Lectures:

1. Intro to point processes, moment measures and the Poisson process
2. Cox and cluster processes
3. The conditional intensity and Markov point processes
4. Likelihood-based inference and MCMC

Aim: overview of stats for spatial point processes - and spatial point process theory as needed.

Not comprehensive: the most fundamental topics and our favorite things.

Data example (Barro Colorado Island Plot)

Observation window $W = [0, 1000] \times [0, 500]$ m$^2$

Sources of variation: elevation and gradient covariates and clustering due to seed dispersal.
What is a spatial point process?

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

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

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

In practice distribution specified by an explicit construction (this and second lecture) or in terms of a probability density (third lecture).

Second-order moments

Second order factorial moment measure:

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

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

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

NB (exercise):

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

Campbell formula (by standard proof)

$$
\mathbb{E} \sum_{u,v \in X} h(u,v) = \int \int h(u,v)\rho^{(2)}(u,v) \, du \, dv
$$

Moments of a spatial point process

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

Intensity measure $\mu$:

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

In practice often given in terms of intensity function

$$
\mu(A) = \int_A \rho(u) \, du
$$

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

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

Pair correlation function and $K$-function

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

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

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

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

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

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

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

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

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

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

$B = [0, 1] \times [0, 0.7]$: Homogeneous: $\rho = 150/0.7$  Inhomogeneous: $\rho(x, y) \propto e^{-10.6y}$

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

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

Exercises (30 minutes)

1. Show that the covariance between counts $N(A)$ and $N(B)$ is given by
   \[
   \text{Cov}[N(A), N(B)] = \mu^{(2)}(A \times B) + \mu(A \cap B) - \mu(A) \mu(B)
   \]
2. Show that
   \[
   K(t) := \int_{\mathbb{R}^2} 1[||u|| \leq t]g(u)du = \frac{1}{|B|} \mathbb{E} \sum_{u \in \mathbf{X} \cap B \neq \emptyset} \frac{1[||u - v|| \leq t]}{\rho(u) \rho(v)}
   \]
   What is $K(t)$ for a Poisson process? 
   (Hint: use the Campbell formula)
3. (Practical spatstat exercise) Compute and interpret a non-parametric estimate of the $K$-function for the spruces data set.
   (Hint: load spatstat using library(spatstat) and the spruces data using data(spruces). Consider then the Kest() function.)

Distribution and moments of Poisson process

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

By definition of a Poisson process

\[
P(\mathbf{X} \in F) = \sum_{n=0}^{\infty} \frac{e^{-\mu(S)}}{n!} \int_{S^n} 1[x_1, x_2, \ldots, x_n] \in F \prod_{i=1}^{n} \rho(x_i)dx_1 \ldots dx_n
\]  

(1)

Similarly,

\[
\mathbb{E} h(\mathbf{X}) = \sum_{n=0}^{\infty} \frac{e^{-\mu(S)}}{n!} \int_{S^n} h(x_1, x_2, \ldots, x_n) \prod_{i=1}^{n} \rho(x_i)dx_1 \ldots dx_n
\]
Proof of independent scattering (finite case)

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

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

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

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

\[
= \sum_{n=0}^{\infty} \frac{e^{-\mu(A \cup B)}}{n!} m! (n-m)! \sum_{m=0}^{n} \int_{A^m} \int_{B^{n-m}} 1[\{x_1, x_2, \ldots, x_m\} \in F] \prod_{i=1}^{n} \rho(x_i) dx_1 \ldots dx_n
\]

( interchange order of summation and sum over \( m \) and \( k = n - m \))

\[
P(\mathbf{X} \cap A \in F)P(\mathbf{X} \cap B \in G)
\]

Superpositioning and thinning

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

Conversely: Independent \( \pi \)-thinning of Poisson process \( \mathbf{X} \): independent retain each point \( u \) in \( \mathbf{X} \) with probability \( \pi(u) \). Thinned process \( \mathbf{X}_{\text{thin}} \) and \( \mathbf{X} \setminus \mathbf{X}_{\text{thin}} \) are independent Poisson processes with intensity functions \( \pi(u)\rho(u) \) and \( (1 - \pi(u))\rho(u) \). (Superpositioning and thinning results most easily verified using void probability characterization of Poisson process, see M & W, 2003)

