

Spatial Models: A *Quick* Overview
Astrostatistics Summer School, 2007

Murali Haran

Department of Statistics
Penn State University

Introduction

- Beginning statistics: Data are assumed to be independent and identically distributed (i.i.d.). Inference is based on theory that relies on this assumption.
- Spatial data: Data are geographically referenced. *Very rarely i.i.d.* Hence there is a need for more realistic models that account for the fact that data are spatially dependent.
- What's more, the dependence may be present in all directions and the relationships may be highly complex.
- Often, data that are further apart are less dependent than data that are located next to each other. However, there are also situations where repulsive behavior between points may need to be modeled.

General reasons to use spatial models

- Utilizing spatial dependence correctly will lead to more accurate estimators, i.e., estimates with lower variability.
- Ignoring dependence can lead to underestimates of variability of estimates.
- Spatial dependence can be a surrogate for unknown and important covariates: Can 'adjust' for these covariates.
- Spatial-dependence component can protect against misspecification of mean structure (by accounting for a variable that is spatially varying) .
- Often learning about spatial dependence is of interest in its own right.

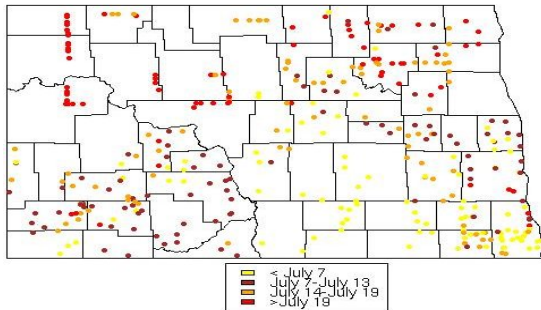
Types of Spatial Data

There are three main categories of spatial data (though it is not always obvious how to classify data into these categories):

- Geostatistical data: When a spatial process that varies continuously is observed only at points.
- Lattice data: Generally associated with an arbitrary division of the area being studied into a regular or irregular lattice. Often used to represent data observed on or aggregated up to arbitrary spatial units such as census tracts, counties etc.
- Spatial point processes: When a spatial process is observed at points and the locations themselves are of interest (and treated as random variables.)

Geostatistical (point-referenced) data: Examples

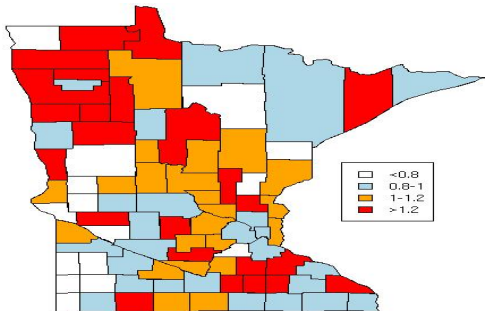
- (1) Concentrations of PM_{2.5} (pollutants) across the U.S.
- (2) Wheat flowering dates by location (below):



Courtesy Plant Pathology, PSU and North Dakota State.

Areal/Lattice Data: Examples

- (1) Pixel values from remote sensing e.g. forest cover in PA.
- (2) Event rates by county (e.g. below).



Courtesy MN Cancer Surveillance System, Dept. of Health

Spatial (linear) model for geostatistics and lattice data

Although geostatistical models and areal/lattice data models are usually talked about separately, they can be viewed in a unified framework.

- Spatial process at location \mathbf{s} is $Z(\mathbf{s}) = \mu(\mathbf{s}) + w(\mathbf{s})$ where:
 - $\mu(\mathbf{s})$ is the mean. Often $\mu(\mathbf{s}) = X(\mathbf{s})\beta$, $X(\mathbf{s})$ are covariates at \mathbf{s} and β is a vector of coefficients.
- Model dependence among spatial random variables by imposing it on the errors (the $w(\mathbf{s})$'s).
- For n locations, $\mathbf{s}_1, \dots, \mathbf{s}_n$, $\mathbf{w} = (w(\mathbf{s}_1), \dots, w(\mathbf{s}_n))^T$ can be jointly modeled via a zero mean Gaussian process (GP), for geostatistics, or Gaussian Markov random field (GMRF), for areal/lattice data.

