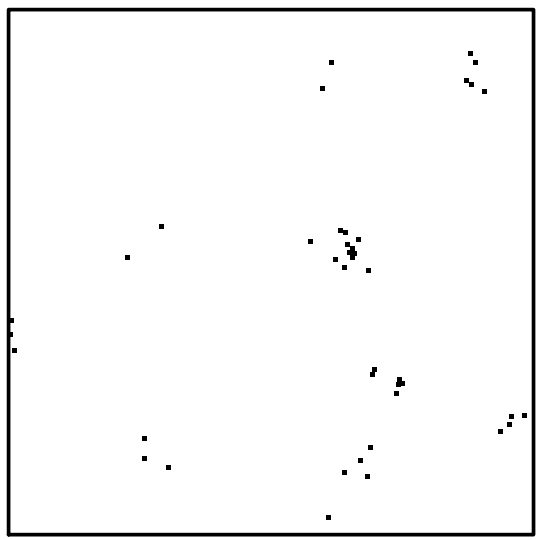
  w.r.t. to the unit-rate Poisson process (more about densities later).
- Void probability:

$$P(\mathbf{X}\cap B = \emptyset) = \mathbb{E}\exp\left(-\int_B Z(u)\mathrm{d}u\right).$$

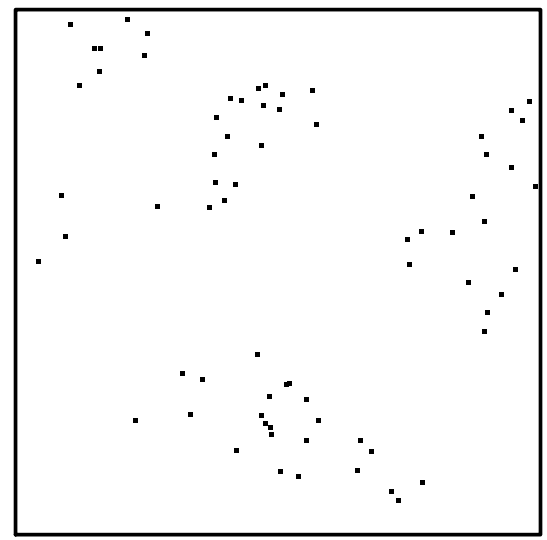
- Neither one can usually be calculated exactly!

Example: Neyman-Scott (1958) process (used in cosmology, ecology, foresty,...)

- Centre process: C is a stat. Poisson process with intensity κ.
- Random intensity: $Z(u) = \alpha \sum_{c \in C} k(u - c)$ for $\alpha > 0$ and some kernel k. So $\mathbf{X}$ is stat. with intensity $\alpha\kappa$.
- Examples:
  - Matérn (1960) cluster process: $k(u) = 1[\|u\| \leq r]/(\omega_d r^d)$.



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



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

- (Modified) Thomas process: k is the density of $N_d(0, \sigma^2 I)$.

Pair correlation function and simulation for a Neyman-Scott process

- Pair correlation function:

$$g(u) = 1 + \int k(v)k(u+v)\mathrm{d}v/\kappa$$

  (special case of result for SNCPs; see later).
- Simulation algorithm:
  - Simulate centre process C as a homogeneous Poisson process with intensity κ.
  - For each $c \in C$, simulate its cluster: Simulate the number of points in the cluster, $N(\mathbf{X}_c) \sim \mathrm{po}(\alpha)$; and simulate each point $x_{c,i} \in \mathbf{X}_c$ using the density $k(\cdot - c)$.
  - Superpose the clusters $\mathbf{X} = \cup_{c \in C} \mathbf{X}_c$.
- Note:
  - the centres $c \in C$ are not part of $\mathbf{X}$;
  - beware of edge-effects (M, 2003).

Example: Shot-noise Cox processes (M, 2003)

- A shot-noise Cox process (SNCP) has random intensity function

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

  where $k(c,\cdot)$ is a kernel (pdf) and $\mathbf{\Phi} \sim \text{Poisson}(\mathbb{R}^d \times]0,\infty[, \zeta)$.
- NB: may be non-stationary.
- A Neyman-Scott process is the special case where
  $k(c,u) = k(u-c)$,
  all $\gamma = \alpha$ are fixed to be a number $\alpha > 0$,
  $\mathbf{\Phi}$ is essentially given by C,
  and $\mathrm{d}\zeta(c,\gamma)$ is essentially given by $\kappa \mathrm{d}c$.

Pair correlation function and simulation for a SNCP

- If $\mathbf{X}$ is a SNCP, then

$$g(u,v) = 1 + \frac{\beta(u,v)}{\rho(u)\rho(v)},$$

where

$$\beta(u,v) = \int\int \gamma^2 k(c,u)k(c,v)\zeta(c,\gamma)\mathrm{d}c\mathrm{d}\gamma.$$

- Proof: Use the so-called Slivnyak-Mecke (omitted here).
- How would you simulate a SNCP?

Special case of SNCP: Shot-noise G Cox processes (Brix, 1999)

- A shot-noise G Cox process has

$$\zeta(c, \gamma) = \beta\gamma^{-\alpha-1} \exp(-\tau\gamma)/\Gamma(1-\alpha),$$

  where $\beta > 0, \alpha < 1, \tau > 0$.
- Stationary if $k(c, u) = k(u - c)$. Then the pair correlation function is

$$g(u) = 1 + \frac{1-\alpha}{\beta\tau^{\alpha}} \int k(v)k(u+v)\mathrm{d}v.$$

  NB: Of the same form as a for a Neyman-Scott process.
  This illustrates that different point process models can share the same first and second order moment properties!

Neyman-Scott processes in R

- Neymann-Scott processes are implemented in the `Spatstat` package.
- `rNeymanScott()` is used to simulate the general case.
- `rMatClust()` and `rThomas()` can be used for the special cases of the Matérn cluster processes and Thomas processes.
- The shot-noise Cox processes are not implemented (as far as I know).
- Have a look at some code... (Rcode3Cox.R).

Log Gaussian Cox process (M, Syversveen & Waagepetersen, 1998)

- Gaussian field: $\mathbf{Y} = \{Y(s) : s \in \mathbb{R}^d\}$ is a Gaussian field if all finite linear combinations of $Y(u_1), \ldots, Y(u_n)$ are normally distributed.
- A log Gaussian Cox process (LGCP) is a Cox process driven by $\mathbf{Z} = \exp(\mathbf{Y})$.
- The distribution of a LGCP is completely determined by the mean and covariance of the Gaussian field:

$$m(u) = \mathbb{E}Y(u) \qquad \text{and} \qquad c(u, v) = \mathrm{Cov}(Y(u), Y(v)).$$

- Mean function $m(u)$:
  - Constant mean: no knowledge about underlying tendencies.
  - Non-constant mean: covariate information (e.g. soil quality when modelling the locations of trees).

