Global Spatial Autocorrelation

SERGIO REY

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

Outline

1. Spatial Autocorrelation Tests: Logic
   - Attribute and Geographical Similarity
   - Gamma Statistics
2. Join Count Statistics
   - Spatial Joins
   - Inference
3. Global Autocorrelation for Continuous Attributes
   - Moran's I
   - Inference

Outline

Autocorrelation Tests: Two Types of Similarity

Geographical Similarity

- How geographically similar are polygons i and j?
- Are they geographical neighbors?
- $w_{i,j}$

Attribute Similarity

- How similar are the attribute values y_i and y_j?

Global Spatial Autocorrelation

What is the nature of these two types of similarity *for all* pairs of neighboring locations?

Standardization

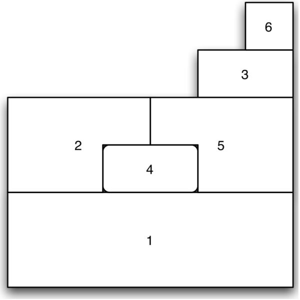
Row Standardization of W

$$w^*_{i,j} = \frac{w_{i,j}}{\sum_j w_{i,j}} \quad (1)$$

Properties

1. $0 \leq w^*_{i,j} \leq 1$
2. $\sum_j w^*_{i,j} = 1 \ \forall i$
3. $\sum_i \sum_j w^*_{i,j} = n$

Example of Row Standardization



Map

$$W = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$W^* = \begin{bmatrix} 0 & 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Spatial Lag

$$yl_i = \sum_j w^*_{i,j} y_j$$

i	y_i	yl_i
1	20	?
2	10	?
3	40	?
4	22	?
5	30	?
6	50	?

Spatial Lag

i	y_i	yl_i
1	20	20.67
2	10	24.00
3	40	40.00
4	22	20.00
5	30	23.00
6	50	40.00

$$W^* = \begin{bmatrix} 0 & 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$\begin{aligned} yl_2 &= \sum_{j=1} w^*_{2,j} y_j \\ &= w^*_{2,1}y_1 + w^*_{2,2}y_2 + w^*_{2,3}y_3 + w^*_{2,4}y_4 + w^*_{2,5}y_5 + w^*_{2,6}y_6 \\ &= 1/3(20) + 0(10) + 0(40) + 1/3(22) + 1/3(30) + 0(50) \\ &= 1/3(20 + 22 + 30) \\ &= 24.00 \end{aligned}$$

Outline

Gamma Statistic

Combines both forms of similarity

$$\Gamma = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{ij}$$

- $a_{i,j}$ is the measure of attribute similarity for i and j
- $b_{i,j}$ is the spatial similarity measure: $b_{i,j} = w_{i,j}$

Γ nests most global spatial autocorrelation tests as special cases.

Γ Inference

Hypotheses

- Two-sided
  - Observed is significantly different from expected
- One-sided
  - Observed is significantly greater than expected
  - Observed is significantly lower than expected

Sampling Distribution Under H_0

- Analytical Distribution
- Permutation Based Distribution

Permutation Based Distribution

$$Prob[\Gamma \geq \Gamma_{obs}|H_0] = \frac{\Phi + 1}{N + 1} \tag{2}$$

where $\Phi = \sum_{r=1}^{N} \Phi_r$ and

$$\Phi_r = \begin{cases} 1 & if\ \Gamma_r \geq \Gamma_{obs} \\ 0 & otherwise. \end{cases} \tag{3}$$

For a two-tailed, or non-directional, alternative hypothesis the probability in (2) would have to be multiplied by 2 to obtain the correct pseudo significance level.

Outline

Joins

Nominal Attributes

- Binary
  - Black (B): $X_i = 1$
  - White (W): $X_i = 0$
- Probabilities
  - $P(X_i = 1) = p_B = P(B)$
  - $P(X_i = 0) = (1 - p_B) = p_w = P(W)$

Join

- When two polygons share a border (edge, vertex)

