

Detection of Hotspots for Geospatial Lattice Data

Sung-Ho Moon¹⁾ · Jong-Duk Kim²⁾

Abstract

Statistical analyses for spatial data are important features for various types of fields. Spatial data are taken at specific locations or within specific regions and their relative positions are recorded. Lattice data are synoptic observation covering an entire spatial region, like cancer rates corresponding to each county in a state. The main purpose of this paper is to detect hotspots for the region with significantly high or low rates. Kulldorff(1997) detected hotspots based on circular spatial scan statistics. We propose a new method to find any shapes of hotspots by use of echelon analysis with spatial scan statistics.

Keywords : Echelon analysis, Hotspot, Spatial scan statistics

1. 서론

지역 데이터나 리모트 센싱(remote sensing) 데이터와 같은 공간 데이터(spatial data)에 관한 통계 분석은 여러 종류(생태학, 도시공학, 천문학, 화상해석 등등)의 분야에서 중요한 역할을 한다. 공간 데이터는 변수나 특정 위치의 값에 대한 정보를 제공하는데, 이 때 데이터는 특정 위치나 특정 지역에서 얻어지며 이들의 상대적인 위치가 기록된다.

lattice(格子) 데이터란 규칙적이거나 불규칙적으로 배치된 영역집단에 속하는 관측 값을 일컫는다. 여기서 영역이란 면적을 지닌 것이라면 무엇이든 상관하지 않으며, 격자상에 제한되는 것만은 아니다. 규칙적인 격자의 예로서는 인공위성에 의한 리모트 센싱 데이터(R^2 상의 규칙적 격자)뿐만 아니라 $N \times M$ 형 분할표도 속한다고 할 수 있다. 또한 불규칙적 격자의 예로서는, 어느 道내의 각 市에서의 암 발생률 등을 들 수 있겠다. 공간 검색 통계량(Kulldorff, 1997)은 질병 발생률과 같이 지역별로 얻어지는 데이터에 있어서 유의하게 높거나 낮은 값을 나타내는 지역인 핫스팟(hotspot)을 발견

1) First Author : Professor, Department of Statistics, Pusan University of Foreign Studies, Busan 608-738, Korea. E-mail : shmoon@pufs.ac.kr

2) Professor, Department of Statistics, Pusan University of Foreign Studies, Busan 608-738, Korea. E-mail : jdkim@pufs.ac.kr

하기 위해 이용된다. 핫스팟을 검출하기 위해서는 검색하는 지역을 선택하는 것이 문제가 된다. 그런데 지역 데이터와 같은 데이터의 경우, 지역군에서 영역을 뽑는 방법(pattern)이 무한히 많이 존재하게 되어 공간 검색 통계량의 계산이 곤란하게 된다. 따라서 통상의 공간 검색 통계량의 분포는 몬테 칼로 방법 등에 의거한 시뮬레이션에 의해 계산되어 왔다. 한편, 분할표나 市區洞 데이터와 같이 cell수가 적은 경우에는 직접 계산에 의해서 확률분포를 계산하는 것이 필요하다.

본 논문에서는 지역 데이터나 분할표와 같이 cell수가 적은 lattice 데이터에 대해, 손쉽게 계산기 지원에 의한 몬테 칼로 방법에 의존하지 않고서 공간 검색 통계량의 확률분포를 직접 계산하는 알고리즘을 개발함과 동시에, 실제 예는 1990년도 인구센서스 조사에 의해 나타난 전국의 각 시도별 외국인 데이터에 적용시켜 보도록 하겠다. 2절에서는 공간 검색 통계량의 산출을, 3절에서는 검색되는 영역의 검출을 그리고, 4절에서는 구체적인 수치 예를 보이고 있다.

2. 공간 검색 통계량

공간 통계학에서의 일반적인 관심사항은 일련의 점(point)들이 랜덤하게 분포되어 있느냐, 아니면 일정한 군집(cluster)성을 가지고 있느냐는 것이다. 이러한 군집의 위치를 찾아내고자 하는 경우에 사용할 수 있는 방법이 바로 공간 검색 통계량이다. 공간 검색 통계량은 질병의 발생률과 같이 지역마다 얻어지는 데이터에 있어서 핫스팟 지역을 발견하는데 이용된다.