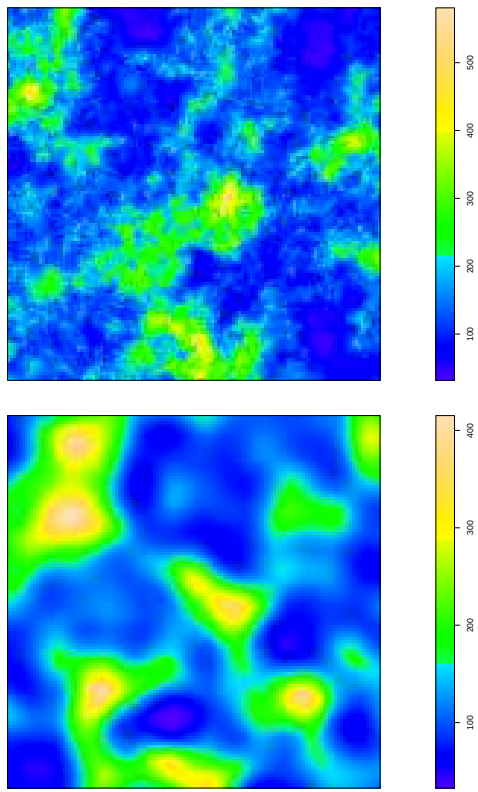
Example: Power exponential covariance function

- The power exponential covariance function is given by

$$c(u, v) = \sigma^2 \exp\left(-\|u - v\|^\delta / \alpha\right),$$

with parameters $\sigma, \alpha > 0,\ 0 \leq \delta \leq 2$.

- $\delta = 1$:
  - The *exponential cov. fct.*
  - Has "jagged" realizations.
- $\delta = 2$:
  - The *Gaussian cov. fct.*
  - Has very smooth realizations.



Summary statistics and simulation

- The intensity and pair correlation function are given by

$$\rho(u) = \exp\left(m(u) + c(u,u)/2\right), \qquad g(u,v) = exp(c(u,v)).$$

- Proof: Use Laplace transforms.
- NB: $(\rho, g) \leftrightarrow (m, c)$.
- Simulation of $\mathbf{Z}_W = \{Z(s) : s \in W\}$ ($W \subset \mathbb{R}^d$ bounded) is somewhat tricky – we can approximate $\mathbf{Z}_W$ on a grid, and simulate this as a multivariate normal variable.
  Small grids: Choleski decomposition.
  Large rectangular grids: extend to a larger rectangular grid wrapped on a torus and use FFT (see M, Syversveen & Waagepetersen, 1998).
- Given a realization $\mathbf{X}_W = \mathbf{x}_W$ of the LGCP within W (bounded), conditional simulation of $\mathbf{Z}_W$ approximated on a grid can be done using INLA (Rue, Martino & Chopin, 2009).

Log Gaussian Cox in R

- LGCPs are implemented into `spatstat`, but only in connection to estimation procedures (not simulation).
- However, there are many other packages which can be used for simulating Gaussian fields...

Overview of Cox process models

Cox process $\rightarrow$

- LGCPs
- SNCPs $\rightarrow$
  - SNGCPs
  - Neyman-Scott $\rightarrow$
    - Thomas process
    - Matérn cluster process

2.2 MARKOV POINT PROCESSES

Densities for Poisson processes

- Recall that $\mathbf{X}_1$ is absolutely continuous w.r.t. $\mathbf{X}_2$ 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 w.r.t. 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 $\rho_1(u) > 0 \Rightarrow \rho_2(u) > 0$. Then Poisson$(S, \rho_2)$ has density

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

  w.r.t. Poisson(S, ρ_1).
- Example: for bounded S, Poisson(S, ρ) has density

$$f(\mathbf{x}) = \exp(|S| - \mu(S)) \prod_{u \in \mathbf{x}} \rho(u)$$

  w.r.t. standard (unit-rate) Poisson process Poisson$(S, 1)$.

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 w.r.t. to the unit rate Poisson process on S:

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

- Often specified by an unnormalized density:

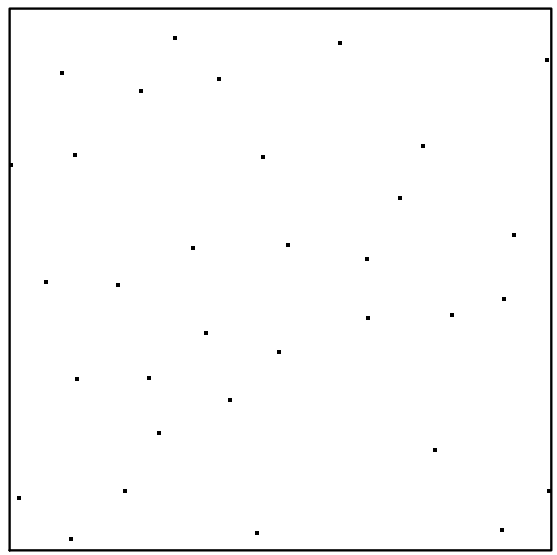
$$h(\mathbf{x}) = c \, f(\mathbf{x}), \quad \mathbf{x} \subset S \text{ finite.}$$

- Problem: calculation of the normalising constant

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

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.
- Kelly & Ripley (1976): Exists and is repulsive if $\gamma \leq 1$; otherwise in general (e.g. if S contains a ball) it does not exist.



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

Stability conditions and existence

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

Proposition:
Local stability $\Rightarrow$ Ruelle stability $\Rightarrow$ integrability.

Example: Multiscale process

- Interaction term:

$$s_i(\mathbf{x}) = \sum_{\{u,v\} \subseteq \mathbf{x}} 1[R_{i-1} < \|u - v\| \leq R_i], \quad i = 1, \ldots, k,$$

  where $0 = R_0 < R_1 < R_2 < \cdots < R_k < \infty$.

- Density:

$$f(\mathbf{x}) \propto \beta^{n(\mathbf{x})} \prod_{i=1}^{k} \gamma_i^{s_i(\mathbf{x})}$$

  where $\beta > 0$ and each $\gamma_i \geq 0$.