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

Density (likelihood) of a finite Poisson process

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

\[
P(\mathbf{X}_1 \in F)
\]

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

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

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

where

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

Hence \( f \) is a density of \( \mathbf{X}_1 \) with respect to distribution of \( \mathbf{X}_2 \).
Data example: tropical rain forest trees

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

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

Inhomogeneous Poisson process

Log linear intensity function

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

Estimate $\beta$ from Poisson log likelihood (spatstat)

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

Model check using edge-corrected estimate of $K$-function

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

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

Implementation in spatstat

```r
> bei=ppp(beilpe$X,beilpe$Y,xrange=c(0,1000),yrange=c(0,500))
> beifit=ppm(bei,~elev+grad,covariates=list(elev=elevim,
> grad=gradim))
> coef(beifit) #parameter estimates
   (Intercept)     elev      grad
   -4.98958664   0.02139856   5.84202684
> fisherinf=vcov(beifit) #Fisher information matrix
> sqrt(diag(fisherinf)) #standard errors
     (Intercept)       elev       grad
   0.01750026 0.00228777 0.25586086
> rho=predict.ppm(beifit)
> Kbei=Kinhom(bei,rho) #warning: problem with large data sets.
> myKbei=myKest(cbind(bei$x,bei$y),rho,100,3,1000,500,F) #my own
#procedure
```

K-functions

Poisson process: $K(t) = \pi t^2$ (since $g = 1$) less than $K$ functions for data. Hence Poisson process models not appropriate.
Exercises (remaining time until 11:30)

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

2. Compute the second order product density for a Poisson process \( X \).
   (Hint: compute second order factorial measure using the Poisson expansion for \( X \cap (A \cup B) \) for bounded \( A, B \subseteq \mathbb{R}^2 \).)

3. (if time) Assume that \( X \) has second order product density \( \rho^{(2)} \) and show that \( g \) (and hence \( K \)) is invariant under independent thinning (note that a heuristic argument follows easy from the infinitesimal interpretation of \( \rho^{(2)} \)).
   (Hint: introduce random field \( R = \{ R(u) : u \in \mathbb{R}^2 \} \), of independent uniform random variables on \([0, 1]\), and independent of \( X \), and compute second order factorial measure for thinned process \( X_{\text{thin}} = \{ u \in X : R(u) \leq \rho(u) \} \).)

Solution: second order product density for Poisson

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

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

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

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

Discussion of exercises: 11.30-12:00

Solution: invariance of \( g \) (and \( K \)) under thinning

Since \( X_{\text{thin}} = \{ u \in X : R(u) \leq \rho(u) \} \),

\[
\mathbb{E} \sum_{u,v \in X_{\text{thin}}} 1[u \in A, v \in B] = \mathbb{E} \sum_{u,v \notin X_{\text{thin}}} 1[R(u) \leq \rho(u), R(v) \leq \rho(v), u \in A, v \in B]
\]

\[
= \mathbb{E} \mathbb{E} \left[ \sum_{u,v \notin X} 1[R(u) \leq \rho(u), R(v) \leq \rho(v), u \in A, v \in B] \mid X \right]
\]

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

\[
= \int_A \int_B p(u)p(v) \rho^{(2)}(u,v) du dv
\]
Cox processes

\( X \) is a Cox process driven by the random intensity function \( \Lambda \) if, conditional on \( \Lambda = \lambda \), \( X \) is a Poisson process with intensity function \( \lambda \).

