

全局空间自相关 Moran 指数和 G 系数对比研究^{*}

张松林¹, 张 昆²

(1. 同济大学 测量与国土信息系, 上海 200092)

(2. 华东师范大学 地理信息科学教育部重点实验室, 上海 200062)

摘 要: 研究了最常用的衡量全局空间自相关的指标: 全局 Moran 指数和全局 G 系数, 基于模拟的空间区域, 设计了一些有代表性的空间聚集方案进行计算, 比较了两种指标的探测结果。得出的初步结论是: G 系数探测高值聚集的能力强于低值聚集; 当研究范围内同时存在高值和低值聚集时, G 系数会受聚集区域规模的影响, 当高值聚集区域和低值聚集区域规模相当时, G 系数往往为正数, 表明 G 系数对高值敏感; Moran 指数主要受聚集区域规模的影响, 随着空间聚集范围的扩展, Moran 指数会明显增大。这些结论对于探测空间自相关模式时选择何种全局指标有指导意义。

关键词: 空间自相关; 全局指标; Moran 指数; G 系数

中图分类号: O212.1 **文献标识码:** A **文章编号:** 0529-6579 (2007) 04-0093-05

空间自相关是指同一个变量在不同空间位置上的相关性, 是空间单元属性值聚集程度的一种度量^[1-2]。空间自相关性使用全局和局部两种指标, 全局指标用于探测整个研究区域的空间模式, 使用单一的值来反映该区域的自相关程度; 局部指标计算每一个空间单元与邻近单元就某一属性的相关程度。

衡量空间自相关的全局指标有全局 Moran 指数、全局 G 系数和 Geary C 系数, 此外还有 Cliff Ord statistic Join-count statistic 等国内统计界不常用的指标^[1-3]。目前应用了全局空间自相关技术的文献, 绝大多数计算的是 Moran 指数或 G 系数。吕安民等采用全局 Moran 指数分析了中国省级人口增长率的空間自相关性^[4]; 颜锋华等^[5]采用全局 G 统计量研究了遥感特征参量的空间分布, Trisajin 等^[6]基于遥感影像采用全局 Moran 指数对成树和幼树进行了精确分类。那么 Moran 指数和 G 系数在进行全局空间自相关分析时是否是等效的呢? 本文对此进行了研究。

1 全局 Moran 指数和全局 G 系数

全局 Moran 指数是由 Moran 于 1948 年提出的, 它反映的是空间邻接或邻近的区域单元属性值的相似程度, 其计算公式为^[3]:

$$I = \left[\frac{n \sum_{i=1}^n \sum_{j=1}^n C_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n \sum_{j=1}^n C_{ij}} \right] /$$

$$\left[\sum_{i=1}^n \sum_{j=1}^n C_{ij} (x_i - \bar{x})^2 \right] \quad (1)$$

其中, x_i 为区域 i 的属性值, C 为空间权重矩阵, C_{ij} 代表空间单元 i 和 j 之间的影响程度, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ 。

Moran 指数的取值范围近似为 $-1 \sim +1$ 之间, 越接近 -1 则代表单元间的差异越大或分布越不集中, 越接近 1 则代表单元间的关系越密切, 性质越相似 (高值聚集或低值聚集), 接近 0 则代表单元间不相关。对于全局 Moran 指数, 可以用标准化统计量 $Z(I)$ 来检验空间自相关的显著性水平。 $Z(I)$ 的计算公式为: $Z(I) = (I - E(I)) / \sqrt{\text{Var}(I)}$, $\text{Var}(I)$ 是 Moran 指数的理论方差, $E(I) = -1/(n-1)$ 为其理论期望。

