Spatial Point Patterns

Characteristics:

- set of $n$ point locations with recorded “events”, e.g., locations of trees, disease or crime incidents $S = \{s_1, \ldots, s_i, \ldots, s_n\}$
- point locations correspond to all possible events or to subsets of them
- attribute values also possible at same locations, e.g., tree diameter, magnitude of earthquakes (marked point pattern) $W = \{w_1, \ldots, w_i, \ldots, w_n\}$

Analysis objectives:

- detect spatial clustering or repulsion, as opposed to complete randomness, of event locations (in space and time)
- if clustering detected, investigate possible relations with nearby “sources”
Mean center of a point pattern:

- point with coordinates $\bar{s} = (\bar{x}, \bar{y})$:

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \quad \text{and} \quad \bar{y} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i}$$

- center of point pattern, or point with average $x$ and $y$-coordinates

Median center of a point pattern:

- **both** of the following two centers are called **median centers**, although they are essentially different (confusing!)
  - the intersection between the median of the $x$ and the $y$ coordinates
  - **center for minimum distance**: $s_c \in \{s_1, \ldots, s_n\} s.t. \min \sum_{i=1}^{n} |s_i - s_c|$  

- the first type of **median center** is not unique, and there is **no** closed form for the second type

- **$p$-median problem** (a typical problem in spatial optimization): the problem of locating $p$ “facilities” relative to a set of “customers” such that the sum of the shortest demand weighted distance between “customers” and “facilities” is minimized
Simple Descriptive Statistics

Changes of population center (year 1790-2000):

<table>
<thead>
<tr>
<th>Year</th>
<th>Median center</th>
<th>Mean center</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Latitude-N</td>
<td>Longitude</td>
</tr>
<tr>
<td>1790</td>
<td>(NA)</td>
<td>(NA)</td>
</tr>
<tr>
<td>1800</td>
<td>40 03 35</td>
<td>84 49 91</td>
</tr>
<tr>
<td>1850</td>
<td>38 52 11</td>
<td>86 56 51</td>
</tr>
<tr>
<td>1880</td>
<td>38 52 46</td>
<td>88 56 32</td>
</tr>
<tr>
<td>1920</td>
<td>38 56 47</td>
<td>90 51 43</td>
</tr>
<tr>
<td>1950</td>
<td>38 52 46</td>
<td>90 51 43</td>
</tr>
<tr>
<td>1980</td>
<td>38 52 46</td>
<td>90 51 43</td>
</tr>
<tr>
<td>2000</td>
<td>38 52 46</td>
<td>90 51 43</td>
</tr>
</tbody>
</table>

NA Not available. "West Virginia was set off from Virginia, Dec. 31, 1862, and admitted as a state, June 19, 1863.

Map showing the movement of the population center from 1790 to 2000.
Descriptive Statistics

Standard distance of a point pattern:

- average squared deviations of \( x \) and \( y \) coordinates from their respective mean:
  \[
  d_{std} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n}(y_i - \bar{y})^2}{n - 2}}
  \]

- related to standard deviation of coordinates, a summary circle (centered at \( \bar{s} \) with radius \( d_{std} \)) of a point pattern

Standard deviational ellipse:

- Taking directional effects into account for anisotropy cases
- Please refer to Levine and Associates, 2004 for calculations
Descriptive Statistics

Examples:

Remarks:

- indicates overall shape and center of point pattern
- do not suffice to fully specify a spatial point pattern
Point Pattern Analysis Methods

1st order (i.e., intensity): absolute location of events on map:

- Quadrat methods
- Density Estimation (KDE)
- Moran’s I and Geary’s C

2nd order (i.e., interactions): interaction of events:

- Nearest neighbor distance
- Distance functions G, K, F, L
- Getis-Ord Gi* and Anselin local Moran’s I
**Quadrat methods**

Consider a point pattern with $n$ events within a study region $A$ of area $|A|$

Global intensity:

$$\hat{\lambda} = \frac{n}{|A|} = \frac{\text{# of events within } A}{|A|}$$

Local intensity via quadrats

1. partition $A$ into $L$ sub-regions $A_l$, $l = 1, \ldots, L$ of equal area $|A_l|$ (also called quadrats)
2. count number of events $n(A_l)$ in each sub-region $A_l$
3. convert sample counts into estimated intensity rates as:

$$\hat{\lambda}(A_l) = \frac{n(A_l)}{|A_l|}$$
• estimated rates $\hat{\lambda}(A_I)$ over set of quadrats
• reveal large-scale patterns in intensity variation over $A$
• larger quadrats yield smoother intensity maps; smaller quadrats yield ‘spiky’ intensity maps
• size, origin, and shape of quadrats is critical (recall: MAUP)
• only first-order effects are captured
Dependence of intensity on a covariate (Inhomogeneous cases)

\[ \rho(Z) \quad \rho_{hi}(Z) \quad \rho_{lo}(Z) \]