Calculation of intensity and product density:

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

\[\text{Cov}(\Lambda(u), \Lambda(v)) > 0 \iff g(u, v) > 1 \quad \text{(clustering)}\]

Overdispersion for counts:

\[\text{Var} N(A) = \mathbb{E} \text{Var}[N(A) | \Lambda] + \text{Var} \mathbb{E}[N(A) | \Lambda] = \mathbb{E} N(A) + \text{Var} \mathbb{E}[N(A) | \Lambda]\]

Log Gaussian Cox process (LGCP)

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

\[
\log \Lambda(u) = z(u)\beta^T + \Psi(u)
\]

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

- Often sufficient to use power exponential covariance functions:

\[
c(u, v) \equiv \text{Cov}[\Psi(u), \Psi(v)] = \sigma^2 \exp \left( -\|u - v\|^2/\delta \right),
\]

\( \sigma, \alpha > 0, \ 0 \leq \delta \leq 2 \) (or linear combinations)

- Tractable product densities

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

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

Two simulated homogeneous LGCP’s

- Exponential covariance function
- Gaussian covariance function
Cluster processes

\( M \) ‘mother’ point process of cluster centres. Given \( M, X_m, m \in M \) are ‘offspring’ point processes (clusters) centered at \( m \).

Intensity function for \( X_m \): \( \alpha f(m, u) \) where \( f \) probability density and \( \alpha \) expected size of cluster.

Cluster process:
\[
X = \bigcup_{m \in M} X_m
\]

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

Nice expressions for intensity and product density if \( M \) Poisson on \( \mathbb{R}^2 \) with intensity function \( \rho(\cdot) \) (Campbell):
\[
\mathbb{E} \Lambda(u) = \mathbb{E} \alpha \sum_{m \in M} f(m, u) = \alpha \int f(m, u) \rho(m) \, dm \quad (= \kappa \alpha \text{ if } \rho(\cdot) = \kappa \text{ and } f(m, u) = f(u - m))
\]

Inhomogeneous Thomas process

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

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

Random intensity function:
\[
\Lambda(u) = \alpha \exp(z(u)_{1:p} \beta_{1:p}) \sum_{m \in M} f(u - m; \omega)
\]

Rain forest example:
\[
z_{1:2}(u) = (z_{\text{elev}}(u), z_{\text{grad}}(u))
\]
elevation/gradient covariate.

Example: modified Thomas process

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

Offspring \( X = \bigcup_m X_m \) distributed around mothers according to bivariate isotropic Gaussian density \( f \).

\( \omega \): standard deviation of Gaussian density
\( \alpha \): Expected number of offspring for each mother.

Cox process with random intensity function:
\[
\Lambda(u) = \alpha \sum_{m \in M} f(u - m; \omega)
\]

Density of a Cox process

- Restricted to a bounded region \( W \), the density is
\[
f(x) = \mathbb{E} \left[ \exp \left( |W| \right) \int_W \Lambda(u) \, du \prod_{u \in X} \Lambda(u) \right]
\]
- Not on closed form
- Fourth lecture: likelihood-based inference (missing data MCMC approach)
- Now: simulation free estimation
### Parameter Estimation: regression parameters

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

\[
\rho_\beta(u) = \kappa \exp(z(u)\beta^T_{1,p}) = \exp(z(u)\beta^T)
\]

\[
z(u) = (1, z_1(u)) \quad \beta = (\log(\kappa), \beta_{1,p})
\]

Consider indicators $N_i = 1\{X \cap C_i \neq \emptyset\}$ of occurrence of points in disjoint $C_i$ ($W = \cup C_i$) where $P(N_i = 1) \approx \rho_\beta(u_i) d C_i$, $u_i \in C_i$

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

\[
\prod_{i=1}^{n} (\rho_\beta(u_i) d C_i)^{N_i} (1-\rho_\beta(u_i) d C_i)^{1-N_i} = \prod_{i=1}^{n} \rho_\beta(u_i)^{N_i} (1-\rho_\beta(u_i))^{1-N_i}
\]

is

\[
I(\beta) = \sum_{u \in X \cap W} \log \rho(u; \beta) - \int_W \rho(u; \beta) \, du
\]

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

### Asymptotic distribution of regression parameter estimates

Assume increasing mother intensity: $\kappa = \kappa_{\rho} = n\tilde{\kappa} \to \infty$ and $M = \cup_{i=1}^{p} M_i$, $M_i$ independent Poisson processes of intensity $\tilde{\kappa}$.