由于 Moran 指数不能判断空间数据是高值聚集还是低值聚集, Getis 和 Ord 于 1992 提出了全局 G 系数。G 系数一般采用距离权, 要求空间单元的属性值为正, 其计算公式为^[3]:

$$G(d) = \left[\frac{\sum_{i=1}^n \sum_{j=1}^n W_{ij}(d) x_i x_j}{\sum_{i=1}^n x_i} \right] \left\{ \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2} \right\} \quad (4)$$

d 为距离, $W_{ij}(d)$ 是 i 点之间的距离权。类似于全局 Moran 指数, 也可对 G 系数进行标准化: $Z(G) = (G(d) - E(G(d))) / \sqrt{\text{Var}(G(d))}$, 正的 $Z(G)$ 表示存在高值聚集, 负的 $Z(G)$ 表示存在低值聚集。

* 收稿日期: 2006-10-23

基金项目: 国家自然科学基金资助项目 (40604001); 现代工程测量国家测绘局重点实验室开放基金 (ESSBM0611)

作者简介: 张松林 (1972 年生) 女, 博士, 讲师, 通讯联系人: 张昆; E-mail: kzhang@geo.ecnu.edu.cn

2 模拟计算

为了检测 G系数和 Moran指数在探测空间聚集能力的差异, 本文设计了一个 11×11 的研究区域, 单元属性值如图 1 所示, 均值为 5893。采用 ArcGIS 的 Spatial Statistics 模块 (后同) 计算其标准化全局 Moran 指数为 -0.017 。从前面的定义可以得出结论: 模拟的属性值不存在显著的空间自相关。

2.1 单核聚集

单核聚集分高值聚集和低值聚集两种情况。以上述模拟数据为基础, 设计 3×3 、 5×5 和 7×7 的聚集区域。相对于 5893 的均值而言, 值取 2000 以模拟低值聚集, 值取 9000 模拟高值聚集 (后同)。

算例 1

以上述模拟数据为基础, 设计 3×3 的聚集区

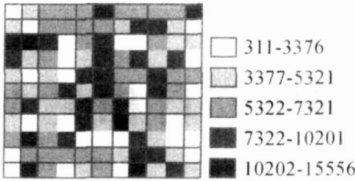


图 1 模拟的 11×11 的空间随机分布的单元属性值
Fig 1 Attribute values of 11×11 spatial random distributed units

域, 用最深色表示 (见表 1 第 2 行)。聚集区域先是位于 11×11 网格的四角和四边上, 然后从角和边向中心逐次移动 1 个网格以构成若干种方案, 计算每种方案的 Z_i 和 Z_G , 计算时均采用距离倒数权。表 1 仅列出了左上角和左边界的模拟计算结果, 其它三个角和边界的计算结果与此类似, 这里不再专门列出。

表 1 3×3 的单核聚集¹⁾
Tab. 1 Single-core clustering of 3×3 spatial units

方案		1	2	3	4	5	6	7	8	9
图形										
低值聚集	Z_G	1.951	0.914	0.092	-0.848	-1.373	0.946	0.359	-0.396	-1.006
	Z_i	1.845 *	1.273	0.196	0.342	0.426	1.278	0.866	0.781	0.453
高值聚集	Z_G	-0.186	1.135	2.114 *	1.873 *	1.469	1.063	1.950 *	1.891 *	1.779 *
	Z_i	0.709	0.636	1.366	0.629	0.211	0.591	1.000	0.420	0.347

1) * 显著性水平高于 0.05, * 显著性水平高于 0.1。

对于低值聚集, 由 Z_G 的定义可知, 其值应为负数。方案 1~3 和 6~7 实际计算的 Z_G 为正, 探测不出低值聚集; 随着聚集区域离开边界向中心移动 (方案 1~5 及方案 6~9), Z_G 逐渐由正数减小到负数, 方案 4、5、8、9 的为负值, 探测出了低值聚集, 但显著水平不高, 其中方案 5 的低值聚集区域位于正中, Z_G 最小, 为 -1.373 。可以有 83% 的把握认为存在低值聚集, 显著水平为 0.17。对于所列 9 个方案, Z_i 均为正值, 表明研究区域可能存在空间聚集; 除方案 1 以 0.05 的显著水平 (Z_i 大于 1.65) 探测出空间聚集外, 其余方案均不显著。

