Spatial Models, Stat 597A
Spatial Point Process Models

Murali Haran

February 24, 2010

The material and examples here are adapted from “Modern statistics for spatial point processes” by J.Møller and R.P.Waagepeterson (2006), P.J.Diggle’s online lecture notes, Møller and Waagepetersen’s monograph on spatial point processes (2004), Baddeley and Turner’s `R spatstat` package and Baddeley et al. “Case Studies in Spatial Point Process Modeling” (2005).

Types of spatial data: Recap/Summary

General spatial process: $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$, D is set of locations.

- **Geostatistics**: D is a fixed subset of $\mathbb{R}^2$ (or $\mathbb{R}^3$ in 3D case).

  $Z(\mathbf{s})$ is a random variable at each location $\mathbf{s} \in D$.

  Basic model: Gaussian process.

- **Areal/lattice data**: $D = \{\mathbf{s}_1, \ldots, \mathbf{s}_N\}$ is a fixed regular or irregular lattice, on $\mathbb{R}^2$ (or $\mathbb{R}^3$).

  $Z(\mathbf{s})$ is a random variable at each location $\mathbf{s} \in D$.

  Basic model: Gaussian Markov random field.

- **Spatial point process**: $D = \{\mathbf{s}_1, \ldots, \mathbf{s}_N\}$ is a random collection of points on the plane.

  Ordinary point process: $Z(\mathbf{s})$ does not exist. For marked point process, $Z(\mathbf{s})$ is a random variable as well. Basic models: Poisson process, Cox process, Markov process.

Spatial Point Processes: Introduction

- **Spatial point process**: The *locations* where the process is observed are random variables and process itself may not be defined; if defined, it is a **marked spatial point process**.
- **Observation window**: the area where points of the pattern can possibly be observed. The observation window specification is vitally important since absence of points in a region where they could potentially occur is also valuable information whereas absence of points outside of an observation window does not tell us anything.

Spatial Point Process: Example 1

Many problems can be formulated as spatial point process problems. Consider a study of tree species biodiversity (from Møller and Waagepetersen):

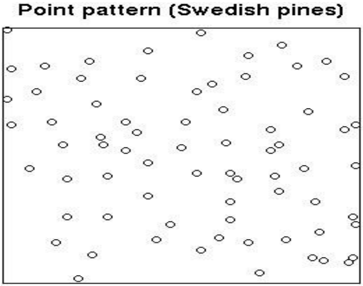
- Information available:
  - Locations of (potentially hundreds of thousands) of trees belonging to potentially thousands of species species.
  - Covariate information such as altitude, norm of altitude gradient etc.
- Some questions of interest:
  - Is the pattern completely random ?
  - If not completely random, can an explanatory point process model be fit to it?
  - How is the point pattern related to the covariates ?

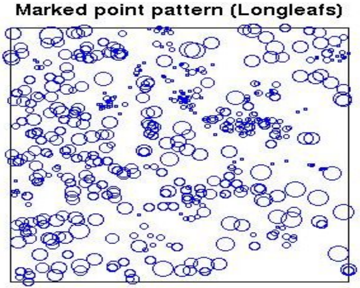
Spatial Point Process Data: Example 2

Locations of pine saplings in a Swedish forest.

Location and diameter of Longleaf pines (marked point process).

Are they randomly scattered or are they clustered?



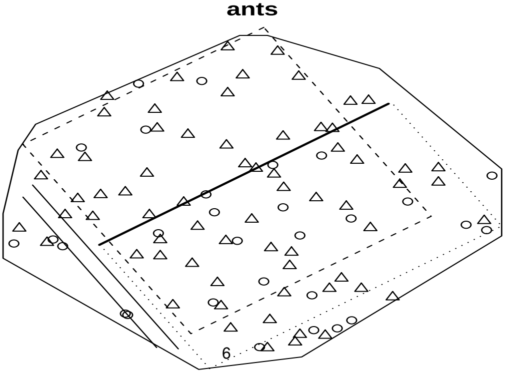


(from Baddeley and Turner `R` package, 2006)

Spatial Point Process: Example 3

Harkness-Isham ants' nests data (1983) from `spatstat` package.

Spatial locations of nests of two species of ants, *Messor wasmanni* and *Cataglyphis bicolor*, recorded at a site in northern Greece.



Spatial Point Process: Example 3 (contd)

- The harvester ant *M. wasmanni* collects seeds for food and builds a nest composed mainly of seed husks. *C. bicolor* is a heat-tolerant desert foraging ant which eats dead insects and other arthropods.
- Question of interest: Is there evidence for intra-species competition between *Messor nests* (i.e. competition for resources) and for preferential placement of *Cataglyphis nests* in the vicinity of *Messor nests*?

