

Analyzing landslide data using Cauchy cluster process

Kise Lee^a · Jeonghwan Kim^a · No-wook Park^b · Woojoo Lee^{a,1}

^aDepartment of Statistics, Inha University;

^bDepartment of Geoinformatic Engineering, Inha University

(Received December 28, 2015; Revised January 26, 2016; Accepted January 26, 2016)

Abstract

Inhomogeneous Poisson process models are widely applied to landslide data to understand how environmental variables systematically influence the risk of landslides. However, those models cannot successfully explain the clustering phenomenon of landslide locations. In order to overcome this limitation, we propose to use a Cauchy cluster process model and show how it improves the goodness of fit to the landslide data in terms of K -function. In addition, a numerical study is performed to select the optimal estimation method for the Cauchy cluster process.

Keywords: landslide, point pattern data, Poisson process, Cauchy cluster process

1. 서론

산사태 취약성 분석(landslide susceptibility analysis)의 목적은 특정 지역에서 미래의 산사태에 취약한 지점을 찾는 것이며, 이는 대상 지역의 토지 이용 및 관리를 위해 필수적이다. 보통 산사태 취약성 분석은 과거 산사태 발생 지역과 산사태 발생과 관련이 있는 환경 변수와의 연관성 분석을 통해 수행되는데, 산사태 발생 지역을 1로 표현하고 발생하지 않은 지역을 0으로 표현한 후 로지스틱 회귀(logistic regression) 모형 등이 널리 사용되어왔다. 그러나 0으로 표현된 지역은 현재까지 산사태가 발생하지 않은 지역일 뿐, 앞으로 영원히 산사태가 발생하지 않는 지역이라는 뜻을 갖는 것은 아니다. 따라서 로지스틱 회귀분석은 잠재적으로 산사태가 발생할 수 있는 많은 곳을 마치 산사태가 발생하지 않는 지역으로 처리하기 때문에 올바른 분석 모형이라 보기 어렵다.

이와는 달리, 산사태가 발생한 곳의 위치만을 확률 과정(stochastic process)의 실현된 값으로 보고 분석하는 것이 가능한데, 특히 포아송 과정(Poisson process)과 그것의 변형이 널리 사용되고 있다. 최근에는 산사태 발생 위험이 해당 지역에서 얻어지는 환경변수에 따라 달라진다고 보고, 이를 모형화하는 비동질적 포아송 과정(inhomogeneous Poisson process)이 여러 연구자들에게 관심을 받았다 (Tonini 등, 2013).

The research was supported by a grant [MPSS-NH-2015-79] through the Natural Hazard Mitigation Research Group funded by Ministry of Public Safety and Security of Korean government, and the work by No-Wook Park was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2015R1A1A1A05000966).

¹Corresponding author: Department of Statistics, Inha University, 100 Inha-ro, Nam-Gu, Incheon 22212, Korea. E-mail: lwj221@gmail.com

The first two authors contribute equally.

주어진 산사태 자료에 대해 비동질적 포아송 과정이 적합한지를 판단하기 위해 K -함수를 사용할 수 있다 (Ripley, 1976). 비동질적 포아송 과정으로부터 추정된 K -함수와 모형에 대한 가정 없이 비모수적 방법으로 추정된 K -함수를 비교하여 두 K -함수의 개형이 비슷한 경향을 보인다면 적합한 모형이 자료를 잘 설명한다고 판단한다. 그러나 실제 산사태 자료는 과거 산사태가 발생한 지점을 주변으로 산사태가 집중적으로 발생하는 특징이 기대되므로, 산사태가 발생한 위치에 군집 현상이 나타날 수 있다. 이러한 군집 현상은 비동질적 포아송 과정 모형에 충분한 환경변수가 고려되더라도 해당 군집현상을 설명하는 것에는 한계가 있을 수 있다. 따라서 이러한 경우는 군집된 산사태 발생과정을 모형화할 수 있는 군집 과정(cluster process)을 사용하는 것이 보다 더 타당하다고 볼 수 있다 (Waagepetersen, 2007). 대표적인 군집 과정으로 네이만-스콧 과정(Neyman-Scott process)이 있는데, 코시 군집 과정(Cauchy cluster process)과 토마스 과정(Thomas process) 등을 특수한 경우로 갖는다 (Ghorbani, 2012; Tanaka 등, 2008). 여기서 코시 군집 과정은 군집 과정에 사용되는 커널 함수가 코시 분포(Cauchy distribution)를 따르는 모형으로, 정규 분포를 커널로 사용하는 토마스 과정에 비해 넓은 범위의 군집을 설명하는데 좀 더 효과적으로 알려져 있다.