对于高值聚集, Z_G 的取值应为正数。除方案 1 外, 其他方案的 Z_G 都为正数, 表明探测出了高值聚集, 只是部分方案的显著性水平不高, 方案 2、5、6 的显著水平分别为 0.26、0.14 和 0.29。当 3×3 的空间聚集位于角落时 (方案 1), 低值聚集的 Z_G 为正, 而高值聚集的 Z_G 为负, 与应该出现的结果相反, 探测错误, 表明 G 系数可能对边

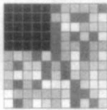
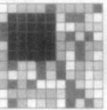
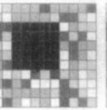
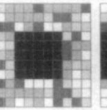
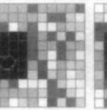
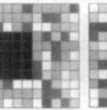
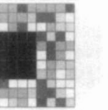
界敏感。对于所列 9 个方案, Z_i 均为正值, 表明探测出了空间聚集, 但均不显著。

算例 2

设计 5×5 的聚集区域, 由于聚集区域的扩大, 计算方案减少了两种 (见表 2 第 2 行)。

对于低值聚集, 位于边界上的方案 1 和 5 实际计算的 Z_G 为正, 探测不出低值聚集; 随着聚集区域离开边界向中心移动 (方案 1~4 及方案 5~7), Z_G 由正数减小到负数, 对应的 Z_G 为: $2.627\rightarrow -0.556\rightarrow -2.102\rightarrow -2.718$ 和 $0.235\rightarrow -1.581\rightarrow -2.566\rightarrow -2.718$ 。方案 2、3、4、6、7 的 Z_G 为负值, 探测出了低值聚集, 方案 3、4、7 的显著水平高, 其中低值聚集区域位于正中的方案 5 的 Z_G 最小, 显著水平最高, 和算例 1 一致。对于所列 7 个方案, Z_i 均为显著的正值, 表明研究区域存在显著的空间聚集, 和算例 1 进行纵向比较, 可看出: 随着聚集范围的扩大, Z_i 显著增加, 表明 Moran 指数能探测出聚集范围的扩张。

表 2 5×5 的单核聚集
Tab. 2 Single-core clustering of 5×5 spatial units

方案		1	2	3	4	5	6	7
图形								
低值	Z_c	2.627	-0.556	-2.102	-2.718	0.235	-1.581	-2.566
聚集	Z_I	6.267	4.478	3.860	5.597	4.157	3.620	5.561
高值	Z_c	1.595	2.936	3.812	3.980	3.276	3.793	6.159
聚集	Z_I	6.431	4.766	3.664	3.497	6.489	5.272	3.668

对于高值聚集, Z_G 均为正数。除位于角落的方案 1 外, 其他方案的 Z_G 都显著, 表明探测出了高值聚集; 随着聚集区域离开边界向研究区域中心移动, G 系数逐渐增加, 再次表明 G 系数可能对边界敏感。 Z_I 值对于所有聚集方案置都显著为正, 能探测出这些方案的空间聚集。可推算例 1 的 Moran 指数不显著的原因可能是模拟的聚集区域的规模相对偏小。

算例 3

设计 7×7 的聚集区域, 由于聚集区域的扩大, 计算方案又减少了两种 (图略)。

对于低值聚集, 和算例 1、2 的趋势一致, Z_G 为正或不显著的方案位于边界上, 说明 Z_G 可能对边界敏感。 Z_I 均为显著的正值, 表明研究区域存在显著的空间聚集, 和算例 1、2 进行纵向比较, 随着聚集范围的扩大, Z_I 进一步增加, 表明 Moran 指数能探测出聚集范围的扩张。

对于高值聚集, Z_G 均为显著的正数, 表明 G 系数能正确探测出这些方案中的高值聚集。 Z_I 非常显著, 探测出了所列方案的空间聚集, 显著性较算例 2 增强, 说明 Moran 指数能探测出聚集规模的扩张。

