Applied Spatial Statistics in R, Section 4
Spatial Point Processes

Yuri M. Zhukov
IQSS, Harvard University
January 16, 2010
Outline

1 Introduction
   - Why use spatial methods?
   - The spatial autoregressive data generating process

2 Spatial Data and Basic Visualization in R
   - Points
   - Polygons
   - Grids

3 Spatial Autocorrelation

4 Spatial Weights

5 Point Processes

6 Geostatistics

7 Spatial Regression
   - Models for continuous dependent variables
   - Models for categorical dependent variables
   - Spatiotemporal models
During World War II, Germany launched 1,358 V-2 Rockets at London.

The V-2’s speed and trajectory made it invulnerable to anti-aircraft guns and fighters.

But its guidance systems were thought to be too primitive to hit specific targets.

After the strikes began in 1944, bomb damage maps were interpreted by some analysts as showing that impact sites were clustered.

This evidence appeared to contradict existing intelligence on the V-2 program.

If the rocket strikes were spatially clustered, the guidance systems must have been more advanced than previously thought.
Point Pattern Analysis: Aerial Bombardment

Figure: Distribution of V-2 Rocket Strikes on Central London, 1944
R.D. Clarke (1946) decided to apply a statistical test to assess whether any support could be found for the clustering hypothesis. He selected an area of 144 km$^2$ in south London, which he divided into 576 squares of 1/4 km$^2$. For each square, Clark recorded the total number of observed bomb hits. There were 537 total in the study area. He then recorded the number of squares with $k = 1, 2, 3, \ldots$ hits. The expected number of squares with $k$ hits was derived from the Poisson distribution $\sum_{k=1}^{n} \frac{e^{-\lambda} \lambda^k}{k!}$, with $\lambda = \frac{537}{576}$ and $n = 576$. 
Point Pattern Analysis: Aerial Bombardment

<table>
<thead>
<tr>
<th>No. of bombs per square</th>
<th>Expected</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>226.74</td>
<td>229</td>
</tr>
<tr>
<td>2</td>
<td>211.39</td>
<td>211</td>
</tr>
<tr>
<td>3</td>
<td>98.54</td>
<td>93</td>
</tr>
<tr>
<td>4</td>
<td>7.14</td>
<td>7</td>
</tr>
<tr>
<td>5+</td>
<td>1.57</td>
<td>1</td>
</tr>
</tbody>
</table>

$\chi^2 = 1.17, p = 0.88$

- It is clear from the cross-tabulation that the distribution of V-2 hits conforms quite closely to the Poisson distribution.

- The occurrence of clustering would have been reflected in an excess number of squares with either a high number of bombs or none at all, and fewer squares in the intermediate classes.

- The closeness of fit suggested that V-2 impact sites were random, rather than clustered.
Point patterns have first- and second- order properties:

1. **First-order properties** measure the distribution of events in a study region: intensity and spatial density.

2. **Second-order properties** measure the tendency of events to appear clustered, independently, or regularly-spaced.
The most basic test which can be performed is that of Complete Spatial Randomness (CSR). Under CSR, events are distributed independently and uniformly over a study area.

A point process which is CSR point process is formally defined as a homogeneous Poisson process (HPP).

- Under HPP, the location of one point in space does not affect the probabilities of other points’ appearing nearby. The intensity of the point process in area $A$ is a constant $\lambda(y) = \lambda > 0$, $\forall y \in A$.

A generalization of HPP which allows for non-constant intensity $\lambda(y)$ is called an inhomogeneous Poisson process (IPP).
Point Pattern Processes: Complete Spatial Randomness

- Let’s explore conformity to CSR among three point patterns: (1) real data on crime locations in Baltimore, (2) points drawn from uniform distribution over the same study area, (3) regularly-spaced point pattern.
Point Pattern Processes: $G$ Function

- The $G$ Function measures the distribution of distances from an arbitrary event to its nearest neighbors.

$$\hat{G}(r) = \frac{1}{n} \sum_{i=1}^{n} l_i$$

$$l_i = \begin{cases} 
1 & \text{if } d_i \in \{d_i : d_i \leq r, \forall i\} \\
0 & \text{otherwise}
\end{cases}$$

- where $d_i = \min_j \{d_{ij}, \forall j \neq i \in S\}, i = 1, \ldots, n$.