- Well-defined (locally stable) if
  - all $\gamma_i \leq 1$
  - or if $\gamma_1 = 0$ (hard core condition).

Hereditary condition and Papangelou conditional intensity

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

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

  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 $\mathbf{x}$.
- If f is hereditary: $f \leftrightarrow \lambda$.
- If $\mathbf{X}$ is a Poisson process with intensity $\rho(u)$: $\lambda(\mathbf{x}, u) = \rho(u)$.

Attractive or repulsive point process

- Consider any finite point configurations $\mathbf{x} \subseteq \mathbf{y}$ and any point u in S.
- If $\mathbf{X}$ is Poisson, then $\lambda(\mathbf{x}, u) = \lambda(\mathbf{y}, u)$.
- $\mathbf{X}$ is *attractive* if
$$\lambda(\mathbf{x}, u) \leq \lambda(\mathbf{y}, u).$$
- $\mathbf{X}$ is *repulsive* if
$$\lambda(\mathbf{x}, u) \geq \lambda(\mathbf{y}, u).$$

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

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

- If $\varphi(\{u,v\}) \leq 1$, then $\mathbf{X}$ exists and is repulsive.
- If $\varphi(\{u,v\}) \geq 1$, then usually $\mathbf{X}$ does not exist.
- *Range of interaction*:

$$R = \inf\{r > 0 : \text{ for all } \{u,v\} \subset S, \varphi(\{u,v\}) = 1 \text{ if } \|u - v\| > r\},$$

i.e. "the smallest r so that $\varphi(\{u,v\}) = 1$ if distance between u and v is larger than r".

Example: Multiscale process (continued)

Recall that for this process,

$$f(\mathbf{x}) \propto \beta^{n(\mathbf{x})} \prod_{i=1}^{k} \gamma_i^{s_i(\mathbf{x})}$$

where

$$s_i(\mathbf{x}) = \sum_{\{u,v\} \subseteq \mathbf{x}} 1[R_{i-1} < \|u - v\| \leq R_i], \quad i = 1, \ldots, k,$$

with $\beta > 0$, $0 = R_0 < R_1 < R_2 < \cdots < R_k < \infty$, all $\gamma_i \geq 0$, and either all $\gamma_i \leq 1$ or $\gamma_1 = 0$.
This is a pairwise interaction process with $\varphi(u) = \beta$ and

$$\varphi(\{u, v\}) = \begin{cases} \gamma_i & \text{if } R_{i-1} < \|u - v\| \leq R_i \\ 1 & \text{if } \|u - v\| > R_k \end{cases}$$

Summary statistics, likelihoods and simulation

- Summary statistics are difficult to obtain, e.g. it can be proved that $\rho(u) = \mathbb{E}\lambda(X, u)$, but in general this mean value cannot be calculated (except for the Poisson process).
- As we have specified the density, we can deal with likelihoods (more later)!
- Likelihoods and simulation procedures usually require MCMC methods – here the Papangelou conditional density is very useful for constructing MCMC algorithms (more later)!

Neighbour relations

- *Neighbour relation*: Let $\sim$ be a reflexive ($u \sim u$) and symmetric ($u \sim v \Rightarrow v \sim u$) relation on points $u, v \in S$.
- The *neighbourhood* of $u \in S$ is

$$N_u = \{v \in S : v \sim u\}.$$

- We will mainly consider the *R-close neighbourhood*:

$$u \sim v \text{ iff } \|u - v\| \leq R.$$

Markov point processes

- **DEF:** *Local Markov property*: $\lambda(\mathbf{x}, u) = h(\mathbf{x} \cup u)/h(\mathbf{x})$ depends only on $\mathbf{x}$ through $\mathbf{x} \cap N_u$.
- **DEF:** Suppose that $h(\cdot) \geq 0$ is hereditary and satisfies the local Markov property. Then h is *Markov* w.r.t. $\sim$.
  If also h is integrable (w.r.t. the unit rate Poisson process on S) then $\mathbf{X} \sim h$ is called a *Markov point process*.
- **Hammersley-Clifford theorem** (Ripley & Kelly, 1977):
  The following are equivalent.
  1. h is Markov w.r.t. $\sim$.
  2. $$h(\mathbf{x}) = \prod_{\mathbf{y} \subseteq \mathbf{x}} \varphi(\mathbf{y})$$
     where $\varphi(\mathbf{y}) = 1$ whenever $u \not\sim v$ for some $u, v \in \mathbf{y}$.
- Then the *Papangelou conditional intensity* becomes
$$\lambda(\mathbf{x}, u) = \prod_{\mathbf{y} \subseteq \mathbf{x} \cap N_u} \varphi(u \cup \mathbf{y}).$$

Examples of Markov point processes

- *Pairwise interaction process*:

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

  This is Markov when $u \sim v$ iff $\varphi(\{u, v\}) \neq 1$.
  (Clearly, $\varphi(\mathbf{y}) = 1$ whenever $n(\mathbf{y}) > 2$.)
- Hence the multiscale process (including the Strauss process) is Markov.

Examples of Markov point processes (cont.)

- *Geyer's triplet process* (Geyer, 1999):

$$h(\mathbf{x}) = \beta^{n(\mathbf{x})} \gamma^{s(\mathbf{x})} \delta^{t(\mathbf{x})},$$

  where $s(\mathbf{x})$ is the number of R-close pairs and $t(\mathbf{x})$ is the number of R-close triplets of points in $\mathbf{x}$. Here,
  - $\beta > 0$, $\gamma \geq 0$, either $\delta \in [0, 1)$ or both $\delta = 1$ and $\gamma \leq 1$;
  - $\varphi(\mathbf{y}) = 1$ if $n(\mathbf{y}) > 3$ or $\|u - v\| > R$ for some $u, v \in \mathbf{y}$.
- *Area-interaction process* (Widom & Rowlingson, 1970; Baddeley & Van Lieshout, 1995):

$$h(\mathbf{x}) = \beta^{n(\mathbf{x})} \gamma^{-|U(\mathbf{x})|}$$

  where $|U(\mathbf{x})| = |\cup_{u \in \mathbf{x}} b(u, R)|$ (area/volume of this union).
  - Poisson if $\gamma = 1$; attractive if $\gamma > 1$; repulsive if $0 < \gamma < 1$.
  - Easily seen to be Markov, but $\varphi(\cdot)$ is complicated...
  - Obviously $\varphi(\mathbf{y}) = 1$ if $\|u - v\| > 2R$ for some $u, v \in \mathbf{y}$, so the range of interaction is $2R$.

