

## K-function Test for the Spatial Randomness among the Earthquakes in the Korean Peninsula

Jangsun Baek<sup>1)</sup>, Jong Sung Bae<sup>2)</sup>

### Abstract

Kim and Baek (2000) tested the spatial randomness for the earthquake occurrence in the Korean Peninsula by using the nearest-neighbor test statistics and empirical distribution functions. The K-function, however, has obvious advantages over the methods used in Kim and Baek (2000), such as it does not depend on the shape of the study region and is an effective summary of spatial dependence over a wide range of scales. We applied the K-function method for testing the randomness to both of the historical and the instrumental seismicity data. It was found that the earthquake occurrences for historical and instrumental seismicity data are not random and clustered rather than scattered.

*Keywords* : Spatial point process, Spatial randomness, K-function, Seismicity data, Intensity.

### 1. 서론

지진의 발생원인으로서 Reid에 의하여 제창된 탄성 반발설이 현재로서는 가장 유력하다. 이 가설에 의하면 지하에 응력이 축적되어 암석의 파괴강도보다 커지면 파괴가 발생하게 되어 단층운동이 일어나며, 이러한 단층운동의 결과가 바로 지진이라고 설명하고 있다. 암석의 파괴실험이나 현재의 지진활동의 양상을 고찰한 바에 의하면, 단층과 같은 구조적인 약선대에는 응력의 집중이 상대적으로 쉽게 일어나 지진발생이 용이하다는 것이 알려져 있다. 단층 중에서 이러한 특성을 가지고 있어 장래의 지진 발생 원인이 될 가능성이 있는 단층을 활성단층(active fault)이라 하며, 지진재해 예측에 있어서 활성단층의 평가는 대단히 중요하다. 본 연구에서는 한반도에서 발생한 지진활동 상태를 역사지진 자료와 계기지진 자료를 사용하여 공간통계학 방법으로 분석함으로써 활성단층의 추출을 위한 기초 자료를 제공하는 것을 목적으로, 한반도에서의 지진이 공간상 무작위적으로 발생하고 있는지, 아니면 군집적인 형태를 취하는지를 검정한다.

---

1) Associate Professor, Information and Telecommunication Research Institute, Department of Statistics, Chonnam National University, Kwang-ju, 500-757, Korea

E-mail : jbaek@chonnam.ac.kr

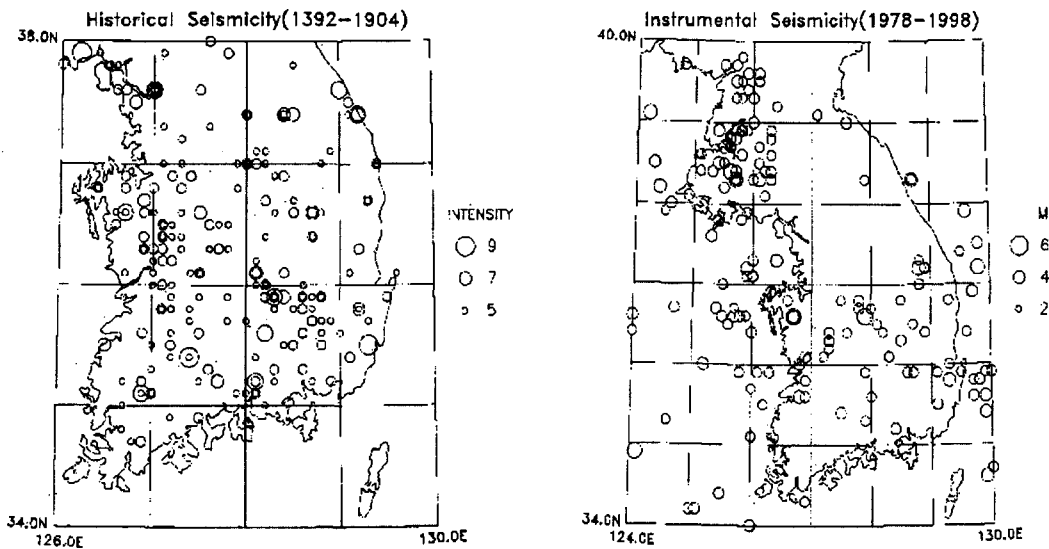
2) Professor, Department of Statistics, Research Institute for Basic Sciences, Chonnam National University, Kwang-ju, 500-757, Korea