그러나 아직까지 산사태 자료에 대해 군집 과정을 이용하여 분석하는 접근은 거의 이루어지지 않고 있다. 이에 본 논문에서는 충청북도 보은 지역의 산사태 자료를 이용하여 군집 과정을 적용하는 사례연구를 수행하였다. 특히 산사태 취약성 분석을 위한 도구로 코시 군집 과정에 대해 소개하고, K -함수를 이용하여 모형의 적합 정도를 판단하는 방법을 설명하고자 한다. 또한 코시 군집 과정에 사용되는 여러가지 모수 추정방법에 대해 아직 체계적인 비교연구가 이루어지지 않았기 때문에 본 연구에서는 시뮬레이션된 자료를 통해 수치연구를 진행하였다.

본 논문의 구성은 다음과 같다. 2절에서는 비동질적 포아송 과정의 특성과 K -함수에 대해 살펴보고, 이를 바탕으로 3절에서 보은 지역 산사태 자료를 분석하였다. 비동질적 포아송 과정이 갖는 모형의 한계점을 논의한 후, 이를 극복하기 위한 방법으로 코시 군집 과정을 제안하고 세부적인 내용을 4절에서 살펴본다. 코시 군집 과정의 모수를 추정하기 위한 다양한 방법의 성능을 비교하기 위해 모의 실험을 5절에서 진행한 후, 이를 기반으로 실제 산사태 자료에 코시 군집 과정을 적용한 결과를 6절에서 살펴보고, 최종 결론은 7절에 제시하였다.

2. 비동질적 포아송 과정

먼저 비동질적 포아송 과정의 특성에 대해서 살펴보고, 모형의 적합과 관련하여 K -함수의 정의와 의미에 대해 알아본다.

2.1. 비동질적 포아송 과정의 특성

먼저 s 는 2차원 상의 위치정보(지점)를 나타내고, W 는 관심있는 전체 영역을, A 는 W 내의 특정 영역을 뜻한다고 하자. Baddeley (2008)는 $N(A)$ 를 영역 A 에서의 사건 발생 횟수를 나타내는 확률변수로 정의하고, 이 변수가 $\int_A \lambda(s)ds$ 를 기대값으로 갖는 포아송 분포를 따른다고 가정하였다. 이를 기호로 나타내면

$$N(A) \sim \text{Poisson}\left(\int_A \lambda(s)ds\right), \quad A \subseteq W$$

이 된다. 여기서 $\lambda(s)$ 는 사건 발생의 세기를 나타내는 강도 함수(intensity function)로 생각할 수 있다. 또한 서로 겹치지 않는 W 내의 특정 영역 A 와 B 의 사건 발생 횟수는 서로 독립으로 가정된다. 즉 $A \cap B = \phi$ 라면, $N(A)$ 와 $N(B)$ 는 독립이다. 환경변수들이 사건의 발생 빈도에 영향을 주는 부분을 모

형화하기 위해서 일반적으로 $\log(\lambda(s))$ 를 사건 발생에 영향을 주는 환경변수들의 선형결합으로 정의하게 되는데, $z(s)$ 를 s 에서 얻어진 환경변수들이라 하고, β 를 그와 연관된 회귀 계수라 할 때,

$$\log \lambda(s) = z(s)^T \beta$$

로 모형화 된다. 특히 $\lambda(s) = \lambda$, 즉 s 에 따라 값이 변화하지 않는 상수인 경우 동질적 포아송 과정(homogeneous Poisson process)이 된다.