모든 영역을 G 로 두고 그 부분집합 영역을 Z 라 하자. 영역 Z 의 내부에서 개인은 어떤 속성을 확률 p , 영역 Z 밖에서는 확률 q 를 취하며 속성을 지닐 확률은 서로 독립이다. 이 때 가설은 다음과 같다.

$$H_0: p = q \quad v.s. \quad H_1: p \neq q$$

$n(G)$ 는 모든 영역 G 에서의 모집단의 수, $n(Z)$ 는 영역 Z 내부의 모집단의 수, $c(G)$ 는 모든 영역 G 에서 속성을 지닌 개체 수, $c(Z)$ 는 영역 Z 내부에서 속성을 지닌 개체 수로 둔다.

여기서는 포아송 분포에 기인한 모델을 고려하며, 전체 영역 G 에서 속성을 지닌 개체 수가 $c(G)$ 로 될 확률은

$$\frac{\exp[-pn(Z) - q(n(G) - n(Z))][pn(Z) + q(n(G) - n(Z))]^{c(G)}}{c(G)!}$$

전체 영역 안에서의 지점 x 에서의 밀도는

$$\begin{cases} \frac{pn(x)}{pn(Z) + q(n(G) - n(Z))} & , x \in Z \\ \frac{qn(x)}{pn(Z) + q(n(G) - n(Z))} & , x \notin Z \end{cases}$$

가 되며, 이 때 포아송 모델에 대한 우도함수는 다음과 같이 주어진다.

$$L(Z, p, q) = \frac{\exp[-pn(Z) - q(n(G) - n(Z))]}{c(G)!} p^{c(Z)} q^{c(G) - c(Z)} \prod_{x_i} n(x_i)$$

$x(Z)$ 를 영역 Z 내에서 속성을 지니는 개체 수를 나타내는 확률변수라 할 때, 임의의 영역 A 에 있어서

$$x(A) \sim \text{Poisson}(pn(A \cap Z) + qn(A \cap Z^c))$$

귀무가설 하에서는 임의의 영역 A 에 있어서

$$x(A) \sim \text{Poisson}(pn(A))$$

우도함수를 최대로 하기 위해서, 영역 Z 가 주어졌을 때의 최대우도함수를 계산한다. 이 때 최우추정량

$$L(Z) = \begin{cases} \frac{e^{-n(G)}}{c(G)!} \left(\frac{c(Z)}{n(Z)}\right)^{c(Z)} \left(\frac{c(G) - c(Z)}{n(G) - n(Z)}\right)^{c(G) - c(Z)} \prod_{x_i} n(x_i), & \frac{c(Z)}{n(Z)} > \frac{c(G) - c(Z)}{n(G) - n(Z)} \\ \frac{e^{-n(G)}}{c(G)!} \left(\frac{c(G)}{n(G)}\right)^{c(G)} \prod_{x_i} n(x_i) = L_0, & \text{그외에} \end{cases}$$

은 $\hat{p} = c(Z)/n(Z)$, $\hat{q} = (c(G) - c(Z))/(n(G) - n(Z))$ 을 대입하여 구한다.

또한 우도비 λ 는 핫스팟을 발견하기 위해 전체영역의 부분집합인 영역 Z 에서 최대인 것으로 한다.

$$\lambda = \begin{cases} \frac{\text{Max}_Z L(Z)}{L_0} = \frac{\left(\frac{c(Z)}{n(Z)}\right)^{c(Z)} \left(\frac{c(G) - c(Z)}{n(G) - n(Z)}\right)^{c(G) - c(Z)}}{\left(\frac{c(G)}{n(G)}\right)^{c(G)}}, & \frac{c(Z)}{n(Z)} > \frac{c(G) - c(Z)}{n(G) - n(Z)} \\ 1, & \text{그외에} \end{cases}$$

여기서 L_0 는 귀무가설이 참일 때의 우도함수 값이다.