E-mail : jsbae@chonnam.ac.kr

한반도의 지진활동과 관련된 연구는 활발하게 진행되어오고 있으며 과거, 현재, 미래에 대한 한반도 지진활동의 개요는 김성균(1997)에 잘 나타나있다. 오충량 외 2인(1995) 역시 과거 및 현재의 지진자료를 활용하여 한반도 지진활동과 응력상태를 연구하였다. 이봉주 외 2인 (1997)은 한반도 특정지역의 선구조선을 영상자료를 분석함으로써 파악하였다. 통계적인 이론을 접목시킨 지진 관련 연구로는 Baag and Shin (1992)의 지진발생에 대한 모의실험 연구가 있고, 확률론적으로 지진위험분포를 추정한 김성균과 송미정(1995)의 연구결과가 있다.

한반도에서 발생한 지진의 진앙지 분포가 다른 지구과학적 정보가 없이 공간적으로 무작위적인지를 통계학적 기법만으로 규명한 연구로는 김성균과 백장선 (2000)이 있다. 그들은 공간자료의 거리에 기반을 둔 검정통계량과 경험분포를 이용하여, 한반도 지진발생이 무작위적이지 않고 군집적임을 밝히고, 또한 지진발생의 공간적 집중도를 비모수적인 방법으로 추정하였다. 김성균과 백장선에서 사용한 방법은 공간 점 과정(spatial point process)의 1차 집중도에 기반을 둔 것이다. K-함수로 표현되는 공간 점 과정의 2차 집중도는 공간 점들 사이의 상호작용 혹은 공간적 의존성이 공간적으로 어떻게 변하는지를 나타내주며, 특히 공간 점들이 무작위적이면 그 함수 값이 알려져 있다. 따라서 본 연구에서는 K-함수를 이용하여 한반도 지진발생이 공간적으로 무작위적인지 혹은 다른 특별한 형태를 가지고 있는지를 밝히고자 한다.

본 연구에서 활용할 자료는 역사지진 자료와 계기지진 자료 두 가지이다. 역사지진 자료는 1392년부터 1904년까지 진도 5이상의 268개 자료이며, 계기지진 자료는 1978년부터 1998년까지 한반도에서 발생한 규모 2 이상의 152개이다. 두 자료 모두 위치를 나타낸 좌표들로 구성되어 있으며 이를 분석에서 활용할 것이다. 두 자료의 진앙지 분포를 지도에 표시하면 [그림 1.1]과 같다. 제 2 장에서는 공간적인 자료에 대한 K-함수를 이용하여 한반도 지진 진앙지 분포가 무작위적인지에 대한 검정을 수행하고, 제 3 장에서는 결론 및 토의를 기술하였다.