Join Types

BB

$$BB = \frac{1}{2}\sum_i \sum_j w_{i,j} x_i x_j \tag{4}$$

WW

$$WW = \frac{1}{2}\sum_i \sum_j w_{i,j}(1 - x_i)(1 - x_j) \tag{5}$$

BW

$$BW = S_J - BB - WW \tag{6}$$

$$S_J = \frac{1}{2}\sum_i \sum_j w_{i,j} \tag{7}$$

Join Count Statistics: Analytical Inference

Expected Values Under H_0

$$E[BB] = S_j p_B^2 \tag{8}$$

$$E[WW] = S_j p_W^2 \tag{9}$$

$$E[BW] = S_j p_B p_W \tag{10}$$

Variances Under H_0

$$\begin{aligned} V[BB] &= S_J p_B^2 + 2mp_B^3 - (S_J + 2m)p_B^4 && (11)\\ V[WW] &= S_J p_W^2 + 2mp_W^3 - (S_J + 2m)p_W^4 && (12)\\ V[BW] &= 2(S_J + m)p_B p_W - 4(S_J + 2m)p_B^2 p_W^2. && (13) \end{aligned}$$

where $m = \sum_{i=1}^{n} k_i(k_i - 1)$ and $k_i = \sum_{j=1}^{n} w_{i,j}$.

Join Count Statistics: Analytical Inference - z-transform

$$z_{BB} = \frac{BB - E[BB]}{\sqrt{V[BB]}} \tag{14}$$

$$z_{WW} = \frac{WW - E[WW]}{\sqrt{V[WW]}} \tag{15}$$

$$z_{BW} = \frac{BW - E[BW]}{\sqrt{V[BW]}} \tag{16}$$

Example Data

39	41	38	45
39	42	41	42
48	49	48	51
47	51	50	55

$n = 16$

W

$$\mathbf{W} = \begin{bmatrix}
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0
\end{bmatrix}.$$

Join Counts: Calculations

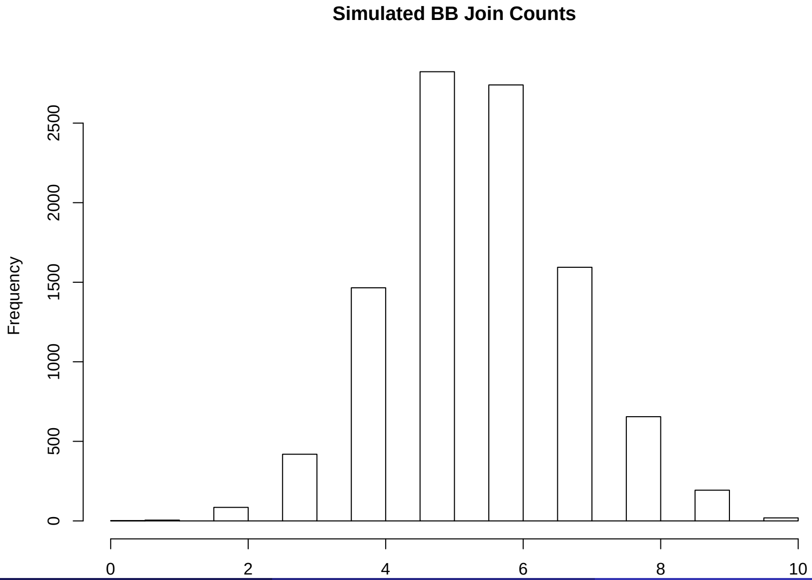
i	y_i	x_i	$\sum_j w_{i,j} x_i x_j$	$\sum_j w_{i,j}(1-x_i)(1-x_j)$
1	39	0	0	2
2	41	0	0	3
3	38	0	0	3
4	45	0	0	2
5	39	0	0	2
6	42	0	0	3
7	41	0	0	3
8	42	0	0	2
9	48	1	2	0
10	49	1	3	0
11	48	1	3	0
12	51	1	2	0
13	47	1	2	0
14	51	1	3	0
15	50	1	3	0
16	55	1	2	0
$\sum_i$	726	8	20	20
			S_J	24
			m	52
			p_B	0.5
			p_W	0.5

Outline

Joint Count Statistics: Inference

Join	Count (J)	$E[J]$	$\sqrt{V[J]}$	z
BB	10	6	3.316625	1.206045
WW	10	6	3.316625	1.206045
BW	4	12	3.316625	-2.412091

Join Counts: Permutation Based Inference



Join Counts: Permutation Based Inference

Join	Count (J)	$\bar{J}$	s_J	Norm. $z(j)$	Norm. $p(J)$	Perm. $p(J)$
BB	10	5.6172	1.371233	3.196254	0.0007	0.0007
WW	10	5.6056	1.361556	3.227484	0.0006	0.0007
BW	4	12.7772	2.325446	-3.774415	0.0001	0.0006

Table: Join Counts Inference Using Random Spatial Permutations

Outline

Moran's I

$$I = \left(\frac{n}{S_0}\right) \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} z_i z_j}{\sum_{i=1}^{n} z_i^2} \tag{17}$$

where $z_i = y_i - \bar{y}$ and

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \tag{18}$$

with w_{ij} an element of a spatial weights matrix.

