1 Introduction to Spatial Point Processes

1.1 Introduction

Modern point process theory has a history that can trace its roots back to Poisson in 1837. However, much of the modern theory, which depends heavily on measure theory, was developed in the mid 20th century. Three lectures is not nearly enough time to cover all of the theory and applications. My goal is to give a general introduction to the topics I find instructive and those that I find useful for spatial modeling (at least from the Bayesian framework). Several good texts books are out there for further study. These include, but not limited to, Daley and Vere-Jones (2003), Daley and Vere-Jones (2008), Møller and Waagepetersen (2004) and Illian et al. (2008). Most of my material comes from Møller and Waagepetersen (2004), Illian et al. (2008), Møller, Syversveen, and Waagepetersen (1998) and Møller and Waagepetersen (2007). Most theorems and propositions will be stated without proof.

One may think of a spatial point process as a random countable subset of a space S. We will assume that $S \subseteq \mathbb{R}^d$. Typically, S will be a d-dimensional box or all of $\mathbb{R}^d$. However, it could also be $\mathcal{S}^{d-1}$, the $(d-1)$-dimensional unit sphere.

As an example, we may be interested in the spatial distribution of weeds in a square kilometer field. In this case we can assume $S = [0,1]^2 \subset \mathbb{R}^2$. As another example, we may be interested in the distribution of all major earthquakes that occurred during the last year. Then $S = \mathcal{S}^2 \subset \mathbb{R}^3$.

In many instances, we may only observe points in a bounded subset (window) $W \subseteq S$. For example we may be interested in the spatial distribution of a certain species of tree in the Amazon basis. For obvious reasons, it is impractical to count and note the location of every tree, and so we may concentrate our efforts to several square windows $W_i = [w_{i0}, w_{i1}]^2$, $w_{i1} > w_{i0}$, $W_i \cap W_j = \emptyset$, $j \neq i$ and $W = \cup_i W_i \subset S$.

The definition of a spatial point process includes both finite and countable processes. We will restrict attention to point processes, X whose realizations, x, are locally finite subsets of S.

Definition 1 *Let $n(x)$ denote the cardinality of a subset $x \subseteq S$. If x is not finite, set $n(x) = \infty$. Let $x_B = x \cap B$ for $B \subseteq S$. Now, x is said to be* locally finite *if $n(x_B) < \infty$ whenever B is bounded.*

Hence, X takes values in the space

$$N_{lf} = \{x \subseteq S : n(x_B) < \infty, \forall \text{ bounded } B \subseteq S\}.$$

Elements of N_{lf} are called *locally finite point configurations* and will be denoted by $x, y, \ldots$, while $\xi, \eta, \ldots$ will denote singletons in S.

Before continuing, we will provide the formal definition of a point process. Assume S is a complete, separable metric space (c.s.m.s.) with metric d and equip it with the Borel sigma algebra $\mathcal{B}$ and let $\mathcal{B}_0$ denote the class of bounded Borel sets. Equip N_{lf} with the sigma algebra

$$\mathcal{N}_{lf} = \sigma(\{x \in N_{lf} : n(x_B) = m\} : B \in \mathcal{B}_0,\, m \in \mathbb{N}_0).$$

That is N_{lf} is the smallest sigma algebra generated by the sets

$$\{x \in N_{lf} : n(x_B) = m\} : B \in \mathcal{B}_0,\, m \in \mathbb{N}_0$$

where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.

Definition 2 *A* point process X *defined on* S *is a measurable mapping defined on some probability space* $(\Omega, \mathcal{F}, \mathcal{P})$ *taking values in* $(N_{lf}, \mathcal{N}_{lf})$*. This mapping induces a distribution* P_X *of* X *given by* $P_X(F) = P(\{\omega \in \Omega : X(\omega) \in F\})$*, for* $F \in \mathcal{N}_{lf}$*. In fact, the measurability of* X *is equivalent to the number, N(B), of points in* $B \in \mathcal{B}$ *being a random variable.*

In a slight abuse of notation we may write $X \subseteq S$ when we mean $X \in \mathcal{B}$ and $F \subseteq N_{lf}$ when we mean $F \in \mathcal{N}_{lf}$. When $S \subseteq \mathbb{R}^d$ the metric $d(\xi, \eta) = ||\xi - \eta||$ is the usual Euclidean distance.

1.1.1 Characterizations of Point Processes

There are three characterizations of a point process. That is the distribution of a point process X is completely specified by each characterization. The three characterizations are by its finite dimensional distributions, its void probabilities and its generating functional.

Definition 3 *The family of finite-dimensional distributions of a point process* X *on c.s.m.s.* S *is the collection of the joint distributions of* $(N(B_1), \ldots, N(B_m))$*, where* $(B_1, \ldots, B_m)$ *ranges over the bounded borel sets* $B_i \subseteq S$*,* $i = 1, \ldots, m$ *and* $m \in \mathbb{N}$*.*

Theorem 1 *The distribution of a point process* X *on a c.s.m.s.* S *is completely determined by its finite-dimensional distributions.*

In other words, if two point processes share the same finite-dimensional distributions, then they are equal in distribution.

Definition 4 *A point process on S is called a* simple point process *if its realizations contain no coincident points. That is $N(\{\xi\}) \in \{0, 1\}$ almost surely for all $\xi \in S$.*

The void probabilities of $B \subseteq S$ is $v(B) = P(N(B) = 0)$.

Theorem 2 (Rényi, 1967) *The distribution of a simple point process X on S is uniquely determined by its void probabilities of bounded Borel sets $B \in \mathcal{B}_0$.*

The probability generating functional plays the same role for a point process as the probability generating function plays for a non-negative integer-valued random variable. For a point process, X, the probability generating functional is defined by

$$G_X(u) = \mathbb{E}\left[\exp\left\{\int_S \ln u(\xi) dN(\xi)\right\}\right] = \mathbb{E}\left[\prod_{\xi \in X} u(\xi)\right]$$

for functions $u : S \to [0, 1]$ with $\{\xi \in S : u(\xi) < 1\}$ bounded. As a simple example, for $B \in \mathcal{B}_0$ take $u(\xi) = t^{\mathrm{I}(\xi \in B)}$ with $0 \leq t \leq 1$. Then $G_X(u) = \mathbb{E}\left[t^{N(B)}\right]$, which is the probability generating function for $N(B)$.

Theorem 3 *The distribution of a point process X on S is uniquely determined by its generating functional.*

For the remaining lectures we assume X is a simple point process with $S \subseteq \mathbb{R}^d$, with $d \leq 3$ typically.

1.2 Spatial Poisson Processes

Poisson point processes play a fundamental role in the theory of point processes. They possess the property of "no interaction" between points or "complete spatial randomness". As such, they are practically useless as a model for a spatial point pattern as most spatial point patterns exhibit some degree of interaction among the points. However, they serve as reference processes when summary statistics are studied and as a building block for more structured point process models.

We start with a space $S \subseteq \mathbb{R}^d$ and an *intensity function* $\rho : S \to [0, \infty)$ that is *locally integrable*: $\int_B \rho(\xi) d\xi < \infty$ for all bounded $B \subseteq S$. We also define the *intensity measure* by $\mu(B) = \int_B \rho(\xi) d\xi$. This measure is *locally finite*: $\mu(B) < \infty$ for bounded $B \subseteq S$ and *diffuse*: $\mu(\{\xi\}) = 0$, for all $\xi \in S$.

We first define a related process, *the binomial point process*:

Definition 5 *Let f be a density function on a set $B \subseteq S$ and let $n \in \mathbb{N}$. A point process X consisting of n independent and identically distributed points with common density f is called a* binomial point process *of n points in B with density f:*

$$X \sim \text{binomial}(B, n, f).$$

Now we give a definition of a (spatial) Poisson process.

Definition 6 *A point process X on $S \subseteq \mathbb{R}^d$ is a* Poisson point process *with intensity function ρ (and intensity measure μ) if the following two properties hold:*

- *for any $B \subseteq S$ such that $\mu(B) < \infty$, $N(B) \sim \text{Pois}(\mu(B))$—the Poisson distribution with mean $\mu(B)$.*
- *For any $n \in \mathbb{N}$ and $B \subseteq S$ such that $0 < \mu(B) < \infty$*

$$[X_B \mid N(B) = n] \sim \text{binomial}\left(B, n, \frac{\rho(\xi)}{\mu(B)}\right),$$

that is, the density function of the binomial point process is $f(\cdot) = \rho(\cdot)/\mu(B)$.