从单核聚集模拟计算可知: 对于低值聚集, 随着聚集区域离开边界向研究区域中心移动, G 值逐渐减小, 表明 G 系数对低值聚集的边界敏感; 对于高值聚集, 随着聚集区域离开边界向研究区域中心移动, Z 值和 G 系数逐渐增加, 表明 Z 指数和 G 系数对高值聚集都受边界的影响。由表 1 和表 2 可知, 随着聚集区域规模的扩大, G 系数探测出低值聚集的正确率提高。对于高值聚集, 当模拟的聚集区域位于研究区域的边界时, 探测结果很可能有误, 尤其是聚集区域的规模偏小时。因此, 受模拟的聚集区域四周随机分布结构变化的影响, 各方案尽管并不完全等效, 但是具有可比性, 仍然可以认

为 G 系数在探测低值聚集时, 聚集的规模会影响探测的结果, 而且聚集区域靠近研究区域的边界时, 会受到极大的影响; Moran 指数 主要受聚集区域规模的影响。

2.2 多核聚集

若研究区域不只存在一处空间聚集, 本文就称之为多核聚集。以上述模拟数据为基础, 在每种方案中模拟一处低值聚集 (以白色表示) 和一处高值聚集 (以黑色表示) 同一方案中高值和低值聚集的规模相当。

算例 4

设计 3×3 的聚集区域 (一个低值和一个高值), 见表 4 聚集区域分别位于 11×11 网格的四角和四边上, 然后从角和边向中心逐次移动 1 个网格以构成若干种方案, 计算每种方案的 $Z(I)$ 和 $Z(G)$, 计算时采用距离倒数权, 表 3 是部分计算结果。

3×3 的 20 个多核聚集方案中, 对于 G 系数, 除方案 14、16、18 和 20 外, 均为正数或显著的正数 (G 大于 1.65 对应的显著水平为 0.1), 鉴于显著的正值代表高值聚集而显著的负值代表低值聚集, 说明 G 系数对高值聚集比低值聚集敏感; 对于高值聚集区域在中心, 低值聚集区域在四角的方案 (方案 13、15、17 和 19), G 系数均为显著的正数, 也说明 G 系数对高值聚集更敏感, 而且聚集区域在中心比在边界对 G 系数影响大。20 个方案中 G 系数为负数的只有 4 个, 它们正好是高值聚集区域在中心、低值聚集区域在四角的方案, 说明位于中心的低值聚集区域对 G 系数的影响较位于边界的高值聚集区域大, 进一步证实了全局 G 系数对边界和高值聚集敏感的初步结论。

Moran 指数均为很显著的正值, 较单核聚集相应方案的 Moran 值有所增大, 说明 Moran 指数能探测出空间聚集范围的扩展。

表 3 3×3 的多核聚集
Tab. 3 Multi-core clustering of 3×3 spatial units

方案	1	2	3	4	5	6	7	8	9	10
图形										
Z_G	1.662	1.578	1.450	1.707	1.492	1.196	1.371	1.395	1.025	0.934
Z_I	3.691	3.226	3.257	4.688	4.376	2.712	2.942	3.075	3.667	3.159

方案	11	12	13	14	15	16	17	18	19	20
图形										
Z_G	1.928	2.256	2.738	-2.048	3.153	-1.545	3.297	-1.802	2.475	-2.205
Z_I	3.252	5.066	2.958	2.424	2.721	2.476	4.079	2.030	2.432	2.232

算例 5

在 11×11 网格内设计一处 5×5 的低值聚集区域和一处 5×5 的高值聚集区域，见表 4

5×5 的 20 个多核聚集方案中，G 系数全部为正数，说明 G 系数对高值聚集敏感；其中方案 14、16、18 和 20 与其他 16 个方案相比差异较大，G 系数明显小于其它方案的 G 系数，这 4 个方案有共同的特点：低值聚集区域靠近中心位置，高值聚集区域在边界上，说明 G 系数对边界敏感；当高、低值聚集区域位置互换时，总是高值聚集区域靠近中心时的 G 系数比低值聚集区域靠近中心时的 G

