Multivariate Point Patterns

Bailey and Gatrell

Chapter 4

SIE 512 Lecture 11

October 6, 2009

1

Analysis of Multiple Types of Events

A multivariate spatial point process is any stochastic mechanism that generates events classified as type *j* for *j= 1,2,…,k.*

The *k* univariate point patterns are referred to as the components of the multivariate process

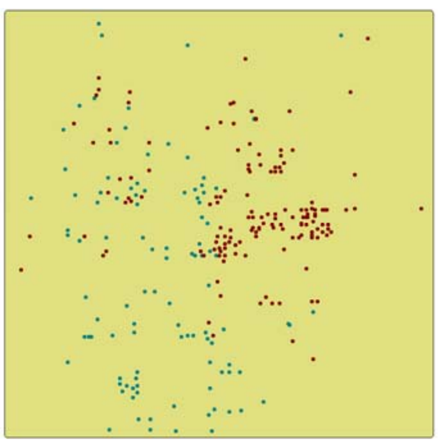
Analysis Objectives

- Detection of relationships in the pattern of one type of event relative to another type of event
- Identification of independence among types of events or patterns as opposed to attraction or repulsion

2

Is the distribution of one set of events related to the distribution of the other?

Black (red dots) and white (blue dots) crimes in Oklahoma



3

Independence

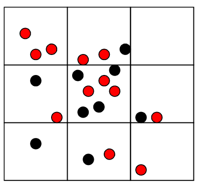
- Independence is analogous to CSR for univariate point patterns
- Independence forms a dividing hypothesis to determine positive or negative dependence among the event types
- Independence implies that the overall pattern in the events is made up from independent component processes, one for each type of event
- Independence does not imply that any of the component processes need be CSR

4

Quadrat Count Analysis

Count presence of events of each type in quadrats:

Test using χ_1^2 test of independence



Type 1 ●
Type 2 ●

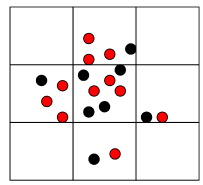
		Type 1 absent	Type 1 present	
Type 2	absent	1	1	2
	present	2	5	7
		3	6	

$$X^2 = \frac{(c_{11}c_{22} - c_{12}c_{21})^2 \sum_i \sum_j c_{ij}}{\sum_j c_{1j} \sum_j c_{2j} \sum_i c_{i1} \sum_i c_{i2}}$$

$$X^2 = \frac{(1*5-2*1)^2 9}{2*7*3*6} = .3214 \qquad \text{Compare to: } \chi_1^2 = 3.84$$

5

Quadrat Count Analysis



Type 1 ●
Type 2 ●

		Type 1 absent	Type 1 present	
Type 2	absent	4	0	4
	present	0	5	5
		4	5	

$$X^2 = \frac{(c_{11}c_{22} - c_{12}c_{21})^2 \sum_i \sum_j c_{ij}}{\sum_j c_{1j} \sum_j c_{2j} \sum_i c_{i1} \sum_i c_{i2}}$$

$$X^2 = \frac{(4*5-0*0)^2 *9}{4*5*4*5} = 9.0 \qquad \text{Compare to: } \chi_1^2 = 3.84$$

6

Nearest Neighbor Tests for Multivariate Patterns

$G_{ij}(h)$ The probability that the distance from a randomly chosen type i event to the nearest event of type j is less than or equal to h

$F_j(h)$ The probability that the distance from a randomly chosen point to the nearest event of type j is less than or equal to h

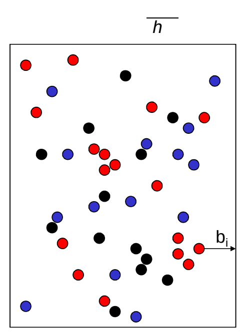
For an independent pattern

$$G_{ij}(h) = F_j(h) \qquad i \neq j$$

The distribution of nearest neighbor distances to events of type *j* from a randomly selected point or event type *i* should be the same

