Point Pattern Interevent Methods

Sergio Rey

Geographic Information Analysis
School of Geographical Sciences and Urban Planning
Arizona State University

Geographic Information Analysis by Sergio Rey
is licensed under a Creative Commons Attribution 4.0 International License.

Outline

Outline

Nearest Neighbor Distance Methods

Limitations

- Considers only the nearest neighbor distances
- Close scale of variation
- Distances to higher order neighbors ignored

Consider full distance distribution

- Distance between all pairs of event points
- Not just nearest neighbors
- How does the cumulative distance distribution behave

Outline

Ripley, 1976

K-function

- Provides and estimate of spatial dependence over a wider range of scales than the nearest neighbor approaches
- Based on all distances between event points in study area
- Assumes stationarity over the region

K function (Intuition)

$$K(d) = \frac{1}{\lambda} n_d \quad (1)$$

where n_d is the number of events within distance d of randomly chose event.
$\lambda = n/|A|$, our intensity

K-function

Estimation

- Construct a circle of radius *d* around each event point *i*
- Count the number of other events (*j*) that fall inside this circle
- Repeat steps 1 and 2 for all points *i* and sum the results
- Increment *d* by small amount and repeat steps 1-3

Interpreting the k-function

- Plot $\hat{K}(d)$ against different values of d
- What should this look like for no spatial dependence?
- Use CSR as our benchmark (again)
    1. The probability of an event at any point in A is independent of what other events have occurred in A
    2. Equally likely anywhere in A

Interpreting the k-function

- Under the assumption of *CSR*:

$$K(d) = \pi d^2 \tag{2}$$

- $K(d) < \pi d^2$ points to regularity
- $K(d) > \pi d^2$ points to clustering

Ripley's K function

K

$$\hat{K}(d) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \psi_{ij}(d)}{n\hat{\lambda}} \quad (3)$$

where:

$$\psi_{ij}(d) = \begin{cases} 1 & \text{if } d_{ij} \leq d \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Circle centered on each point s_i

$$\sum_{j=1}^{n} \psi_{ij}(d) \quad (5)$$

is the number of events within a circle of radius d centered on even s_i.

Ripley's K function

K

$$\hat{K}(d) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \psi_{ij}(d)}{n\hat{\lambda}} \quad (3)$$

where:

$$\psi_{ij}(d) = \begin{cases} 1 & if\ d_{ij} \leq d \\ 0 & otherwise \end{cases} \quad (4)$$

Circle centered on each point s_i

$$\sum_{j=1}^{n} \psi_{ij}(d) \quad (5)$$

is the number of events within a circle of radius d centered on even s_i.

Outline

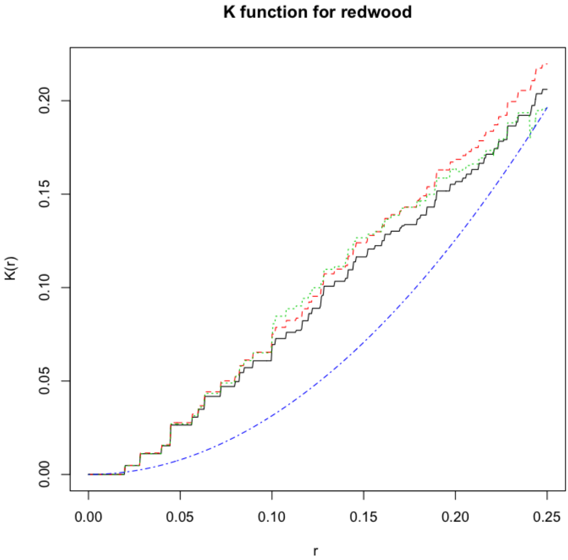
1. Interevent Distances
   - Motivation
   - Ripley's K function
2. Application of K-function
   - R code
   - L function
   - Simulation Envelopes
3. Extensions
   - Non-homogeneous base

R code: K Function

```
> rk=Kest(redwood)
> plot(rk,main="K function for redwood")
       lty col
iso      1   1
trans    2   2
border   3   3
theo     4   4
```

- *iso*: Ripley's isotropic correction.
- *trans*: Translation correction.
- *border*: reduced sample estimator.
- *theo*: the theoretical value of K

K function



K function for redwood
K(r)
0.00
0.05
0.10
0.15
0.20
0.00
0.05
0.10
0.15
0.20
0.25
r

Outline

L function

Scaling of K

$$L(d) = \sqrt{K(d)/\pi} - d \tag{6}$$

Useful since:

$$E[K(d)] = \frac{\pi \lambda d^2}{\lambda} \tag{7}$$

which can get large with d^2 and obscures small differences between expected and observed values.