Spatial Markov property

- For $B \subseteq S$, define the neighbourhood
  $N_B = \{u \in S : u \sim v \text{ for some } v \in B\}$
  and the boundary
  $\partial B = N_B \setminus B$.
- *Theorem*: If $A, B \subset S$ so that $A \cap N_B = \emptyset$, then $\mathbf{X}_A$ and $\mathbf{X}_B$ are conditionally independent given $\mathbf{X}_C$ where $C = S \setminus (A \cup B)$.
- Proof: follows from the Hammersley-Clifford theorem.

Edge correction using the spatial Markov property

- Suppose we observe $\mathbf{X}_W = \mathbf{x}$ where $W \subset S$.
- Problem: density (likelihood) $f_W(\mathbf{x}) = \mathbb{E} f(\mathbf{x} \cup \mathbf{Y}_{S \setminus W})$ unknown (here $\mathbf{Y} \sim$Poisson$(S, 1)$).
- Border method: let

$$W_{\ominus R} = \{u \in W : b(u, R) \subseteq W\}$$

and base inference on

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

$$\prod_{\emptyset \neq \mathbf{y} \colon \mathbf{y} \subseteq \mathbf{x} \cap W_{\ominus R}} \left[\varphi(\mathbf{y}) \prod_{\mathbf{z} \colon \emptyset \neq \mathbf{z} \subseteq \mathbf{x} \cap (W \setminus W_{\ominus R})} \varphi(\mathbf{y} \cup \mathbf{z}) \right]$$

since conditional on $\mathbf{X} \cap (W \setminus W_{\ominus R})$, we have that $\mathbf{X} \cap W_{\ominus R}$ and $\mathbf{X} \setminus W$ (the unobserved part of the process) are independent.

Inhomogeneous Markov processes

- First order inhomogeneity: $\varphi(\mathbf{y})$ is non-constant for $n(\mathbf{y}) = 1$, but translation invariant for $n(\mathbf{y}) > 1$.
- Example (inhomogeneous Strauss process):

$$h(\mathbf{x}) = \left(\prod_{u \in \mathbf{x}} \beta(u) \right) \gamma^{s(\mathbf{x})}.$$

- Inhomogeneity can also be obtained by independent thinning of a homogeneous Markov process.
  - The thinned process is not Markov, but the spatial Markov property still holds.

Overview of Markov point processes

Markov/Gibbs process $\rightarrow$

- Pairwise interaction process $\rightarrow$
  - Strauss process $\rightarrow$
    - (Gibbs) hard core process
  - Multiscale process
- Geyer's triplet process
- Area-interaction process

There are MANY other models for Markov point processes – see M & Waagepetersen (2004) and the references therein.

Finite Markov/Gibbs processes can be extended to $\mathbb{R}^d$, but the theory is technical – see again M & Waagepetersen (2004) and the references therein.

3 SIMULATION AND INFERENCE PROCEDURES

3.1 MARKOV CHAIN MONTE CARLO (MCMC) METHODS

When do we need MCMC?

We already know how to simulate

- Poisson processes;
- Cox processes.

But what about (finite) Markov point processes?

- Then we consider a finite point process $\mathbf{X}$ with points in a bounded set $S \subset \mathbb{R}^d$ and specified by a hereditary density $f(\mathbf{x}) \propto h(\mathbf{x})$ or equivalently by its Papangelou conditional intensity $\lambda(\mathbf{x}, u) = h(\mathbf{x} \cup u)/h(\mathbf{x})$.
- For this we use MCMC methods... noting that $n(\mathbf{X})$ is random.

(See M & Waagepetersen (2004) if we condition on $n(\mathbf{X}) = n$.)

Birth-death Metropolis-Hastings algorithm (Geyer & M, 1994)

Generate Markov chain $\mathbf{Y}_0, \mathbf{Y}_1, \ldots$: For $m = 0, 1, \ldots$,
given $\mathbf{Y}_m = \mathbf{x}$, make either a birth proposal w.p. $p(\mathbf{x})$
or a death proposal w.p. $1 - p(\mathbf{x})$;

- if birth: generate a new point V from a density $q_b(\mathbf{x}, \cdot)$;
  accept $\mathbf{Y}_{m+1} = \mathbf{x} \cup V$ w.p. $\min\{1, r_b(\mathbf{x}, V)$;
- if death: select a point $V \in \mathbf{x}$ w.p. $q_d(\mathbf{x}, V)$;
  accept $\mathbf{Y}_{m+1} = \mathbf{x} \setminus V$ w.p. $\min\{1, r_d(\mathbf{x}, V)$;
  (note: if $\mathbf{Y}_m = \emptyset$, then $\mathbf{Y}_{m+1} = \emptyset$);
- else $\mathbf{Y}_{m+1} = \mathbf{x}$.
- Here r_b and r_d are so-called Hastings ratios:

$$r_b(\mathbf{x}, V) = \frac{\lambda(\mathbf{x}, V)(1 - p(\mathbf{x} \cup V))q_d(\mathbf{x} \cup V, V)}{p(\mathbf{x})q_b(\mathbf{x}, V)}$$

$$r_d(\mathbf{x}, V) = 1/r_b(\mathbf{x} \setminus V, V)$$

Convenient initial state: $\mathbf{Y}_0 = \emptyset$ or $\sim$Poisson process.

Example: Birth-death simulation of a Strauss process

- Proposals distributions:

$$p(\mathbf{x}) = 1/2, \qquad q_d(\mathbf{x}, v) = \frac{1}{n(\mathbf{x})}, \qquad q_b(\mathbf{x}, v) = \frac{1}{|S|}.$$

- Hastings ratios:

$$r_b(\mathbf{x}, v) = \frac{\beta\gamma^{\sum_{u\in\mathbf{x}} 1[\|v-u\|\leq R]}|S|}{n(\mathbf{x}\cup v)},$$

$$r_d(\mathbf{x}, v) = \frac{n(\mathbf{x})}{\beta\gamma^{\sum_{u\in\mathbf{x}\setminus v} 1[\|v-u\|\leq R]}|S|}.$$

- What happens if we ignore that $\gamma \leq 1$ is required for the Strauss process to exist?

Detailed balance condition

- Let $a_b = \min\{1, r_b\}$, $a_d = \min\{1, r_d\}$ denote the acceptance probabilities for a birth or a death.
- The Hastings ratios r_b and r_d are specified such that a detailed balance condition is satisfied:

$$f(\mathbf{x})p(\mathbf{x})q_b(\mathbf{x}, v)a_b(\mathbf{x}, v) = f(\mathbf{x} \cup v)(1 - p(\mathbf{x} \cup v))q_d(\mathbf{x} \cup v, v)a_d(\mathbf{x} \cup v, v)$$

- This implies reversibility and f is the equilibrium density!