[그림 1.1] 역사지진 (a)와 계기지진 (b)의 진앙지 분포

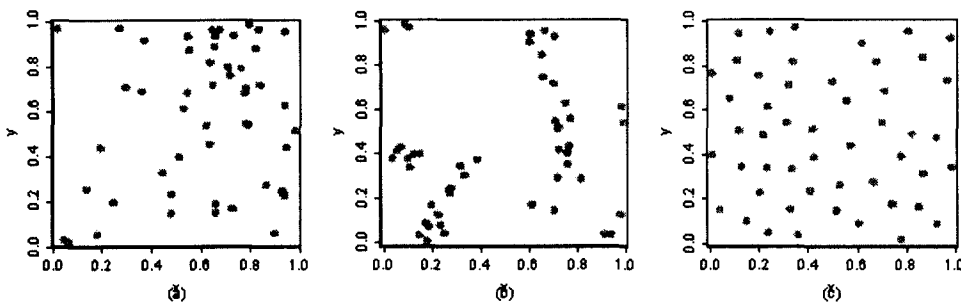
## 2. 진앙지 분포의 공간적 무작위성에 대한 K-함수 검정

일반적으로 공간통계학 (spatial statistics)에서  $s \in R^d$ 을  $d$ 차원 유클리드 공간내의 자료위치라 하고, 공간위치  $s$ 에서의 관심확률변수를  $Z(s)$ 라 할 때,  $R^d$ 의 부분집합  $D$ 에 속해 있는  $s$ 에 대하여  $\{Z(s) : s \in D\}$ 을 공간과정(spatial process)이라 한다. 특히,  $D$ 가  $R^d$ 내의 점 과정(point process)이고 모든  $s \in D$ 에 대하여  $Z(s) = 1$  이면, 위의 공간과정을 공간 점 과정(spatial point process)이라 부른다. 이러한 공간 점 과정은 발생사건(event)의 위치들이 주요연구 대상일 때 주로 적용되는 모형이다. 이때, 우리의 주요 관심사는 발생사건들의 위치가 어떤 패턴을 취하고 있는가 이다. 즉, 사건들이 관심공간상에서 무작위적으로 발생하는지, 혹은 군집적이거나 규칙적으로 발생하는지 밝혀보고자 하는 것이다.

[그림 2.1]에는 여러 가지 형태의 2차원 공간 자료가 예시되어 있다. (a)는 무작위적이며, (b)는 군집적이고 (c)는 규칙적이다. 우리는 지진의 진앙지들이 공간 점 과정을 따를 때 이러한 공간 점 패턴들이 무작위적인지, 아닌지를 과거 지진에 대한 자료를 가지고 통계적으로 검정하고자 한다. 공간자료의 무작위성은 완전 공간 무작위성(complete spatial randomness)을 말한다. 관심지역  $A$ 에 대하여  $N(A)$ 를  $A$ 지역에서 발생한 사건(event)의 갯수라 하고  $s_1, s_2, \dots, s_n$ 을 공간 점 패턴, 즉 진앙지들이라 하면,  $B_1, B_2, \dots, B_n \subset A$ 에 대하여,

$$\Pr(s_1 \in B_1, s_2 \in B_2, \dots, s_n \in B_n \mid N(A) = n) = \prod_{i=1}^n (|B_i| / |A|)$$

이면 공간 점 패턴들은 완전공간 무작위성을 갖는다고 정의한다. 이때,  $|B| = \int_B ds$ 이다. 공간무작위성은 공간점 패턴이 일정한 집중도  $\lambda$ (단위 면적 당 공간 점의 수)를 갖는 동질적 포아송 과정(homogeneous Poisson process)을 따르는 것을 의미한다. 직관적으로 말해서 사건(지진)들이 지역  $A$ 내의 어느 곳에서나 발생할 확률이 모두 같고 사건들끼리 상호작용이 없으면 지진 진앙지 분포는 무작위적인 것이다.



[그림 2.1] (a) 완전 무작위적 공간자료 (b) 군집적 공간자료 (c) 규칙적 공간자료

공간 점들 사이의 상호작용이나 공간적 의존성이 공간을 통해 어떻게 변하는지를 나타내는 방법으로 K-함수 (Ripley, 1976)를 사용한다. K-함수는 거리  $h \geq 0$ 에 대하여 다음과 같이 정의된다:

$K(h) = \lambda^{-1}E[\text{임의의 공간점으로 부터 거리 } h\text{이내에 존재하는 사건들의 수}]$ .

이 때  $\lambda$ 는 집중도이며,  $E[\ ]$ 는 기대값을 나타낸다. K-함수는 공간 점들의 공간적 의존성을 나타내는 공간 점 과정의 2차 집중도(second-order intensity)  $\lambda_2^0(h)$ 와 밀접한 관계를 갖고 있다(2차 집중도의 정의는 Cressie(1991)의 8.3절 참조.) 실제로 정상 등방 점 과정(stationary, isotropic point process)의 2차 집중도  $\lambda_2^0(h)$ 에 대하여 K-함수와의 관계는

$$\lambda_2^0(h) = \frac{\lambda^2 \Gamma(1 + \frac{d}{2})}{d \pi^{\frac{d}{2}} h^{d-1}} K'(h) \quad , h > 0$$

이다.