系数大，也说明 G 系数对高值聚集更敏感。

对于多核聚集，Moran 指数较 3×3 的多核聚集方案显著增加，证实 Moran 指数能探测出空间聚集范围的扩展。

从多核聚集模拟计算可知：G 系数对高值聚集和边界敏感；尽管 Moran 指数只能探测出空间聚集，对判断高值还是低值聚集无能为力，但随着空间聚集范围的扩展，Moran 指数会准确地增加，而 G 系数对高值和低值聚集共存时，有明显的判断为高值聚集的倾向。

表 4 5×5 的多核聚集
Tab. 4 Multi-core clustering of 5×5 spatial units

方案	1	2	3	4	5	6	7	8	9	10
图形										
Z_G	3.731	4.058	3.326	3.921	3.160	3.012	3.000	3.681	3.426	4.145
Z_I	13.098	14.747	13.181	16.248	16.298	15.510	13.037	16.577	15.903	19.721

方案	11	12	13	14	15	16	17	18	19	20
图形										
Z_G	4.040	4.532	5.327	1.767	5.587	1.914	5.326	1.159	5.277	0.503
Z_I	16.228	18.712	14.249	17.085	14.438	15.925	16.249	15.808	17.760	14.760

3 结 论

由上节的模拟计算可知，G 系数探测高值聚集和低值聚集的能力并不相同，检测出高值聚集的正

确率远大于检测出低值聚集的正确率，当研究范围内同时存在高值和低值聚集时，G 系数会受聚集区域规模的影响，当高值聚集区域和低值聚集区域规模相当时，G 往往为正数。尽管 G 具有区分高值

和低值聚集的优势, 但是判断一个区域是否存在空间聚集, 尤其是估计聚集区域位于区域的边缘时, 采用 Moran 指数统计的结果更为可靠, 要区分是高值聚集还是低值聚集, 可通过属性值的相对大小直接判断。

参考文献:

[1] GETIS A, ORD J K. The analysis of spatial association by use of distance statistics[J]. Geographical Analysis, 1992, 24(3): 89—206

[2] GOODCHILD M. Spatial Autocorrelation: Concepts and Techniques in Modern Geography[M]. Norwich, UK: Geo Books, 1986: 3—6

[3] SAWADA M. Global Spatial Autocorrelation Indices -

Moran's I, Geary's C and the General Cross-Product Statistic[J]. <http://www.geogebra.org/publications/Moran'sI/Moran.html>, 2006—05—16

[4] 吕安民, 李成名, 林宗坚, 等. 中国省级人口增长率及其空间关联分析[J]. 地理学报, 2002, 57(2): 143—150

[5] 颜锋华, 金亚秋. 尺度分布的 Getis统计对遥感图像特征参量空间自相关性的研究[J]. 中国图象图形学报, 2006, 11(6): 191—196

[6] TRISALYN N, BARRY B, MICHAEL A W. Techniques for accuracy assessment of tree locations extracted from remotely sensed imagery[J]. Journal of Environmental Management, 2005, 74(33): 265—271

Comparison between General Moran's Index and
Getis-Ord General G of Spatial Autocorrelation

ZHANG Song-lin¹, ZHANG Kun²

(1. Key Laboratory of Geographic Information Science, East China Normal University, Shanghai 200062, China;
2. Department of Surveying and Geomatics, Tongji University, Shanghai 200092, China)

Abstract: Moran's Index and Getis-Ord general G are among the most commonly used indices for the analysis of general spatial autocorrelation. This paper compared Moran's Index and Getis-Ord General G based on simulating computation. The results showed that Getis-Ord general G is more sensitive to high clusters than low ones. When high and low cluster coexist at the study area, Getis-Ord general G will be affected by the scale of cluster. G will more likely to be positive to indicate high cluster when the scale of high and low cluster are equal. Moran's Index is mainly affected by the scale of clusters, it will increase when the scale is extended. These conclusions are useful in selecting general index to detect the spatial autocorrelation pattern.

Key words: spatial autocorrelation; general indices; Moran's Index; Getis-Ord general G