We write

$$X \sim \text{Poisson}(S, \rho).$$

Note that for any bounded $B \subseteq S$, μ determines the expected number of points in B:

$$\mathbb{E}(N(B)) = \mu(B).$$

So, if S is bounded, this gives us a simple way to simulate a Poisson process on S. 1) Draw $N(B) \sim \text{Pois}(\mu(B))$. 2) Draw $N(B)$ independent points uniformly on S.

Definition 7 *If $X \sim \text{Poisson}(S, \rho)$, then X is called a* homogeneous Poisson process *if ρ is constant, otherwise it is said to be* inhomogeneous. *If $X \sim \text{Poisson}(S, 1)$, then X is called the* standard Poisson process *or the* unit rate Poisson process *on S.*

Definition 8 *A point process X on $\mathbb{R}^d$ is* stationary *if its distribution is invariant under translations. It is* isotropic *if its distribution is invariant under rotations about the origin.*

The void probabilities of a Poisson process, for bounded $B \subseteq S$, are

$$v(B) = P(N(B) = 0) = \frac{\exp(-\mu(B))\mu^0(B)}{0!} = \exp(-\mu(B)),$$

since $N(B) \sim \text{Pois}(\mu(B))$.

Proposition 1 *Let $X \sim \text{Poisson}(S, \rho)$. The probability generating functional of X is given by*

$$G_X(u) = \exp\left\{-\int_S (1 - u(\xi))\rho(\xi)d\xi\right\}.$$

We now prove this for the special case when ρ is constant. Consider a bounded $B \subseteq S$. Set $u(\xi) \equiv 1$ for $\xi \in S \setminus B$. Then

$$\begin{aligned}
\mathbb{E}\left[\exp\left\{\int_S \ln u(\xi) dN(\xi)\right\}\right] &= \mathbb{E}\left[\prod_{\xi \in X} u(\xi)\right] = \mathbb{E}\left\{\mathbb{E}\left[u(x_1)\cdots u(x_n) \mid N(B) = n\right]\right\} \\
&= \sum_{n=0}^{\infty} \frac{\exp(-\rho|B|)(\rho|B|)^n}{n!} \int_B \cdots \int_B \frac{u(x_1)\cdots u(x_n)}{|B|^n} dx_1, \ldots, dx_n \\
&= \exp(-\rho|B|) \sum_{n=0}^{\infty} \left\{\int_B \rho u(\xi) d\xi\right\}^n / n! \\
&= \exp\left\{-\rho \int_B [1 - u(\xi)] d\xi\right\} \\
&= \exp\left\{-\rho \int_S [1 - u(\xi)] d\xi\right\}.
\end{aligned}$$

where the last equality holds since $u \equiv 1$ on $S \setminus B$.

There are two basic operations for a point process: thinning and superposition.

Definition 9 *A disjoint union $\cup_{i=1}^{\infty} X_i$ of point processes $X_1, X_2, \ldots$ is called a* superposition.

Definition 10 *Let $p : S \to [0,1]$ be a function and X a point process on S. The point process $X_{thin} \subseteq X$ obtained by including $\xi \in X$ in X_{thin} with probability $p(\xi)$, where the points are included/excluded independently on each other, is said to be an* independent thinning *of X with retention probabilities $p(\xi)$.*

Proposition 2 *If $X_i \sim \text{Poisson}(S, \rho_i)$, $i = 1, 2, \ldots$, are mutually independent and $\rho = \sum \rho_i$ is locally integrable, then with probability one, $X = \cup_{i=1}^{\infty} X_i$ is a disjoint union and $X \sim \text{Poisson}(S, \rho)$.*

That $X \sim Poisson(S, \rho)$ is easy to show using void probabilities:

$$\begin{aligned} P(N(B) = 0) &= P\left(\sum N_i(B) = 0\right) = P(\cap(N_i(B) = 0)) \\ &= \prod_i P(N_i = 0) = \prod_i \exp(-\mu_i(B)) = \exp(-\mu(B)). \end{aligned}$$

Proposition 3 *Let $X \sim \text{Poisson}(S, \rho)$ and suppose it is subject to independent thinning with retention probabilities p_{thin} and let $\rho_{thin}(\xi) = p_{thin}(\xi)\rho(\xi)$. Then X_{thin} and $X \setminus X_{thin}$ are independent Poisson processes with intensity functions $\rho_{thin}(\xi)$ and $\rho - \rho_{thin}(\xi)$, respectively.*

The density of a Poisson process does not exist with respect to Lesbesgue measure. However the density does exist with respect to another Poisson process under certain conditions. And if the space S is bounded, the density of any Poisson process exists with respect to the unit rate Poisson process. We need the following proposition in order to define the density of a Poisson process.

Proposition 4 *Let $h : N_{lf} \to [0, \infty)$ and $B \subseteq S$. If $X \sim Poisson(S, \rho)$ with $\mu(B) < \infty$, then*

$$\mathbb{E}[h(X_B)] = \sum_{n=0}^{\infty} \frac{\exp(-\mu(B))}{n!} \int_B \cdots \int_B h(\{x_i\}_{i=1}^n) \prod_{i=1}^n \rho(x_i) dx_1, \ldots dx_n. \tag{1}$$

In particular, when $h(\{x_i\}_{i=1}^n) = \text{I}(\{x_i\}_{i=1}^n \in F)$ for $F \subseteq N_{lf}$ we get

$$P(X_B \in F) = \sum_{n=0}^{\infty} \frac{\exp(-\mu(B))}{n!} \int_B \cdots \int_B \text{I}(\{x_i\}_{i=1}^n \in F) \prod_{i=1}^n \rho(x_i) dx_1, \ldots dx_n. \tag{2}$$

Note that (2) also follows directly from Definition 6.

Definition 11 *If X_1 and X_2 are two point processes defined on the same space S, then the distribution of X_1 is* absolutely continuous *with respect to the distribution of X_2 if and only if $P(X_2 \in F) = 0$ implies that $P(X_1 \in F) = 0$ for $F \subseteq N_{lf}$. Equivalently, by the Radon-Nikodym theorem, if there exists a function $f : N_{lf} \to [0, \infty]$ so that*

$$P(X_1 \in F) = \mathbb{E}\left[\mathrm{I}(X_2 \in F) f(X_2)\right], \quad F \subseteq N_{lf}. \tag{3}$$

We call f a density *for X_1 with respect to X_2.*

Proposition 5 *Let $X_1 \sim \text{Poisson}(\mathbb{R}^d, \rho_1)$ and $X_2 \sim \text{Poisson}(\mathbb{R}^d, \rho_2)$. Then the distribution of X_1 is absolutely continuous with respect to the distribution of X_2 if and only if $\rho_1 = \rho_2$.*

Thus two Poisson processes are not necessarily absolutely continuous respect to one another. The next proposition gives conditions when the distribution of one Poisson process is absolutely continuous with respect to another.

Proposition 6 *Let $X_1 \sim \text{Poisson}(S, \rho_1)$ and $X_2 \sim \text{Poisson}(S, \rho_2)$. Also suppose $\mu_i(S) < \infty$, $i = 1, 2$ so that S is bounded and that $\rho_2(\xi) > 0$ whenever $\rho_1(\xi) > 0$. Then the distribution of X_1 is absolutely continuous with respect to X_2 with density*

$$f(x) = \exp\left[\mu_2(S) - \mu_1(S)\right] \prod_{\xi \in x} \frac{\rho_1(\xi)}{\rho_2(\xi)}$$

for finite point configurations $x \subset S$ with $0/0 = 0$.

The proof of this follows easily from Proposition 4. We need to show that with this f, (3) is satisfied. But this follows immediately from (1). Now for bounded S the distribution of any Poisson process, $X \sim \text{Poisson}(S, \rho_1)$, is absolutely continuous with respect to the distribution of the unit rate Poisson process. For we always have that $\rho_2(\xi) \equiv 1 > 0$ whenever $\rho_1(\xi) > 0$.

1.3 Spatial Cox Processes

As mentioned above, the Poisson process is usually too simplistic to be of much value in a statistical model of point pattern data. However, it can be used to construct more complex and flexible models. The first class of model is the Cox process. We will spend the remaining time today discussing the Cox process, in general. Tomorrow we will discuss several specific Cox processes. Cox processes are models for aggregated or clustered point patterns.