Questions related to spatial randomness of process

Ecology example:

- Is there regular spacing? Explanation: Possibly from competition for limited resources.
- Clustering: possibly from non-homogenous environmental conditions that are related to the presence/absence (e.g. conditions affect the habitability of a region).
- Clustering may occur by virtue of how species spreads (e.g. clustering of offsprings near parents).

Disease mapping example:

- Do cases show a surprising tendency to cluster together?
- Does risk vary spatially? *How* does it vary?
- Is risk elevated near a location? By how much?

  Hypothesis testing alone is inadequate.

Spatial Point Processes: Notes

- There appear to be many important problems where spatial point process modeling may be the most appropriate approach.
- However, the complexity of the theory along with computational difficulties have made it much less ‘friendly’ to applications than geostatistical models or areal models.
- Recent methodological developments and software such as the `R` library `spatstat` (A.Baddeley and Turner) are slowly opening up greater possibilities for practical modeling and analyses.

Classical Approaches

- Relatively small spatial point patterns.
- Assumption of stationarity is central and non-parametric methods based on summary statistics play a major role.
- Focus on hypothesis testing, Monte Carlo tests.
- Lack of software that works for classes of problems (software has been tailored to specific problems).
- In recent years, fast computing resources and better algorithms have allowed for analyses of larger point pattern data sets.

Some definitions for spatial point processes

- A spatial point process is a stochastic process, a realization of which consists of a countable set of points $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ in a bounded region $S \in \mathbb{R}^2$
- The points $\mathbf{s}_i$ are called **events.**
- For a region $A \in S$, $N(A) = \#(\mathbf{s}_i \in A)$.
- The **intensity measure** $\Lambda(A) = E(N(A))$ for any $A \in S$.
- If measure $\Lambda(A)$ has a density with respect to Lebesgue measure (we will typically assume this holds), then it can be written as:

$$\Lambda(A) = \int_A \lambda(\mathbf{s}) d\mathbf{s} \quad \text{for all A} \in S.$$

$\lambda(\mathbf{s})$ is called the **intensity function.**

Some definitions for spatial point processes (contd.)

- The process is **stationary** if for any integer k and regions $A_i,\ i = 1, \ldots, k$, the joint distribution of $N(A_1), \ldots, N(A_k)$ is translation-invariant,i.e., the joint distribution of $N(A_1), \ldots, N(A_k)$=joint distribution of $N(A_1 + \mathbf{y}), \ldots, N(A_k + \mathbf{y})$ for arbitrary **y.**
- The process is **isotropic** if for any integer k and regions $A_i,\ i = 1, \ldots, k$, the joint distribution of $N(A_1), \ldots, N(A_k)$ is invariant to rotation through an arbitrary angle, i.e., there is no directional effect.

Spatial point process modeling

Spatial point process models can be specified by :

- A deterministic intensity function (analogous to generalized linear model framework)
- A random intensity function (analogous to random effects models).
- Two classes of models:
  - Poisson Processes $\approx$ provide models for no interaction patterns.
  - Cox processes $\approx$ provide models for aggregated point patterns.
- Poisson process: Fundamental point process model — basis for exploratory tools and constructing more advanced point process models.

Homogeneous Poisson Process

Poisson process on $\mathbf{X}$ defined on S with intensity measure Λ and intensity function λ, satisfies for any bounded region $B \in S$ with $\Lambda(B) > 0$:

1. $N(B) \sim$Poisson($\Lambda(B)$).
2. Conditional on $N(B)$, the points (event locations) $\mathbf{X}_B = \{X_1, \ldots, X_{N(B)}\}$ in the bounded region are (i.i.d.) and each uniformly distributed in the region B.

- **Homogeneous Poisson process**: The intensity function, $\lambda(\mathbf{s})$ is constant for all $\mathbf{s} \in S$.
- Poisson process is a model for complete spatial randomness since $\mathbf{X}_A$ and $\mathbf{X}_B$ are independent for all $A, B \in S$ that are disjoint.

Poisson Process (contd.)

- The intensity $\lambda(\mathbf{s})$ specifies the mean number of events per unit area as a function of location $\mathbf{s}$.
- Intensity is called “density” in ecology (this term would be confused with a probability density, which is why it is not used in statistics).
- It is important as a ‘null model’ and as a simple model from which to build other models.
- Homogeneous Poisson process is model for complete spatial randomness against which spatial point patterns are compared.

Poisson Process (contd)

Some notes:

1. Stationarity $\Rightarrow \lambda(\mathbf{s})$ is constant $\Rightarrow \mathbf{X}$ is isotropic.
2. **Random thinning** of a point process is obtained by deleting the events in series of mutually independent Bernoulli trials. Random thinning of Poisson process results in another Poisson process.