Score function asymptotically normal:

\[
\frac{1}{\sqrt{n}} \frac{dI(\beta)}{d\alpha} = \frac{1}{\sqrt{n}} \left( \sum_{u \in X \cap W} z(u) - n\tilde{\kappa} \alpha \int_W z(u) \exp(z(u)\beta^T_{1,p}) \, du \right)
\]

\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \sum_{m \in M_i} \sum_{u \in X_{m} \cap W} z(u) - \tilde{\kappa} \alpha \int_W \exp(z_1(u)\beta^T_{1,p}) \, du \right] \approx N(0, V)
\]

where $V = \text{Var} \left( \sum_{m \in M_i} \sum_{u \in X_{m} \cap W} z(u) \right)$ ($X_m$ offspring for mother $m$).

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

\[
\sqrt{n} \left[ (\log(\hat{\alpha}), \hat{\beta}_{1,p}) - (\log \alpha, \beta_{1,p}) \right] \approx N(0, J^{-1} V J^{-1})
\]

### Parameter Estimation: clustering parameters

Theoretical expression for (inhomogeneous) $K$-function:

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

Estimate $\kappa$ and $\omega$ by matching theoretical $K$ with semi-parametric estimate (minimum contrast)

\[
\hat{K}(t) = \sum_{u,v \in X \cap W} 1[\|u - v\| \leq t] \lambda(u; \beta) \lambda(v; \beta) |W \cap W_{u-v}|
\]

### Results for Beilschmiedia

Parameter estimates and confidence intervals (Poisson in red).

<table>
<thead>
<tr>
<th>Elevation</th>
<th>Gradient</th>
<th>$\kappa$</th>
<th>$\alpha$</th>
<th>$\omega$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.02</td>
<td>5.84</td>
<td>0.02-0.06</td>
<td>0.89-10.80</td>
<td>8e-05</td>
</tr>
</tbody>
</table>

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

Evidence of positive association between gradient and Beilschmiedia intensity.
Generalisations

- Shot noise Cox processes driven by \( \Lambda(u) = \sum_{(c, \gamma) \in \Phi} \gamma \kappa(c, u) \)
  where \( c \in \mathbb{R}^2, \gamma > 0 \) (\( \Phi = \) marked Poisson process)

- Generalized SNCP's... (Møller & Torrisi, 2005)

Exercises (remaining time until 13:00)

1. For a Cox process with random intensity function \( \Lambda \), show that

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

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

\[
\Lambda(u) = \alpha \sum_{m \in \mathcal{M}} f(m, u)
\]

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

3. Compute the intensity and second-order product density for an inhomogeneous Thomas process.

   (Hint: interpret the Thomas process as a Cox process and use the Campbell formula)

Lunch: we return at 14:00

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

2. Cox and cluster processes

3. The conditional intensity and Markov point processes

4. Likelihood-based inference and MCMC
Density with respect to a Poisson process

\( X \) on bounded \( S \) has density \( f \) with respect to unit rate Poisson \( Y \) if

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

Example: Strauss process

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

A Strauss process \( X \) on \( S \) has density

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

with respect to a unit rate Poisson process \( Y \) on \( S \) and

\[
c = \mathbb{E} \exp(\beta n(Y) + \psi s(Y))
\]

is the normalizing constant (unknown).

Note: only well-defined (\( c < 1 \)) if \( \psi \leq 0 \).

Intensity and conditional intensity

Suppose \( X \) has hereditary density \( f \) with respect to \( Y \):

\( f(x) > 0 \Rightarrow f(y) > 0, y \subset x \).

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

Instead consider conditional intensity

\[
\lambda(u, x) = \frac{f(x \cup \{u\})}{f(x)}
\]

(does not depend on normalizing constant !)

Note

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

and

\[
\rho(u) dA \approx P(X \text{ has a point in } A) = \mathbb{E}P(X \text{ has a point in } A|X \sim A), u \in A
\]

Hence, \( \lambda(u, X)dA \) probability that \( X \) has point in very small region \( A \) given \( X \) outside \( A \).

Markov point processes

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

Thm (Hammersley-Clifford) The following are equivalent.

1. \( f \) is Markov.
2. 

\[
f(x) = \exp(\sum_{y \in x} U(y))
\]

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

Pairwise interaction process: \( U(y) = 0 \) whenever \( n(y) > 2 \).