A Cox process is a natural extension of a Poisson process. It is obtained by considering the intensity function of a Poisson process as a realization of a random field. These processes were first studied by Cox (1955) under the name *doubly stochastic Poisson processes*.

Definition 12 *Suppose that $Z = \{Z(\xi) : \xi \in S\}$ is a non-negative random field such that with probability one, $\xi \to Z(\xi)$ is a locally integrable function. If $[X \mid Z] \sim \text{Poisson}(S, Z)$, then X is said to be a* Cox process *driven by Z.*

Example 1 *A simple Cox process is the* mixed Poisson process. *Let $Z(\xi) \equiv Z_0$ be constant on S. Then $[X \mid Z_0] \sim \text{Poisson}(S, Z_0)$ (homogeneous Poisson). A special tractable case is when $Z_0 \sim G(\alpha, \beta)$. Then the counts $N(B)$ follows a negative binomial distribution. It should be noted that for disjoint bounded $A, B \subset S$, $N(A)$ and $N(B)$ are positively correlated.*

Example 2 *Random independent thinning of a Cox process X results in a new Cox process X_{thin}. Suppose X is a Cox process driven by Z. Let $\Pi = \{\Pi(\xi) : \xi \in S\} \subseteq [0, 1]$ be a random field that is independent of both X and Z. Let X_{thin} denote the point process obtained by independent thinning of the points in X with retention probabilities Π. Then X_{thin} is a Cox process driven by $Z_{thin}(\xi) = \Pi(\xi)Z(\xi)$. This follows immediately from the definition of a Cox process and the thinning property of a Poisson process.*

Typically Z is not observed and so it is impossible to distinguish a Cox process X from the Poisson process $X \mid Z$ when only one realization of X is available. Which of the two models is most appropriate (whether Z should be random or deterministic) depends on

- prior knowledge. In the Bayesian setting, one can incorporate prior knowledge of the intensity function into the model. See Example 1 above.
- if we want to investigate the dependence of certain covariates associated with Z. The covariates can then be treated as systematic terms, while unobserved effects may be treated as random terms.
- It may be difficult to model an aggregated point pattern with a parametric class of inhomogeneous Poisson processes (for example, a class of polynomial intensity functions). Cox processes allow more flexibility and perhaps a more parsimonious model. For example, the coefficients in the polynomial intensity functions could be random, as opposed to fixed.

The properties of a Cox process X are easily derived by conditioning on the random intensity Z and exploiting the properties of the Poisson process $X \mid Z$.

The intensity function of X is $\rho(\xi) = \mathbb{E}[Z(\xi)]$. If we restrict a Cox process X to a set $B \subseteq S$, with $|B| < \infty$, then the density of $X_B = X \cap B$ with respect to the distribution of a unit-rate Poisson process is, for finite point configurations $x \subset B$,

$$\pi(x) = \mathbb{E}\left[\exp\left(|B| - \int_B Z(\xi)d\xi\right)\prod_{\xi\in x} Z(\xi)\right].$$

For a bounded $B \subseteq S$, the void probabilities are given by

$$v(B) = P(N(B) = 0) = \mathbb{E}[P(N(B) = 0 \mid Z)] = \mathbb{E}\left[\exp\left(-\int_B Z(\xi)d\xi\right)\right].$$

For $u : S \to [0, 1]$, the generating functional is

$$G_X(u) = \mathbb{E}\left[\exp\left(-\int_S (1 - u(\xi))Z(\xi)d\xi\right)\right].$$

1.3.1 Log Gaussian Cox Processes

We now consider a special case of a Cox process that is analytically tractable. Let $Y = \ln Z$ be a Gaussian random field in $\mathbb{R}^d$. A Gaussian random field is a random process whose finite dimensional distributions are Gaussian distributions.

Definition 13 (Log Gaussian Cox process (LGCP)) *Let X be a Cox process on $\mathbb{R}^d$ driven by $Z = \exp(Y)$ where Y is a Gaussian random field. Then X is said to be a* log Gaussian Cox process*.*

The distribution of (X, Y) is completely determined by the mean and covariance function

$$m(\xi) = \mathbb{E}(Y(\xi)) \quad \text{and} \quad c(\xi, \eta) = \mathrm{Cov}(Y(\xi), Y(\eta)).$$

The intensity function of a LGCP is

$$\rho(\xi) = \exp(m(\xi) + c(\xi, \xi)/2).$$

We will assume that $c(\xi, \eta) = c(||\xi - \eta||)$ is translation invariant and isotropic with form

$$c(||\xi - \eta||) = \sigma^2 r(||\xi - \eta||/\alpha)$$

where $\sigma^2 > 0$ is the variance and r is a known correlation function with $\alpha > 0$ a correlation parameter.

The function $r : \mathbb{R}^d \to [-1, 1]$ with $r(0) = 1$ for all $\xi \in \mathbb{R}^d$ is a correlation function for a Gaussian random field if and only if r is a positive semidefinite function:

$$\sum_{i,j} a_i a_j r(\xi_i, \xi_j) \geq 0$$

for all $a_1, \ldots, a_n \in \mathbb{R}$ and $\xi_1, \ldots, \xi_n \in \mathbb{R}^d$ and $n \in \mathbb{N}$.

A useful family of correlation functions is the *power exponential family*:

$$r(||\xi - \eta||/\alpha) = \exp(-||\xi - \eta||^\delta/\alpha) \quad \delta \in [0, 2].$$

The parameter δ controls the amount of smoothness of the realizations of the Gaussian random field and typically fixed based (hopefully) on a priori knowledge. Special cases are the *exponential correlation function* when $\delta = 1$, the *stable correlation function* when $\delta = 1/2$ and the *Gaussian correlation function* when $\delta = 2$.

We will discuss Bayesian modeling of LGCPs along with a fast method to simulate a Gaussian field in Lecture 3.

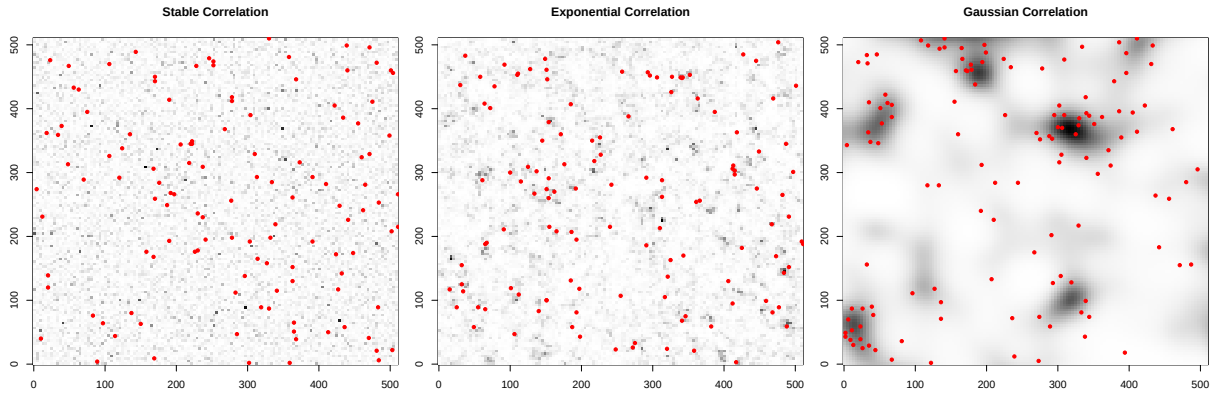

Table 1: Realizations of LGCPs on $S = [0, 512]^2$ with $m(\xi) = 4.25$, $\sigma^2 = 1$ and $\alpha = 0.01$.

2 Aggregative, Repulsive and Marked Point Processes

2.1 Aggregative Processes

The identification of centers of clustering is of interest in many areas of applications including archeology, mining, forestry and astronomy. For example, young trees in a natural forest and galaxies in the universe form clustered patterns. In clustered patterns the point density varies strongly over space. In some regions the density is high and other regions it is low.

The definition of a cluster is subjective. But here is one definition:

Definition 14 *A* cluster *is a group of points whose intra-point distance is below the average distance in the pattern as a whole.*

This local aggregation is not simply the result of random point density fluctuations. There exists a "fundamental ambiguity between heterogeneity and clustering, the first corresponding to spatial variation of the intensity function $\lambda(x)$, the second to stochastic dependence amongst the points of the process...[and these are]...difficult to disentangle" Diggle (2007).