Outline

Expected Value under H_0

$$E(I) = \frac{-1}{n-1}. \tag{19}$$

Variance under H_0 assuming normality

$$V[I] = E[I^2] - E[I]^2 \tag{20}$$

$$E[I_N^2] = n^2 S_1 - nS_2 + 3S_0^2 / \left[S_0^2 \left(n^2 - 1 \right) \right] \tag{21}$$

with

$$S_1 = (1/2) \sum_{i=1}^{n} \sum_{j=1}^{n} \left(w_{i,j} + w_{j,i} \right)^2 \tag{22}$$

and

$$S_2 = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} w_{i,j} + \sum_{j=1}^{n} w_{j,i} \right)^2 . \tag{23}$$

Substituting (18), (22), and (23) into (21), and then substituting the latter into (20) results in

$$V(I_N) = \frac{n^2 S_1 - nS_2 + 3S_0^2}{(n-1)(n+1)S_0^2} - (1/(n-1))^2 . \tag{24}$$

Variance under H_0 assuming randomization

$$V[I] = E[I^2] - E[I]^2 \tag{25}$$

$$E[I_R^2] = \frac{n\left[\left(n^2 - 3n + 3\right)S_1 - nS_2 + 3S_0^2\right] - b_2\left[\left(n^2 - n\right)S_1 - 2nS_2 + 6S\right.}{(n-1)(n-2)(n-3)S_0^2} \tag{26}$$

with

$$b_2 = \frac{\sum_{i=1}^{n} z_i^4}{\left(\sum_{i=1}^{n} z_i^2\right)^2}. \tag{27}$$

The variance becomes:

$$V(I_R) = E[I_R^2] - \left(1/(n-1)\right)^2. \tag{28}$$

Now distribution of I is sensitive to both properties of Y and W.

Moran's I Example

i	y_i	$z_i = y_i - \bar{y}$	z_i^2	z_i^4	$\sum_j w_{ij} z_i z_j$
1	39	-6.375	40.640625	1.651660e+03	68.531250
2	41	-4.375	19.140625	3.663635e+02	74.921875
3	38	-7.375	54.390625	2.958340e+03	67.296875
4	45	-0.375	0.140625	1.977539e-02	4.031250
5	39	-6.375	40.640625	1.651660e+03	45.421875
6	42	-3.375	11.390625	1.297463e+02	38.812500
7	41	-4.375	19.140625	3.663635e+02	50.312500
8	42	-3.375	11.390625	1.297463e+02	-2.953125
9	48	2.625	6.890625	4.748071e+01	-2.953125
10	49	3.625	13.140625	1.726760e+02	27.187500
11	48	2.625	6.890625	4.748071e+01	24.937500
12	51	5.625	31.640625	1.001129e+03	49.921875
13	47	1.625	2.640625	6.972900e+00	13.406250
14	51	5.625	31.640625	1.001129e+03	55.546875
15	50	4.625	21.390625	4.575588e+02	82.671875
16	55	9.625	92.640625	8.582285e+03	98.656250
$\sum_i$	726	0	403.75	18570.61	695.75
				$\bar{y}$	45.375
				$\sum_i \sum_j w_{i,j}$	48

$$\sum_j w_{5j} z_5 z_j = z_5(z_1 + z_6 + z_9) = -6.375(-6.375 - 3.375 + 2.625) = 45.421875. \quad (29)$$

Morans'I Example

Component	Equation	Value
n		16
S_0	(18)	48
S_1	(22)	96
S_2	(23)	608
$E(I)$	(19)	-0.06667
$V(I_N)$	(24)	0.0326
b_2	(27)	0.1139203
$E(I_R^2)$	(26)	0.04404776
$V(I_R)$	(28)	0.0396

Table: Intermediate Results for Moran's I

Moran's I Example

$$E(I) = \frac{-1}{16-1} = -0.06667, \tag{30}$$

and

$$V(I_N) = \frac{16^2(96) - 16(608) + 3(48)^2}{(16-1)(16+1)48^2} - (1/(16-1))^2 = 0.03259. \tag{31}$$

$$z(I_N) = \frac{I - E(I)}{\sqrt{V(I_N)}} = \frac{0.5744066 + 0.0667}{\sqrt{0.03259}} = 3.568112, \tag{32}$$

$$z(I_R) = \frac{I - E(I)}{\sqrt{V(I_R)}} = \frac{0.5744066 + 0.0667}{\sqrt{0.03960331}} = 3.221380. \tag{33}$$