Goal in point pattern data analysis is to assess whether there is a spatial pattern in occurrences of an event.

Distinguish between a point and an event location.

In geostatistics our points were locations in a domain that we made a measurement. These points make up a set of spatial random variables for which we wanted to determine the spatial relationships (via the covariance function).

Point patterns consist of event locations where we are concerned with the presence/absence of an event rather than the value of the measurement at a point.

We ask the question: are the events that we observe in our domain from a completely random spatial process or are they exhibiting some type of pattern or clustering?
Point Pattern Data

Bramble canes

Finpines

Spruces

bramblecanes.pts[,1]  
bramblecanes.pts[,2]

finpines.pts[,1]  
finpines.pts[,2]

spruces.pts[,1]  
spruces.pts[,2]
Point Pattern Data
Questions of interest are:

- Are points closer together than they would be by chance?
- Are the points more regularly spaced than they would be by chance?
- What model might reproduce our observed pattern?
Point pattern data

Point pattern notation:

- Spatial location in \((x,y)\) denoted as \(s\).
- \(Y(s)\) represents the presence or absence of \(Y\) where \(Y(s) = 1\) if there is an observed case at location \(s\), and \(Y(s) = 0\) otherwise.
- Spatial domain of observed cases: \(D, D = s\); \(Y(s) = 1\)
- The null hypothesis: no spatial pattern (complete spatial randomness)
- Find a statistic to test whether the data is clustered, or regular
- Develop a model to generate spatial pattern (PCP, IPP, Cox, SIP)
Spatial Randomness

- Typical terms that are used are spatial randomness, random pattern, at random or by chance
- Complete spatial randomness (CSR): events are uniformly distributed across a domain $D$ and are independent of each other
- CSR means an event is equally likely to occur at any location or region within $D$
Spatial Randomness

- Particular set of CSR point processes arise from a stationary **homogeneous spatial Poisson point process**
- Homogeneous Poisson process have the following characteristics:
  1. For any $n$ events in region $D$, the events are an independent random sample from a uniform distribution where each point is a location where an event could occur is equally likely to be chosen as an event
  2. The number of events $n$ occurring within region $D$ is a random variable following a Poisson distribution with mean $\lambda|D|$ where $\lambda$ is a constant and $|D|$ is the area of the region
Properties of homogeneous Poisson processes

- The Poisson distribution allows the total number of observed events to vary from realization to realization while maintaining a fixed but unknown number of events per unit area.
- The expected number of events per unit area is the intensity, \( \lambda \).
- \( \frac{n}{|D|} = \lambda \).
Point Pattern Data

intensity = 100, unit square

intensity = 1, 10 x 10 square
Properties of homogeneous Poisson processes

- Subregions of a homogeneous Poisson process, \( A \in D \)
- Can think of this similarly to stationarity

\[
\lim_{|A| \to 0} \frac{P[\text{exactly one event in } A]}{|A|} = \lambda > 0
\]

This implies that the probability of a single event in an increasingly small area is constant, independent of the location of the region \( A \) within \( D \)

\[
\lim_{|A| \to 0} \frac{P[\text{two or more events in } A]}{|A|} = 0
\]

This implies that the probability of a two or more events in the same location is zero
Properties of homogeneous Poisson processes

- if we let \( n \) be the number of events in \( A \), then \( n \) following a Poisson distribution means

\[
P(n = k) = \frac{\exp(-\lambda|A|)(\lambda|A|)^k}{k!}
\]

Again, homogeneity is similar to stationarity and isotropy for geostatistical data. The intensity of the point process \( \lambda \) does not vary as a function of spatial location within our domain. Homogeneity means that the intensity is constant across the study area.
Testing for CSR

- Many tests of CSR use Monte Carlo methods
- Compare the observed value of a test statistic to its distribution under the null hypothesis of CSR
- Simulate a large number of CSR processes and compare the test statistic from $N_{sim}$ to test statistic from observed
Testing for homogeneous CSR

- Ripley’s $K$, $K(h) = \lambda^{-1} E(N_0(h))$
- Where $(N_0(h))$ is the number of events within a distance $h$ of an arbitrary event
- $K(h)$ tests the expected number of events within distance $h$ from an arbitrary event (excluding the chosen event itself) divided by the average number of events per unit area
- $K(h)$ is equivalent to showing the variance of the number of events occurring in subregion $A$ (Ripley 1977) so is a second order property of the point process.
Testing for homogeneous CSR

- Ripley’s $K$, $K(h) = \lambda^{-1} E(N_0(h))$
- Under CSR $K(h) = \pi h^2$, the area of a circle of radius $h$
Testing for homogeneous CSR
Point Pattern Data