Clustering can occur by a variety of mechanisms which cannot be distinguished based on statistical approaches below. Subject specific knowledge is also required. For example, the objects represented by the points were originally scattered completely at random in the region of interest, but remain only in certain subregions that are distributed irregularly in space. For example, plants with wind dispersed seeds which germinate only in areas of with suitable soil conditions. As another example, the pattern results from a mechanism that involves *parent points* and *daughter points*, where the daughter points are scattered around the parent points. An example was given above: young trees cluster around parent trees, the daughters arise from the seeds of the parent plant. Often, the parent points are unknown or even fictitious.

We begin we the most general type of clustering process

Definition 15 *Let X be a finite point process on $\mathbb{R}^d$ and conditional on X associate with each $x \in X$ a finite point process Z_x of points 'centered' at x and assume that these processes are independent of one another. Then $\mathcal{Z} = \cup_{x \in X} Z_x$ is an* independent cluster process. *The data consist of a realization of $Y = \mathcal{Z} \cap W$ in a compact window $W \subseteq S \subseteq \mathbb{R}^d$, with $|W| < \infty$.*

Note that there is nothing in the definition that precludes the possibility that $x \in X$ lies outside of W. In any analysis, we would want to allow for this possibility to reduce bias from

edge effects. Also, the definition is flexible enough in that characteristics of the daughter clusters Z_x, $x \in X$, such as the intensity or spread of the cluster, are allowed to be randomly and spatially varying.

However, in order to make any type of statistical inference, we will require that the distribution of Y is absolutely continuous with respect to the distribution of a unit rate Poisson process on S compact and of positive volume. Table 2 summarizes standard nomenclature for various cluster processes—all of which are special cases of the independent cluster process.

Name of Process	Clusters	Parents
Independent cluster process	general	general
Poisson cluster process	general	Poisson
Cox cluster process	Poisson	general
Neyman-Scott process	Poisson	Poisson
Matérn cluster process	Poisson (uniform in ball)	Poisson (homog.)
Modified Thomas Process	Poisson (Gaussian)	Poisson (homog.)

Table 2: Nomenclature for independent cluster processes.

2.1.1 Neyman-Scott Processes

Definition 16 *A* Neyman-Scott process *is a particular case of a Poisson cluster process (the daughter clusters are assumed to be Poisson).*

We will now consider the case when a Neyman-Scott process becomes a Cox cluster process. Let C be a homogeneous Poisson process of cluster centers (the parent process) with intensity $\lambda > 0$. Let $C = \{c_1, c_2, \dots\}$, and conditional on C associate with each c_i a Poisson process X_i centered at c_i and that these processes are independent of one another. The intensity measure of X_i is given by

$$\mu_i(B) = \int_B \alpha f(\xi - c_i; \theta) d\xi$$

where $\alpha > 0$ is a parameter and f is the density function of a continuous random variable in $\mathbb{R}^d$ parameterized by θ. We further assume that α and θ are known. By the superposition property of independent Poisson processes, $X = \cup_i X_i$ is a Neyman-Scott process with intensity

$$Z(\xi) = \sum_{c_i \in C} \alpha f(\xi - c_i; \theta).$$

Further, by the definition of a Cox process, X is also a Cox process driven by

$$Z(\xi) = \sum_{c_i \in C} \alpha f(\xi - c_i; \theta).$$

This process is stationary and is isotropic if $f(\xi)$ only depends on $||\xi||$. The intensity function of the Cox process then

$$\rho(\xi) = \mathbb{E}[Z(\xi)] = \alpha\lambda. \tag{4}$$

Example 3 (Matérn cluster process) *A Matérn cluster process is a special case of a Neyman-Scott process (and also of a Cox cluster process) where the density function is*

$$f(\xi - c; \theta) = \frac{dr^{d-1}}{R^d}, \quad \textit{for} \quad r = ||\xi - c|| \leq R.$$

That is, conditional on C and $c_i \in C$, the points $x_i \in X_i$ are uniformly distributed in the ball $b(c_i, R)$ centered at c_i of radius R. Here $\theta = R$.

Example 4 (Modified Thomas process) *A modified Thomas process is also a special case of a Neyman-Scott process (and also of a Cox cluster process) where the density function is*

$$f(\xi - c; \theta) = \left(2\pi\sigma^2\right)^{-d/2} \exp\left[-\frac{1}{2\sigma^2}||\xi - c||^2\right].$$

That is $\xi \sim N_d(c, \sigma^2 I)$. Here $\theta = \sigma^2$.

Both of these examples can be further generalized by allowing R and σ^2 to depend on c.

The Neyman-Scott can be generalized in one other way.

Definition 17 *Let X be a Cox process on $\mathbb{R}^d$ driven by*

$$Z(\xi) = \sum_{(c,\gamma)\in Y} \gamma f(\xi - c)$$

where Y is a Poisson point process on $\mathbb{R}^d \times (0, \infty)$ with a locally integrable intensity function λ. Then X is called a shot noise Cox process (SNCP).

The SNCP is different from a Neyman-Scott process in two regards. 1) γ is random and 2) the parent process Y is not homogeneous.

Now, the intensity function of X is given by

$$\rho(\xi) = \mathbb{E}[Z(\xi)] = \int \gamma f(\xi - c)\lambda(c, \gamma) dc d\gamma. \tag{5}$$

provided the integral is finite. The proofs of (4) and (5) rely on the Slivnyak-Mecke Theorem.

Theorem 4 (Slivnyak-Mecke) *Let $X \sim Poisson(S, \rho)$. For any function $h : S \times N_{\ell f} \to [0, \infty)$ we have*

$$\mathbb{E}\left[\sum_{\xi \in X} h(\xi, X \setminus \xi)\right] = \int_S \mathbb{E}\left[h(\xi, X)\right] \rho(\xi) d\xi.$$

Consult Møller and Waagepetersen (2004) for a proof.

2.1.2 Edge effects

Edge effects occur in these aggregative processes when the sampling window W is a proper subset of the space of interest S. In this case it can happen that the parent process has points outside of the sampling window and that this parent point has daughter points that lie within W. This causes a bias in estimation of the intensity function. One way to reduce the bias, when modeling, is to increase the search for parents outside the sampling window W. For example, if we are using a Matérn cluster process with a radius of R, then we can increase our search outside of W by this radius.

2.2 Repulsive Processes

Repulsive processes occur naturally in biological applications. For example, mature trees in a forest may represent a repulsive process as they compete for nutrients and sunlight. Locations of male crocodiles along a river basin also represent a repulsive process as males are territorial. Repulsive process are typically modeled through Markov point processes. The setup for Markov point processes will differ slightly than the setup for the point processes we have studied thus far.

Consider a point process X on $S \subset \mathbb{R}^d$ with density f defined with respect to the unit rate Poisson process. We will require $|S| < \infty$ so that f is well-defined. The density is then concentrated on the set of finite point configurations contained in S (I am not sure why):

$$N_f = \{x \subset S : n(x) < \infty\}.$$

(Contrast this with $N_{lf} = \{x_B \subset S : n(x_B) < \infty, \text{ for all bounded } B \subseteq S\}$, where $x_B = x \cap B$).