- Independence properties of Poisson process makes it unrealistic for most applications. However, it is mathematically tractable and hence easy to use/study.
- For modeling, usually consider log model of intensity function (to preserve non-negativity of intensity): $\log \lambda(\mathbf{s}) = z(\mathbf{s})\beta^T$

Intensity of Poisson point process

Let $d\mathbf{s}$ denote a small region containing location **s**.

- First-order intensity function of a spatial point process:

$$\lambda(\mathbf{s}) = \lim_{d\mathbf{s}\to 0} \frac{E(N(d\mathbf{s}))}{|d\mathbf{s}|}.$$

- Second-order intensity function of a spatial point process:

$$\lambda^{(2)}(\mathbf{s}_1, \mathbf{s}_2) = \lim_{d\mathbf{s}_1\to 0} \lim_{d\mathbf{s}_2\to 0} \frac{E\{N(d\mathbf{s}_1)N(d\mathbf{s}_2)\}}{|d\mathbf{s}_1||d\mathbf{s}_2|}.$$

- Covariance density of a spatial point process

$$\gamma(\mathbf{s}_1, \mathbf{s}_2) = \lambda^{(2)}(\mathbf{s}_1, \mathbf{s}_2) - \lambda(\mathbf{s}_1)\lambda(\mathbf{s}_2).$$

Intensity of Poisson point process (contd.)

Assuming stationarity and isotropy:

- Constant intensity: If $\mathbf{s} \in A$, $\lambda(\mathbf{s}) = \lambda = E(N(A))/|A|$, constant for all A.
- Second order intensity depends only on distance between locations $\mathbf{s}_1, \mathbf{s}_2$: $\lambda^{(2)}(\mathbf{s}_1, \mathbf{s}_2) = \lambda^{(2)}(\|\mathbf{s}_1 - \mathbf{s}_2\|)$.
- $\gamma(d) = \lambda^{(2)}(d) - \lambda^2$, where $d = \|\mathbf{s}_1 - \mathbf{s}_2\|$.

Hard to interpret $\lambda^{(2)}$. Instead, consider the *reduced second moment function*, the K-function:

$$K(d) = 2\pi \frac{1}{\lambda^2} \int_0^d \lambda^{(2)}(r) dr.$$

Intensity of Poisson point process (contd.)

Still assuming stationarity and isotropy:

$K(d) = \frac{1}{\lambda}E$(number of events within distance d of an arbitrary event).

- Easier to interpret than second-order intensity and by dividing by λ, eliminate dependence on the intensity.
  - If process is clustered: Each event is likely to be surrounded by more events from the same cluster. $K(d)$ will therefore be *relatively large* for small values of d.
  - If process is randomly distributed in space: Each event is likely to be surrounded by empty space. For small values of d, $K(d)$ will be *relatively small*.
- Can obtain an intuitive estimator for $K(d)$ for a given data set.

Ripley's K Function

Let λ be the intensity of the process.

- Effective method for seeing whether the processs is completely random in space.

$$K(d) = \frac{\text{Mean number of events within distance d of an event}}{\lambda}$$

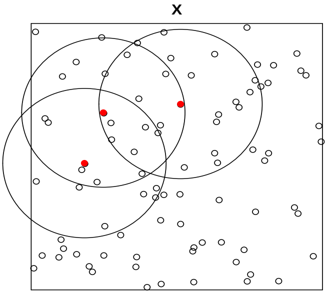
- This can be estimated by

$$\hat{K}(d) = \frac{\sum_{i \neq j} w_{ij} I(d_{ij} \leq d)}{\hat{\lambda}}$$

where $\hat{\lambda} = N/|A|$ with $|A|$ as the total area of the observation window and N is the observed count.

Estimating Ripley's K Function

Three circles of radius $d = 0.2$ each have been drawn with centers located at 3 locations where the process was observed. Note that they may overlap and also part of the circle may be outside the obervation window. Circles are drawn for every point, number of points within each circle is counted.



Ripley's K Function (contd.)

- What are the weights (w_{ij}s) ?
- Just a way to account for edge effects: For events close to the edge of the observation window, we cannot observe the events within radius d.
- When we are estimating the $K(d)$ corresponding to a circle centered at location of an event at $\mathbf{s}_i$, and we are looking at an event at location $\mathbf{s}_j$, the weight w_{ij} is the reciprocal of the portion of the circle of radius d that is inside the region. If circle is completely contained in the region, w_{ij} is 1; the smaller the portion contained in the region, the larger the weight w_{ij} assigned (to 'correct' for the fact that the count was only for an area smaller than πd^2).

Ripley's K Function (contd.)

- Under complete spatial randomness (homogeneous spatial Poisson point process):

$$E(K(d)) = \pi d^2.$$

- Easy to see why (simple proof):
  1. Location of events in a Poisson process are independent so occurrence of one event does not affect other events.
  2. Since E(number of events in a unit area)=λ, E(number of events in area within radius d)=$\lambda \pi d^2$.
  3. $E(K(d)) = \frac{1}{\lambda} \lambda \pi d^2 = \pi d^2$.
- Once we have obtained $\hat{K}(d)$, we can plot $\hat{K}(d)$ versus d.
- Compare it to the plot we would have obtained under complete spatial randomness.

