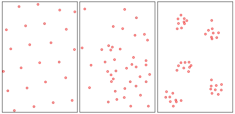
Ripley's K Function

Mickey Smith

April 21, 2012

Spatial Point Patterns

- A data set consisting of point locations, $(\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_n)$, in some study area $\mathcal{R}$.
- Each vector $\mathbf{s}_i$ consists of spatial coordinates and is called an "event."
- "points"
- regularity, randomness, or clustering



First order vs. Second order effects

- First order effects relate to variation in the mean value of the process in space.
- Second order effects result from the spatial correlation of events. (spatial dependence)
- Iron shavings on paper example.

Intensity, $\lambda(s)$

- Defined as the mean number of events per unit area at the point **s**.
- $\lambda(s) = \lim_{ds \to 0} \left\{ \frac{E[Y(\mathbf{ds})]}{ds} \right\}$
- **ds** is a small region around event **s**.
- ds is the area of **ds**.
- $Y(\cdot)$ is the number of events in a region.
- Second order intensity exists in a limit form also but a better characterization of second order properties is Ripley's K function.

K function

- Provides a more effective summary of spatial dependence over a wide range of scales.
- Need to assume our process to be homogeneous or isotropic over such scales.

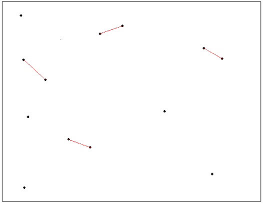
$\lambda K(h) = \mathrm{E}(\text{number of events within distance } h \text{ of an arbitrary event})$

- Again where the intensity, λ, is the mean number of events per unit area, assumed constant throughout $\mathcal{R}$.

Derivation of Estimated K function, $\widehat{K}(h)$

- The expected number of events in $\mathcal{R}$ is λR, where the area of $\mathcal{R}$ is R.
- Using the definition of the K function,

$$\begin{aligned}
\lambda K(h) &= E(\#\text{ of events within distance } h \\
&\qquad\qquad\qquad\qquad \text{of an arbitrary event}) \\
\lambda R \cdot \lambda K(h) &= \lambda R \cdot E(\#\text{ of events within distance } h \\
&\qquad\qquad\qquad\qquad \text{of an arbitrary event}) \\
\lambda^2 R K(h) &= E(\#\text{ of ordered pairs of events at most h apart}).
\end{aligned}$$



Indicator Function

- Let d_{ij} be the distance between the ith and jth observed event in $\mathcal{R}$.

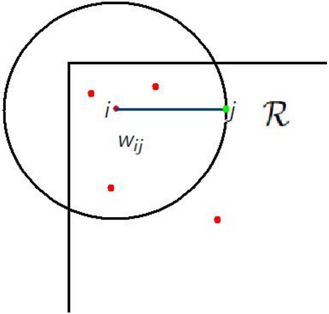
$$I_h(d_{ij}) = \begin{cases} 1 & if \quad d_{ij} \le h \\ 0 & otherwise \end{cases}$$

  - Thus $\sum\sum_{i \neq j} I_h(d_{ij})$ is the observed number of such ordered pairs.
  - Therefore a suitable estimate of $K(h)$ is given by

$$\widehat{K}(h) = \frac{1}{\lambda^2 R} \sum\sum_{i \neq j} I_h(d_{ij})$$

Edge Effect

- Consider a circle centered on event i, passing through the point j. Let w_{ij} be the proportion of the area of this circle which lies within $\mathcal{R}$.



$$\widehat{K}(h) = \frac{1}{\lambda^2 R} \sum \sum_{i \neq j} \frac{I_h(d_{ij})}{w_{ij}}$$

Estimation of Intensity

- Lastly we must estimate λ with the obvious choice

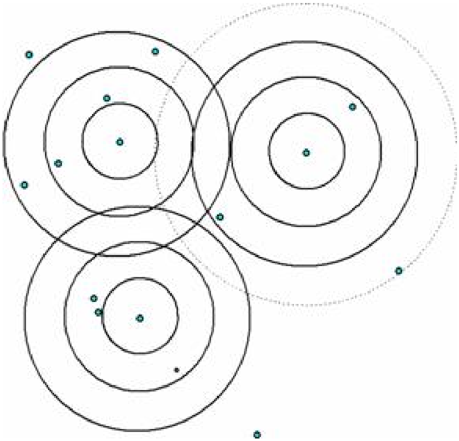
$$\hat{\lambda} = \frac{n}{R}$$

where n is the number of events in $\mathcal{R}$.

$$\widehat{K}(h) = \frac{R}{n^2} \sum_{i \neq j} \sum \frac{I_h(d_{ij})}{w_{ij}}$$

Graphical Notion of K function

- Imagine that an event is "visited" and that around this event we create circles of radius h, then count number of events within the circle.