Gaussian Processes

- Gaussian Process (GP): Let Θ be the parameters for covariance matrix $\Sigma(\Theta)$. Then:

$$\mathbf{w}|\Theta \sim N(\mathbf{0}, \Sigma(\Theta)).$$

This implies:

$$\mathbf{Z}|\Theta, \beta \sim N(\mathbf{X}\beta, \Sigma(\Theta))$$

- We have used the simplest multivariate distribution (the multivariate normal). We will specify $\Sigma(\Theta)$ so it reflects spatial dependence.
- Need to ensure that $\Sigma(\Theta)$ is positive definite for this distribution to be valid, so we assume some valid parametric forms for specifying the covariance.

Gaussian Processes: Example

- Consider the popular **exponential** covariance function.
- Let $\Sigma(\Theta) = \kappa I + \psi H(\phi)$ where I is the $N \times N$ identity matrix. Note that $\Theta = (\kappa, \psi, \phi)$ and $\kappa, \psi, \phi > 0$.
- The i, j th element of the matrix H ,
$$H(\|\mathbf{s}_i - \mathbf{s}_j\|; \phi)_{ij} = \exp(-\phi \|\mathbf{s}_i - \mathbf{s}_j\|).$$
- Note: covariance between i, j th random variables depends only on distance between \mathbf{s}_i and \mathbf{s}_j , and does not depend on the locations themselves (implying *stationarity*) and only depends on the magnitude of the distance, not on direction (implying *isotropy*).
- Extremely flexible models, relaxing these conditions, can be easily obtained though fitting them can be more difficult.

Gaussian Processes: Inference

- The model completely specifies the likelihood, $\mathcal{L}(\mathbf{Z}|\Theta, \beta)$.
- This means we can do likelihood-based inference:
 - If we observe \mathbf{Z} , can find maximum likelihood estimates of Θ, β by maximizing $\mathcal{L}(\mathbf{Z}; \Theta, \beta)$ with respect to Θ, β .
 - Using the MLEs of Θ, β , and conditioning on the observed values \mathbf{Z} , we can easily estimate the value of this process at other locations ('kriging' with Gaussian processes.)
- If we place priors on Θ, β , we can do Bayesian inference:
 - Simulate from the posterior distribution, $\pi(\Theta, \beta | \mathbf{Z})$ via Markov chain Monte Carlo (tutorial tomorrow!)
 - Using sampled values of Θ, β , conditioning on \mathbf{Z} , can easily simulate value of this process at other locations.
 - Bayesian version incorporates variability due to uncertainty about Θ, β .

Gaussian Processes: Computing

- For likelihood based inference: R's `geoR` package by Ribeiro and Diggle.
- For Bayesian inference:
 - R's `spBayes` package by Finley, Banerjee and Carlin.
 - `WINBUGS` software by Spiegelhalter, Thomas and Best.
- Very flexible packages: can fit many versions of the linear Gaussian spatial model. Also reasonably well documented.
- Warning: With large datasets (>1000 data points), matrix operations (of order $O(N^3)$) become very slow. Either need to be clever with coding or modeling. Above software will not work.

Modeling areal/lattice data

- Recall that we are specifying dependence on the spatial data \mathbf{Z} via \mathbf{w} where $Z(\mathbf{s}) = \mu(\mathbf{s}) + w(\mathbf{s})$.
- We could stick to geostatistical modeling (using, say, distances between centroids of subregions.)
- This is often not reasonable. For e.g. centroids of subregions may lie outside the subregion.
- Modeling the process through its adjacencies may make more sense.
- Adjacent (neighboring) regions/pixels are thought to be more strongly related than those further away so, again, concerned with incorporating dependence into modeling.