7

Nearest Neighbor Tests for Multivariate Patterns



● Type i event
● Type j event
● Random point

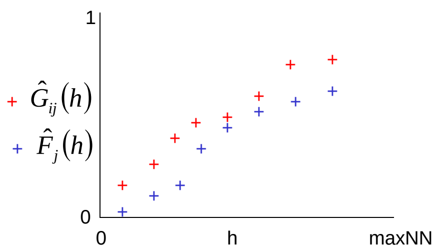
$$\hat{G}_{ij}(h) = \frac{\#(b_i > h \geq w_{ij})}{\#(b_i > h)}$$

$$\hat{F}_j(h) = \frac{\#(b_j > h \geq x_j)}{\#(b_j > h)}$$

b_i is the distance to the nearest point on the border

8

Nearest Neighbor Tests for Multivariate Patterns



Plot on same graph with values of h from 0 to max nearest neighbor distance

Bivariate or Cross K- Function

$K_{ij}(h), i \neq j$ is cross K function

$$\lambda_j K_{ij}(h) = E(\#(\text{type j events} \leq h \text{ from an arbitrary type } i \text{ event}))$$

λ_j Intensity of type j events

Under assumed independence, type j events should be random with respect to type i events

Then the expected number of type j events within a distance h of a randomly chosen type i event is $\lambda_j \pi h^2$

Theoretically under independence $K_{ij}(h) = \pi h^2$

Estimation of Bivariate K Function

n_1 is number of type i events and n_2 is number of type j events

d_{ij} is distance between ith and jth event in R

$I_h(d_{ij})$ is an indicator function – $1 \; if \; d_{ij} \leq h, \quad 0 \; otherwise$

The observed number of ordered pairs is $\Sigma\Sigma \, I_h(d_{ij})$

Estimate for K_{12} is

$$\hat{K}_{12}(h) = \frac{R}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \frac{I_h(d_{ij})}{w_{ij}}$$

edge correction

The Bivariate K Function

Theoretically $K_{12}(h) = K_{21}(h)$

But the corresponding estimated values will not necessarily be equivalent

So estimate is

$$\hat{K}_{12}(h) = \frac{\left(n_2 \tilde{K}_{12}(h) + n_1 \tilde{K}_{21}(h)\right)}{n_1 + n_2}$$

The Bivariate K Function

As for univariate case compute $\hat{L}_{ij}(h)$

$$\hat{L}_{ij}(h) = \sqrt{\frac{\hat{K}_{ij}(h)}{\pi}} - h$$

$\hat{L}_{11}(h)$
$\hat{L}_{22}(h)$
$\hat{L}_{12}(h)$

Plotted on the same axis, these values show the tendency of individual components to deviate from CSR as well as attraction or repulsion between the component patterns

13

Testing the Significance of Cross K Function

Separate component event patterns need to be preserved in their observed form under any simulations

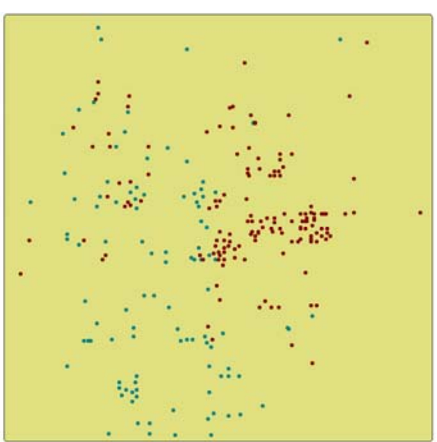
Assume a torus and compute ***m*** random shifts of one pattern against the other

The envelop created from these simulations can be used to test significance of peaks or troughs in $\hat{K}_{ij}(h)$ or $\hat{L}_{ij}(h)$

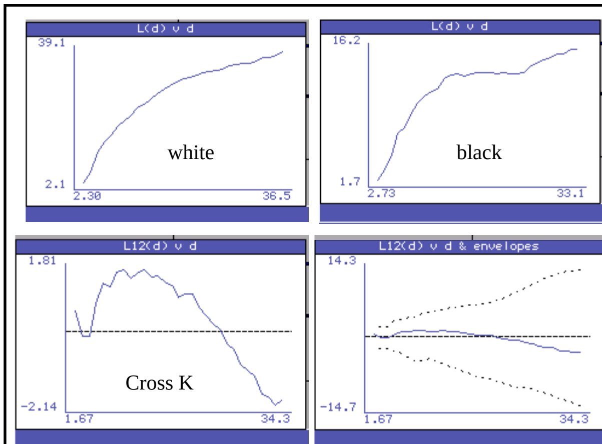
$$\Pr(\hat{L}(h) > U(h) = \Pr(\hat{L}(h) < L(h)) = \frac{1}{(m+1)}$$