검정통계량 λ 는 다음과 같이 나타낼 수 있다.

$$\lambda = \frac{\text{Max}_Z L(Z)}{L_0} = \frac{(c(Z)/n(Z))^{c(Z)} \{(c(G) - c(Z))/(n(G) - n(Z))\}^{c(G) - c(Z)}}{(c(G)/n(G))^{c(G)}}$$

또한 주어진 통계량에 대해 영역 Z 에 있어서 어떠한 속성을 지닌 것의 수는 포아송 모형에 기인한 확률분포 $\text{Poisson}(n(Z) \cdot (c(G)/n(G)))$ 을 따른다.

3. 핫스팟 검출

종래에 공간 검색 통계량의 확률분포는 직접 계산이 곤란하여 몬테 칼로 방법이 주로 이용되어 왔다. 하지만 분할표나 市區洞 데이터 등의 지역 데이터와 같이 cell수가 적은 경우에는, 직접 확률분포를 계산해야할 필요가 있다. lattice 데이터에 공간적인 순서적 연결이 있을 경우에는 모든 cell 데이터에 대해서 개개의 검정을 행하는 것은 적절하지 않으며, 인접한 공간의 위치 정보를 고려한 데이터 군을 발견하는 것이 중요하다.

공간 검색 통계량을 직접 계산할 경우에 lattice 데이터 중에서 인접해 있는 공간 영역의 검출을 하지 않으면 안 된다. 즉, lattice 데이터의 임의의 인접한 cell 패턴을 발견해야하는 필요성이 있다. 소수 cell에 의해 구성되는 lattice 데이터라면 인접한 cell 패턴 Z 를 검색하는 것은 가능하다. 하지만 지금 그 패턴 수가 막대하므로 패턴 Z 를 효율적으로 검색하는 것이 중요하다.

에셀론(echelon) 해석(Myers et al., 1997, Myers and Patil, 2002)은 공간적인 위치를 표면상의 데이터 고저를 바탕으로 분할해서, 공간 데이터의 위상적인 구조를 계통적이고 개관적으로 발견하기 위해 개발된 것이다. 이 방법에 의한 상위 계층 영역을 검색 영역으로 하는 공간 검색 통계량을 이용해서 핫스팟 검색을 행한다. 에셀론 해석을 이용하는 방법은 실제로 공간 검색 통계량을 계산하는 알고리즘을 구축해서 실시한다.

예를 들어 지금 3×3 lattice 데이터에 있어서, 각각의 모집단의 수 $n(Z)$ 가 <표 1>, 속성을 지닌 개체의 수 $c(Z)$ 가 <표 2>와 같이 주어졌다고 하자. 모든 패턴 Z 를 구해서(cell 1-9) 검정 통계량 λ 의 확률분포를 계산한다.

<표 1> 모집단의 수 $n(Z)$

| | | |
|------|------|------|
| 800 | 1300 | 1400 |
| 700 | 900 | 1100 |
| 1200 | 1000 | 1600 |

<표 2> 속성을 지닌 수 $c(Z)$

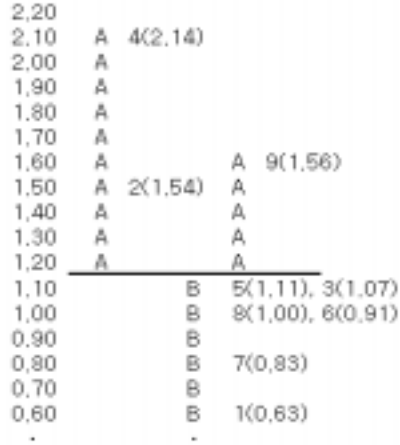
| | | |
|----|----|----|
| 5 | 20 | 15 |
| 15 | 10 | 10 |
| 10 | 10 | 25 |

<표 3> cell ID

| | | |
|---|---|---|
| 1 | 2 | 3 |
| 4 | 5 | 6 |
| 7 | 8 | 9 |

<표 3>은 <표 1>과 <표 2>의 lattice 데이터의 cell ID이며 $n(Z)$ 와 $c(Z)$ 는 각각 1000, 120이다. 각 cell당 모집단에서 속성을 지니는 개체 수의 비율을 구해서 작성한 에셀론 덴드로그램이 <그림 1>이다. 검색된 각 영역에서 속성을 지니는 개체의 수는 초기하분포를 따르는 확률분포 $P = \left(\frac{c(G)}{c(Z)} \right) \left(\frac{n(G) - c(G)}{n(Z) - c(Z)} \right) \left| \left(\frac{n(G)}{n(Z)} \right) \right.$ 를 가지는 것으로 한다.