Areal/lattice data: conditionally specified models

- Dependence in such cases can be imposed by a **conditionally specified modeled**. Idea is as follows: What is the distribution of the random variable at this location *given* that I know the values of the random variable at neighboring locations?
- Caution: Need to ensure that conditional specification of distribution results in valid joint distribution. Theory relies on Hammersley-Clifford theorem and Gibbs distributions (cf. Besag, 1974).
- Best to use well studied conditionally specified models. For example, we could use Gaussian Markov random field models.

Areal/lattice data: conditionally specified models

- We model the conditional distribution of $Z(\mathbf{s}_i) \mid Z(\mathbf{s}_{-i})$ where $Z(\mathbf{s}_{-i})$ denotes all $Z(\mathbf{s}_j)$ except $Z(\mathbf{s}_i)$.
- Markov property: $Z(\mathbf{s}_i) \mid Z(\mathbf{s}_{-i})$ is the same as $Z(\mathbf{s}_i) \mid Z(\mathbf{s}_{j \sim i})$ where $j \sim i$ indicates that \mathbf{s}_j is a neighbor of \mathbf{s}_i . The distribution of $Z(\mathbf{s}_i)$ is *conditionally independent* of all the other values, *given* its neighboring values.
- This is therefore a local specification, although the model indirectly implies a global specification, i.e., a joint distribution (all the $Z(\mathbf{s})$'s will still be dependent on one another.)

Gaussian Markov random field (contd.)

- If we assume all conditional distributions are Normal (with appropriate conditions on variance parameters), resulting distribution is a multivariate normal.
- See any standard references on GMRFs for (slightly messy) details.
- If we let Θ be the parameters for the *precision matrix* $Q(\Theta)$. Then:

$$\mathbf{Z}|\Theta, \beta \sim N(\mathbf{X}\beta, Q^{-1}(\Theta))$$

- Since we have a likelihood, we can (as before) find an MLE or place priors on the parameters and do Bayesian inference and estimation.

Models for areal/lattice data: computing

- GMRF models (and more generally, conditionally specified models) have an important advantage: The matrices involved tend to be quite sparse and hence can yield considerable computational advantages over a Gaussian Process specification.
- GeoDa package at <https://www.geoda.uiuc.edu/> (free) by Luc Anselin
- R's `spdep` package by Roger Bivand et al.
- Bayesian inference: WINBUGS includes GeoBUGS which is useful for fitting such models.

Spatial generalized linear model

What if data are non-Gaussian? (Diggle, Tawn, Moyeed, 1998)

- Stage 1: Model $Z(\mathbf{s}_i)$ conditionally independent with distribution f given parameters β, Θ , spatial errors $w(\mathbf{s}_i)$

$$f(Z(\mathbf{s}_i)|\beta, \Theta, w(\mathbf{s}_i)),$$

where $g(E(Z(\mathbf{s}_i))) = \eta(\mathbf{s}_i) = \mathbf{x}(\mathbf{s}_i)\beta + w(\mathbf{s}_i)$, η is a canonical link function (for example the logit link).

- Stage 2: Again $\mathbf{w} = (w(\mathbf{s}_1), \dots, w(\mathbf{s}_n))^T$. Model \mathbf{w} as spatially dependent either via a GP or GMRF.
- Stage 3 (if Bayesian inference): Priors for Θ, β .
- Inference based on either $\mathcal{L}(\mathbf{Z}; \mathbf{w}, \Theta, \beta)$ or $\pi(\Theta, \beta, \mathbf{w} | \mathbf{Z})$.

Spatial glm (contd)

To get an idea of how this is useful, consider $Z(\mathbf{s}_i)$'s that are spatially dependent count data.

- Stage 1: Model $Z(\mathbf{s}_i) \sim \text{Poisson}(\mu(\mathbf{s}_i))$, conditionally independent.
- Stage 2: Now model $\log(\mu(\mathbf{s}_i)) = X(\mathbf{s}_i)\beta + w(\mathbf{s}_i)$.
- Let: $\mathbf{w} = (w(\mathbf{s}_1), \dots, w(\mathbf{s}_n))^T$. Model \mathbf{w} as spatially dependent either via a GP or GMRF, with parameters Θ .
- Stage 3 (if Bayesian inference): Priors for Θ, β .
- Inference based on either $\mathcal{L}(\mathbf{Z}; \mathbf{w}, \Theta, \beta)$ or $\pi(\Theta, \beta, \mathbf{w} \mid \mathbf{Z})$.