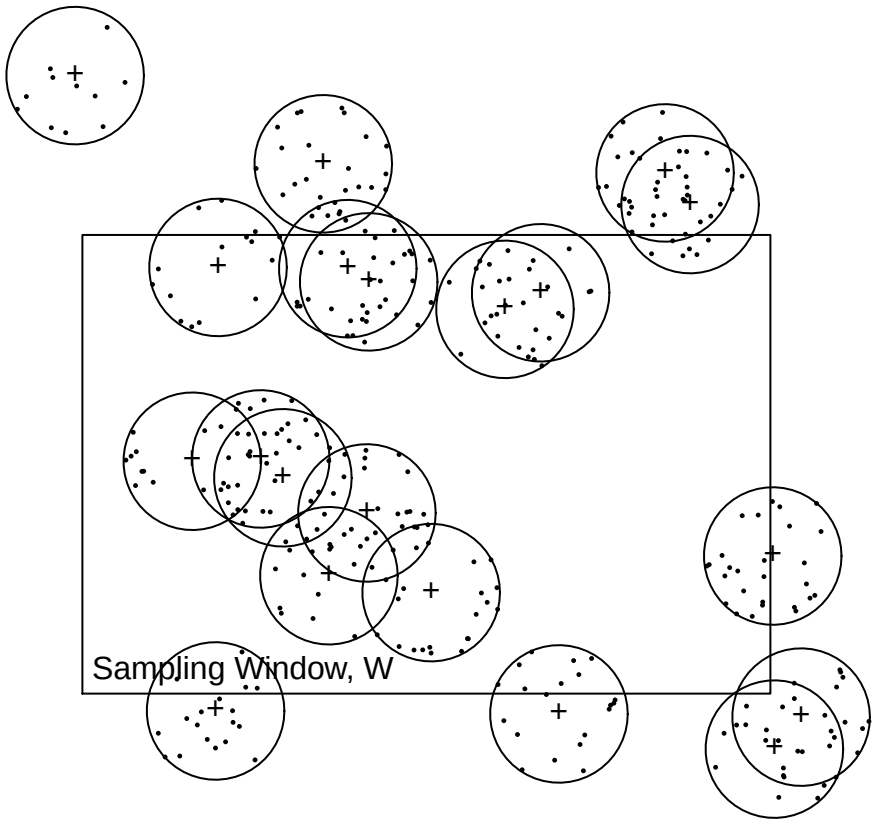

Figure 1: Edge effects for a Matérn process. Parent process is homogeneous Poisson with a constant rate of 20 in a 20×20 square field. The cluster processes are homogeneous Poisson centered at the parents (+) with a radius of 3 and a constant rate of 10.

Now, by Proposition 4, for $F \subseteq N_f$

$$P(X \in F) = \sum_{n=0}^{\infty} \frac{\exp(-|S|)}{n!} \int_S \cdots \int_S I(\{x_1, \ldots, x_n\} \in F) f(\{x_1, \ldots, x_n\}) dx_1 \ldots dx_n.$$

An example we have already seen is when $X \sim \text{Poisson}(S, \rho)$, with $\mu(S) = \int_s \rho(\xi) d\xi < \infty$. Then

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

(see Proposition 6).

In most cases, especially for Markov point processes, the density f is only known up to a normalizing constant: $f = ch$, where $h : N_f \to [0, \infty)$ and the normalizing constant is

$$c^{-1} = \sum_{n=0}^{\infty} \frac{\exp(-|S|)}{n!} \int_S \cdots \int_S h(\{x_1, \ldots, x_n\}) dx_1 \ldots dx_n.$$

Markov models (and Gibbs distributions) originated in statistical physics where c is called the *partition function*.

2.2.1 Papangelou Conditional Intensity

As far as likelihood inference goes, Markov point process densities are intractable as the normalizing constant is unknown and/or extremely complicated to approximate. However, we can resort to MCMC techniques to compute approximate likelihoods and posterior estimates of parameters using the following conditional intensity.

Definition 18 *The* Papangelou conditional intensity *for a point process X with density f is defined by*

$$\lambda^*(x, \xi) = \frac{f(x \cup \{\xi\})}{f(x)}, \quad x \in N_f, \, \xi \in S \setminus x,$$

where we take $a/0 = 0$ for $a \geq 0$.

The normalizing constant for f cancels, thus $\lambda^*(x, \xi)$ does not depend on it. For a Poisson process, $\lambda^*(x, \xi) = \rho(\xi)$ does not depend on x either (since the points are spatially independent).

Definition 19 *We say that X (or f) is* attractive *if*

$$\lambda^*(x,\xi) \leq \lambda^*(y,\xi) \text{ whenever } x \subset y$$

and repulsive *if*

$$\lambda^*(x,\xi) \geq \lambda^*(y,\xi) \text{ whenever } x \subset y.$$

2.2.2 Pairwise interaction point processes

In the above discussion, f is a quite general function where all the $x \in X$ can interact. However, statistical experience has shown that often a pairwise interaction is sufficient to model many types of point pattern data. Thus we restrict attention to densities of the form

$$f(x) = c \prod_{\xi \in x} g(\xi) \prod_{\{\xi,\eta\} \subseteq x} h(\{\xi,\eta\}) \tag{6}$$

where h is an *interaction function*. That is h is a non-negative function for which the right hand side of (6) is integrable with respect to the unit rate Poisson process.

The range of interaction is defined by

$$R = \inf\{r > 0 : \forall\, \{\xi,\eta\} \subset S, h(\{\xi,\eta\}) = 1 \text{ if } ||\xi - \eta|| > r\}.$$

Pairwise interaction processes are analytically intractable because of the unknown normalizing constant c. There is one exception: the Poisson point process where $h(\{\xi,\eta\}) \equiv 1$, so that $R = 0$.

The Papangelou conditional intensity, for $f(x) > 0$ and $\xi \notin x$, is given by

$$\lambda^*(x,\xi) = g(\xi) \prod_{\eta \in x} h(\{\xi,\eta\}).$$

By the definition of a repulsive process, f is repulsive if and only if $h(\{\xi,\eta\}) \leq 1$. Most pairwise interaction processes are repulsive. For the attractive case, when $h(\{\xi,\eta\}) \geq 1$ for all distinct $\xi, \eta \in S$, f is not always well defined (one must check that it is).

Now we consider a special case where $g(\xi)$ is constant and $h(\{\xi,\eta\}) = h(||\xi - \eta||)$. In this case, the pairwise interaction process is called homogeneous.

Example 5 (The Strauss Process) *A simple, but non-trivial, pairwise interaction process is the* Strauss process *with*

$$h(||\xi - \eta||) = \gamma^{I(||\xi-\eta|| \leq R)}$$

where we take $0^0 = 1$. *The interaction parameter* $\gamma \in [0, 1]$ *and* $R > 0$ *is the range of interaction. The density of the Strauss process is*

$$f(x) \propto \beta^{n(x)} \gamma^{s_R(x)}$$

where $\beta > 0$ *and*

$$s_R(x) = \sum_{\{\xi,\eta\} \subseteq x} I(||\xi - \eta|| \leq R)$$

is the number of pairs of points in x *that are a distance of* R *or less apart.*

There are two limiting cases of the Strauss process. When $\gamma = 1$ *we have*

$$f(x) \propto \beta^{n(x)}$$

which is the density of a Poisson(S, β) *(with respect to the unit rate* Poisson). When $\gamma < 1$ the process is repulsive with range of interaction R. When $\gamma = 0$ points are prohibited from being closer than the distance R apart and is called a *hard core process.*

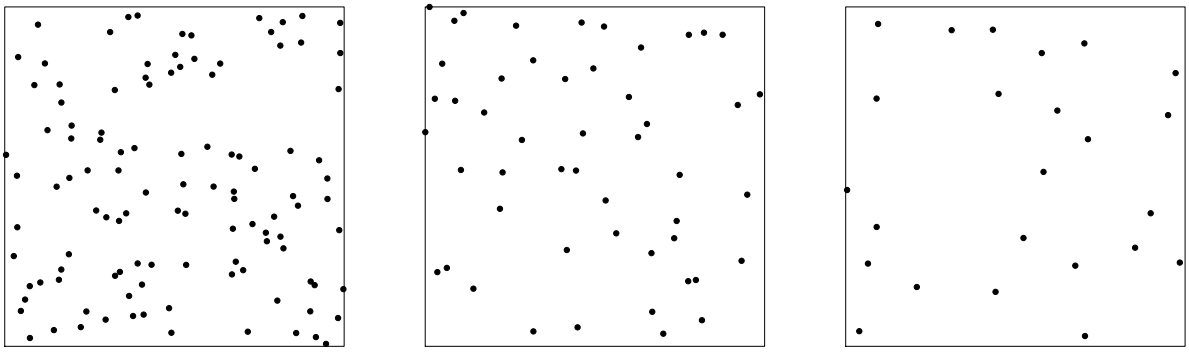

Table 3: Realizations of Strauss processes on $S = [0, 1]^2$. Here $\beta = 100$, $R = 0.1$ and $\gamma = 1.0$, 0.5, 0.0 from left to right. Processes generated from the function rStrauss in the R package spatstat by Adrian Baddeley.

2.3 Marked Processes

Let Y be a point process on $T \subseteq \mathbb{R}^d$. Given some space $\mathcal{M}$, if a random 'mark' $m_\xi \in \mathcal{M}$ is attached to each point $\xi \in Y$, then

$$X = \{(\xi, m_\xi : \xi \in Y\}$$

is called a *marked point process* with points in T and *mark space* $\mathcal{M}$. Marks can be thought of as extra data or information associated with each point in Y.

