

A generalized significance testing method for global measures of spatial association: an extension of the Mantel test

Environment and Planning A, 36(9), 1687-1703

이상일*

이 연구의 목적은 공간 연관성 척도(spatial association measures; 이하 SAM)에 대한 일반화된 유의성 검정 방법을 제시하는 것이다. 공간 연관성 척도는 공간적 패턴이 보여주고 있는 공간적 의존성(spatial dependence) 혹은 공간적 자기상관(spatial autocorrelation)을 측정하는 것이다. 공간데이터분석(spatial data analysis) 혹은 공간통계학(spatial statistics)의 최근 진보에서 SAM이 차지하는 비중은 매우 크다. 왜냐하면 일반 데이터 분석이 아닌 공간데이터분석, 일반 통계학이 아닌 공간통계학의 핵심적인 존재 근거가 공간 데이터의 본질적인 특성, 즉 공간적 의존성에 있기 때문이다. 예를 들어 회귀분석의 잔차에 공간적 자기상관이 존재할 경우 일반 회귀분석에 의한 통계학적인 결과는 더 이상 신뢰할 수 없게 된다. 이러한 측면에서 지리학과 유관 학문 분야에서는 1990년대 중반 이후 일변량(univariate) SAM의 개발과 통계적 검정 절차의 제시에 많은 노력을 경주해 왔고 일반 방법론으로서의 토대를 굳건히 마련하고 있다.

그런데, 상관관계에서 공간적 의존성을 다루는 문제, 즉, 이변량(bivariate) SAM의 개발은 거의 주목을 받지 못했다. 두 공간적 패턴의 유사성을 검토하기 위해 흔히 사용되는 일반 통계학의 방법은 상관계수, 특히 피어슨의 상관계수(Pearson's correlation coefficient)를 적용하는 것인데, 이는 두 공간 패턴 간에서 존재하는 이변량 공간적 의존성(bivariate spatial dependence)을 측정하지 못하는 한계가 있다. 이변량 공간적 의존성이란 유사한 이변량 연관성을 보여주는 케이스들이 공간적으로 집중/분산하는 경향을 의미한다. 만일 이러한 이변량 공간적 의존성이 존재한다면, 피어슨 상관계수에 의한 측정과 유의성 검정은 더 이상

* 서울대학교 사범대학 지리교육과 조교수

신뢰할 수 없는 것이 된다. 요약해서 말하면, 이변량 SAM은 피어슨 상관계수의 공간통계학적 대안이다. 이러한 의미에서 본 연구 이전에 수행된 연구(Lee, 2001)는 새로운 이변량 SAM을 제시한 것이었으며, 지리학뿐만 아니라 생태학, 식물학, 기상학 등 다양한 영역에서 주목을 받았다(Haining(2003)과 Wong and Lee(2005)의 공간데이터 분석 개론서에 소개된 바 있다).

문제는 바로 이러한 이변량 SAM에 대한 통계적 검정 절차를 제시하는 것이다. 일변량 SAM에 대해서는 정확 검증(exact test), 정규성 가정(normality assumption)에 근거한 검증, 랜덤화 가정(randomization assumption)에 근거한 검증, 시뮬레이션에 기반한 검증 등 다양한 방식의 검정법이 존재하고 있다. 그러나 이변량 SAM의 검정 절차를 제시하는 것은 일변량 SAM에 비해 훨씬 더 복잡한 문제를 안고 있다. 또한 일변량 SAM에 대한 검정의 경우도 공간 근접성 행렬(spatial proximity matrix)의 정의에 대한 제약, 즉, 행렬의 대각선 요소는 모두 0이어야 한다는 인위적인 제약으로부터 탈피할 필요가 있는 것이다. 따라서 본 연구의 주된 과제는 랜덤화 가정에 기반하여, 공간 근접성 행렬의 정의에 대한 어떠한 인위적 제약도 없이, 일변량과 이변량 SAM 모두에 적용될 수 있는 일반 검정 절차를 제시하는 것이다.

일반화된 검정 절차를 제시하기 위해서는 현재 사용되고 있는 일변량 SAM 뿐만 아니라 제안된 이변량 SAM이 구조적으로 동일한 형태를 지니고 있다는 사실을 밝혀야 한다. 본 연구자는 3개의 일변량 SAM과 두 개의 이변량 SAM의 기본 형태가 피어슨의 상관계수와 같은 교차곱(cross-product)의 형태를 취하고 있으며, 행렬의 형태로 변환되었을 때, 모두 이차형태(quadratic form)로 표현될 수 있음을 발견했다. 다시 말해 모든 SAM은 공간 근접성 행렬과 수치적 연관성 행렬의 케이스별 곱의 합산 형태를 보인다는 점을 밝힌 것이다. 이러한 발견은 모든 SAM이 동일한 토대 위에서 가설 검정될 수 있음을 의미하는 것이다. 이 연구는 Mantel(1967), Siemiatycki(1978), Mielke(1979), Heo and Gabriel(1998) 등 통계학자에 의한 연구들을 참고하고, 일변량 SAM에 대한 Cliff and Ord(1981)의 논의를 확장함으로써 일변량과 이변량 SAM에 공통적으로 적용가능하고, 공간 근접성 행렬의 대각선에 0이 아닌 요소가 존재할 경우도 포괄할 수 있는 일반화된 검정법을 제시할 수 있게 되었다.