<그림 1>의 상위 에셀론 순으로 영역 검색을 수행하여 핫스팟을 구한다. 각각의 영역에 있어서의 통계량 λ 와 p-value가 <표 4>이다. 핫스팟의 후보로서 영역 (2, 4, 5, 9)가 구해지며 이 때의 p-value가 0.003이다.



<그림 1> 비율에 대한 에셜론 덴드로그램

<표 4> $\log \lambda$ 와 p-value

| cell ID | $n(Z)$ | $c(Z)$ | $\log \lambda$ | p-value |
|-------------------|--------|--------|----------------|---------|
| 4 | 700 | 15 | 2.296 | 0.031 |
| 9 | 1600 | 25 | 0.969 | 0.162 |
| 2,4 | 2000 | 35 | 2.861 | 0.016 |
| 2,4,5,9 | 4500 | 70 | 4.284 | 0.003 |
| 2,3,4,5,9 | 5900 | 85 | 3.618 | 0.007 |
| 2,3,4,5,8,9 | 6900 | 95 | 3.122 | 0.012 |
| 2,3,4,5,6,8,9 | 8000 | 105 | 2.359 | 0.029 |
| 2,3,4,5,6,7,8,9 | 9200 | 115 | 1.432 | 0.089 |
| 1,2,3,4,5,6,7,8,9 | 10000 | 120 | . | . |

일반 공간 데이터의 경우에 Z 의 패턴은 무한히 많이 존재하므로, 검정 통계량 λ 의 계산은 통상 몬테 카를로 방법에 의존해 왔다. cell수가 적은 lattice 데이터라면 Z 의 패턴을 직접 계산하는 것이 가능하게 되어 λ 의 분포를 직접 계산할 수 있다. 하지만 cell수가 적은 lattice 데이터라 해도 패턴 수가 막대해서 모든 것을 검출하는 것은 매우 힘든 작업이 된다. 따라서 계산기를 이용함으로써 효율적인 검출을 행하게 된다. 조건으로서는 영역 Z 가 될 수 있는 lattice 데이터 내의 cell은 같은 영역 Z 로 될 수 있는 다른 cell에 반드시 인접하도록 취한다. 이 방식에 따라 핫스팟을 검출하는데 사용되는 SatScanTM(Kulldorff, 2005)과 같은 소프트웨어를 활용하면 탐지해내지 못하는 복잡한 영역 Z 의 검출도 가능하게 된다.

4. 수치예

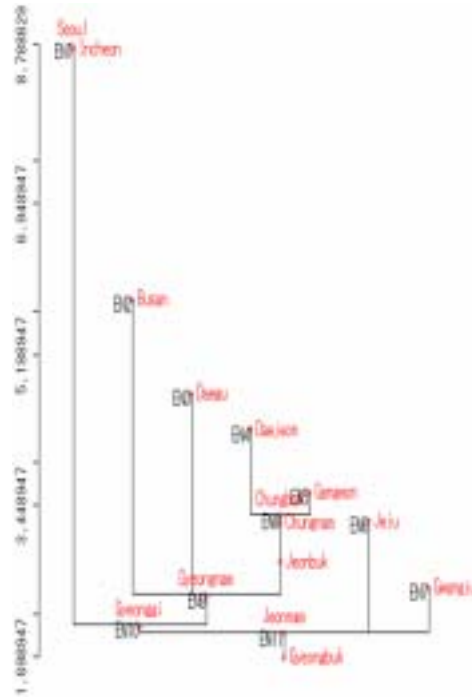
<표 5>는 1990년도 인구센서스 조사 결과에 따른 15개 시도별 총인구, 외국인 수와 총인구수에 대한 그 비율을 나타낸 것이다. 이 데이터에 대해 우리는 이들 도시에 있어서의 외국인 인구의 핫스팟을 검출하고자 한다.