If 0-1 data, replace Stage 1 by Bernoulli($p(\mathbf{s}_i)$) and model $\log\left(\frac{p(\mathbf{s}_i)}{1-p(\mathbf{s}_i)}\right)$ as in Stage 2 and the rest follows in similar fashion.

Spatial Point Processes: Introduction

Have so far discussed the first two major categories of spatial data. The third category is also equally important and perhaps of particular interest to astronomers.

- **Spatial point process:** The *locations* where the process is observed are random variables, 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 ?

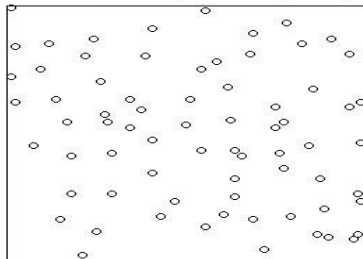
Spatial Point Process Data: Example 2

Locations of pine saplings in a Swedish forest.

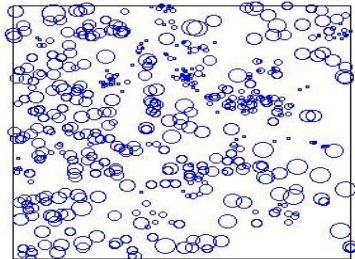
Location and diameter of Longleaf pines (marked point process).

Are they randomly scattered or are they clustered?

Point pattern (Swedish pines)



Marked point pattern (Longleafs)



(from Baddeley and Turner R package, 2006)

Questions related to spatial randomness of process

Some examples:

- Is there regular spacing between locations where process was observed or do locations show a tendency to cluster together? Need to fit clustering models and perhaps do some hypothesis testing.
- Does the probability of observing the event vary according to some factors? (Need to relate predictors to observations in a regression type setting.)
- Can we estimate the overall count from only partial observations? Need to fit a model to observations and make estimates/predictions based on fitted model.

Questions related to spatial randomness (contd)

Assume multiple (sometimes competing) models for the process. For instance, when studying point patterns of observations:

- Perhaps non-homogenous environmental conditions (associated with locations) are related to the presence/absence.
- Maybe the pattern arose by virtue of how the process spreads (e.g. clustering of 'offspring' near 'parents')?
- Note that **hypothesis testing alone is inadequate** for most of these questions. Can try to resolve these by fitting appropriate models where intensity of the process is modeled according to one of the models above.

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.
- 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 $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ 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 , $i = 1, \dots, k$, the joint distribution of $N(A_1), \dots, N(A_k)$ is translation-invariant, i.e., the joint distribution of $N(A_1), \dots, N(A_k)$ = joint distribution of $N(A_1 + \mathbf{y}), \dots, N(A_k + \mathbf{y})$ for arbitrary \mathbf{y} .
- The process is **isotropic** if for any integer k and regions A_i , $i = 1, \dots, k$, the joint distribution of $N(A_1), \dots, N(A_k)$ 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 $N(B) \sim \text{Poisson}(\Lambda(B))$.
 - 2 Conditional on $N(B)$, the points (event locations) $\mathbf{X}_B = \{X_1, \dots, X_{N(B)}\}$ 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 sometimes called the 'density' in other fields such as 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 \mathbf{s} .

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

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

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

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

- 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:

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

- 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 process is completely random in space.

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

- This can be estimated by

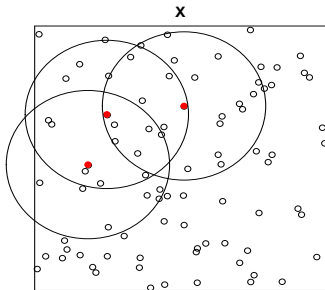
$$\hat{K}(d) = \frac{\sum_{i \neq j} w_{ij} I(d_{ij} \leq d)}{\hat{\lambda}}$$

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

- Note: K can also be viewed as an integral of the two point correlation function as used by astronomers (cf. Martinez and Saar, 2002).