2.2. 비동질적 포아송 과정의 K -함수

Ripley (1976)는 K -함수를 공간 내의 점들 사이에 공간적 의존성이 어떻게 변하는지 나타내기 위해 도입하였으며, 정상 과정(stationary process)에 대해 거리 r 에 따라 다음의 K -함수($K(r)$)

$$K(r) = \lambda^{-1} E[\text{임의의 사건 발생 지점으로부터 거리 } r(> 0) \text{ 내에 존재하는 사건의 수}]$$

로 정의하였다. 여기서 λ 는 동질적 포아송 과정의 강도 함수를 E 는 기댓값을 의미한다. 2차원 공간 상에서 포아송 과정인 경우 이론적인 K -함수의 형태가 계산 가능하며, 동질적인지 비동질적인지에 관계없이 K -함수는 다음과 같이 표현됨이 알려져 있다 (Waagepetersen, 2007).

$$K_{IPP}(r) = \pi r^2.$$

이는 비동질적 포아송 과정이 동질적 포아송 과정과 마찬가지로 교집합이 없는 두 영역에서의 사건 발생 빈도는 서로 독립을 가정하고 있기 때문에 얻어지는 결과이다.

적합도 비교를 위해 사용되는 비모수 K -함수에 관한 추정량으로는 정상과정에 대해 경계효과를 보정한 추정량이 처음 Ripley (1976)에 의해 제안되었고, 이후 Baddeley 등 (2000)에 의해 비정상 과정(non-stationary process)으로 확장되었다. 경계효과는 관측점이 관심 지역 W 내에서 수집된 것으로 관심 지역 경계선 근처의 자료가 관심 지역 밖의 자료와 더 가까울 수 있다는 점을 고려하여 이를 보정한 것이다.

먼저, (s_1, s_2, \dots, s_n) 을 사건이 발생한 n 개의 지점이라고 할 때, $e(s_i, s_j)$ 를 경계효과를 고려한 가중치라고 하고, $\|s_i - s_j\|$ 를 두 사건 발생 지점의 유클리디언 거리(Euclidean distance)라 하자. 이 때, 비모수 K -함수의 추정치는 다음과 같이 주어진다.

$$\hat{K}(r) = \sum_i \sum_j \frac{1[0 < \|s_i - s_j\| \leq r] e(s_i, s_j)}{\hat{\lambda}_{Ker}(s_i) \hat{\lambda}_{Ker}(s_j)}.$$

경계 효과 보정치에 대한 자세한 사항은 R의 spatstat 패키지의 Kinhom이라는 함수의 도움말에 자세히 설명되어있다. $\hat{\lambda}_{Ker}(s_i)$ 는 커널을 이용한 비모수 추정기법으로부터 얻어진 강도 함수의 추정치이다. 포아송 과정은 서로 교집합이 없는 두 지역 사이에 사건 발생 횟수는 독립이라고 가정하였기 때문에, 만약 $\hat{K}(r) > \pi r^2$ 이게 되면 가까운 거리 내에 독립을 가정한 경우보다 많은 점들이 분포한 것이라 볼 수 있고 따라서 점들이 서로 군집해 있음을 알 수 있다. 이와는 반대로, $\hat{K}(r) < \pi r^2$ 인 경우에는 공간 상의 점들이 서로 밀어내고 있는 상황이 된다.