검색 방식으로는 에셜론 해석에 있어서의 상위 에셜론 순으로 영역을 검색하고서 그 영역에서의 핫스팟을 구한다.

Step 1. 비율과 시도별 인접정보를 바탕으로 에셜론 덴드로그램을 작성. <표 5>에서의 비율과 각 시도별 지리적 인접정보를 이용하여 상위 에셜론 순으로 영역을 검색한 결과는 <그림 2>와 같다.

<표 5> 전국 15개 시·도별 인구현황(1990년도)

| 시·도명 | 총인구 | 외국인 인구 | 비율 (×10,000) |
|-----------|------------|-----------|-----------------|
| Seoul | 10,612,577 | 9,327 | 8.79 |
| Busan | 3,798,113 | 2,221 | 5.85 |
| Daegu | 2,229,040 | 1,061 | 4.76 |
| Incheon | 1,817,919 | 1,591 | 8.75 |
| Gwangju | 1,139,003 | 286 | 2.51 |
| Daejeon | 1,049,578 | 456 | 4.34 |
| Gyeonggi | 6,155,632 | 1,273 | 2.07 |
| Gangwon | 1,580,430 | 571 | 3.61 |
| Chungbuk | 1,389,686 | 464 | 3.34 |
| Chungnam | 2,013,926 | 656 | 3.26 |
| Jeonbuk | 2,069,960 | 582 | 2.81 |
| Jeonnam | 2,507,439 | 495 | 1.97 |
| Gyeongbuk | 2,860,595 | 486 | 1.70 |
| Gyeongnam | 3,672,396 | 887 | 2.42 |
| Jeju | 514,605 | 169 | 3.28 |
| Total | 43,410,899 | 20,525 | |



<그림 2> 에셜론 덴드로그램

Step 2. <그림 2>의 에셜론 덴드로그램에서 상위 계층 순으로 공간 검색 통계량 $\log \lambda$ 를 계산한 결과가 <표 6>이다.

$$\lambda = \frac{(c(Z)/n(Z))^{c(Z)} \{ (c(G) - c(Z)) / (n(G) - n(Z)) \}^{c(G) - c(Z)}}{(c(G)/n(G))^{c(G)}}$$

이 때, $n(G) = 43,410,899$, $c(G) = 20,525$

<표 6> 공간 검색 통계량의 계산결과

| 에설론 | 시·도명 | 총인구 | 외국인 인구 | $n(z)$ | $c(z)$ | $\log \lambda$ |
|------|-----------|------------|--------|------------|--------|----------------|
| EN1 | Seoul | 10,612,577 | 9,327 | 10,612,577 | 9,327 | 2,136.39 |
| | Incheon | 1,817,919 | 1,591 | 12,430,496 | 10,918 | 2,709.62 |
| EN2 | Busan | 3,798,113 | 2,221 | 3,798,113 | 2,221 | 51.65 |
| EN3 | Daegu | 2,229,040 | 1,061 | 2,229,040 | 1,061 | 0.03 |
| EN4 | Daejeon | 1,049,578 | 456 | 1,049,578 | 456 | 1.72 |
| EN5 | Gangwon | 1,580,430 | 571 | 1,580,430 | 571 | 23.43 |
| EN6 | Jeju | 514,605 | 169 | 514,605 | 169 | 12.86 |
| EN7 | Gwangju | 1,139,003 | 286 | 1,139,003 | 286 | 73.12 |
| EN8 | Chungbuk | 1,389,686 | 464 | 4,019,694 | 1,491 | 52.16 |
| | Chungnam | 2,013,926 | 656 | 6,033,620 | 2,147 | 109.45 |
| | Jeonbuk | 2,069,960 | 582 | 8,103,580 | 2,729 | 212.10 |
| EN9 | Gyeongnam | 3,672,396 | 887 | 17,803,129 | 6,898 | 237.82 |
| EN10 | Gyeonggi | 6,155,632 | 1,273 | 36,389,257 | 19,089 | 779.99 |
| EN11 | Jeonnam | 2,507,439 | 495 | 40,550,304 | 20,039 | 388.38 |
| | Gyeongbuk | 2,860,595 | 486 | 43,410,899 | 20,525 | |