Ergodicity of birth-death chain

- Let $E = \{\mathbf{x} \subset S : n(\mathbf{x}) < \infty, f(\mathbf{x}) > 0\}$.
- Suppose initial state $\mathbf{Y}_0 \in E$, $p(\emptyset) < 1$, and for all $\mathbf{x} \in E$ with $\mathbf{x} \neq \emptyset$, there is a $v \in \mathbf{x}$ so that

$$(1 - p(\mathbf{x}))q_d(\mathbf{x}, v) > 0, \qquad f(\mathbf{x} \setminus v)p(\mathbf{x} \setminus v)q_b(\mathbf{x} \setminus v, v) > 0.$$

  Then birth-death chain is aperiodic and irreducible.
- Example (Strauss process – continued): $E = \{\mathbf{x} \subset S : n(\mathbf{x}) < \infty\}$ and for all $\mathbf{x} \in E$ with $\mathbf{x} \neq \emptyset$,

$$(1 - p(\mathbf{x}))q_d(\mathbf{x}, v) = \frac{1}{2n(\mathbf{x})} > 0,$$

$$f(\mathbf{x} \setminus v)p(\mathbf{x} \setminus v)q_b(\mathbf{x} \setminus v, v) = \frac{1}{2|S|} f(\mathbf{x}) > 0.$$

- Geometric ergodicity is implied by local stability (Geyer & M, 1994; Geyer, 1999).

MCMC in practice

- Looking at plots of realizations of $\mathbf{Y_m}$ (for large m), it may be hard to say whether the chain has converged approximately.
- Trace plots of various statistics can be useful.
- For example, in the case of the Strauss process, $n(\mathbf{Y}_m)$ and $s(\mathbf{Y}_m)$ are obvious choices.
- Should comparing trace plots for $\mathbf{Y}_m$ started at different $\mathbf{Y}_0$, e.g. $\mathbf{Y}_0 = \emptyset$ or $\mathbf{Y}_0 \sim$Poisson(S, β) process (see next item).
- Local stability: if we have local stability, $\lambda(\mathbf{x}, u) \leq K(u)$, then there exists $\mathbf{Y} \sim$Poisson(S, K) such that $\mathbf{X} \subseteq \mathbf{Y}$, i.e. $\mathbf{Y}$ dominates $\mathbf{X}$ (Preston 1977).
- `Spatstat` contains the `rmh` function for simulating point patterns using the Metropolis-Hastings algorithm. Have a look at some code... (Rcode4Markov.R)

Perfect simulation

- Perfect simulation means an algorithm so that the output is following exactly the target density f (at least in theory, i.e. ignoring the fact that a random number generator may not produce exactly uniformly distributed numbers).
- There exists a perfect simulation algorithm for locally stable point processes based on so-called *dominating coupling from the past (DCFTP)* (Kendall & M, 2000; Berthelsen & M 2002, 2003).
- The DCFTP algorithm is implemented in `Spatstat` for the Strasuss process as the `rStrauss` function... (Rcode4Markov.R)

3.2 LIKELIHOOD AND MOMENT-BASED ESTIMATION PROCEDURES

Likelihood-based inference for point processes

- Consider a point process on a bounded region S and specified by unnormalized density $h_\theta(\mathbf{x})$, so

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

- Problem: Usually c_θ is unknown. Then log likelihood

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

  is also unknown.
- Both maximum likelihood inference and Bayesian inference need this constant (it does not cancel out as it did for the birth-death Metropolis-Hastings algorithm).

Normalising constants and expectations

- If $\mathbf{X} \sim f_\theta$ and $\mathbf{Y} \sim$Poisson$(S, 1)$, then

$$\mathbb{E}_\theta[k(\mathbf{X})] = \mathbb{E}[k(\mathbf{Y})f_\theta(\mathbf{Y})] = \mathbb{E}[k(\mathbf{Y})h_\theta(\mathbf{Y})]/c_\theta$$

  for any non-negative (measurable) function k.

- Thus, for $k = 1$,

$$c_\theta = \mathbb{E}\left[h_\theta(\mathbf{Y})\right] = \sum_{n=0}^{\infty} \frac{\mathrm{e}^{-|S|}}{n!} \int_{S^n} h_\theta(\{x_1, \ldots, x_n\}) \mathrm{d}x_1 \ldots \mathrm{d}x_n.$$

- In general impossible to calculate unless $\mathbf{X}$ is a Poisson process, i.e. when

$$h_\theta(\{x_1, \ldots, x_n\}) = \prod_{i=1}^{n} \rho(x_i)$$

  and so

$$c_\theta = \exp\left(\int_S \rho(u)\mathrm{d}u - |S|\right).$$

Importance sampling

- c_θ / c_{θ_0} can be approximated using importance sampling, where θ_0 is a fixed reference parameter:

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

- Importance sampling formula: If $f_{\theta_0}(\mathbf{x}) > 0$ whenever $f_\theta(\mathbf{x}) > 0$, then

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

- Hence

$$\frac{c_\theta}{c_{\theta_0}} \approx \frac{1}{n}\sum_{i=1}^{n} \frac{h_\theta(\mathbf{X}_i)}{h_{\theta_0}(\mathbf{X}_i)}$$

where $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is a sample from f_{θ_0} (ideally i.i.d. simulations, but simulations taken after burn-in and at regular spacing (or no spacing at all!) in MCMC will do just fine).