3. 비동질적 포아송 과정을 이용한 보은 지역 산사태 자료 분석

본 연구에서는 1998년 여름 집중호우의 피해를 받아 산사태가 많이 발생한 충청북도 보은 지역을 대상으로 분석을 시행하였다. 보은 지역 산사태 자료에 비동질적 포아송 과정을 적합하여 보고 이를 K -함수의 관점에서 평가해보고자 한다.

Table 3.1. Model selection using AIC

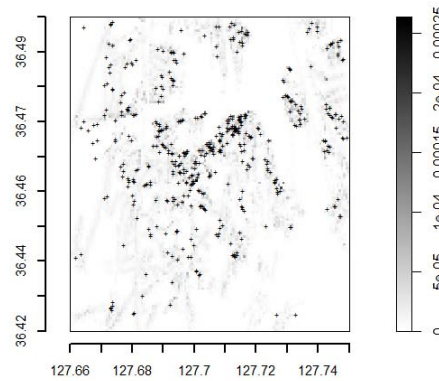
Formula	AIC
~ 1	12333.02
$\sim \text{distance}$	11829.95
$\sim \text{distance} + \text{slope}$	11590.39
$\sim \text{distance} + \text{slope} + \text{elevation}$	11534.32
$\sim \text{distance} + \text{slope} + \text{elevation} + \text{slope:elevation}$	11496.33
$\sim \text{distance} + \text{slope} + \text{elevation} + \text{slope:elevation} + \text{distance:slope}$	11494.55

AIC = Akaike information criterion

Table 3.2. The regression coefficient estimates for the selected inhomogeneous Poisson process model

	$\hat{\beta}$	s.e.($\hat{\beta}$)	Z-value	Z-test
(Intercept)	-13.43752	0.52863	-25.41955	***
Slope	0.22116	0.02387	9.26495	***
Elevation	0.00705	0.00253	2.79103	**
Distance	-0.25911	0.03749	-6.91096	***
Slope: Elevation	-0.00069	0.00011	-6.18008	***
Slope: Distance	0.00297	0.00153	1.94300	.

(Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1)

**Figure 3.1.** The estimated intensities of inhomogeneous Poisson process: Boeun data

보은 지역 자료의 총 격자 수는 290,068개이고, 481곳에서 산사태가 발생한 것으로 관측되었다. 보은 지방의 산사태 자료를 분석하기 위하여 선행 연구 (Park, 2015)를 고려하여 경사(slope), 고도(elevation) 그리고 선구조로부터의 거리(distance)를 산사태 발생 환경 요인, 즉 설명변수로 사용하였다. 또한 이들의 모든 2차 교호작용을 고려하였다. 일반적으로 경사가 급할수록, 고도가 높을수록 그리고 선구조로부터 가까운 지역일수록 산사태에 더 취약한 것으로 알려져 있다 (Park 등, 2003; Park, 2015).

비동질적 포아송 과정을 이용하여 보은 지역의 산사태 자료를 분석한 후 Akaike information criterion(AIC)을 이용하여 모형을 비교하였다. 변수 선택 결과는 Table 3.1과 같다. AIC를 기준으로 단계적 방법을 통해 선택된 변수는 경사, 고도, 선구조로부터의 거리 그리고 경사와 고도의 교호작용, 경사와 선구조로부터의 거리의 교호작용이다. 선택된 변수로 적합된 계수 결과는 Table 3.2에 나타나 있다.

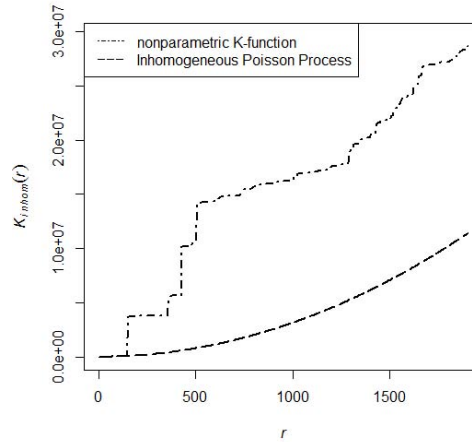


Figure 3.2. Comparison of two K -functions: nonparametric and inhomogeneous Poisson process.