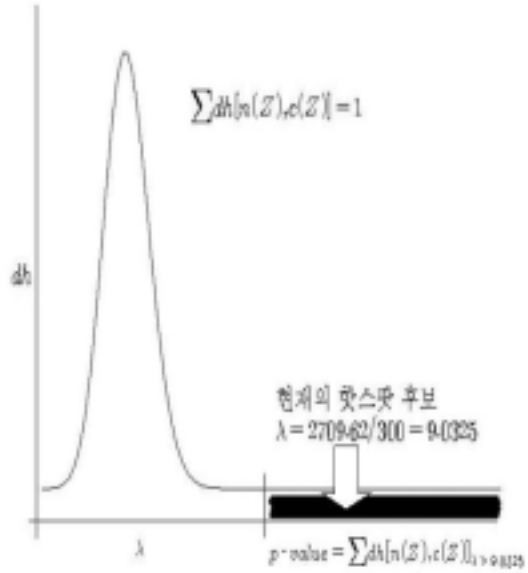
- Step 3. <표 5>에서의 에설론의 구조를 바탕으로 $\log \lambda$ 가 가장 높은 값을 취하는 곳 까지를 핫스팟으로 한다. 여기서는 서울과 인천의 두 도시에 $\log \lambda$ 가 가장 높은 값(2,709.62)을 취한다는 사실을 알 수 있다.
- Step 4. p-value의 추정. Step 3에서 핫스팟 후보로 된 두 도시의 p-value를 몬테 칼로법을 이용해서 추정한다. 여기서 공간 검색 통계량 $\lambda[n(Z), c(Z)]$ 는 초기분포 $dh[n(Z), c(Z)]$ 에 대응하는 것으로 한다.

$$dh[n(z), c(z)] = \frac{\binom{c(G)}{c(Z)} \binom{n(G)-c(G)}{N(Z)-c(Z)}}{\binom{n(G)}{n(Z)}}, \quad (n(G) = \sum n(Z), c(G) = \sum c(Z))$$

여기서는 원 데이터의 자리수가 너무 많아 $dh[n(Z), c(Z)]$ 의 계산이 곤란하므로 데이터를 1/300으로 줄이기로 한다. 한국의 15개 도시에 있어서, 서로 인접하고 있는 영역에 의해 이루어지는 패턴의 종류는 전부 4,742 가지가 존재한다. 다시 말해서 4,742 개의 $n(Z)$ 가 존재할 가능성이 있다. 그 각각의 $n(Z)$ 에 있어서 지금 외국인 인구 ($c(Z)$)가 0명에서 48명까지 존재할 가능성이 있는 것이다(<표 5-1>). 따라서 실제로 계산하려면, $4,742 \times 69 = 327,198$ 가지의 $dh[n(Z), c(Z)]$ 와 $\lambda[n(Z), c(Z)]$ 를 구해야만 한다. 하지만 이 계산이 실로 막대하므로 랜덤으로 10,000개의 $n(Z)$ 와 $c(Z)$ 를 추출, 그 10,000개로부터 분포를 작성해서 p-value를 추정한다.