L function

Scaling of K

$$L(d) = \sqrt{K(d)/\pi} - d \tag{6}$$

Useful since:

$$E[K(d)] = \frac{\pi \lambda d^2}{\lambda} \tag{7}$$

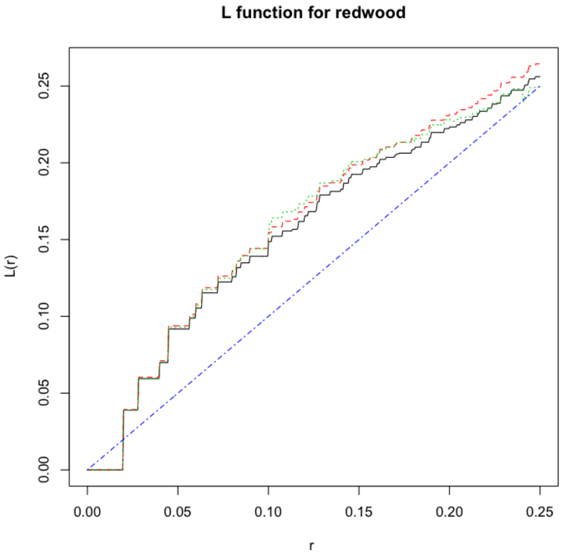
which can get large with d^2 and obscures small differences between expected and observed values.

R code: L Function

```
>  plot(rk, sqrt(./pi) ~ r, ylab="L(r)",
     + main="L function for redwood")
        lty col
iso       1   1
trans     2   2
border    3   3
theo      4   4
>
```

- *iso*: Ripley's isotropic correction.
- *trans*: Translation correction.
- *border*: reduced sample estimator.
- *theo*: the theoretical value of K

L function



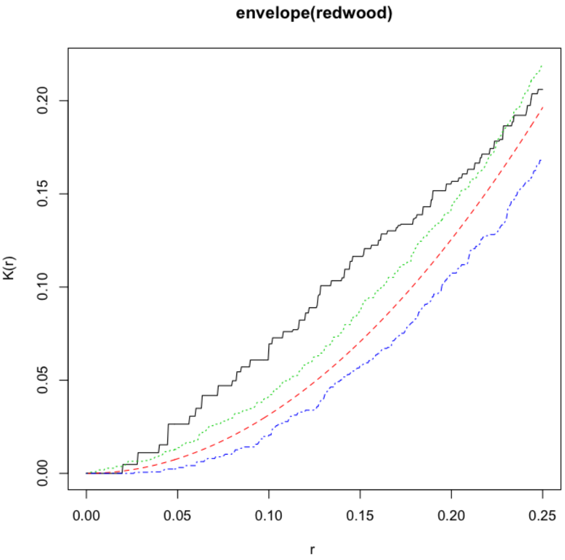
Outline

Simulation envelopes for K

```
> plot(envelope(redwood))
Generating 99 simulations of CSR ...
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,
16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30,
31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45,
46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60,
61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75,
76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90,
91, 92, 93, 94, 95, 96, 97, 98, 99.

Done.
     lty col
obs    1   1
theo   2   2
hi     3   3
lo     4   4
```