Importance sampling formula

- General *importance sampling formula*:

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

  for any non-negative (measurable) function k.
- Examples: $k = 1$; $k(\mathbf{x}) = n(\mathbf{x})$ or $k(\mathbf{x}) = s(\mathbf{x})$ (the sufficient statistics of a Strauss process).
- Using a sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from f_{θ_0}, then $\mathbb{E}_\theta\left[k(\mathbf{X})\right]$ can be approximated by

$$\frac{\frac{1}{n}\sum_{m=1}^n k(\mathbf{X}_m)\frac{h_\theta(\mathbf{X}_m)}{h_{\theta_0}(\mathbf{X}_m)}}{\frac{1}{n}\sum_{m=1}^n \frac{h_\theta(\mathbf{X}_m)}{h_{\theta_0}(\mathbf{X}_m)}} = \sum_{m=1}^n k(\mathbf{X}_m) w_{\theta,\theta_0,n}(\mathbf{X}_m)$$

  where we have importance weights (probabilities)

$$w_{\theta,\theta_0,n}(\mathbf{X}_m) = \frac{h_\theta(\mathbf{X}_m)/h_{\theta_0}(\mathbf{X}_m)}{\sum_{i=1}^n h_\theta(\mathbf{X}_i)/h_{\theta_0}(\mathbf{X}_i)}, \qquad m = 1, \ldots, n.$$

Importance sampling in practice

- Although

$$\frac{c_\theta}{c_{\theta_0}} \approx \frac{1}{n} \sum_{m=1}^{n} \frac{h_\theta(\mathbf{X}_m)}{h_{\theta_0}(\mathbf{X}_m)}$$

  is an unbiased estimate of $\frac{c_\theta}{c_{\theta_0}}$, it may be a very bad estimate if h_θ and h_{θ_0} are not close: as most $\mathbf{X}_m$ will be located where h_{θ_0} is high, then $\frac{h_\theta(\mathbf{X}_m)}{h_{\theta_0}(\mathbf{X}_m)}$ may be typically low, and so most terms will count very little in the sum, while a few may count a lot!
- Typically this is a problem when θ and θ_0 are far away from each other, so we need a way of making a path between them (path sampling or bridge sampling; see M & Waagepetersen, 2004).

Example: Exponential families

- Consider an *exponential family* (on canonical) form,

$$h_\theta(\mathbf{x}) = \exp(t(\mathbf{x})\theta^\top).$$

- Example: The Strauss process (with fixed R) has

$$t(\mathbf{x}) = (n(\mathbf{x}), s(\mathbf{x})), \qquad \theta = (\log\beta, \log\gamma).$$

- Ratio of normalizing constants used in importance sampling:

$$\frac{c_\theta}{c_{\theta_0}} = \mathbb{E}_{\theta_0}\left[\exp(t(\mathbf{X})(\theta-\theta_0)^\top)\right].$$

- If $\theta - \theta_0$ is 'large', $\exp(t(\mathbf{X})(\theta-\theta_0)^\top)$ has very large variance in many cases, so small steps needed for finding the MLE.

Maximisation of likelihood

- Score and observed information in exp. family case:

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

  can be approximated by importance sampling.
- Since j is a covariance matrix, it is positive semi-definite, and thus the log-likelihood is concave. In fact, also the approximate log-likelihood is concave.
- To find the MLE we need to solve $u(\theta) = 0$.
- Newton-Raphson iterations:

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

- Approximate MLE: Use the importance sampling approximation of score and observed information in the Newton-Raphson iterations.

Missing data

- With a slight abuse of notation, assume f_θ is the joint density of point processes $\mathbf{X}$ and $\mathbf{Y}$ with respect to independent Poisson processes $\mathbf{X}_1$ and $\mathbf{X}_2$ but we only observed $\mathbf{X} = \mathbf{x}$.
- Typically, $(\mathbf{X}, \mathbf{Y})$ corresponds to $(\mathbf{X}_W, \mathbf{X}_{S \setminus W})$, where $W \subseteq S$; then $\mathbf{X}_1$ and $\mathbf{X}_2$ are defined on W and $S \setminus W$, resp.
- The likelihood is given by

$$f_\theta(\mathbf{x}) = \mathbb{E}\left[f_\theta(\mathbf{x}, \mathbf{X}_2)\right] = \mathbb{E}_\theta\left[h_\theta(\mathbf{x}, \mathbf{X}_2)\right] / c_\theta.$$

- 'Missing data problem': Typically, $f_\theta(\mathbf{x}, \mathbf{y})$ or at least or $h_\theta(\mathbf{x}, \mathbf{y})$ has a simple expression but $f_\theta(\mathbf{x})$ does not.
- Note $f_\theta(\mathbf{x})$ is normalizing constant of a conditional density:

$$f_\theta(\mathbf{y}|\mathbf{x}) \propto f_\theta(\mathbf{x}, \mathbf{y}) = h_\theta(\mathbf{x}, \mathbf{y}) / c_\theta.$$

  Thus importance sampling for estimation of the normalizing constants c_θ and $f_\theta(\mathbf{x})$ applies.

Missing data and importance sampling

- Recall previous importance sampling formula:

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

- Adapting to the missing data case:

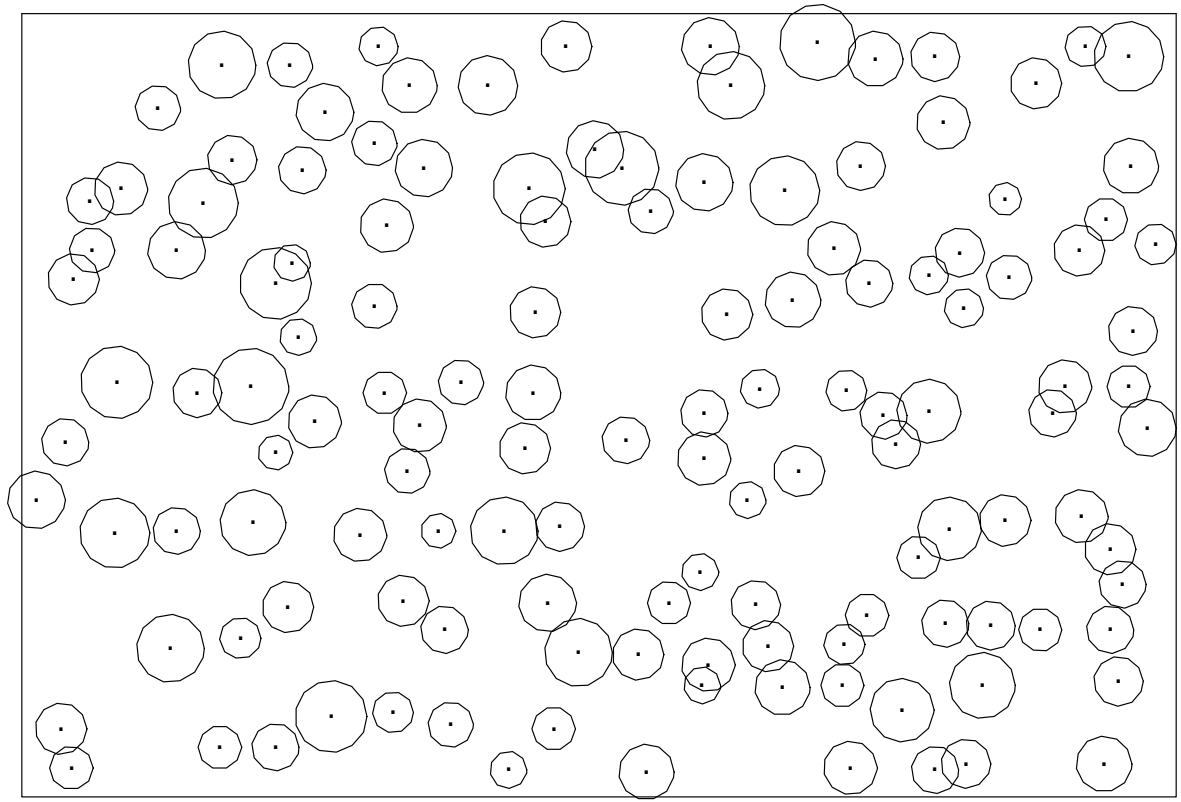
$$\frac{f_\theta(\mathbf{x})}{f_{\theta_0}(\mathbf{x})} = \mathbb{E}_{\theta_0}\left[\frac{f_\theta(\mathbf{X},\mathbf{Y})}{f_{\theta_0}(\mathbf{X},\mathbf{Y})}\middle|\mathbf{X}=\mathbf{x}\right]$$