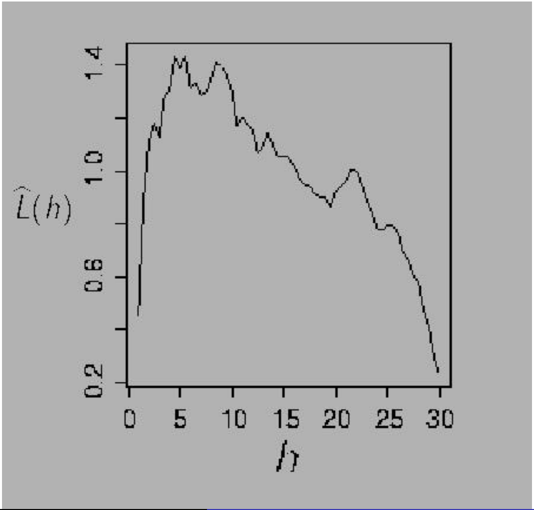
Complete Spatial Randomness (CSR) and $\widehat{L}(h)$

- The standard model for CSR is that events follow a homogeneous Poisson process over the study area.
- Under CSR, the expected number of events within a distance h of a randomly chosen event would be $\lambda\pi h^2$.
- Thus, under CSR, $K(h) = \pi h^2$
- One could simply compare $\widehat{K}(h)$ to πh^2, or plot $\widehat{K}(h) - \pi h^2$ against h, but in practice we consider the $\widehat{L}(h)$.

$$\widehat{L}(h) = \sqrt{\frac{\widehat{K}(h)}{\pi}} - h$$

$\widehat{L}(h)$

- Positive peaks indicate clustering and negative troughs indicate regularity.

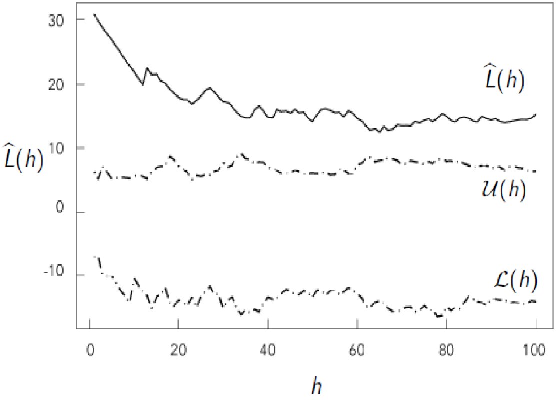


Simulation Envelopes

- First we simulate $\widehat{L}^s(h)$ under CSR.
- Our envelopes will then be percentiles of $\widehat{L}^s(h)$ according to some α level.

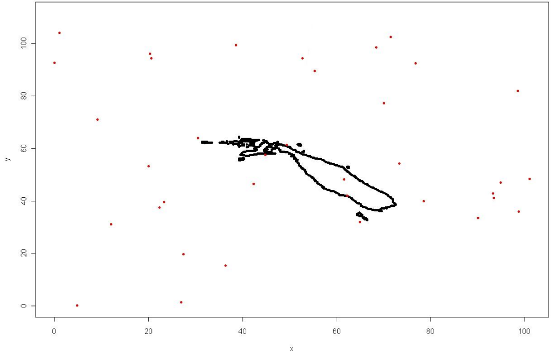
$$\mathcal{U}(h) = \widehat{L}^s_{1-\alpha/2}(h)$$

$$\mathcal{L}(h) = \widehat{L}^s_{\alpha/2}(h)$$



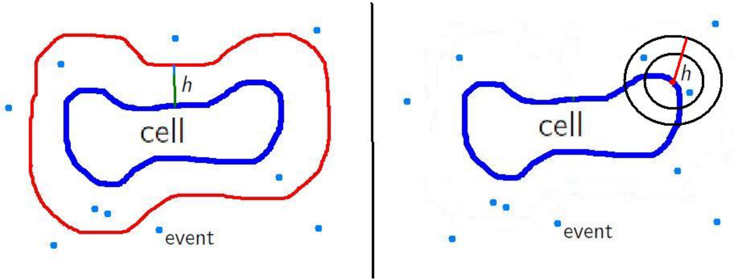
What are we interested in?

- Application has events (release points), and a volume (cell).
- K function considers event to event distances.
- Interested in how release points are arranged about cell boundary.



Modification of K function

- Consider a distance h from the cell boundary and count number of events.
- Visit each cell boundary point and consider circles with radius h and count number of events within each circle.
- Consider cell boundary as an edge.



References

- Bailey, Trevor C., and Anthony C. Gatrell. Interactive Spatial Data Analysis. Harlow: Longman, 1995. Print.
- Cressie, Noel A. C. Statistics for Spatial Data. New York: Wiley, 1993. Print.