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 observation 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 $E(\text{number of events in a unit area}) = \lambda$, $E(\text{number of events in area within radius } d) = \lambda \pi d^2$.
 - 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

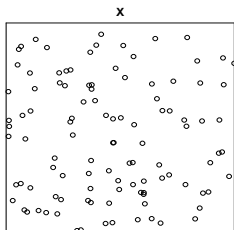
$$\Lambda(A) = \int_A \lambda(\mathbf{s}) d\mathbf{s},$$

equivalently, $P(N(A) = N) = \frac{1}{N!} e^{-\Lambda(A)} (\Lambda(A))^N$.

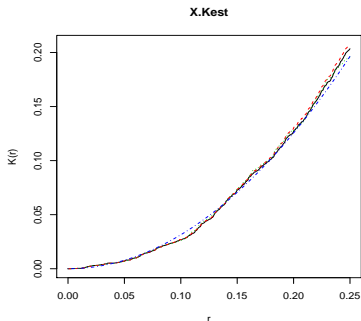
- Conditional on $N(A)$, event locations are independently sampled from a probability density function proportional to $\lambda(\mathbf{s})$.

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

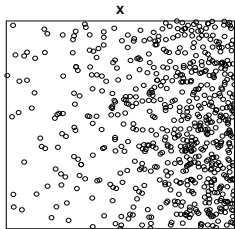
black (and red and green) are various versions of estimates of the K function

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

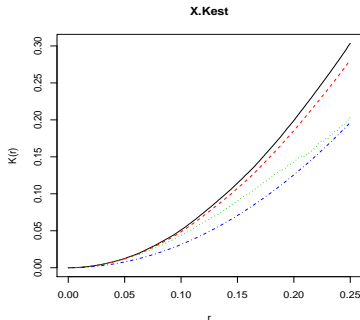
Process was simulated with intensity function

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

Inhomogeneous Poisson Process



Ripley's K



blue=K function under complete spatial randomness

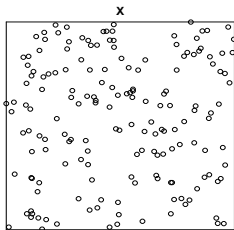
black (and red and green) are various versions of estimates of the K function

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

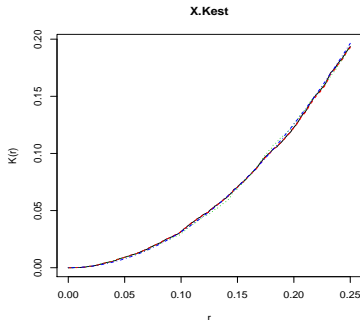
Process was simulated with intensity function

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

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: $E(L(d)) < 0$.

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 $\Lambda(B) > 0$:

- $N(B) \sim \text{Poisson}(\Lambda(B))$, 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) $\mathbf{X}_B = \{X_1, \dots, X_{N(B)}\}$ in the bounded region are (i.i.d.) and each uniformly distributed in the region B :

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

Inference for Poisson process (contd.)

- The joint distribution is then:

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

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

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

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

Space-varying covariates: modulated Poisson process

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

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

- Inhomogeneous Poisson process with this intensity is a *modulated Poisson process*.
- Examples of $\mathbf{X}(\mathbf{s})$: 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 \mathbf{X} related to the spatial point process intensity, i.e., what is β ?

Parameter estimation for modulated Poisson process

Maximum likelihood estimation using observed \mathbf{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

- Impractical to assume covariates are observed for every observed event and all locations in 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 (a form of spatial interpolation, falling under 'Geostatistics') 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.
- Maximize this approximate likelihood to obtain coefficient estimates, $\tilde{\beta}$.

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: $N(A)|\Lambda \sim \text{Poisson}(\Lambda(A))$.
 - 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. \mathbf{X} is a homogeneous Poisson process with intensity λ_0 .
- Special case of Mixed Poisson process: $\lambda_0 \sim \text{Gamma}$. 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).