or

$$\frac{f_\theta(\mathbf{x})}{f_{\theta_0}(\mathbf{x})} = \mathbb{E}_{\theta_0}\left[\frac{h_\theta(\mathbf{X},\mathbf{Y})}{h_{\theta_0}(\mathbf{X},\mathbf{Y})}\middle|\mathbf{X}=\mathbf{x}\right] \Big/ \frac{c_\theta}{c_{\theta_0}}.$$

- We can then use Monte Carlo approximations to approximate this expectation but w.r.t. the conditional distribution of $\mathbf{Y}$ given $\mathbf{X}=\mathbf{x}$ and $\theta=\theta_0$; just as for the normalising constants c_θ/c_{θ_0}.

Example: Norwegian spruces



$W = [0, 56] \times [0, 38]$. Suppose $\tilde{R} = 10 \geq$ the interaction radius. Should we take $S = W$ (i.e. no missing data) or e.g. $S = [-10, 66] \times [-10, 48]$ (missing data) or something else, and which kind of model/likelihood should we consider?

Likelihood-functions for Markov point processes

Consider 5 types of models/likelihoods corresponding to different situations:

1. $l(\theta)$ (the simple likelihood – no missing data):
   $S = W = [0, 56] \times [0, 38]$.
2. $l_{\text{mis}}(\theta)$ (the missing data likelihood): $W = [0, 56] \times [0, 38]$ and
   $S = [-10, 66] \times [-10, 48]$.
3. $\tilde{l}_{\text{mis}}(\theta)$ (the torodial missing data likelihood):
   $S = [-10, 66] \times [-10, 48]$ is wrapped on a torus and
   $W = [0, 56] \times [0, 38]$.
4. $\tilde{l}(\theta)$ (the torodial likelihood – no missing data):
   $S = W = [0, 56] \times [0, 38]$ is wrapped on a torus.
5. $l_{\ominus \tilde{R}}(\theta)$ (the conditional likelihood – based on the border method for missing data):
   $S = [0, 56] \times [0, 38]$, $W = S_{\ominus \tilde{R}} = [10, 46] \times [10, 28]$, and the likelihood is based on $\mathbf{X}_W | \mathbf{X}_{S \setminus W}$.

Table 9.1 in M & Waagepetersen (2004)

- Ignoring the radii, we fit a multiscale process with parameters β and $0 < \gamma_i \leq 1$, $R_i = 1.1 \times i$, $i = 1, \ldots, 4$.
- Parameter-estimates with different likelihood functions:

	$\log \beta$	$\log \gamma_1$	$\log \gamma_2$	$\log \gamma_3$	$\log \gamma_4$
$l(\theta)$	-1.78	-3.24	-1.03	-0.27	0.00
$l_{\text{mis}}(\theta)$	-0.84	-3.58	-1.38	-0.55	-0.12
$\tilde{l}_{\text{mis}}(\theta)$	-0.86	-3.63	-1.35	-0.55	-0.13
$\tilde{l}(\theta)$	-0.95	-3.53	-1.34	-0.55	-0.11
$l_{\ominus \tilde{R}}(\theta)$	-0.64	-3.26	-1.46	-0.64	-0.14

- -2 log likelihood ratio test for Strauss model ($\gamma_1 = \ldots = \gamma_4$) based on toroidal missing data likelihood: 47 (path sampling).
- Wald test for Strauss process: 29.

Briefly about Bayesian statistics

Given a realization $\mathbf{X}_W = \mathbf{x}_W$ of a spatial point process.

- 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 SNCP, as the centre process $\mathbf{C}$ is not observed, include this into the posterior. MCMC: Alternate between updating the unknown parameters and $\mathbf{C}$ (using the MH birth-death algorithm in the latter case).
  - For a LGCP, as the Gaussian process on the observation window, $\mathbf{Z}_W$, is not observed, include this (approximated on a grid) into the posterior and use INLA (Rue, Martino & Chopin, 2009), i.e. no MCMC!
- Markov point processes: Here the problem is the intractable normalizing constant of the likelihood which also enters in the posterior. Use the auxiliary variable method (M, Pettitt, Berthelsen & Reeves, 2006) – based on perfect simulations!

Alternative to likelihood: pseudo likelihood (Besag, 1975, 1977)

- Maximum likelihood estimation can be done approximatively for most Markov point processes.
- But the normalising constant can be time-consuming to approximate, since it needs importance (or path or bridge) sampling based on long MCMC runs.
- An alternative to the likelihood function is the pseudo likelihood function, which does not depend on unknown normalising constants.

Pseudo likelihood function

Definition of pseudo likelihood function $PL_A(\theta; \mathbf{x})$ on a region $A \subseteq S$ (Besag, 1977; Jensen & M, 1991):

- For $i = 1, 2, \ldots$, let $\{B_{ij}\}_{j=1,\ldots,m_i}$ be a partition of A into m_i cells, where partitions become finer as i increases.
- Then

$$PL_A(\theta; \mathbf{x}) = \exp(|A|) \lim_{i \to \infty} \prod_{j=1}^{m_i} f_{\theta,A}(\mathbf{x}_{B_{ij}} | \mathbf{x}_{S \setminus B_{ij}}).$$

- Fact: If $\mathbf{X} \sim f_\theta$ is Ruelle stable, then

$$PL_A(\theta; \mathbf{x}) = \exp\left(- \int_A \lambda_\theta(\mathbf{x}, u) \mathrm{d}u \right) \prod_{u \in \mathbf{x}_A} \lambda_\theta(\mathbf{x} \setminus u, u).$$

- Note that the normalising constants have disappeared in the last expression.