NB: in H-C, \( R \)-close neighbourhood relation can be replaced by an arbitrary symmetric relation between pairs of points.
Modelling the conditional intensity function

Suppose we specify a model for the conditional intensity. Two questions:
1. does there exist a density $f$ with the specified conditional intensity?
2. is $f$ well-defined (integrable)?

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

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

Some examples

**Strauss** (pairwise interaction):

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

**Overlap** process (pairwise interaction marked point process):

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

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

**Area-interaction** process:

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

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

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

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

$$\mathbb{E} \sum_{u \in X} k(u, X \setminus \{u\}) = \int_S \mathbb{E}[\lambda(u, X)k(u, X)] \, du = \int_S \mathbb{E}[k(u, X) \mid u] \rho(u) \, du$$

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

Density of reduced Palm distribution (easily shown):

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

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

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

Statistical inference based on pseudo-likelihood

$x$ observed within bounded $S$. Parametric model $\lambda_0(u, x)$.

Let $N_i = 1[x \cap C_i \neq \emptyset]$ where $C_i$ disjoint partitioning of $S = \bigcup_i C_i$.

$$P(N_i = 1 \mid X \cap S \setminus C_i) \approx \lambda_0(u_i, X) \, dC_i$$

where $u_i \in C_i$. Hence composite likelihood based on the $N_i$:

$$\prod_{i=1}^n (\lambda_0(u_i, x) \, dC_i)^{N_i} (1 - \lambda_0(u_i, x) \, dC_i)^{1 - N_i} \equiv \prod_{i=1}^n \lambda_0(u_i, x)^{N_i} (1 - \lambda_0(u_i, x) \, dC_i)^{1 - N_i}$$

which tends to pseudo likelihood function

$$\prod_{u \in X} \lambda_0(u, x) \exp \left( - \int_S \lambda_0(u, x) \, du \right)$$

Score of pseudo-likelihood: unbiased estimating function by GNZ.
Pseudo-likelihood estimates asymptotically normal but asymptotic variance must be found by parametric bootstrap.

Flexible implementation for log linear conditional intensity (fixed R) in spatstat

Estimation of interaction range R: profile likelihood (§)

The spatial Markov property and edge correction

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

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

Factorization (Hammersley-Clifford):

$$f(x) = \prod_{y \in x \cap (B \cup \partial B)} \exp(U(y)) \prod_{y \in x \cap B, y \setminus S \setminus (B \cup \partial B) \neq \emptyset} \exp(U(y))$$

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

$$f_B(z|y) \propto f(z \cup y)$$

depends on $y$ only through $\partial B \cap y$.

Example: spruces

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

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

Edge correction using the border method

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

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

Border method: base inference on

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

i.e. conditional density of $X \cap W \cap R$ given $X$ outside $W \cap R$.
Strauss model for spruces

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

Exercises (remaining time until 15:30)

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

   (Hint: $\sum_{n=0}^{\infty} (\frac{n\epsilon}{R})^n \exp(n\beta + \psi(n(n-1)/2) = \infty$ if $\psi > 0$.)

2. Show that local stability for a spatial point process density ensures integrability. Verify that the area-interaction process is locally stable.

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

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

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

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

4. (if time) Verify the Georgii-Nguyen-Zessin formula for a finite point process.

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

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

Exercises (cntd).

4. (if time) Verify the Georgii-Nguyen-Zessin formula for a finite point process.

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

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

Concentrate on point processes specified by unnormalized density $h_0(x)$,

$$f_0(x) = \frac{1}{c(\theta)} h_0(x)$$

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

$$l(\theta) = \log h_0(x) - \log c(\theta)$$

Importance sampling

Importance sampling: $\theta_0$ fixed reference parameter:

$$l(\theta) \equiv \log h_0(x) - \log \frac{c(\theta)}{c(\theta_0)}$$

and

$$\frac{c(\theta)}{c(\theta_0)} = \mathbb{E}_{\theta_0} \frac{h_0(X)}{h_0(X)}$$

Hence

$$\frac{c(\theta)}{c(\theta_0)} \approx \frac{1}{m} \sum_{i=0}^{m-1} \frac{h_0(X_i)}{h_0(X_i)}$$

where $X_0, X_1, \ldots$, sample from $f_0$ (later).