K function simulation

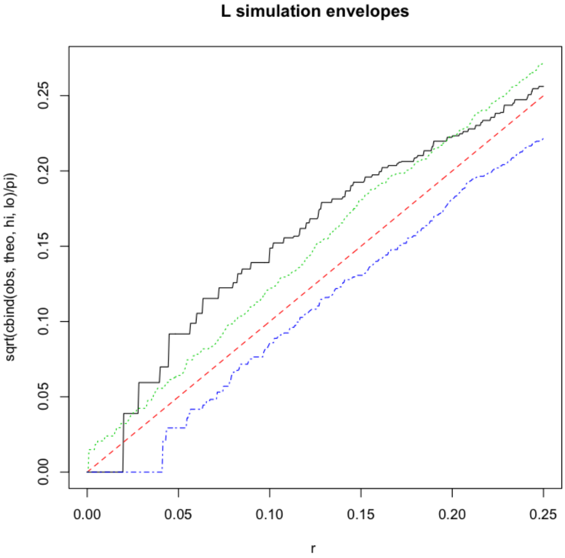


Simulation envelopes for L

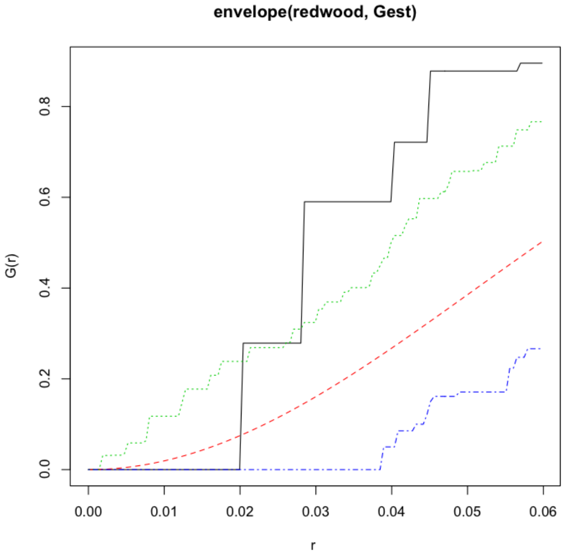
```
> E=envelope(redwood,Kest)
Generating 99 simulations of CSR ...
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,
16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30,
31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45,
46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60,
61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75,
76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90,
91, 92, 93, 94, 95, 96, 97, 98, 99.

Done.
> plot(E,sqrt(./pi)~r,main="L simulation envelopes")
     lty col
obs    1   1
theo   2   2
hi     3   3
lo     4   4
```

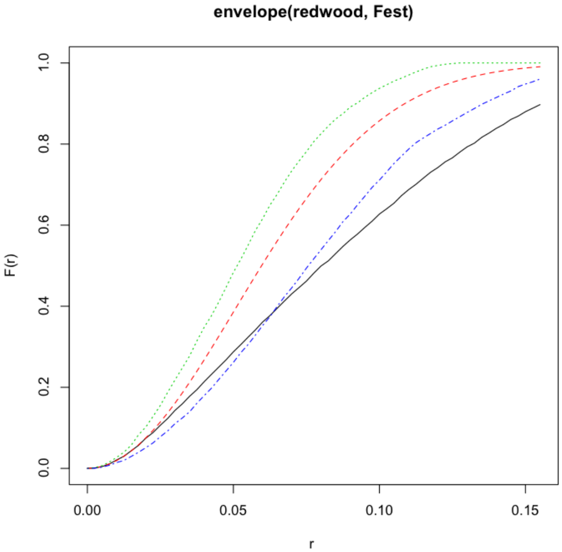
L function simulation



G function simulation



F function simulation



Outline

1. Interevent Distances
   - Motivation
   - Ripley's K function
2. Application of K-function
   - R code
   - L function
   - Simulation Envelopes
3. Extensions
   - Non-homogeneous base

Non-homogenous base

Variations in density of population at risk

- With constant risk
- More event points where base density is higher
- Clustering of events due to clustering of population

Alternative comparison

- Not CSR under the null
- Control for variation in the “control” intensity
- Are the events (“cases”) still clustered (or uniform) after the intensity of controls are taken into account?

Compare two k-functions

$$K_{i,j}(d) = \lambda_j^{-1} E[n_{i,j}(d)] \tag{8}$$

where j is the intensity of the controls, E is the expected value, and $n_{i,j}(d)$ is the number of further controls within distance d of an arbitrary event.

$$K_{1,1} = K_{2,2} = K_{1,2} \tag{9}$$

holds when the cases (events) are neither more or less clustered than the set of controls. This gives the motivation for the Diggle and Chetwynd (1991) statistic:

$$D(s) = K_{1,1} - K_{2,2} \tag{10}$$

Feser and Sweeney (2000)

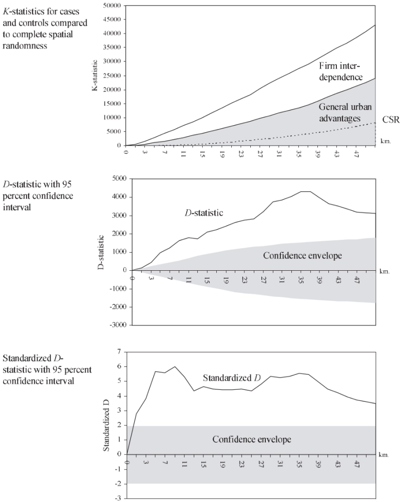


Fig. 1. Logic and interpretation of spatial clustering test