K-함수를 이용하는 장점은 몇 가지 공간 점 과정 모형에 대하여  $K(h)$ 의 이론적인 값이 알려져 있다는 것이다. 즉  $d=2$ 차원의 동질적 과정이 공간적인 의존성이 없으면  $K(h) = \pi h^2$ 이다. 또한 공간 점 과정이 군집적이라면 서로간 짧은 거리로 많은 점들이 분포하리라 예상되며, 따라서 작은  $h$ 에 대하여  $K(h) > \pi h^2$ 이다. 마찬가지로 과정이 규칙적으로 분포되었다면  $K(h) < \pi h^2$ 임을 예상할 수 있다. 일반적으로 K-함수에 대한 추정량은 기대값을 경험적인 평균치로 대체하여 구성된다.  $(s_1, s_2, \dots, s_n)$ 을 공간 점 즉 진앙지들이라 하면, 관심지역  $A$ 에 대하여  $n = N(A)$ 는 그 지역에서 발생한 진앙지들의 총 개수이다.  $(d_1, d_2, \dots, d_n)$ 을 각 진앙지로부터 지역  $A$ 의 가장 가까운 경계까지의 거리들이라 하자.

K-함수에 대하여 첫 번째로 고려해 볼 수 있는 추정량으로 다음과 같은  $\widehat{K}_1$ 이 있다.

$$\widehat{K}_1(h) = \lambda^{-1} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n I(\|s_i - s_j\| \leq h) / n.$$

경계효과를 보정할 수 있는 다른 추정량으로는

$$\widehat{K}_2(h) = \lambda^{-1} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n I(\|s_i - s_j\| \leq h, d_j > h) / \sum_{j=1}^n I(d_j > h).$$

이 있다. 경계효과라 함은 관측점이 관심지역  $A$ 내에서만 수집된 것이므로  $A$  지역 내 경계선 근처의 자료와 더 가까운 자료가  $A$ 지역 밖에 존재할 수 있으므로  $A$  지역 내 경계선 근처의 최근접거리가  $A$ 지역의 중앙에 위치한 그것보다 평균적으로 더 크게되는 상황을 의미한다. 마지막으로 경계효과도 보정하고  $d_j \leq h$ 인 공간 점들에 대한 정보도 활용한 추정량  $\widehat{K}_3$ 이 Ripley(1976)

에 의해 제안되었다:

$$\widehat{K}_3(h) = \lambda^{-1} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \omega(s_i, s_j)^{-1} I(\|s_i - s_j\| \leq h) / n.$$

이 때  $\omega(s_i, s_j)$ 는  $s_i$ 을 중심으로  $s_j$ 을 지나는 원의 원주가 관심지역  $A$ 이내에 속한 비율을 나타낸 것이다. 위의 모든 추정량들은 집중도  $\lambda$ 의 추정 값  $\hat{\lambda}$ 이 필요하며, 보통  $\hat{\lambda} = N(A)/|A| = n/|A|$ 로 계산된다. 본 연구에서는 Ripley (1976)의 추정량  $\hat{K}(h) = \hat{K}_3(h)$ 을 이용하여 한반도 지진자료의 공간적 무작위성을 검정하였다.

완전 공간 무작위성 하에서  $K(h)$ 의 분산을 안정화시키기 위해,  $K(h)$ 을 제곱근 변환하여 구한  $L(h) = \sqrt{\frac{K(h)}{\pi}}$ 을 이용할 수도 있다. 동질적 포아송 과정 하에서는  $K(h) = \pi h^2$ 이므로  $L(h)$ 는