이 연구가 가지는 파급효과는 대단히 크다. 첫째, 공간데이터를 이용한 다양한 경험적 연구에서 피어슨 상관계수를 대체하면서, 통계학적 근거 하에 두 공간적

패턴의 연관성을 논할 수 있게 되었다. 다른 연구자와 공동으로 진행된 연구(Brown *et al.*, 2005)는 미국의 노동시장에서 여성이 차지하는 비율이 1970년과 1990년 사이에 카운티 수준에서 얼마나 변했는지를 정량적으로 측정하고 그에 대한 통계적 결론을 제시하는데 있어 제안된 방법론이 얼마나 유용한지를 잘 보여주고 있다.

둘째, 제안된 일반화된 유의성 검정법이 일단 전역적(global) SAM을 위해 개발된 것이지만 약간의 수정 작업을 거치면 국지적(local) SAM의 유의성 검정에도 적용될 수 있다. *Geographical Analysis* 지에 실릴 예정인 연구(Lee, 2008)는 이러한 측면을 명확히 보여주고 있다. 특히 이변량 SAM에 대한 국지적 통계량과 그에 대한 유의성 검정이 확립되면, 이변량 탐색적 공간데이터분석(exploratory spatial data analysis; 이하 ESDA)에서 새로운 장을 열 수 있을 것으로 기대된다. ESDA의 발전은 GIS(geographic information systems)와의 관련성 속에서 보다 중요하다. 최근 GIS에서 시스템 개발 위주의 연구를 지양하고 분석 도구로서의 GIS를 강화함으로써 지리정보과학(GIScience)으로 나아가려는 움직임이 두드러진다. 이런 분석-지향적 GIS에서 핵심적인 것이 ESDA이며, 이러한 ESDA를 활성화하는데 국지적 이변량 SAM과 유의성 검정법의 발달은 핵심적인 역할을 담당할 것으로 보인다.

셋째, 연구가 SAM에 대한 일반화된 가설 검정 방법을 제시한 것이기 때문에, 향후 개발될 다양한 종류의 SAM에 대해 즉각적으로 이용 가능한 프레임워크를 제공할 수 있다. 실질적으로 또 다른 연구(이상일, 2007, 2008)는 새로이 개발된 거주지 분리 측도의 통계학적 분포 속성을 이끌어 내는데 일반화된 가설 검정 방법이 얼마나 유용한지를 보여주고 있다.

참고문헌

- Brown, L. A., Lee, S.-I., Lobao, L., and Chung, S.-Y., 2005, Continuity amidst restructuring: the U.S. gender division of labor in geographic perspective, 1970 and 1990, *International Regional Science Review*, 28(3), 271-301.
- Cliff, A. D. and Ord, J. K., 1981, *Spatial Processes: Models and Applications*, London: Pion Limited.
- Haining, R., 2003, *Spatial Data Analysis: Theory and Practice*, New York: Cambridge University Press.
- Heo, M. and Gabriel, K. R., 1998, A permutation test of association between configurations by means of the RV coefficient, *Communications in Statistics: Simulation and Computation*, 27(3), 843-856.
- Lee, S.-I., 2001, Developing a bivariate spatial association measure: an integration of Pearson's r and Moran's I , *Journal of Geographical Systems*, 3(4), 369-385.
- Lee, S.-I., 2004, A generalized significance testing method for global measures of spatial association: an extension of the Mantel test, *Environment and Planning A*, 36(9), 1687-1703.
- Lee, S.-I., 2008, A generalized randomization approach to local measures of spatial association, *Geographical Analysis*, forthcoming.
- Mantel, N., 1967, The detection of disease clustering and a generalized regression approach, *Cancer Research*, 27(2), 209-220.
- Mielke, P. W., 1979, On asymptotic non-normality of null distributions of MRPP statistics, *Communications in Statistics: Theory and Methods*, A8(15), 1541-1550; 1981, Erratum, A10(17), 1795; 1982, Erratum, 11(7), 847.
- Siemiatycki, J., 1978, Mantel's space-time clustering statistic: computing higher moments and a composition of various data transforms, *Journal of Statistical Computation and Simulation*, 7(1), 13-31.
- Wong, D. W. S. and Lee, J., 2005, *Statistical Analysis of Geographic*

Information with ArcView GIS and ArcGIS, Hoboken, NJ: John Wiley & Sons.

이상일, 2007, “거주지 분화에 대한 공간통계학적 접근(I): 공간 분리성 측도의 개발,” *대한지리학회지*, 42(4), 616-631.

이상일, 2008, “거주지 분화에 대한 공간통계학적 접근(II): 국지적 공간 분리성 측도를 이용한 탐색적 공간데이터 분석,” *대한지리학회지*, 43(1), 134-153.