STAT 534 - Lecture 21: Key

Point Process Data

Point Process Motivation

With point process data, the defining characteristic is that the location is random. Sketch two hypothetical point processes: one with a pattern and one without any spatial pattern.

How did you select the number of points to add to your figure?

- *With point process data, assume a bounded region $\mathcal{D}$ with realizations $\boldsymbol{s}_i$, where $i = 1, ..., n$.*

- Both n *and the* $\boldsymbol{s}_i$ *are random.*

There are many interesting questions related to point process data. Consider, for example, locations of Ponderosa Pines and Aspen Trees. With location data for these trees, we can answer the following questions:

- *What do the occurence patterns look like?*
- *Do the different species have different patterns?*
- *Can environmental factors explain the occurence pattern and do they differ by species?*

Point process data is also prevalent in public health settings; in this context the goal is to understand disease patterns.

- *Are there disease outbreaks?*

- *Do locations of disease cluster?*

- *Is there an evolution of patterns over time?*

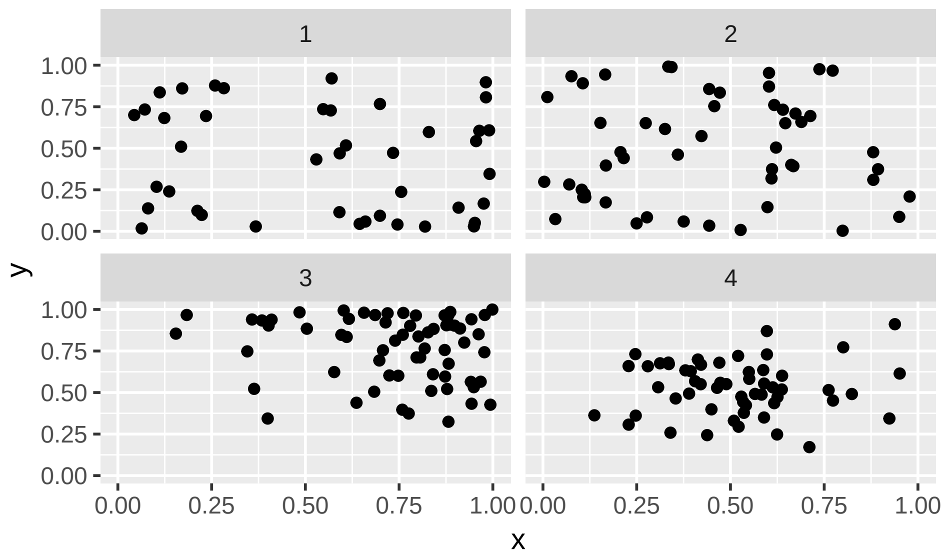
If there is additional meta-information associated with a point pattern, this can be used in what is known as a *marked* point process. For instance, the location of a tree could be considered the point process with the marks associated with tree species.

Discuss the interpretation of marked point process data set with continuous data. For instance, suppose that rather than tree species that the height of the tree was collected. What questions would be of interest?

Exploratory Point Process Analysis

With point process data, there are a few things to assess.

1. *Is there complete randomness or spatial homogeneity?*



2. *How does the process differ from spatial homogeneity?*

3. *Can the spatial heterogeneity be explained by environmental factors that can be used in a regression model?*

4. *Is there clustering (attraction) or inhibition (repulsion) of points?*

Theoretical Details

- Our focus will be largely on $\mathcal{R}^2$, but the same principles apply for other dimensions.

- A random realization of a point pattern is denoted $\mathcal{S}$ with elements $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$. Hence, $\mathcal{S}$ is random and inference focuses on a probabilistic model for $\mathcal{S} \in \mathcal{D}$.

To specify a probabilistic model for $\mathcal{S}$, there are two essential elements.

1. *A distribution for $\mathcal{N}(D)$, the number of points in $\mathcal{D}$ and*

2. *A multivariate density over $\mathcal{D}$ for $f(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)$, which is often referred to as the location density.*

- Then the likelihood of $\mathcal{S}$ can be written as

$$\mathcal{L}(\mathcal{S}) = Pr(\mathcal{N}(d) = n)n!f(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)$$

- Given the likelihood, *the point pattern model is stationary if*

$$f(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n) = f(\boldsymbol{s}_1 + \boldsymbol{h}, \ldots, \boldsymbol{s}_n + \boldsymbol{h})$$

Counting Measure

- The first part of a point process is the number of points.

- *Let $\mathcal{B}$ be a set, then $N(\mathcal{B})$ is the number of points in $\mathcal{B}$. This can be computed as* $\sum_{\boldsymbol{s}_i \in \mathcal{S}} 1(\boldsymbol{s}_i \in \mathcal{B})$

- One common approach for the counting measure is a Poisson Process.

- *Thus $N(\mathcal{B}) \sim Pois(\lambda(\mathcal{B}))$, where $\lambda(\mathcal{B}) = \int_B \lambda \boldsymbol{s} d\boldsymbol{s}$ is the intensity function.*

- If $\mathcal{B}_1$ and $\mathcal{B}_2$ are disjoint, then $N(\mathcal{B})_1$ and $N(\mathcal{B})_2$ are independent.

- Sketch $\mathcal{D}$ such that $\mathcal{B}_1$ and $\mathcal{B}_2$ are disjoint.

- What would $\lambda(\mathcal{B}_1)$ and $\lambda(\mathcal{B}_2)$ be if we had spatial homogeneity?

- When $\lambda(\boldsymbol{s}) = \lambda$, this is referred to as a homogenous Poisson process (HPP).

- *Then $\lambda(\mathcal{B}) = \lambda|\mathcal{B}|$, where $|\mathcal{B}|$ is the area of $\mathcal{B}$.*

- What is the expected number of points for a HPP in $\mathcal{B}$?

- Note $\lambda(\mathcal{D}) = \int_D \lambda \boldsymbol{s} d\boldsymbol{s}$ is an intensity measure, not a probability distribution. What, in general, is necessary to normalize the location density?

Nonhomogenous Poisson process

- A HPP is the simplest form of a point process, sketch the intensity function or a nonhomogenous Poisson process (NHPP). Label the intensity $\lambda(\boldsymbol{s})$ on your figure.

- *When $\lambda(\boldsymbol{s})$ is random, this NHPP is referred to as a Cox process.*

- *With a Cox process, $\lambda(\boldsymbol{s})$ is a realization from a stochocastic process - such as a Gaussian process*

- *The Cox process can clearly be stated as a hierarchical model: 1. draw $\lambda(\boldsymbol{s})$ and 2. draw $\boldsymbol{S}|\lambda(\boldsymbol{s})$.*