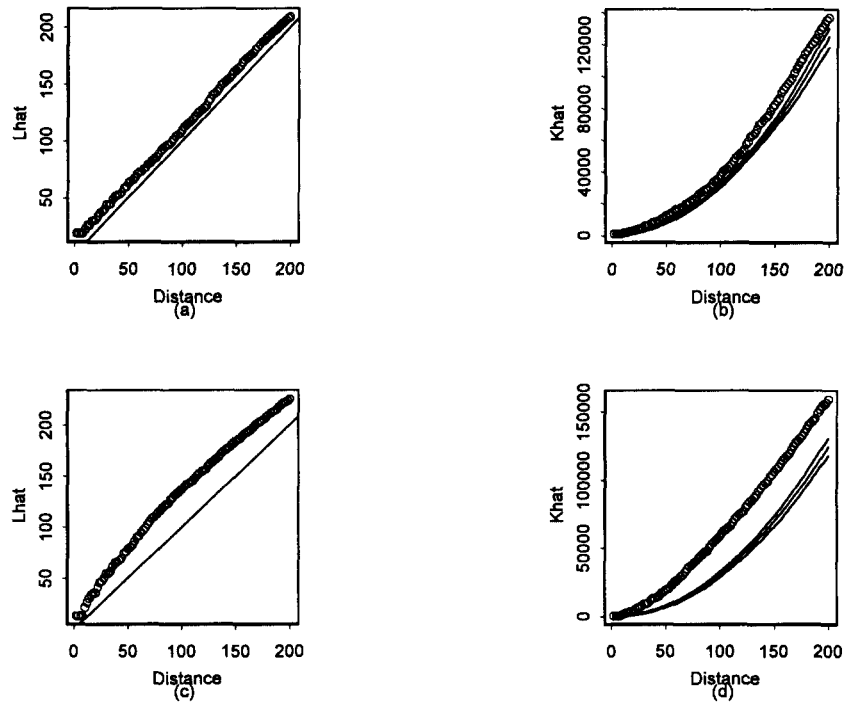
$L(h) = h$ , 즉 원점을 통과하는 기울기 1의 직선이다. 따라서  $\hat{L}(h) = \sqrt{\frac{\hat{K}(h)}{\pi}} > h$ 이면 공간자료는 균 집적임을 나타내고, 반대로  $\hat{L}(h) = \sqrt{\frac{\hat{K}(h)}{\pi}} < h$ 이면 규칙적임을 알 수 있다.

분석대상 자료인 역사지진과 계기지진의 진앙지 분포의 지도가 [그림 1.1]에 나타나 있으나, 실제 분석에서는 위도와 경도 좌표로 표시된 [그림 1.1]의 자료를 각각 좌측 하단 모서리를 원점으로 하여 실제거리(km)로 환산된 좌표 자료로 바꾼 후 사용하였다. 또한 진앙지가 기록되지 않은 지역을 최대한 배제하기 위하여 모든 진앙지 좌표자료를 포함하는 최소한의 사각형을 분석대상지역  $A$ 로 확정하였다.

먼저 역사지진의 경우  $N(A) = 268$ 개의 진앙지 자료에 대하여 집중도는  $\hat{\lambda} = N(A)/|A|$

$= 0.0024$ 로 계산되었다. 200 km까지 등 간격의 100개의  $h$ 값에 대하여  $\hat{L}(h)$ 를 계산하여 [그림 2.2]의 (a)에 절편 0, 기울기 1인 직선과 함께 그렸다.  $\hat{L}(h)$ 가  $h$  전 영역에 걸쳐  $\hat{L}(h) > h$ 임을 확인할 수 있고, 따라서 역사지진의 진앙지 분포는 균집적임을 알 수 있다. [그림 2.2]의 (b)에는  $\hat{K}(h)$ 과 함께 실제자료를 이용하여 무작위성 하에 모의실험한 모의실험 외피(simulation envelopes)를 그려서 비교하였다. 즉 실제자료와 같은 집중도를 갖는  $n = 268$ 개의 무작위적인 공간 점들을  $A$  지역 내에 여러 번 생성하여 각각  $\hat{K}(h)$ 들을 계산한 후 그것들의 평균을 실제자료에 대한  $\hat{K}(h)$ 과 비교를 위한 기준으로 삼고, 또한 최대, 최소를 구하여 그것을 외피(envelope)로 활용한다. 본 연구대상 자료들에 대하여 100번의 모의실험을 실시하였다. 실제자료에 대한  $\hat{K}(h)$ 는 모의실험을 통한 평균치보다 상위에 위치하고 있으며, 더욱이 모의실험 외피 밖에 대부분 위치함으로써 자료가 무작위적이지 않음을 확실히 보여준다.  $\hat{L}(h)$ 와  $\hat{K}(h)$ 의 계산은 Kaluzny 등 (1998)의 S+SPATIAL STATS의 Lhat, Khat module을 사용하였다.