Inhomogeneous Poisson processes

Useful for modeling spatial process that varies in intensity over space. An inhomogeneous Poisson process with intensity λ satisfies:

- Number of events $N(A)$ in an observation window A is Poisson with mean

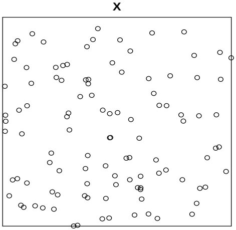
  $$\Lambda(A) = \int_A \lambda(\mathbf{s}) d\mathbf{s},$$

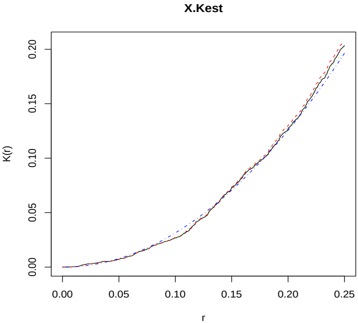
  equivalently, $P(N(A) = N) = \frac{1}{N!} e^{-\Lambda(A)} (\Lambda(A))^N$.
- Conditional on N(A), event locations are independently sampled from a probability density function proportional to $\lambda(\mathbf{s})$.

Ripley's K for homogeneous Poisson Process

Process was simulated with intensity function $\lambda(x, y) = 100$.

homogeneous Poisson Process Ripley's K





blue=K function under complete spatial randomness

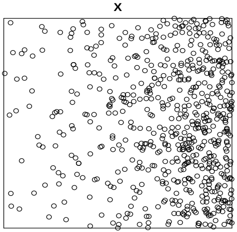
black (and red and green) are various versions of estimates of the K function

Ripley's K for inhomogeneous Poisson Process (Eg.1)

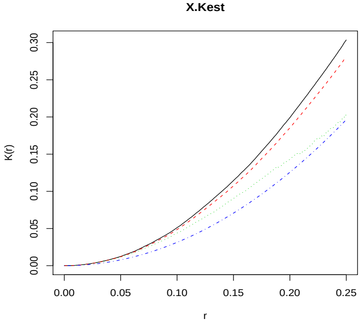
Process was simulated with intensity function

$\lambda(x, y) = 100\exp(3x)$.

Inhomogeneous Poisson Process



Ripley's K



blue=K function under complete spatial randomness

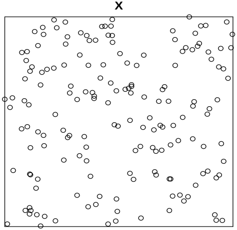
black (and red and green) are various versions of estimates of the K function

Ripley's K for inhomogeneous Poisson Process (Eg.2)

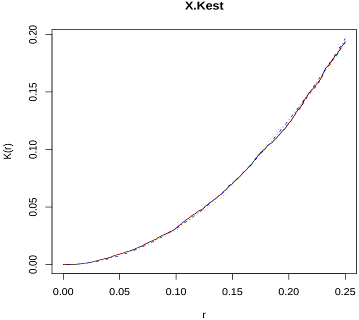
Process was simulated with intensity function

$\lambda(x, y) = 100 \exp(y)$.

Inhomogeneous Poisson Process



Ripley's K



blue=K function under complete spatial randomness

black (and red and green) are various versions of estimates of the K function

Ripley's K function: transformation

- As can be seen from Eg.2, even for strong departures from complete spatial randomness, difference between Ripley's K and its expectation under complete spatial randomness can be small.
- Plot of K function may not suffice. Instead, consider a linearizing transformation:

$$L(d) = \sqrt{K(d)/\pi} - d.$$

  - Complete spatial randomness: $E(L(d)) = 0$.
  - Clustering: $E(L(d)) > 0$.
  - Regular spacing: $E(L(d)) < 0$.
- Easy to interpret.

Ripley's K function: robustness to thinning

- As pointed out before, random thinning of a Poisson process results in a Poisson process.
- Also, random thinning reduces the intensity and the number of events within a distance d of a location by the same multiplicative factor.
- Since $K(d)$ is the ratio of the number of events within a distance d and the intensity of the process, it is robust to incomplete ascertainment (random thinning).
- Hence, $K(d)$ does not change as long as missing cases are missing at random (missingness does not depend on location).