14

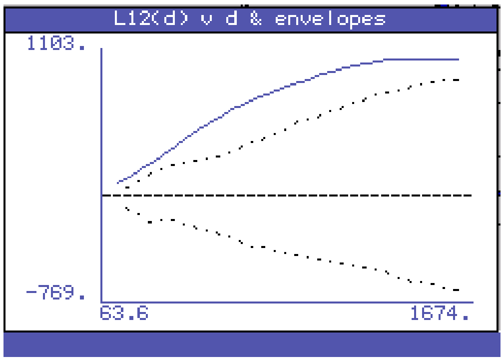
Black (red dots) and white (blue dots) crimes in Oklahoma



15



16



17

Space-Time Interaction

Assume time has been recorded for the set of events

Objective:

Determine if events are clustered in space as well as time

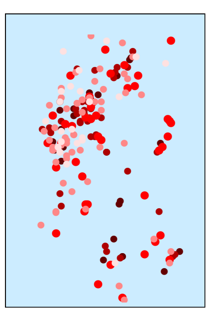
18

Space-Time Interaction

Given events that are close in time, are there events in space closer than would be expected by chance

- 413 - 1714
- 1715 - 2888
- 2889 - 3985
- 3986 - 5002
- 5003 - 5775

Onset of disease in days



19

Space-Time Interaction

- Complete spatio-temporal randomness (CSTR) is the absence of structure in time as well as space.
- CSTR is the natural null hypothesis against which space time point patterns can be tested.
- Space-time clustering is the alternative to CSTR
- Space time clustering is said to exist if, among those events that are close in time, there are events that are closer in space than would be expected due to chance alone.

20

Knox Test for space time interaction

Quantifies space-time interaction based on critical space and time distances

Uses the n(n-1) ordered pairs from all n events

n_s = pairs close in space

n_t = pairs close in time

A test statistic X is a count of pairs of events that are separated by less than critical space and time distances

“Close” is defined by previous knowledge or expertise

Pairs of events will be near to one another when interaction is present, and the statistic will be large

Knox Test for space time interaction

d is critical space distance

t is critical time distance

$$X = \sum_{i=1}^{N} \sum_{j=1}^{i-1} s_{ij} t_{ij}$$

s_{ij} spatial proximity, 1 if distance between events i and j is less than d, and 0 otherwise

t_{ij} temporal proximity, 1 if time separation between events i and j is less than t, and 0 otherwise

Knox Test for space time interaction

	Close in Space	Not Close in Space	Total
Close in Time	n_{st}	$n_t - n_{st}$	n_t
Not Close in Time	$n_s - n_{st}$	$n(n-1)/2 - n_s - n_t + n_{st}$	$n(n-1)/2 - n_t$
Total	n_s	$n(n-1)/2 - n_s$	$n(n-1)/2$

$X = n_{st}$

Under independence, ***X*** is Poisson distributed with

Expected value = $\dfrac{n_s n_t}{n(n-1)/2} = \dfrac{2 n_s n_t}{n(n-1)}$

Knox Test for space time interaction

30 events

		space		
		close	**not close**	total
time	**close**	163	77	240
	not close	93	102	195
	total	256	179	435

Compare to one-sided critical value

$$Z = \frac{(163-141.24)}{\sqrt{141.24}} = 1.84$$

X

$$\frac{n_s n_t}{n(n-1)/2}$$

(256*240)/(30*29/2)=141.24

If X is unusually large in comparison with theoretical distribution then reject null hypothesis of independence in favor of attraction

Knox Test for space time interaction

The null distribution of X can be constructed under an approximate randomization.

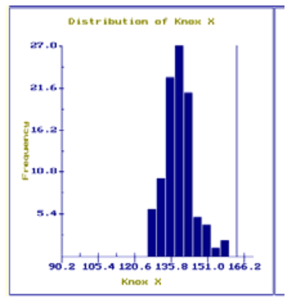
Permute time values while holding the locations constant.

X is calculated for each simulation of randomized data.

Let NGE be the number of X values under simulation that are larger than X obtained from the original (not randomized) data. M is the number of simulations. The P-value is

$$P = \frac{NGE + 1}{M + 1}$$

The frequency distribution of X under the null hypothesis of no association between the space and time adjacencies.



The test statistic is the vertical line on the distribution.

When space-time interaction is present this line will be to the right tail of the distribution.

Mantel Test for space time interaction

The Mantel test avoids problem of determining critical distances for space and time

Test statistic is the sum across all pairs of the time distance multiplied by the spatial distance

$$Z = \sum_{i=1}^{N} \sum_{j=1}^{N} s_{ij} t_{ij}$$

N = number of events

t_{ij} = distances between events i and j in time

s_{ij} = distances between events i and j in space

$\bar{s}, \bar{t}$ average space and time distances

s_s and s_t are standard deviations of the space and time distances.