계기지진의 경우에는  $N(A) = 152$ 개의 진앙지 자료에 대하여 위에서와 동일한 절차에 따라 분석되었다.  $\hat{L}(h)$ 과  $\hat{K}(h)$ 를 계산하기 위한  $h$ 는 역시 200 km까지 등 간격의 100개 값이 선택되었다.  $\hat{L}(h)$ 와  $\hat{K}(h)$  모두 모든  $h$ 에 대하여 무작위성 기준을 뚜렷이 초과하므로 균집적임을 확인할 수 있다. 계기지진의 경우 역사지진에 비해  $\hat{L}(h)$ 와  $\hat{K}(h)$ 가 모두 무작위 기준을 더 초과하므로 더욱 균집적인 분포를 하고 있음을 확인할 수 있다.



[그림 2.2] (a) 역사지진의  $h(\text{Distance})$ 에 대한  $\hat{L}(h)$ , (b) 역사지진의  $\hat{K}(h)$ ,  
 (c) 계기지진의  $h(\text{Distance})$ 에 대한  $\hat{L}(h)$ , (d) 계기지진의  $\hat{K}(h)$ ,  
 \* (b)와 (d)에서 상위실선 : 무작위 모의실험자료의  $\hat{K}(h)$  최대 외피,  
 중간실선 : 무작위 모의실험자료의  $\hat{K}(h)$  평균, 하위실선 : 무작위 모의실험  
 자료의  $\hat{K}(h)$  최소 외피

### 3. 결론 및 토의

본 연구에서는 한반도에서의 지진발생이 무작위적인지를 K-함수라는 공간통계학 방법을 이용하여 검정하였다. 역사지진과 계기지진의 두 가지 분석대상 자료에 대하여 두 자료 모두 무작위적이지 않고 군집적인 분포를 가지고 있는 것으로 판명되었다. 이것은 진앙지 사이의 최 근접 거리에 근거한 검정통계량과 경험분포를 이용하여 무작위성을 검정한 김성균과 백장선의 결과와 동일하다. 그러나 K-함수를 이용한 방법은 그들의 방법에 비해 관심지역  $A$ 의 모양에 의존하지 않고, 또한 폭넓은 거리 측도에 대응하는 공간적 의존성을 효율적으로 요약할 수 있다는 장점을 갖고 있다.

## 감사의 글

본 연구에서 사용된 역사지진과 계기지진 자료를 제공해 주신 전남대 지구환경과학부 김성균 교수님께 감사를 드립니다.

## 참고문헌

- [1] 김성균 (1997). 한반도의 지진활동 : 과거, 현재, 미래, 「콘크리트학회지」, 제 9권 1호, 4-12.
- [2] 김성균, 백장선 (2000). 한반도 지진발생의 무작위성에 대한 통계적 검정과 집중도 추정, 「한국지구과학회지」, 제 21권 2호, 159-167.
- [3] 김성균, 송미정 (1995). 한반도의 확률론적 지진위험분포, 「The Journal of Engineering Geology」, 제 5권 1호, 45-58.
- [4] 오충량, 김소구, 고복춘 (1995). 과거 및 현재 지진 Data로부터 한반도 지진활동과 응력 상태, 「The Journal of Engineering Geology」, 제 5권 3호, 309-329.
- [5] 이봉주, 최위찬, 강필종 (1997). 한반도 동남부지역의 선구조선, 「지질학회지」, 제 33권 1호, 18-26.
- [6] Baag, C. E. and Shin, J. S. (1992). Monte-Carlo simulation of earthquake sequence in the time and magnitude space, *The Journal of Engineering Geology*, Vol. 2, 147-154.
- [7] Cressie, N. (1991). *Statistics for Spatial Data*, John Wiley & Sons, New York.
- [8] Kaluzny, S. P., Vega, S. C., Cadoso, T. P. and Shelly, A. A. (1998). *S+SPATIAL STATS*, Springer-Verlag, New York.
- [9] Ripley, B. D. (1976). The second-order analysis of stationary point processes, *Journal of Applied Probability*, Vol. 13, 255-266.