Example 6 *We have already encountered a marked point process, although we did not identify it as such. In Definition 17, we defined the SNCP. In that definition Y was considered a Poisson point process on $\mathbb{R}^d \times (0, \infty)$. However, we can also consider it a marked point process with points in $\mathbb{R}^d$ and mark space $\mathcal{M} = (0, \infty)$.*

Definition 20 (Marked Poisson processes) *Suppose that Y is* Poisson(T, ρ)*, where ρ is a locally integrable intensity function. Also suppose that conditional on Y, the marks $\{m_\xi : \xi \in Y\}$ are mutually independent. Then $X = \{(\xi, m_\xi) : \xi \in Y\}$ is a* marked Poisson process*. If the marks are i.i.d. with common distribution Q, then Q is called the* mark distribution*.*

Now we will discuss three marking models: 1) independent marks, 2) random field model and 3) intensity-weighted marks.

The simplest model for marks is the independent marks model. In this model, the marks are i.i.d. random variables and are independent of the locations of the point pattern. If $X = \{(\xi, m_\xi) : \xi \in Y\}$ then the density of X can be factorized as

$$\pi(X) = \pi(Y)\pi(m).$$

The random field model, a.k.a. 'geostatistical marking' is the next level of generalization. In the random field model, the marks are assumed to be correlated, but independent of the point process Y. The marks are derived from the random field:

$$m_\xi = Z(\xi)$$

for some (stationary) random field Z such as the Gaussian random field.

The next level of generalization is to assume that there is correlation between the point density and the marks. An example of such a model is the intensity-weighted log-Gaussian Cox model. Y is a stationary LGCP driven by Z. Each point $\xi \in Y$ is assigned a mark m_ξ according to

$$m_\xi = a + bZ(\xi) + \varepsilon(\xi)$$

where $\varepsilon(\xi)$ are i.i.d. $N(0, \sigma_\varepsilon^2)$ and a and b are model parameters. When $b > 0$ the marks are large in areas of high point density and small in areas of low point density. When $b < 0$ the marks are small in regions of high point density and large in regions of low point density. When $b = 0$ the marks are independent of the point density.

3 Simulation and Bayesian Methods

3.1 Fast Simulation for LGCPs

The following is adapted from Møller, Syversveen, and Waagepetersen (1998). Recall that a Gaussian random field is completely specified by its finite dimensional distributions. Therefore, to simulate a stationary Gaussian random field, we discretize it to a uniform grid and define the appropriate finite dimensional Gaussian distribution on this grid. Consider for a moment a Gaussian random field Y on the unit square $S = [0,1]^2$. We discretize the unit square into an $M \times M$ grid I with cells $D_{ij} = [(i-1)/M, i/M] \times [(j-1)/M, j/M]$ where the center of cell D_{ij} is given by $(c_i, c_j) = ((2i+1)/(2M), (2j+1)/(2M))$ and we then approximate the Gaussian random field $\{Y(s)\}_{s\in[0,1]^2}$ by its value $Y((c_i, c_j))$ at the center of each cell D_{ij}. This is easily extended to rectangular lattices.

Now take the lexicographic ordering of the indices i, j: $k = 2Mi + M(2Mj - 1)$ and thus $Y_k = Y((c_i, c_j))$. Let $Y = (Y_1, \ldots, Y_{M^2})^{\mathrm{T}}$ (hopefully without any confusion). Now the problem is to simulate $Y \sim N_{M^2}(0, \Sigma)$ (w.l.o.g. we can assume the mean of the stationary GRF is 0). Here Σ has elements of the form $C(||(c_i, c_j) - (c_i', c_j')||) = \sigma^2 r(||(c_i, c_j) - (c_i', c_j')||/\alpha)$ (or some other valid isotropic covariance function).

Easy, correct? Well in theory this is simple if Σ is a symmetric positive definite matrix. We take the Cholesky decomposition of Σ, say GG^{T}, where G is lower triangular, simulate M^2 normal mean 0, variance 1 random variates, $x = (x_1, \ldots, x_{M^2})^{\mathrm{T}}$, and set $Y = Gx$. We may also need to evaluate the inverse of Σ as well as its determinant.

However, for many grids, M can be moderately large, say 128. Then Σ is a $128^2 \times 128^2$ matrix and it gets even worse if we need to simulate a GRF on a bounded $S \subset \mathbb{R}^d$ where $d \geq 3$. The Cholesky decomposition is an expensive operation $O(n^3)$ and if M is too large we may not even have enough RAM in our computers to compute it.

Now consider Σ to be symmetric non-negative definite, and for simplicity consider simulation of a Y on $[0,1]$ (Note that we are in one-dimensional Euclidean space now). The following method (Rue and Held, 2005) will work for any hyperrectangle in any Euclidean space, with a few minor modifications.

Let's first assume that $M = 2^n$ for some positive constant n. First, we note that Σ is a Toeplitz matrix. Next, we embed Σ into a circulant matrix of size $2M \times 2M$ by the following procedure. Extend $[0,1]$ to $[0,2]$ and map $[0,2]$ onto a circle with circumference 2 and extend the grid from I with M cells to I_{ext} with $2M$ cells. Let d_{ij} denote the minimum distance on this circle between c_i and c_j. Now $\Sigma_{\text{ext}} = (C(d_{ij}))_{(i,j)\in I_{\text{ext}}}$ and is a circulant matrix of size

$2M \times 2M$.

You may be wondering at this point, why extend the matrix and make it 4 times as large when we the original matrix was too unwieldy. The answer to this is that there is a relationship between the eigenvalues and eigenvectors of a circulant matrix and the discrete Fourier transform. Thus we can use the fast Fourier tranform (FFT) to compute the eigenvalues and eigenvectors of a circulant matrix and, once these are in hand, can perform all of necessary matrix operations. In particular we will be able to draw a value from $N_M(0, \Sigma)$ by drawing from $N_{2M}(0, \Sigma_{\text{ext}})$ and taking the first M values.

Definition 21 *A $M \times M$ matrix C is circulant if and only if it has the form*

$$C = \begin{pmatrix} c_0 & c_1 & c_2 & \cdots & c_{M-1} \\ c_{M-1} & c_0 & c_1 & \cdots & c_{M-2} \\ c_{M-2} & c_{M-1} & c_0 & \cdots & c_{M-3} \\ \vdots & \vdots & \vdots & & \cdots \\ c_1 & c_2 & c_3 & \cdots & c_0 \end{pmatrix}.$$

We call $\mathbf{c} = (c_0, c_1, \ldots, c_M)^{\mathrm{T}}$ the base of C. (actually any row or column of C will suffice as the base).

Let λ be any eigenvalue of C with associated eigenvector $\mathbf{e}$. Then $C\mathbf{e} = \lambda\mathbf{e}$. This can be written row by row as M difference equations,

$$\begin{aligned} & \sum_{i=0}^{j-1} c_{M-j+i} e_i + \sum_{i=j}^{M-1} c_{i-j} e_i = \lambda e_j, \quad \text{for} \quad j = 0, \ldots, M-1 \\ \Longrightarrow \quad & \sum_{i=0}^{M-1-j} c_i e_{i+j} + \sum_{i=M-j}^{M-1} c_i e_{i-(M-j)} = \lambda e_j. \end{aligned} \tag{7}$$

This system of M linear difference equation have constant coefficients, and so like with systems of linear differential equations with constant coefficients we guess that the solution has the form $e_j \propto \rho^j$ for some complex scalar ρ. Now (7) can be written as

$$\sum_{i=0}^{M-1-j} c_i \rho^i + \rho^{-M} \sum_{i=M-j}^{M-1} c_i \rho^i = \lambda.$$

Now choose ρ such that $\rho^{-M} = 1$, then

$$\lambda = \sum_{i=0}^{M-1} c_i \rho_i$$

and

$$\mathbf{e} = \frac{1}{\sqrt{M}}(1, \rho, \rho^2, \ldots, \rho^{M-1})^{\mathrm{T}},$$

where we have included the factor $\sqrt{M}$ so that $\mathbf{e}$ is orthonormal: $\mathbf{e}^{\mathrm{T}}\mathbf{e} = 1$. Now, since $\rho^M = 1$ and ρ is complex, we have that the M roots of 1 are $\{\exp(2\pi\imath j/M),\, j = 0, \ldots, M\}$ where $\imath = \sqrt{-1}$. Thus, the M eigenvalues are

$$\lambda_j = \sum_{i=0}^{M-1} c_i \exp(-2\pi\imath\, ij/M), \quad j = 0, \ldots, M-1$$

with corresponding eigenvectors

$$\mathbf{e}_j = \frac{1}{\sqrt{M}}\left(1, \exp(-2\pi\imath\, j/M), \exp(-2\pi\imath\, j2/M), \ldots, \exp(-2\pi\imath\, j(n-1)/M)\right)^{\mathrm{T}},$$

for $j = 0, \ldots, M-1$.