Mantel Test for space time interaction

Mantel test statistic $Z = \sum_{i=1}^{N} \sum_{j=1}^{N} s_{ij} t_{ij}$

The test statistic is the sum, across all case pairs, of the time distance multiplied by the spatial distance.

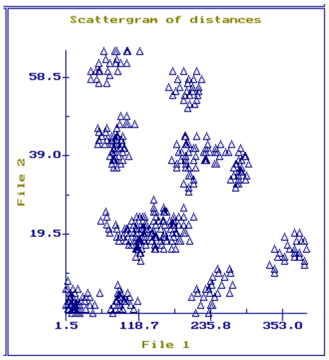
Standardized Mantel statistic

$$r = \frac{1}{N^2 - N - 1} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{(s_{ij} - \bar{s})}{s_s} \frac{(t_{ij} - \bar{t})}{s_t}$$

r is a measure of correlation with range from -1 to 1.

H_0: spatial distance between events is independent of the time distance between those events

Scattergram of the time distances on the space distances.



Points will form a line with positive slope when there is perfect correlation between the space and time distances.

Association between time and space distances appear as patterns in the scattergram.

29

Mantel Test for space time interaction

Significance can be tested using a Monte Carlo method: randomly permute the time elements while holding locations constant.

Equivalent to repeatedly scrambling the time observations across the localities, and calculating *Z* each time.

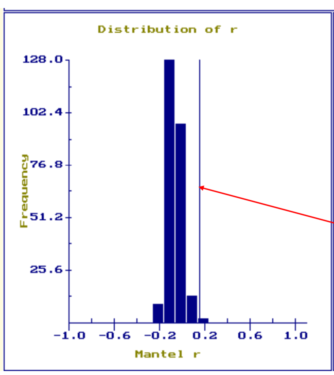
This is done *m* times to generate a distribution of *Z* under the null hypothesis.

A P-value is obtained by comparing the test statistic to this null distribution.

Let *NGE* be the number of times *Z* under simulation was greater than or equal to *Z* (the observed test statistic) obtained from the original data. The P-value is then $P = \frac{NGE+1}{M+1}$

30

Frequency distribution of *r* under the null hypothesis.



Mantel's cross-product is calculated for every randomization, and a distribution of the Mantel statistic results

The standardized test statistic, *r*, is shown by the vertical line.

31

Mantel Test for space time interaction

Both *r* and *Z* become large when the time distances are linearly dependent on the space distances. A non-linear dependence of time on space therefore will not necessarily result in a significant Mantel test.

Mantel recommended the use of a reciprocal transformation to reduce the effect of large space time distances

$$\frac{1}{d+C} \qquad s_{ij} = \frac{1}{d_{ij}+C}$$

C is a constant and *d* is the distance to be transformed

C avoids zero in cases where space time pair values might be identical

selection of the transformation is somewhat subjective.

32

Issues with Knox and Mantel tests

how to determine critical time and distance values?

distance may not be constant as underlying population densities may vary

if one examines a region in which an unexpectedly large number of cases have been reported, testing for space-time clustering may show no significant spatio-temporal effects, which is actually the result of region pre-selection rather than failure to detect a spatio-temporal effect;

changes in the underlying population distribution in the study region over time may have occurred, which will affect the results.

33

K Function for space time interaction

K(h,t) expected number of events within distance *h* of an arbitrary event and time interval *t* of an arbitrary event – scaled by expected number of events per unit area and per unit time

If independent then the estimated *K(h,t)* should be the same as the product of the space and time functions computed independently

$$\hat{D}(h,t) = \hat{K}(h,t) - \hat{K}_S(h)\hat{K}_T(t)$$

If there is space-time interaction *D(h,t)* will be large

34

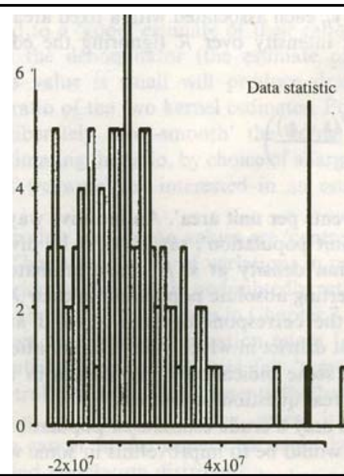
Tests of significance

Perform m simulations in which the n events are randomly labeled with the n time stamps

Generates m estimates of $\hat{D}_i(h,t),\, i = 1,\cdots,m$

An extreme value indicates space time interaction

35



36