Maximization of pseudo likelihood function

- MPLE is found by solving $(\mathrm{d}/\mathrm{d}\theta) \log PL_A(\theta; \mathbf{x}) = 0$, i.e.

$$\int_A \frac{\mathrm{d}}{\mathrm{d}\theta} \lambda_\theta(\mathbf{x}, u) \mathrm{d}u = \sum_{u \in \mathbf{x}_A} \frac{\mathrm{d}}{\mathrm{d}\theta} \log \lambda_\theta(\mathbf{x} \backslash u, u)$$

- Exponential family case:

$$\int_A t(\mathbf{x}, u) \exp(\theta \cdot t(\mathbf{x}, u)) \mathrm{d}u = \sum_{u \in \mathbf{x}_A} t(\mathbf{x} \backslash u, u),$$

where $t(\mathbf{x}, u) = t(\mathbf{x} \cup u) - t(\mathbf{x})$.

Some facts about MPLE

- For Poisson processes, the pseudo likelihood agrees with the likelihood.
- For weak interaction, $PL \approx L$.
- For non-Poisson, the MPLE usually needs to be found numerically, but this is much easier/faster than for MCMC-MLE.
- In the exp. fam. case, the pseudo likelihood function is log concave, and the `spatstat` package contains fast estimation procedures for finding the MPLE.
- Edge effects are handled as for the usual likelihood (e.g. using the border method).
- Consistency: Jensen & M (1991).
- Asymptotic normality: Jensen & Künsch (1994).

Minimum contrast estimation

- The likelihood function or Papangelou conditional intensity is usually not known on closed form for Cox processes, so here we need something else than MLE or MPLE – e.g. minimum contrast estimation based on moments or summary statistics...
- Basic idea: choose parameters such that a selected theoretical summary statistic $T(r)$ is as close as possible to its empirical estimate $\hat{T}(r)$.
- "As close as possible" means that we find the parameters which minimise some "distance" function $d(\hat{T}, T)$, e.g.

$$d(\hat{T}, T) = \int_{a_1}^{a_2} \left(\hat{T}(r) - T(r) \right)^2 \mathrm{d}r$$

for some $0 \leq a_1 < a_2$.

Example of minimum contrast estimation

- Pair correlation function for Thomas process:

$$g(r) = 1 + \exp(-r^2/(4\omega^2))/(4\pi\omega^2\kappa).$$

- Minimum contrast using g-function:

$$\int_{a_1}^{a_2} \left(\hat{g}(r) - 1 - \exp(-r^2/(4\omega^2))/(4\pi\omega^2\kappa)\right)^2 \mathrm{d}r$$

  (here we could use an edge-corrected kernel estimate $\hat{g}$, see M & Waagepetersen (2004)).
- Numerical methods often required for calculating and minimising integral. Note that for fixed ω^2, the contrast is a second order polynomial w.r.t. κ.
- Note that Thomas process has parameters $(\kappa, \alpha, \omega^2)$ – α is not in the pair correlation function, so other methods are needed for estimating this. Use e.g. that $\mathbb{E}n(\mathbf{X}_W) = \alpha\kappa|W|$.

Minimum contrast estimation

- In principle any known summary statistic can be used (e.g. $K(r)$, $g(r)$, $F(r)$), or any function of these (e.g. $K(r)^2$).
- Choosing a summary statistic:
  - It should be easy to calculate.
  - It should reflect important aspects of data.
  - It should be well-defined (e.g. we defined F only for stationary processes).
- If we use a summary statistic for fitting a model, we should not use the same one for model checking!

R and spatstat

Consider some examples of statistical analyses in R using maximum pseudo likelihood estimation and minimum contrast estimation... (Rcode5Estimation.R)

3.3 OVERVIEW AND CONCLUDING REMARKS

Statistical analysis of a point pattern

A rough skeleton for doing a statistical analysis of a point pattern using the theory from this course:

1. Preliminary analysis
2. Model building
3. Parameter estimation (and testing)
4. Model checking
5. Conclusions

Note: the following slides provide some inspiration – there is no "standard way" of analysing a point pattern.

Preliminary analysis

- What do you know about the data before the analysis?
- Plot the data!
- Is the data clustered, regular, Poisson-like or something else? Visual inspection and/or estimation of summary statistics (g, K, L, F, G, J).
- Homogeneous or inhomogeneous? Visual inspection and/or estimation of ρ.
- Does the data have any features that require special attention (e.g. covariates)?

Model

- Choosing a model:
  - Are there scientifically interesting questions that should guide your modelling?
  - A mechanistic model (e.g. a Neyman-Scott process) – choice based on underlying mechanics
  - A statistical model (e.g. a LGCP) – choice based on preliminary analysis
- Model classes:
  - Poisson process – specify intensity function
  - Cox process – specify random intensity function: For a cluster process, specify cluster construction; for a LGCP, specify GRF.
  - Markov process – specify density/Papangelou conditional intensity
- If simulation is easy, , before going on to the much more time consuming task of estimation, simulate realizations for various parameter values to check whether the model has any resemblance to the data.

Parameter estimation

- Maximum likelihood estimation
  - Analytical – simple Poisson process models
  - Numerical (Newton-Raphson) – complicated Poisson process models and Markov point process models
  - Profile likelihoods for difficult parameters (e.g. the interaction radius in the Strauss process)
- Maximum pseudo likelihood estimation
  - Numerical (but no normalizing constants) – apply for Markov point processes
- Bayesian inference
  - MCMC-based – missing data approach for Cox process models and cluster process models – more complicated for Markov point processes...
- Minimum contrast estimation
  - Cox processes

Model checking

- Based on simulation:
  - Using definition directly (Poisson; Cox)
  - MCMC-based (Markov)
- Compare simulations and data visually
- Compare summary statistics for simulations and data
- If something does not fit:
  - Does the model checking tell how to improve the model?
  - Don't expect everything to work perfectly!

Conclusions

- Derive conclusion from the model
- Does the estimated parameters tell something about the data?
- Does the fact the model fits tell anything?
- If it does not fit, does this tell anything?