- reclass of slope
- quadrat based on reclass-ed slope
- intensity vs. slope
Kernel Density Estimation

Procedure of Kernel Density Estimation (KDE)

1. define a kernel $K(s; r)$ of radius (or bandwidth) $r$ centered at any arbitrary location $s$

2. estimate local intensity at $s$ as:

$$\hat{\lambda}(s) = \frac{1}{n} \sum_{i=1}^{n} K(s_i - s; r)$$

3. repeat estimation for all points $s$ in the study region to create a density map
Kernel Density Estimation

An illustration of the KDE procedure in 1D
Kernel Density Estimation

Example for the previous dataset:
Kernel Density Estimation

Example with 2km bandwidth

[Map of a region with a highlighted area]
Kernel Density Estimation

Example with 10km bandwidth
Kernel Density Estimation

Example with 40km bandwidth
Comments

• Choice of kernel function is not critical (Diggle, 1985)
• Choice of bandwidth, or degree of smoothing critical:
  • Small bandwidth $\rightarrow$ spiky results
  • Large bandwidth $\rightarrow$ loss of detail
• Multi-scale analyses can use these bandwidth characteristics to investigate both broad trends and localized variation
• How to choose bandwidth: choose the degree of smoothing subjectively, by eye, or by formula (Diggle)
• could define local bandwidth based on function of presence of events in neighborhood of $s$ (i.e., adaptive kernel estimation)

What does the output of KDE means?
Distance-based Descriptors of Point Patterns

- Distances: accessing second order effects
  - Event-to-event distance: distance $d_{ij}$ between event at arbitrary location $s_i$ and another event at another arbitrary location $s_j$:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$
### Distances

![Graph showing distances between points](image)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00</td>
<td>11947.70</td>
<td>16042.65</td>
<td>3481.22</td>
<td>10742.98</td>
</tr>
<tr>
<td>2</td>
<td>11947.70</td>
<td>0.00</td>
<td>5126.79</td>
<td>15219.58</td>
<td>1599.07</td>
</tr>
<tr>
<td>3</td>
<td>16042.65</td>
<td>5126.79</td>
<td>0.00</td>
<td>19481.59</td>
<td>6720.59</td>
</tr>
<tr>
<td>4</td>
<td>3481.22</td>
<td>15219.58</td>
<td>19481.59</td>
<td>0.00</td>
<td>13913.70</td>
</tr>
<tr>
<td>5</td>
<td>10742.98</td>
<td>1599.07</td>
<td>6720.59</td>
<td>13913.70</td>
<td>0.00</td>
</tr>
</tbody>
</table>

*Table: Euclidean distance matrix*
Nearest-Neighbor Distances

Nearest neighbour distances

- Mean nearest neighbour distance: Average of all $d_{min}(s_i)$ values

$$\bar{d}_{min} = \frac{1}{n} \sum_{i=1}^{n} d_{min}(s_i)$$
The $G$ function

- Definition: nearest neighbour distance function, i.e., proportion of event-to-nearest-neighbor distances $d_{\text{min}}(s_i)$ no greater than given distance cutoff $d$, estimated as:

$$\hat{G}(d) = \frac{\#\{d_{\text{min}}(s_i) < d, i = 1, \ldots, n\}}{n}$$

- alternative definition: cumulative distribution function (CDF) of all $n$ event-to-nearest-neighbor distances; instead of computing average $\bar{d}_{\text{min}}$ of $d_{\text{min}}$ values, compute their CDF

- the $G$ function provides information on event proximity

- example for previous clustering point pattern:
Examples of $G$ function

Expected plot:

- for clustered events, $\hat{G}(d)$ rises sharply at short distances, and then levels off at larger $d$-values
- for randomly-spaced events, $\hat{G}(d)$ rises gradually up to the distance at which most events are spaced, and then increases sharply
K function

Working with pair-wise distances & looking beyond nearest neighbours

Concept

1. construct set of concentric circles (of increasing radius $d$) around each event
2. count number of events in each distance “band”
3. cumulative number of events up to radius $d$ around all events becomes the sample $K$ function $\hat{K}(d)$
Working with pair-wise distances & looking beyond nearest neighbours

• Formal definition:

\[ K(d) = \frac{1}{\lambda} \frac{n}{n} \frac{\#\{d_{ij} \leq d, i, j = 1, \ldots, n\}}{|A| \frac{n}{n} \frac{\#\{d_{ij} \leq d, i, j = 1, \ldots, n\}}{|A|} } = |A| (\text{proportion of event-to-event distance} \leq d) \]

• In other words, the \( \hat{K}(d) \) is the sample cumulative distribution function (CDF) of all \( n^2 \) event-to-event distances, scaled by \(|A|\)
Examples of Event-to-Event Distance Histogram and CDFs

- **Cluster Histogram**: 103 clustering points
  - Frequency distribution shows multiple peaks due to grouping in space.