- 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)

- This is a very natural model when viewed in a hierarchical spatial framework.
- As before, $N(A) | \Lambda \sim \text{Poisson}(\Lambda(A))$.
- 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,

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

where Φ is a Poisson process on $R^2 \times (0, \infty)$, $\mathbf{c} \in R^2$, $\gamma > 0$ and $k(\mathbf{c}, \cdot)$ 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 $\mathbf{Y}_{(\mathbf{c}, \gamma)}$ with intensity function $\gamma k(\mathbf{c}, \cdot)$, and $(\mathbf{c}, \gamma) \in \Phi$.
- $\mathbf{Y}_{(\mathbf{c}, \gamma)}$ is a **cluster** with center \mathbf{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 $\gamma = \alpha > 0$. This process is stationary with intensity $\lambda = \alpha\kappa$.
- 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, $k(c, \cdot)$.
- Other examples: Thomas process, SNCP with covariate information etc.

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 γ s are i.i.d. Gamma r.v.s with mean α and unit scale parameter.
- Assume c s are independent of the γ s.
- $k(c, \cdot)$ 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.

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 referred to as simulated maximum likelihood (SML).
- No 'automatic' methods exist for fitting such models.
- Simulating from a point process model is often easy but inference (estimating a point process model based on observations) is usually more difficult. Challenging to fit flexible new models.

Spatial point processes: computing

- **R command:** `spatstat` function `ppm` fits models that include spatial trend, interpoint interaction, and dependence on covariates, generally using MPL.
- MPL often works well in practice (Baddeley, 2005). Caveat: MPL can work very poorly in some cases, particularly when there is strong dependence.
- MPLE can be used to get a guess for MLE before doing something more elaborate like Markov chain Maximum Likelihood (cf. C.J.Geyer's chapter in "MCMC in Practice", 1996 for a gentle introduction.)
- There is not much in the way of computing resources for fitting Bayesian models, even though they are becoming increasingly common.

Summary: spatial data types and associated models

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 $\mathbf{s} \in D$.

Usual (basic) model: **Gaussian process**.

- **Areal/lattice data:** $D = \{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ 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 $\mathbf{s} \in D$.

Usual (basic) model: **Gaussian Markov random field**.

- **Spatial point process:** $D = \{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ 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. Usual

(basic) models: **Poisson process, Cox process**.

Geostatistics and Lattice/Areal Data:

- Schabenberger and Gotway (2005) "Statistical Methods for Spatial Data Analysis". A fairly comprehensive easy to read book on spatial models for (in order of emphasis): geostatistics, lattice data and point processes.
- Cressie (1994) "Statistics for Spatial Data". This is a comprehensive guide to classical spatial statistics, but it is considerably more technical than the other two references listed here.
- S. Banerjee, B.P. Carlin and A.E. Gelfand "Hierarchical Modeling and Analysis for Spatial Data". This is a textbook on Bayesian models for spatial data.

References: Spatial Point Processes

Spatial Point Processes:

- Møller and Waagepetersen's monograph "Modern statistics for spatial point processes" (2007) To appear in Scandinavian Journal of Statistics.
- Baddeley and Turner's R `spatstat` package.
- Baddeley et al. "Case Studies in Spatial Point Process Modeling" (2005).
- P.J.Diggle's online lecture notes:
<http://www.maths.lancs.ac.uk/~diggle/spatialepi/notes.ps>

References: Spatial Point Processes

- P.J.Diggle “Statistical Analysis of Spatial Point Patterns” (2003)
- “Modern statistics for spatial point processes” by J.Møller and R.P.Waagepeteron (2004).
- V.J.Martinez and E.Sarr “Statistics of the Galaxy Distribution.”

Notes about the references:

- 1 Several of these references also cover spatiotemporal (space-time) process, that may also be of significant interest.
- 2 Acknowledgement: A lot of the material and examples in this tutorial were drawn from several of the listed references.