Testing for homogeneous CSR

- For a process that is more regular than CSR we expect fewer events within distance $h$ of a randomly chosen event.
- For a process that is more clustered than CSR we expect more events within distance $h$ of a randomly chosen event.
- Estimating $K(h)$:
  \[
  \hat{K}(h) = \hat{\lambda}^{-1} \frac{1}{N} \sum_i \sum_j \delta(d(i, j) < h)
  \]
  where $i \neq j$ and $d(i, j)$ is the Euclidean distance between events and $\delta(d(i, j) < h) = 1$ if $d(i, j) < h$ and 0 otherwise.
Testing for homogeneous CSR

- There is an alternate $\hat{K}(h)$ estimator that corrects for edges (boundaries of the region).
- Want to prevent including events that occur outside the boundary but within distance $h$.
- Estimating $K(h)$ accounting for boundaries:
  \[
  \hat{K}_{ec}(h) = \hat{\lambda}^{-1} \frac{1}{N} \sum_{i} \sum_{j} w_{ij} \delta(d(i, j) < h)
  \]

- where $w_{ij} = 1$ if the distance between $i$ and $j$ is less than the distance between event $i$ and the boundary of the region.
Point Pattern Data

- Using the $K(h)$ function and determining p-values to test CSR
- Plot $K(h)$; under CSR $K(h) = \pi h^2$ is a parabola
- $(K(h)/\pi)^{1/2} = h$, so plot $h$ vs $(\hat{K}(h)_{ec}/\pi)^{1/2} - h$
- Under CSR, $(\hat{K}(h)_{ec}/\pi)^{1/2} - h = 0$
- Departures from the horizontal line that defines CSR indicate clustering or regularity
  - Deviations above the horizontal line indicate clustering because there are more events within distance $h$ than expected
  - Deviations below the horizontal line indicate regularity because there are fewer events within distance $h$ than expected
Using the $K(h)$ function and determining p-values to test CSR

Plot $K(h)$; under CSR $K(h) = \pi h^2$ is a parabola

$(K(h)/\pi)^{1/2} = h$, so plot $h$ vs $(\hat{K}(h)_{ec}/\pi)^{1/2} - h$

Under CSR, $(\hat{K}(h)_{ec}/\pi)^{1/2} - h = 0$

Departures from the horizontal line that defines CSR indicate clustering or regularity

Deviations above the horizontal line indicate clustering because there are more events within distance $h$ than expected

Deviations below the horizontal line indicate regularity because there are fewer events within distance $h$ than expected
Testing for homogeneous CSR

Point Pattern Data
Testing for homogeneous CSR
Inhomogeneous (heterogeneous) Poisson processes occurs when the intensity $\lambda$ is not constant over the region.

Many cases homogeneity in intensity is not realistic, for example the locations of trees in a forest may be irregular due to geographic features such as soil, rock, slope or other terrain irregularities.

In the case of IPP, the intensity is a function that varies spatially, $\lambda(s)$.
Inhomogeneous Poisson process

- We can estimate the intensity function in different ways: parametrically by defining a specific function or non-parametrically using kernel smoothing

\[ \hat{\lambda}(s) = \frac{1}{h^2} \sum_i \kappa(\frac{\|s - s_i\|}{h})/q(\|s\|) \]

Where \( \kappa(s) \) is a kernel function and \( q\|s\| \) is a boundary correction. The distance \( h \) is our bandwidth for smoothing.
Inhomogeneous Poisson process

- There are various kernel functions, but a quadratic function is often used

\[ \kappa(s) = \frac{3}{\pi} (1 - \|s\|^2)^2 \]
**Poisson cluster process**

- A spatial point process where each event belongs to a cluster
- There is a parent event that produces a random number of offspring
- Parent events are usually a realization of an Poisson process with intensity $\lambda(s)$
- We have $i$ parents, and each parent produces a random number of offspring, $O_i$
- The $O_i$ are distributed within $h_i$ of the parent and follow a bivariate probability distribution
Poisson cluster process

- Can have homogeneous cluster processes where the intensity of the offspring around a parent is constant $\lambda$
- Or an inhomogeneous cluster process where the intensity of the offspring around a parent is not constant across domain $\lambda(s)$
- Parent events are usually a realization of an inhomogeneous Poisson process with intensity $\lambda(s)$ distribution
Point Pattern Data

Poisson cluster process

PCP, \((P,O,\text{Spread})=(25,4,.0025)\)

PCP, \((P,O,\text{Spread})=(25,4,.005)\)