<표 5-1> 1990년도 데이터(1/300)

| 시·도명 | 총인구 | 외국인 인구 |
|-----------|---------|--------|
| Seoul | 35,375 | 31 |
| Busan | 12,660 | 7 |
| Daegu | 7,430 | 4 |
| Incheon | 6,060 | 5 |
| Gwangju | 3,797 | 1 |
| Daejeon | 3,499 | 2 |
| Gyeonggi | 20,519 | 4 |
| Gangwon | 5,268 | 2 |
| Chungbuk | 4,632 | 2 |
| Chungnam | 6,713 | 2 |
| Jeonbuk | 6,900 | 2 |
| Jeonnam | 8,358 | 2 |
| Gyeongbuk | 9,535 | 2 |
| Gyeongnam | 12,241 | 3 |
| Jeju | 1,715 | 1 |
| Total | 144,703 | 68 |



<그림 3> 작성될 분포의 예상도

이번 연구에서는 10회의 추정을 수행했으며 그 결과는 아래와 같다.

| | | | | | |
|---------|--------------|--------------|--------------|--------------|--------------|
| 회수 | 1회 | 2회 | 3회; | 4회 | 5회 |
| p-value | 2.383504E-05 | 2.248706E-05 | 2.264245E-05 | 2.387876E-05 | 2.129779E-05 |
| 회수 | 6회 | 7회 | 8회 | 9회 | 10회 |
| p-value | 2.231822E-05 | 2.436653E-05 | 2.110853E-05 | 2.39602E-05 | 2.210665E-05 |

위의 추정 결과를 살펴보면 어느 경우에도 p-value가 0에 가까울 만큼 매우 작은 값으로 나타나 유의하다는 사실을 알 수 있다. 따라서 1990년도의 한국 15개 도시에 있어서 유의하게 외국인 인구가 많은 지역군은 Seoul과 Incheon으로 나타났다.

5. 마무리

각종 공간 데이터에 있어서 유의하게 높거나 낮은 값을 나타내는 지역(핫스팟)의 검출은 중요한 과제이다. 공간 검색 통계량(Kulldorff, 1997)은 데이터가 얻어진 영역의 중심을 원의 중심으로 해서 원상(圓狀)에서 영역을 검색해서 핫스팟을 발견하는 방법이다. 하지만 Kulldorff에 의한 방법으로는 원상의 핫스팟 밖에 검출할 수 없다. 따라서 검색 방법으로서 예설론 해석(Myers et al., 1997)을 이용한다. 즉, 데이터의

高低를 바탕으로 한 위상적 구조로 이루어지는 상위 에셀론 순으로 영역 검색을 행함으로써, 원상에만 국한되지 않는 모든 영역에서의 핫스팟 검출이 가능하게 됐다.

본 연구에서 우리는 1990년도 인구 센서스 데이터에 대해 공간 검색 통계량과 에셀론 해석법을 적용해 보았다. 공간 데이터에 관해서도 데이터의 계층구조를 이용하는 에셀론 해석법의 결과가 객관적으로도 충분한 해석이 가능하다는 장점을 수치 예에서 알 수 있었다. 또한 lattice 데이터의 영역 Z 의 패턴을 구함으로써 공간 검색 통계량을 직접 계산할 수 있게 되어 p-value를 구할 수 있었다. 이로 인해 복잡한 범위에서의 핫스팟 검출도 가능하게 되었다. 이처럼 공간 검색 통계량을 이용한 에셀론 해석법에 의한 핫스팟의 발견은 여러 분야에의 적용이 가능하다. 추후 해석 대상의 lattice 데이터를 종래의 일변량 데이터로부터 다변량 공간 데이터로 확장시켜 공간 검색 통계량을 이용한 핫스팟의 검출을 시도해 볼 계획이다. 즉, 공간 데이터를 시계열 구조를 지닌 데이터로 확장해서 시공간 상에서의 구조 해석이 차후의 연구과제 대상이다.

참고문헌

1. Kulldorff, M.(1997). A spatial scan statistics, *Communications in Statistics, Theory and Methods*, 26, 1481-1496.
2. Kulldorff, M.(2005). SaTScan™ User Guide for version 5.11. (<http://www.satscan.org>).
3. Myers, W.L., Patil, G.P.(2002). Echelon analysis, *Encyclopedia of Environmetrics*, Vol. 2, 583-586.
4. Myers, W.L., Patil, G.P. and Joly, K.(1997). Echelon approach to areas of concern in synoptic regional monitoring, *Environmental and Ecological Statistics*, 4, 131-152.

[2005년 12월 접수, 2006년 2월 채택]