Let $\omega = \exp(-2\pi\imath/M)$ and let

$$F = (\mathbf{e}_0|\mathbf{e}_1|\ldots|\mathbf{e}_{M-1}) = \frac{1}{\sqrt{M}}\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega^1 & \omega^2 & \cdots & \omega^{M-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(M-1)} \\ \vdots & \vdots & \vdots & & \cdots \\ 1 & \omega^{M-1} & \omega^{2(M-1)} & \ldots & \omega^{(M-1)(M-1)} \end{pmatrix}$$

be the discrete Fourier transform matrix and define

$$\Lambda = (\lambda_0, \lambda_1, \ldots, \lambda_{M-1}).$$

F is unitary matrix: $F^{-1} = F^H$ where F^H is the complex conjugate transpose of F and

$$\Lambda = \sqrt{M}\mathrm{diag}(F\mathbf{c}).$$

One can verify that $C = F\Lambda F^H$ by direct calculation. Thus, any circulant matrix can be diagonalized by some Λ.

Since F is the discrete Fourier transform (DFT) matrix, Fv can be calculated by taking the DFT of some vector v and $F^H v$ is calculated by taking the inverse DFT (IDFT) of v. Now if M can be factorized into small primes, the fast Fourier transform (FFT) can be used and Fv can be computed in $O(n \ln n)$ operations.

Let X_{ext} be a vector of $2M$ i.i.d. $N(0,1)$ random variables. Then, Y_{ext} is equal in distribution to $\Sigma_{\mathrm{ext}}^{1/2} X_{\mathrm{ext}}$. Now, diagonalize $\Sigma_{\mathrm{ext}} = F\Lambda F^H$, which in turn implies $\Sigma_{\mathrm{ext}}^{1/2} = F\Lambda^{1/2}F^H$ and that

$$Y_{\mathrm{ext}} \stackrel{d}{=} \Sigma_{\mathrm{ext}}^{1/2} X_{\mathrm{ext}} = F\Lambda^{1/2}F^H X_{\mathrm{ext}}.$$

So we compute $F^H X_{\text{ext}} = \text{IDFT}(X_{\text{ext}})$. Compute $\Lambda^{1/2} = \sqrt{\sqrt{M}\,\text{DFT}(c)}$. Let $b = \Lambda^{1/2} F^H X_{\text{ext}}$, then $b = \sqrt{\sqrt{M}\,\text{DFT}(c)} \odot \text{IDFT}(X_{\text{ext}})$ where $\odot$ denotes elementwise mulitplication. Finally take

$$Y_{\text{ext}} \stackrel{d}{=} \Sigma_{\text{ext}}^{1/2} X_{\text{ext}} = \text{DFT}(b) = \text{DFT}\left(\sqrt{\sqrt{M}\,\text{DFT}(c)} \odot \text{IDFT}(X_{\text{ext}})\right).$$

Y is then obtained by taking the first M elements of Y_{ext}.

This is much faster than trying to compute the Cholesky decomposition of Σ when the number of elements is large. When we wish to simulate a GRF in $\mathbb{R}^2$, then we extend covariance matrix by wrapping it around a torus to define the distances. We end up with a block circulant matrix and we use the 2D-DFT. For simulation of a GRF in $\mathbb{R}^3$, the extended covariance matrix is a nested block circulant matrix and we use the 3D-DFT.

3.2 Bayesian modeling of a LGCP

Now back to LGCP. Suppose we have a realization x of a spatial point pattern X in $S \subset \mathbb{R}^2$ for which we wish to estimate the intensity function. We will assume that a LGCP model is appropriate for the data. The intensity we wish to estimate is $Z(\xi) = \exp(Y^\star(\xi) + \mu)$ where $Y^\star$ is a stationary mean zero GRF. Let $Y(\xi) = Y^\star(\xi) + \mu$. Assume that μ is known as well as the σ^2 and α in the isotropic covariance function $c(\cdot) = \sigma^2 r(\cdot/\alpha)$. The first order of business is to write down the density of the process with respect to a unit rate Poisson:

$$\pi(x \mid Y, \mu) \propto \exp\left(-\int_S \exp(Y(\xi))d\xi\right) \prod_{x_i \in x} \exp(Y(x_i)).$$

After discretizing onto a fine grid and then extending we can write the log density as

$$\ln[\pi(x \mid Y = y)] \propto \sum_{(i,j) \in I_{\text{ext}}} (-A_{ij} \exp(y_{ij}) + n_{ij} y_{ij})$$

where A_{ij} is the area of cell D_{ij} and n_{ij} is the number of points of x contained in cell D_{ij} and y_{ij} is the value of the Gaussian process at the center of cell D_{ij}. We set $A_{ij} = n_{ij} = 0$ if $(i,j) \notin I$ (thus $(i,j) \in I_{\text{ext}} \setminus I$). As shown in Møller, Syversveen, and Waagepetersen (1998), it is computationally more efficient to work with $\Sigma^{-1/2} W_{\text{ext}}$ where $W_{\text{ext}} \sim N_d(0, I)$ where $d = (2M)^2$. Now $[W \mid x]$ has log density

$$\ln[\pi(w \mid x)] \propto -1/2||w||^2 + \sum_{(i,j) \in I_{\text{ext}}} \left[-A_{ij} \exp((\Sigma^{1/2} w)_{ij}) + n_{ij}((\Sigma^{1/2} w)_{ij})\right].$$

Since this posterior does not have a form which can be easily sampled from, we must resort to MCMC. We need to update w which is an extremely long vector. Using the Metropolis-Hastings algorithm mixes very slowly, so instead Møller, Syversveen, and Waagepetersen

(1998) proposed the use of the Metropolis adjusted Langevin algorithm (MALA), suggested initially by Besag (1994) and further studied Roberts and Tweedie (1997).

MALA requires the gradient of the posterior (which is strictly log-concave). Let the gradient of the posterior of w be denote $\nabla(w)$,

$$\nabla(w) \equiv \frac{\partial \ln[\pi(w \mid x)]}{\partial w} = -w + \Sigma^{1/2}\left(n_{ij} - \exp(\Sigma^{1/2}w)A_{ij}\right)_{(i,j)\in I_{\text{ext}}}.$$

MALA is specified in two steps. First, if $w^{(t)}$ is the current draw, we propose a new draw $\omega^{(t+1)}$ from the an independent multivariate normal distribution with mean $m(w^{(t)}) = w^{(t)} + (h/2)\nabla(w^{(t)})$ and common variance h. Second, with probability

$$1 \wedge \frac{\pi(\omega^{(t+1)} \mid x)\exp(-||w^{(t)} - m(\omega^{(t+1)})||^2/(2h))}{\pi(w^{(t)} \mid x)\exp(-||\omega^{(t+1)} - m(w^{(t)})||^2/(2h))}$$

$w^{(t+1)} = \omega^{(t+1)}$, otherwise $w^{(t+1)} = w^{(t)}$.

We can also assign priors to μ, σ^2 and α. The full conditionals of these parameters do not have closed form and so we must use the Metropolis-Hastings algorithm to update these parameters.

3.3 Bayesian Analysis of Cluster Processes & the Spatial Birth and Death Process Algorithm

We will use the notation from last lecture and assume that X is a finite point process on $\mathbb{R}^d$ and conditional on X we associate with each $\xi \in X$ a finite point process Y_ξ of points centered on ξ and that these processes are independent of one another. Then $Y = \cup_{\xi\in X} Y_\xi$ is an independent cluster process. We will assume that the parent process X is not observed. To simplify the exposition, we will assume that the processes X and Y are both only found on a bounded subset S of $\mathbb{R}^d$ and that we observe the process Y on all of S. This avoids edges effects.

