Point Pattern Analysis
Nearest Neighbor Statistics

Luc Anselin

http://spatial.uchicago.edu
principle

G function

F function

J function
Principle
Terminology

events and points

event: observed location of an event

point: reference point (e.g., point on a grid)

distances

event-to-event distance

point-to-event distance
• Nearest Neighbor Statistic

principle

under CSR the nearest neighbor distance between points has known mathematical properties

testing strategy = detect deviations from these properties
• Nearest Neighbor Statistic (2)

implementation

event to nearest event

point to nearest event

characterize this distribution relative to CSR

many nearest neighbor statistics
G function
• **Event-to-Event Distribution**

  cumulative distribution of nearest neighbor distances

  \[ G(r) = n^{-1} \#(r_i \leq r) \]

  proportion of nearest neighbor distances that are less than \( r \)

  plot estimated \( G(r) \) against \( r \)

  implementation: many types of edge corrections
G under CSR

nearest neighbor at distance $r$ implies that no other points are within a circle with radius $r$

$P[y=0]$ is $\exp(-\lambda \pi r^2)$ under Poisson distribution

the probability of finding a nearest neighbor is then the complement of this

$P[r_i < r] = 1 - \exp(-\lambda \pi r^2)$

reference function, plot $1 - \exp(-\lambda \pi r^2)$ against $r$
G function with reference curve for CSR
• Inference

analytical results intractable or only under unrealistic assumptions

mimic CSR by random simulation

random pattern for same \( n \)

compute \( G(r) \) for each random pattern

create a simulation envelope
G function with randomization envelope using min and max for each r
• Interpretation

clustering

G(r) function above randomization envelope

inhibition

G(r) function below randomization envelope
G for Poisson Clustered Process
G for Matern II Inhibition Process
F function
Point-to-Event Distribution

cumulative distribution of nearest neighbor distances from reference points to events

F(r) = m^{-1} \#(r_i \leq r)

proportion of nearest neighbor distances that are less than r

also referred to as empty space function

plot estimated F(r) against r

implementation: many types of edge corrections
F under CSR

same approach as for G function

under CSR

\[ P(r_i < r) = 1 - \exp(-\lambda \pi r^2) \]

reference function, plot \(1 - \exp(-\lambda \pi r^2)\) against \(r\)
F function with reference curve for CSR
• Inference

same logic as for G function

randomization envelope

interpretation is opposite from G

\[ F(r) \text{ below envelope implies clustering} \]

\[ F(r) \text{ above envelope implies inhibition} \]
F function with randomization envelope using min and max for each r
F for CSR
F for Poisson Clustered Process
F for Matern II Inhibition Process
J function
• J is Combination of G and F Functions

connect to models for spatial processes

specific form for different processes

Van Lieshout and Baddeley (1996)

• \[ J(r) = \frac{[1 - G(r)]}{[1 - F(r)]} \]

various edge corrections
• Inference and Interpretation

for CSR $J(r) = 1$

inference based on randomization envelope

$J(r) < 1$, or below envelope implies clustering

$J(r) > 1$, or above envelope implies inhibition
J function with reference line for CSR
J function with randomization envelope using min and max for each r
$J$ for CSR
J for Poisson Clustered Process
Point Pattern Analysis

Advanced Distance Statistics

Luc Anselin

http://spatial.uchicago.edu
principles

pair correlation function

K function

L function

Kd function
Principles
Beyond Nearest Neighbor Statistics

nearest neighbor distances do not fully capture the complexity of point processes

instead, take into account all the pair-wise distances

as a density function or as a cumulative density function
First Order Intensity Function

first order moment of a point process is the intensity $\lambda(x)$, similar to the notion of mean

the number of points over a given area $dx$ as the area gets infinitesimally small is $\lambda dx$

for a stationary process, the intensity is constant

$\lambda(x) = \lambda$
• **Second Order Intensity Function**

  similar to the notion of covariance

  with reference areas $dx$ and $dy$ becoming infinitely small, it is the expected cross product of the number of points in $dx$ and $dy$ over the product of the (very small) areas

  $$\lambda_2(x,y) = \frac{[E(N(dx)N(dy))]}{|dx||dy|} \text{ as } |dx|,|dy| \to 0$$

  for a stationary and isotropic process $\lambda_2(x,y)$ only depends on the distance between $x$ and $y$
Second Order Statistics

second order statistics exploit the notion of covariance

based on the number of other points within a given radius of a point

pair correlation function, or g-function

Ripley’s K and Besag’s L function
Pair Correlation Function
• g Function

  correlation-like function

• $g = \frac{\lambda_2(x,y)}{\lambda(x) \cdot \lambda(y)}$

  complex estimation procedure
• Inference and Interpretation

for CSR $g = 1$

g(r) > 1$ implies a cluster process

$g(r)$ can take large values, especially for small $r$, decreases as $r$ increases

$g(r) < 1$ implies a regular process (inhibition)

$g(r) = 0$ for $r_i < r$ implies a hard core (no point pairs within this distance)

randomization envelope for inference
g function with reference line for CSR
Chicago Supermarkets - g Function Envelope

$g(r)$

$g_{obs}(r)$
$\bar{g}(r)$
$\hat{g}_{hi}(r)$
$\hat{g}_{lo}(r)$

$g$ function with randomization envelope using min and max for each $r$
K function
• Ripley’s K Function

best known second order statistic

so-called reduced second order moment

\[ \lambda K(r) = E[N_0(r)] \]

\( E[N_0(r)] \) is the expected number of events within a distance \( r \) from an arbitrary event

\[ K(r) = \lambda^{-1} E[N_0(r)] \] is the K function
Estimating the K Function

expected events within distance $r$

$$E[N_0(r)] = n^{-1} \sum_i \sum_{j \neq i} 1_{h}(r_{ij} < r)$$

for each event, sum over all other events within the given distance band, for increasing distances

cumulative function

dge corrections
• Inference and Interpretation

for CSR, $K(r) = \pi r^2$

$K(r) > \pi r^2$ implies clustering

$K(r) < \pi r^2$ implies inhibition (regular process)

use randomization envelope for inference
K function with reference line for CSR
K function with randomization envelope using min and max for each r
K for CSR
K for Poisson Cluster Process
K for Matern II Inhibition Process
• Cross-K Function

extension to bivariate (multi-type) point patterns

count events of type j within a distance band r from every event i

cumulative function
• Interpretation and Inference

absence of correlation

\[ K_{11} = K_{22} = K_{12} = \pi r^2 \ (CSR) \]

clustering of two patterns

\[ K_{12} > CSR \]

inhibition between two patterns

\[ K_{12} < CSR \]

inference based on randomization envelope

*GeoDa*
Cross K function with randomization envelope using min and max for each r
L function
• Besag’s L Function

variance of $K(r)$ increases with $r$

variance-stabilizing transformation

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

plot $L(r)$ against $r$ (diagonal)

plot $L(r) - r$ against $r$
• Interpretation and Inference

$L(r) > r$ implies clustering

$L(r) < r$ implies inhibition (regular process)

inference based on randomization envelope
L function with reference line for CSR
L - r function with randomization envelope using min and max for each r
Kd function
• Density instead of Cumulative Distances

  Kd function of Duranton and Overman (2005)

  density function of all pairwise distances

  computed using a kernel smoothing procedure
Inference and Interpretation

randomization envelope

local and global

reference distribution other than CSR

density above envelope suggests clustering

density below envelope suggests inhibition
Kd density function
Kd density function with randomization envelope