Exponential family case

$$h_0(x) = \exp(t(x)\theta^T)$$

$$l(\theta) = t(x)\theta^T - \log c(\theta)$$

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

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

Path sampling (exp. family case)

Derivative of cumulant transform:

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

Hence, by integrating over differentiable path $\theta(t)$ (e.g. line) linking $\theta_0$ and $\theta_1$:

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

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

NB Monte Carlo approximation on log scale more stable.
Maximisation of likelihood (exp. family case)

Score and observed information:
\[ u(\theta) = t(x) - E_{\theta} t(X), \quad j(\theta) = \text{Var}_{\theta} t(X), \]

Newton-Raphson iterations:
\[ \theta^{m+1} = \theta^m + u(\theta^m) j(\theta^m)^{-1} \]

Monte Carlo approximation of score and observed information: use importance sampling formula
\[ E_{\theta} k(X) = E_{\theta_0} \left[ k(X) \exp \left( t(X) (\theta - \theta_0)^T \right) \right] / (c_\theta / c_{\theta_0}) \]
with \( k(X) \) given by \( t(X) \) or \( t(X)^T t(X) \).

MCMC simulation of spatial point processes

Birth-death Metropolis-Hastings algorithm for generating ergodic sample \( X^0, X^1, \ldots \) from locally stable density \( f \) on \( S \):

Suppose current state is \( X^i, i \geq 0 \).
1. Either: with probability 1/2
   - (birth) generate new point \( u \) uniformly on \( S \) and accept \( X^{\text{prop}} = X^i \cup \{ u \} \) with probability
     \[ \min \left\{ 1, \frac{f(X^i \cup \{ u \} \mid S)}{f(X^i) \mid S} \right\} \]
   - (death) select uniformly a point \( u \in X^i \) and accept \( X^{\text{prop}} = X^i \setminus \{ u \} \) with probability
     \[ \min \left\{ 1, \frac{f(X^i \setminus \{ u \} \mid n)}{f(X^i) \mid S} \right\} \]
     (if \( X^i = \emptyset \) do nothing)
2. if accept \( X^{i+1} = X^{\text{prop}} \); otherwise \( X^{i+1} = X^i \).

Missing data

Suppose we observe \( x \) realization of \( X \cap W \) where \( W \subset S \).
Problem: likelihood (density of \( X \cap W \))
\[ f_{W, \theta}(x) = E_{\theta}(x \cap Y_{S \setminus W}) \]
not known - not even up to proportionality! (\( Y \) unit rate Poisson on \( S \))

Possibilities:
- Monte Carlo methods for missing data.
- Conditional likelihood
  \[ f_{W_{\subset R}, \theta}(x \cap W_{\subset R} | x \cap (W \setminus W_{\subset R})) \propto \exp(t(x) \theta^T) \]
  (note: \( x \cap (W \setminus W_{\subset R}) \) fixed in \( t(x) \))
Likelihood-based inference for Cox/Cluster processes

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

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

observed within $W$ ($\mathcal{M}$ Poisson with intensity $\kappa$).

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

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

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

$$f(x, m; \theta) = f(m; \theta) f(x|m; \theta) = e^{\lambda \left[1 - \kappa \right]} e^{-\int_{\tilde{W}} \lambda(u) \, du} \prod_{u \in X} \Lambda(u)$$

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

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

(data augmentation) using birth-death MCMC.

Exercises

1. Check the importance sampling formulas

$$E_\theta k(X) = E_{\theta_0} \left[ k(X) \frac{h_\theta(X)}{h_{\theta_0}(X)} \right] / (c_\theta / c_{\theta_0})$$

and

$$c(\theta) / c(\theta_0) = E_{\theta_0} \frac{h_\theta(X)}{h_{\theta_0}(X)} \quad (3)$$

2. Show that the formula

$$L(\theta) / L(\theta_0) = E_{\theta_0} \left[ \frac{f(x, M \cap \tilde{W}; \theta)}{f(x, M \cap \tilde{W}; \theta_0)} \bigg| X \cap W = x \right]$$

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

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