- So, the $G$ function represents the number of elements in the set of distances up to some threshold $r$, normalized by the total number of points $n$ in point pattern $S$.

- Under CSR, the value of the $G$ function becomes:

$$G(r) = 1 - e^{\lambda \pi r^2}$$

- where $\lambda$ is the mean number of events per unit (intensity).
The comparability of a point process with CSR can be assessed by plotting the empirical function $\hat{G}(r)$ against the theoretical expectation $G(r)$.

For a clustered pattern, observed locations should be closer to each other than expected under CSR. A regular pattern should have higher nearest-neighbor distances than expected under CSR.

This is shown below for the Baltimore crime locations dataset.

![G Function Graph](image-url)
The $\mathcal{F}$ Function measures the distribution of all distances from an arbitrary point $k$ in the plane to the nearest observed event $j$.

$$\hat{\mathcal{F}}(r) = \frac{\sum_{k=1}^{m} I_k}{m}$$

$$I_k = \begin{cases} 
1 & \text{if } d_k \in \{d_k : d_k \leq r, \forall k\} \\
0 & \text{otherwise}
\end{cases}$$

where $d_k = \min_j \{d_{kj}, \forall j \in S\}$, $k = 1, \ldots, m$, $j = 1, \ldots, n$.

Under CSR, the expected value is also

$$\mathcal{F}(r) = 1 - e^{\lambda \pi r^2}$$
Point Pattern Processes: $\mathcal{F}$ Function

- As before, we can plot the empirical function $\hat{\mathcal{F}}(r)$ against its theoretical expectation $\mathcal{F}(r)$.
- For a clustered pattern, observed locations $j$ should be farther away from random points $k$ than expected under CSR. In a regular pattern, random locations should be closer to observed points.
- This is again shown below for the Baltimore crime locations dataset.
Point Pattern Processes: Intensity

- For an HPP point process, intensity is a constant $\lambda(x) = \lambda = \frac{n}{|A|}$, where $n$ is the number of points observed in region $A$, and $|A|$ is the area of region $A$.
- For an IPP point process, intensity is non-constant and can be estimated non-parametrically with kernel smoothing (Diggle 1985, Berman and Diggle 1989, Bivand et al. 2008).
Point Pattern Processes: Kernel Density

- The kernel density estimator is:

\[
\hat{\lambda}(x) = \frac{1}{h^2} \sum_{i=1}^{n} \frac{\kappa\left(\frac{||x-x_i||}{h}\right)}{q(||x||)}
\]

where \(x_i \in \{x_1, \ldots, x_n\}\) is an observed point, \(h\) is the bandwidth, \(q(||x||)\) is a border correction to compensate for observations missing due to edge effects, and \(\kappa(u)\) is a bivariate and symmetrical kernel function.

- \(R\) currently implements a two-dimensional quartic kernel function:

\[
\kappa(u) = \begin{cases} 
\frac{3}{\pi} (1 - ||u||^2)^2 & \text{if } u \in (-1, 1) \\
0 & \text{otherwise}
\end{cases}
\]

- where \(||u||^2 = u_1^2 + u_2^2\) is the squared norm of point \(u = (u_1, u_2)\)
Point Pattern Processes: Kernel Density

- There is no general rule for selecting the bandwidth $h$, which governs the level of smoothing.
- Small bandwidth $\rightarrow$ spiky map; large bandwidth $\rightarrow$ smooth map.
- Berman and Diggle (1989) propose a criterion based on minimization of mean square error (MSE) of the kernel smoothing estimator.
- The plot below implements this approach for the Baltimore crime dataset. The “optimal” bandwidth here is 0.01.
The plot below shows kernel density estimates for the Baltimore crime locations at different values of the bandwidth $h$.

Lighter values indicate greater intensity of the point process.

Clearly, different bandwidths tell very different stories about the spatial intensity of crime in Baltimore...
Examples in R

Switch to R tutorial script. Section 4.