Inference for Poisson process

We would like to be able to perform statistical inference for a point process. By definition, Poisson process on $\mathbf{X}$ defined on S with intensity measure Λ and intensity function λ satisfies for any bounded region $B \in S$ with $\Lambda(B) > 0$:

- $N(B) \sim$Poisson($\Lambda(B)$), i.e.

$$f(N(B)|\Lambda(B)) = \frac{\exp(-\Lambda(B))\Lambda(B)^{N(B)}}{N(B)!}$$

- Conditional on $N(B)$, the points (event locations) $\mathbf{X}_B = \{X_1, \ldots, X_{N(B)}\}$ in the bounded region are (i.i.d.) and each uniformly distributed in the region B:

$$f(X_1, \ldots, X_{N(B)}|N(B)) = \prod_{i=1}^{N(B)} f(X_i|N(B)) = \prod_{i=1}^{N(B)} \frac{\lambda(X_i)}{\int_B \lambda(\mathbf{s})ds}$$

Inference for Poisson process (contd.)

- The joint distribution is then:

$$f(X_1, \ldots, X_{N(B)}, N(B)) = \frac{\exp(-\Lambda(B))\Lambda(B)^{N(B)}}{N(B)!} \prod_{i=1}^{N(B)} \frac{\lambda(X_i)}{\int_B \lambda(\mathbf{s})ds}$$

$$= \frac{\exp(-\Lambda(B))\Lambda(B)^{N(B)}}{N(B)!} \prod_{i=1}^{N(B)} \frac{\lambda(X_i)}{\Lambda(B)} = \frac{\exp(-\Lambda(B))}{N(B)!} \prod_{i=1}^{N(B)} \lambda(X_i).$$

- For instance, this means that for a region $F \in S$ and a point process $\mathbf{X}$:

$$P(\mathbf{X} \in F, N(S) = n) = \int_S 1(\mathbf{X} \in F) \frac{\exp(-\Lambda(S))}{n!} \prod_{i=1}^{n} \lambda(X_i) d\mathbf{X}$$

and $P(\mathbf{X} \in F) = \sum_{n=0}^{\infty} P(\mathbf{X} \in F, N = n)$.

Space-varying covariates: modulated Poisson process

- As before, denote covariates at a location **s** by **X**(**s**). Impact of spatially varying covariates on a spatial point pattern may be modeled through the intensity function

$$\lambda(\mathbf{s}) = \exp(\beta X(\mathbf{s}))$$

- Inhomogeneous Poisson process with this intensity is a *modulated Poisson process*.
- Examples of **X**(**s**): spatially varying environmental variables such as elevation, precipitation etc., known functions of the spatial coordinates or distances to known environmental features (e.g. distance to nearest road).
- Important question: How is **X** related to the spatial point process intensity, i.e., what is β?

Parameter estimation for modulated Poisson process

Maximum likelihood estimation using observed **X** on a region S:

- The likelihood for the simple linear model is (from before):

$$\mathcal{L}(\mathbf{X}, N; \beta) = \frac{\exp(\Lambda(S))}{N(B)!} \prod_{i=1}^{N(B)} \lambda(X_i).$$

$$\mathcal{L}(\mathbf{X}, N; \beta) = \frac{\exp(-\int_S \exp(\mathbf{X}(\mathbf{s})\beta))}{N(B)!} \prod_{i=1}^{N(B)} \exp(\beta X_i).$$

- MLE for β: Find $\hat{\beta}$ that maximizes likelihood. This may be difficult, need to use Newton-Raphson or other optimization algorithm.
- Note that an assumption above is that covariates are available everywhere.

Modulated Poisson process with missing covariates

- In practice: Impractical to assume covariates are observed for every observed event and all locations in the observation window.
- Need to turn to other approaches. Natural approach is to estimate covariate information based on observed covariate information (cf. Rathbun, 1996).
- Use kriging to predict the values of the covariates at locations of observed events and at unsampled locations.
- Substitute predicted values of the covariates into the likelihood, though Rathbun (1996) suggests a bias correction term to add to the log-likelihood.
- Maximize this approximate likelihood to obtain coefficient estimates, $\tilde{\beta}$.

Cox Process

The Cox process or the *doubly stochastic Poisson process* (Cox, 1955) is a more flexible and realistic class of models than the Poisson process model.

- Natural extension of a Poisson process: Consider the intensity function of the Poisson process as a realization of a random field. We assume $\Lambda(A) = \int_A \lambda(\mathbf{s}) d\mathbf{s}$.
  - Stage 1: $N(A)|\Lambda \sim$Poisson$(\Lambda(A))$.
  - Stage 2: $\lambda(\mathbf{s})|\Theta \sim f(\cdot;\Theta)$ so that λ is stochastic, a nonnegative random field parametrized by Θ.
- Simple case: If $\lambda(\mathbf{s})$ is deterministic, $\mathbf{X}$ is a Poisson process with intensity $\lambda(\mathbf{s})$.