- **Uniform Histogram**: 103 uniform points
  - Frequency distribution is more uniform and spread out.

- **Cluster CDF**: Proportion of distances <= x, n=10609, m=0

- **Uniform CDF**: Proportion of distances <= x, n=10609, m=0

Examples of $K$ functions

- the sample $K$ function $\hat{K}(d)$ is monotonically increasing and is a scaled (by area $|A|$) version of the CDF of E2E distances
Recap

Spatial point patterns

- set of $n$ point locations with recorded “events”

Describing the first-order effect

- overall intensity
- local intensity (quadrat count and kernel density estimation)

Describing the second-order effect

- nearest neighbour distances
  - the G function
- pair-wise distances
  - the K function
Caveats:

- theoretical G, K functions are defined and estimated under the assumption that the point process is stationary (homogeneous)
- these summary functions do not completely characterise the process
- if the process is not stationary, deviations between the empirical and theoretical functions (e.g. \( \hat{K} \) and K) are not necessarily evidence of interpoint interaction, since they may also be attributable to variations in intensity

Example
Descriptive vs Statistical Point Pattern Analysis

Descriptive analysis:

- set of quantitative (and graphical) tools for characterizing spatial point patterns
- different tools are appropriate for investigating first- or second-order effects (e.g., kernel density estimation versus sample G function)
- can shed light onto whether points are clustered or evenly distributed in space

Limitation:

- no assessment of how clustered or how evenly-spaced is an observed point pattern
- no yardstick against which to compare observed values (or graph) of results
Statistical analysis:

- assessment of whether an observed point pattern can be regarded as one (out of many) realizations from a particular spatial process
- measures of confidence with which the above assessment can be made (how likely is that the observed pattern is a realization of a particular spatial process)

Are daisies randomly distributed in your garden?
Complete Spatial Randomness (CSR)

- yardstick, reference model that observed point patterns could be compared with, i.e., null hypothesis
- homogeneous (uniform) Poisson point process
- basic properties:
  - the number of points falling in any region $A$ has a Poisson distribution with mean $\lambda |A|$
  - given that there are $n$ points inside region $A$, the locations of these points are i.i.d. and uniformly distributed inside $A$
  - the contents of two disjoint regions $A$ and $B$ are independent

Example:
Nearest Neighbour Index (NNI) test under CSR

Nearest neighbour index

- Compares the mean of the distance observed between each point and its nearest neighbor ($\bar{d}_{min}$) and the expected mean distance under CSR $E(d_{min})$

\[
NNI = \frac{\bar{d}_{min}}{E(d_{min})}
\]

- Under CSR, we have:

\[
E(d_{min}) = \frac{1}{2\sqrt{\lambda}}
\]

\[
\sigma(d_{min}) = \frac{0.26136}{\sqrt{n^2 / A}}
\]
The K Function under CSR

- The K function is a function of pair-wise distances
- For a homogeneous Poisson point process of intensity $\lambda$, the pair-wise distance distribution (the K function) is known to be:

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

- A commonly-used transformation of K is the L-function:

$$L(d) = \sqrt{\frac{K(d)}{\pi}} = d$$

Example

![Graphs showing K and L functions for different distributions](image-url)
Monte Carlo test

- because of random variability, we will never obtain perfect agreement between sample functions (say the K function) with theoretical functions (the theoretical K functions), even with a completely random random pattern

Example
Monte Carlo test

- A *Monte Carlo* test is a test based on simulations from the null hypothesis
- Basic procedures:
  - generate $M$ independent simulations of CSR inside the study region $A$
  - compute the estimated K functions for each of these realisations, say $\hat{K}(j)(r)$ for $j = 1, \ldots, M$
  - obtain the pointwise upper and lower envelopes of these simulated curves
- not a confidence interval

Example
Recap

Statistical analysis of spatial point patterns:

- allows to quantify departure of results obtained via exploratory tools, e.g., $\hat{G}(d)$, from expected such results derived under specific null hypotheses, here CSR hypothesis
- can be used to assess to what extent observed point patterns can be regarded as realizations from a particular spatial process (here CSR)
- Same concepts can be applied for hypothesis of other types of point processes (e.g., Poisson cluster process, Cox process)

Sampling distribution of a test statistics

- lies at the heart of any statistical hypothesis testing procedure, and is tied to a particular null hypothesis
- simulation and analytical derivations are two alternative ways of computing such sampling distributions (the latter being increasingly replaced by the former)

Edge Effects
Recap

Scale effects

- Wolf pack example

- Nearest neighbour distance (NN distance, G functions) vs K function

Edge effects
Recap

Extended into line processes

- Line density