Assume that each Y_ξ is a Poisson process with known intensity $h(\cdot;\xi)$. The observed data will be denoted $y = \{y_1, y_2, \dots\}$. We will also assume that some of the points in y don't cluster with other points and that these points follow a homogeneous Poisson process with intensity ϵ. Then the intensity function for Y, conditional on $X = x = \{x_1, \cdots, x_n\}$, is given by

$$\lambda(\cdot \mid x) = \epsilon + \sum_{i=1}^{n} h(\cdot; x_i),$$

and the conditional density of Y given x with respect to a unit rate Poisson is

$$\pi(y \mid x) = \exp\left(\int_S [1 - \lambda(\eta \mid x)] d\eta\right) \prod_{\eta \in y} \lambda(\eta; x).$$

The goal of the analysis is to estimate $\lambda(\cdot \mid x)$. However, x is not observed and are latent data and so must be estimated. To this end, we need to estimate the posterior of X given y:

$$\pi(x \mid y) \propto \pi(y \mid x)\pi(x) = \pi(x) \exp\left(\int_S [1 - \lambda(\eta \mid x)] d\eta\right) \prod_{\eta \in y} \lambda(\eta; x),$$

where $\pi(x)$ is the prior density for X. For concreteness, let's assume that $X \sim \text{Poisson}(S, \rho)$ (not necessarily homogeneous).

The posterior Papangelou conditional intensity is

$$\lambda^*_{X|Y}(x \mid y, \xi) = \lambda^*_X(x, \xi) \exp\left[- \int_S h(s \mid \xi) ds \right] \prod_{\eta \in y} \left[1 + \frac{h(\eta; \xi)}{\lambda(\eta \mid x)}\right]$$

Standard MCMC algorithms (such as Metropolis-Hastings) will not work for this problem as not only are the locations of x random, but the number of points $n(X)$ is random as well. RJMCMC (Green, 1995) is one option (there are at least four algorithms that are possible). However, we will only discuss one, developed by Preston (1977).

The spatial birth and death process (Preston, 1977; Møller and Waagepetersen, 2004) is a continuous time Markov process whose transitions are either births or deaths which can be used to simulate spatial point processes.

3.3.1 The Spatial Birth and Death Algorithm

We wish to construct a spatial birth-and-death process to simulate a latent parent point process X from its posterior $\pi(x \mid y)$, given it offspring, or daughters. If the birth and death rates satisfy the detailed balance equation (Preston, 1977)

$$\pi(x \mid y) b(x, \xi) = \pi(x \cup \{\xi\} \mid y) d(x \cup \{\xi\}, \{\xi\}), \tag{8}$$

then the chain is time reversible and that the spatial birth-and-death process has a unique equilibrium distribution $\pi(x \mid y)$ to which it converges in distribution from any initial state (with a few extra conditions imposed on the total birth and death rates discussed below). In (8) $b(x, \xi)$ is the birth rate for adding a new point ξ to the current configuration, x, of the point process X, and $d(x, \xi)$ denotes the death rate for removing a point ξ from x. A

common strategy is to assume a constant death rate and use a birth rate proportional to the posterior Papangelou conditional intensity. However, the total birth rate (see below) may be difficult to compute and their may be a large number of terms in the product. An alternative birth rate, suggested by van Lieshout and Baddeley (2002), is given by

$$b(x,\xi) = \lambda_X^*(x,\xi)\left[1 + \sum_{\eta\in y}\frac{h(\eta;\xi)}{\epsilon}\right].$$

To satisfy the detailed balance equation, the death rate for removing ξ from $x\cup\{\xi\}$ is

$$d(x\cup\{\xi\},\xi) = \frac{\exp\left[\int_S h(s;\xi)ds\right]}{\prod_{\eta\in y}\left[1+\frac{h(\eta;\xi)}{\lambda(\eta|x)}\right]}\left[1+\sum_{\eta\in y}\frac{h(\eta;\xi)}{e}\right].$$

The total birth rate is given by

$$B(x) = \int_S b(x,\xi)d\xi = \int_S \lambda_X^*(x,\xi)\left[1+\sum_{\eta\in y}\frac{h(\eta;\xi)}{\epsilon}\right]d\xi$$

and the total death rate is

$$D(x) = \sum_{\xi\in x} d(x,\xi).$$

The conditions on the total birth rate and total death rate are that the birth rate must be bounded above by a constant B and the death rate must be bounded below by a constant $D \geq 0$. The total birth rate $B(x)$ may be difficult to compute here as well as it may be difficult to integrate $\lambda_X^*(x,\xi)$ w.r.t. ξ over S. Thus, we resort rejection sampling. If $\lambda_X^*(x,\xi) \leq \lambda$ uniformly in x and ξ and $h(\eta;\xi) \leq H$ uniformly in η and ξ, then the total birth rate is bounded:

$$B(X) \leq \lambda\left[|S| + \frac{1}{\epsilon}\sum_{\eta\in y}\int_S h(\eta;\xi)d\xi\right] \leq \lambda|S|\left[1+\frac{n(y)H}{\epsilon}\right] \equiv B.$$

The total death rate is bounded below by $n(x)(1+H/\epsilon)^{-n(y)}$ (see van Lieshout and Baddeley (2002)).

Suppose that we can integrate $h(\eta;\xi)$ over S easily. This is the case for the Matérn process and the modified Thomas process. Let

$$B = \lambda\left[|S| + \frac{1}{\epsilon}\sum_{\eta\in y}\int_S h(\eta;\xi)d\xi\right].$$

If the current state is x, after an exponentially distributed waiting time with rate $B+D(X)$, a death of a point in x occurs with probability $D(X)/[B + D(x)]$. If a death is to occur, the point, ξ, is deleted from x with probability $d(x,\xi)/D(X)$. A birth is proposed with complementary probability $B/[B + D(x)]$. Sample a candidate ξ from the mixture density

$$\frac{\lambda}{B}\left[1+\sum_{\eta\in y}\frac{h(\eta \mid \xi)}{\epsilon}\right] \tag{9}$$

and accept the candidate ξ as a new point in the parent process with probability

$$\frac{\lambda_X^*(x,\xi)}{\lambda}.$$

To draw a candidate ξ from (9) note that we can rewrite it as

$$\frac{\lambda|S|}{B}\frac{1}{|S|}+\sum_{\eta\in y}\frac{\lambda\int_S h(\xi \mid \eta)d\xi}{\epsilon B}\frac{h(\xi \mid \eta)}{\int_S h(\xi \mid \eta)d\xi}.$$

Therefore, with probability $\lambda|S|/B$ we draw a point uniformly over S and with probability $\lambda\int_S h(\eta \mid \xi)d\xi/(\epsilon B)$ we draw a point ξ from $h(\xi \mid \eta)/\int_S h(\xi \mid \eta)d\xi$

References

Besag, J. E. (1994). Discussion of the paper by grenander and miller. *Journal of the Royal Statistical Society, Series B* **56**, 591–592.

Cox, D. R. (1955). Some statistical models related with series of events. *Journal of the Royal Statistical Society, Series B* **17**.

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods*. 2 edition. Springer.

Daley, D. J. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes, Volume II: General Theory and Structure*. 2 edition. Springer.

Diggle, P. J. (2007). Spatio-temporal point processes: methods and applications. In Finkenstad, B., Held, L., and Isham, V., editors, *Statistical Methods for Spatio-temporal Systems*, pages 1–45. Chapman & Hall/CRC.

Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika* **82**, 711–732.

Illian, J., et al. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley & Sons.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998). Log gaussian cox processes. *Scand. J. Statist.* **25**, 451–482.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC.

Møller, J. and Waagepetersen, R. P. (2007). Modern statistics for spatial point processes. *Scandinavian Journal of Statistics* **34**, 643–684.

Preston, C. J. (1977). Spatial birth-and-death processes. *Bulletin of the International Statistical Institute* **46**, 371–391.

Roberts, G. O. and Tweedie, R. L. (1997). Exponential convergence of langevin diffusions and their approximations. *Bernoulli* **2**, 341–363.

Rue, H. and Held, H. (2005). *Gaussian Markov Random Fields*. Chapman & Hall/CRC.

van Lieshout, M. N. M. and Baddeley, A. J. (2002). Extrapolating and interpolating spatial patterns. In Lawson, A. B. and Denison, D. G. T., editors, *Spatial Cluster Modelling*, chapter 4, pages 61–86. Chapman & Hall/CRC.