Cox Process: Examples

- Mixed Poisson process: $\lambda(\mathbf{s}) = \lambda_0$, a common positive *random variable* for all locations. **X** is a homogeneous Poisson process with intensity λ_0.
- Special case of Mixed Poisson process: $\lambda_0 \sim$Gamma. Then $N(A)$ (number of points in region A) follows a negative binomial distribution.
- Thinning of a Cox process: Random thinning of a Cox process results in a Cox process (same as in the Poisson process scenario).

Cox Process: Notes

- Important: For a single realization of the process, it is not possible to distinguish a Cox process from its corresponding Poisson process. Hence, the decision to use a Cox process model may be due to a variety of different reasons:
  - Desire to incorporate prior knowledge in a Bayesian setting.
  - Scientific questions to be investigated (e.g. fixed and random effects that influence intensity).
  - Features of particular problem — flexibility of Cox process may provide a better fit to the data.
- If Λ is stationary, then $\mathbf{X}$ is stationary.

Log Gaussian Cox Process (LGCP)

Møller, Syversveen and Waagepetersen (1996).

- This is a very natural model when viewed in a hierarchical spatial framework.
- As before, $N(A)|\Lambda \sim$Poisson$(\Lambda(A))$.
- Linear model for log-intensity, $\log(\lambda;\Theta)(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta} + \psi(\mathbf{s};\Theta)$.
- Now model the error, $\psi(\mathbf{s};\Theta)$ as a Gaussian process with mean 0 and parameters Θ.
- As before, any of the usual covariance function forms (exponential, Matern etc.) can be selected for the Gaussian process covariance.

Shot noise Cox Process (SNCP)

- A shot noise Cox process $\mathbf{X}$ has random intensity function,

$$\lambda(\mathbf{s}) = \sum_{(c,\gamma)\in\Phi} \gamma k(c, \mathbf{s}),$$

  where Φ is a Poisson process on $\mathbb{R}^2 \times (0, \infty)$, $c \in \mathbb{R}^2, \gamma > 0$ and $k(c, \cdot)$ is a density for a two-dimensional continuous random variable.
- Interpretation: $\mathbf{X}$ is distributed as the superposition (i.e. union) of independent Poisson processes $\mathbf{Y}_{(c,\gamma)}$ with intensity function $\gamma k(c, \cdot)$, and $(c, \gamma) \in \Phi$.
- $\mathbf{Y}_{(c,\gamma)}$ is a **cluster** with center c, mean number of points γ.
- $\mathbf{X}$ is therefore a **Poisson cluster process** (Bartlett, 1964).

Special cases of SNCP

- Neyman-Scott process: The center points c form a stationary Poisson process with intensity κ and the γs are all equal to $\alpha > 0$. This process is stationary with intensity $\lambda = \alpha\kappa$.
- Thomas process: If, in addition, $k(c, \cdot)$ is a bivariate normal density with mean c and covariance $\omega^2 I$, it is a modified Thomas process. Process is isotropic.
- Extension of SNCP: Incorporate covariate information:

$$\lambda(\mathbf{s}) = \exp(\mathbf{X}(\mathbf{s})\boldsymbol{\beta}) \sum_{(c,\gamma)\in\Phi} \gamma k(c, \mathbf{s})$$

Neyman-Scott process: Explanation

Useful to see how a Neyman-Scott process is constructed:

1. 'Parent' events (cluster centers, c): From a Poisson process with intensity κ.
2. Each parent (located at c) produces a random (Poisson) number of 'offspring' with expectation α. Each offspring is i.i.d. for each parent.
3. Positions of offspring relative to parents are i.i.d. distributed according to a bivariate density, $k(c, \cdot)$.

Shot noise Cox Process (SNCP): Example

An example from Waagepetersen and Schweder (2006) illustrates the use of an SCNP model:

- Consider the problem of modeling positions of 55 minke whales observed in a part of the North Atlantic.
- The whales are observed visually from a ship sailing along predetermined transect lines.
- The point pattern is an incomplete observation of all whale positions since it is only possible to observe whales within the vicinity of the ship, visibility may be poor and whales may be diving.

Shot noise Cox Process (SNCP): Example

- Probability of observing a whale is a decreasing function of the distance from the whale to the ship and is effectively 0 beyond 2kms. from the ship.
- Whales tend to cluster around locations of high prey intensity.
- Spatial point process model should take clustering (around some clustering centers) into account.
- Ultimate goal: estimate whale abundance around the region (including areas that are beyond observation window).
- Model that estimates intensity function of the whales will accomplish this.

Shot noise Cox Process (SNCP): Example

- Denote probability of observing a whale at location $\mathbf{s}$ by $p(\mathbf{s})$.
- Intensity, $\lambda(\mathbf{s}) = p(\mathbf{s}) \sum_{(c,\gamma)\in\Phi} \gamma k(c, \mathbf{s})$.
- Cluster centers, c form a stationary Poisson process with intensity κ.
- Assume γs are i.i.d. Gamma r.v.s with mean α and unit scale parameter.
- Assume cs are independent of the γs.
- $k(c, \cdot)$ is the density of a $N_2(c, \omega^2 I)$ restricted to $c + [-3\omega, 3\omega]^2$.

Markov Point Processes

- Point patterns may require a flexible description that allows for the points to interact.
- Markov point processes are models for point processes with interacting points (attractive or repulsive behavior can be modeled).
- 'Markovian' in that intensity of an event at some location $\mathbf{s}$, given the realization of the process in the remainder of the region, depends only on information about events within some distance of $\mathbf{s}$.
- Origins in statistical physics, used for modeling large interacting particle systems.

Strauss Process definition

The (stationary) Strauss process with interaction radius *r* and parameters β, γ is the pairwise interaction point process in which each point contributes a factor $\beta > 0$, and each pair of points closer than r units apart contributes a factor $\gamma \in [0, 1]$ to the probability density of the point pattern.

$$f(x_1, \ldots, x_n) = \alpha \beta^{n(\mathbf{x})} \gamma^{s(\mathbf{x})},$$

where $x_1, \ldots, x_n$ represent the points of the pattern, $N(\mathbf{x})$ is the number of points in the pattern, $s(\mathbf{x})$ is the number of distinct unordered pairs of points that are closer than r units apart, and α is the normalizing constant.

Strauss Process: notes

- The interaction parameter $\gamma \in (0, 1]$ so that this model describes an "ordered" or "inhibitive" pattern.
- Nonstationary Strauss process is similar except that the contribution of each individual point x_i is a function $\beta(x_i)$:

$$f(x_1, \ldots, x_n) = \alpha \left\{ \prod_{i=1}^{n(\mathbf{x})} \beta(x_i)^{n(\mathbf{x})} \right\} \gamma^{s(\mathbf{x})},$$

- Note that a Strauss process is both a Markov point process and an exponential family (the latter allows for the use of theory and computation based on general exponential family ideas).

Inference for spatial point process models

- Maximum likelihood for all but the simplest spatial point process model is analytically intractable. Maximum pseudolikelihood (MPL) is a useful approximation to maximum likelihood.
- For some models, can use Newton-Raphson or some variant but often need (Markov chain) Monte Carlo maximum likelihood (MCML), also refered to as simulated maximum likelihood (SML).
- No ‘automatic’ methods exist for fitting such models.
- Challenging to fit flexible new models.
- Simulating point processes: often easy. Inference: usually difficult.

Pseudolikelihood: A first ('quick and dirty') approach

Pseudolikelihood (due to Besag, 1974, 1977) is a general method for approximating the likelihood (not just for spatial point processes). However, it is particularly useful for relatively easy spatial point process model fitting.

- Let $\mathbf{X} = (X1, \ldots, X_N)$ be a vector of N discrete random variables.
- Pseudolikelihood for this vector is defined as follows:

$$PL(\theta, \mathbf{x}) = \prod_{i=1}^{N} P_\theta(X_i = x_i | X_j = x_j, j \neq i),$$

  that is, the product of conditional likelihoods of each X_i given the other X_is.
- Maximize $PL(\theta, \mathbf{x})$ to obtain MPLE $\tilde{\theta}$.

Pseudolikelihood: Notes

- PL works as an approximation by exploiting the conditional independence structure of $(X_1, \ldots, X_N)$.
- PL is a good approximation to the likelihood if $X_1, \ldots, X_N$ are approximately independent.
- PL is analytically tractable (unlike the likelihood for most spatial point processes).
- For exponential families, MPLE is consistent and asymptotically normal under certain conditions.
- General principle (folklore): MPLE is efficient if interaction is weak; inefficient for strong interactions.

Pseudolikelihood for spatial point processes

- Advantage of PL: possible to have an automated algorithm for a general class of problems.
- R command: `spatstat` function `ppm` fits models that include spatial trend, interpoint interaction, and dependence on covariates.
- Often works well in practice (Baddeley, 2005) and is, at the very least, a good first attempt at fitting a model to a spatial point pattern.
- MPLE can be used to get a guess for MLE before doing something more elaborate like Markov chain Maximum Likelihood.

Markov chain Maximum Likelihood

- Although we are discussing Markov chain Maximum Likelihood in the context of spatial point processes, it is a very general approach for Maximum likelihood estimation.
- Consider the problem of finding the MLE,

$$\hat{\theta} = arg \max_{\theta \in \Theta} f_\theta(x)$$

  but we only have

$$h_\theta(x) \text{ where } f_\theta(x) = \frac{h_\theta(x)}{c(\theta)}$$

  with $c(\theta) = \int_x h_\theta(x)dx$ and unknown.
- $c(\theta)$ is the **normalizing function** (often referred to as the normalizing constant) and may be very difficult to evaluate.
- How can we find $\hat{\theta}$ in this situation?

MCML

- Markov chain Maximum Likelihood (MCML), also called Simulated maximum likelihood (SML) is a clever method for estimating the MLE in such situations.
- Pick $\psi \in \Theta$, a fixed point in the parameter space such that, $h_\psi(x) = 0 \Rightarrow h_\theta(x) = 0$ for any (almost all) $\theta \in \Theta$.
- Now, notice that $\hat{\theta} = arg \max_{\theta \in \Theta} \log\left(\frac{f_\theta(x)}{f_\psi(x)}\right)$ since $f_\psi(x)$ is just a non-negative constant in θ and log is a monotone transformation.
- Let $\ell(\theta) = \log\left(\frac{f_\theta(x)}{f_\psi(x)}\right)$ so that

$$\hat{\theta} = arg \max_{\theta \in \Theta} \ell(\theta).$$

MCML

- This likelihood ratio, $\ell(\theta) = \log\left(\frac{h_\theta(x)}{c(\theta)}\right) - \log\left(\frac{h_\psi(x)}{c(\psi)}\right)$ which can be rewritten as:

$$\ell(\theta) = \log\left(\frac{h_\theta(x)}{h_\psi(x)}\right) - \log\left(\frac{c(\theta)}{c(\psi)}\right).$$

- Notice that the first term above is completely known. The second term is unknown and intractable.
- But, $\frac{c(\theta)}{c(\psi)} = \frac{1}{c(\psi)} \int_x h_\theta(x) dx = \frac{1}{c(\psi)} \int_x \frac{h_\theta(x)}{h_\psi(x)} h_\psi(x) dx$.
- Hence, $\frac{c(\theta)}{c(\psi)} = \int_x \frac{h_\theta(x)}{h_\psi(x)} \frac{h_\psi(x)}{c(\psi)} dx = \int_x \frac{h_\theta(x)}{h_\psi(x)} f_\psi(x) dx$.
- This implies:

$$\frac{c(\theta)}{c(\psi)} = E_{f_\psi}\left(\frac{h_\theta(x)}{h_\psi(x)}\right).$$

MCML

$\frac{c(\theta)}{c(\psi)}$ is an expectation with respect to a distribution f_ψ. Can be estimated by MCMC methods.

1. Draw samples $X_1, \ldots, X_N \sim f_\psi$. We only need h_ψ to construct a Metropolis-Hastings algorithm to simulate the X_is.
2. Compute the estimate of the log-likelihood ratio for any θ,

$$\hat{\ell}(\theta) = \log\left(\frac{h_\theta(x)}{h_\psi(x)}\right) - \log\left(\sum_{i=1}^{N} \frac{h_\theta(x_i)/h_\psi(x_i)}{N}\right) \to \ell(\theta).$$

- Can estimate $\ell(\theta)$ for any $\theta \in \Theta$ using a **single set of samples** $X_1, \ldots, X_N \sim f_\psi$.

MCML

Let simulated maximum likelihood estimate,

$$\tilde{\theta} = \mathit{arg} \max_{\theta \in \Theta} \hat{\ell}(\theta).$$

- Under fairly general conditions, $\tilde{\theta} \to \hat{\theta}$, the MLE (cf. Geyer, 1992).
- If θ is of low dimensionality (say 1 or 2-D), we can find θ by direct search.
- For higher dimensions, will need a more sophisticated approach. For example: take the derivative of $\hat{\ell}(\theta)$, $\bigtriangledown \hat{\ell}(\theta)$, and set it to 0. Can also estimate s.errors.

MCML notes

- Important to select ψ well. MCML works best when ψ is reasonably close to $\hat{\theta}$.
- Flexibility of MCML: Enormously expands the range of models we can consider. For instance, exponential family, $f_\theta(x) \propto h_\theta(x) = \exp(t(x)\theta)$ where $t(x)$ may be some vector valued statistic, θ is a parameter. Just a few examples:
  - Spatial point processes
  - Spatial generalized linear models.
  - Network models.
- See C.J.Geyer (1996, MCMC in Practice) for an easy to read description and references.

MCML for a simple spatial point processes

- Simple example: Stationary Strauss process $f_{\beta,\gamma}(x_1, \ldots, x_n) \propto \beta^{n(\mathbf{x})}\gamma^{s(\mathbf{x})}$, which can be written as an exponential family with vector valued statistic $t(x) = (N(x), S(x))$.
- MLE involves maximizing a function with an unknown normalizing function $c(\beta, \gamma)$.
- Easy to do with MCML: Find β^*, γ^* reasonable guesses for $\hat{\beta}, \hat{\gamma}$ (MLEs).
- Use M-H algorithm to simulate $X_1, \ldots, X_N \sim f_{\beta^*,\gamma^*}$.
- Using $X_1, \ldots, X_N$, can maximize estimated log-likelihood ratio to find MCML estimate of MLE.