Figure 3.1은 보은 지역의 산사태 자료에 대해 선택된 비동질적 포아송 과정의 강도 함수 추정치를 실제 산사태 발생지점과 같이 표시한 것이다. 그림에서 강도 함수의 값이 높아질수록 색깔이 진하게 나타나는데, 실제 산사태가 발생한 지역 주변에서 강도 함수의 값이 높은 경향이 관측된다.

K -함수를 이용하여 모형의 적합도를 살펴보기 위해 비모수적 방법으로 추정된 K -함수와 비동질적 포아송 과정으로부터 얻어지는 이론적 K -함수를 비교하였다. 이 두 K -함수가 비슷하다면, 해당하는 비동질적 포아송 과정은 두 산사태 발생지역 사이의 공간적 의존성을 잘 설명한다고 말할 수 있다 (Baddeley, 2008). K -함수의 비교결과는 Figure 3.2에서 주어지는데, 비모수 K -함수와 비동질적 포아송 과정의 K -함수가 큰 차이를 보이며, 비모수 K -함수가 위쪽에 위치하는 것을 알 수 있다. 따라서 산사태 발생 지점 사이에는 비동질적 포아송 과정으로는 설명되지 않는 강한 군집 현상이 존재한다고 볼 수 있으므로, 군집 과정을 고려할 필요성이 있는 것으로 생각된다.

4. 네이만-스콧 과정과 코시 군집 과정

이 절에서는 먼저 대표적인 군집과정인 네이만-스콧 과정에 대해 살펴본다. 그 후에, 코시 커널을 커널 함수로 사용하는 네이만-스콧 과정인 코시 군집 과정에 대해 자세히 살펴보고자 한다.

군집 과정을 생성하는 기본 아이디어는 두 개의 과정, 부모 과정(parent process)과 자손 과정(offspring process)을 고려하는 것이다. 먼저 부모 과정이 실현되고, 각 부모 과정의 실현된 곳 주변에 자손 과정을 생성함으로써 군집 현상을 자연스럽게 설명하는 것이 가능하다. 그 중 실제 분석에 널리 사용되는 네이만-스콧 과정을 Waagepetersen (2007)은 아래와 같이 정의하였다.

- 1) 부모 과정 C 는 강도 함수가 κ 인 동질적 포아송 과정을 따른다.
- 2) 부모 과정으로부터 생성된 각 점이 주어졌을 때, 자손 과정 S_c 는 그 점들을 중심으로 (1) 자손의 수를 평균이 $M\alpha$ 인 포아송 분포로부터 생성한 후, (2) 부모로부터 떨어진 거리는 독립이고 동일한 분포를 따르며, 밀도함수는 $k_\sigma(s - c)$ 으로 주어진다고 가정한다. 따라서, S_c 는 강도함수가 $M\alpha k_\sigma(s - c)$ 인 확률 과정이다.
- 3) 각 자손에 대해 $\exp(z^T(s)\beta)/M$ 의 확률로 보유하고, 나머지 확률로는 사라지게 한다.

이러한 방식으로 얻어진 살아남은 자손 과정의 집합을 네이만-스콧 과정이라고 한다. 여기서 $k_\sigma(s - c)$ 를 네이만-스콧 과정의 커널 함수라 하며 부모 과정으로부터 생성된 각 점인 c 를 중심으로 하면서 σ 를 모수로 가지고 있다. 보통 σ 는 중심으로부터 떨어진 거리에 따라 커널 함수 값이 얼마나 빨리 감소하는지 결정해 주는 역할을 한다. 이 때 커널의 정의에 따라 다양한 네이만-스콧 과정이 정의된다. 특히 커널함수가 정규 분포일 경우 토마스 과정이라고 하며, 코시 분포일 경우 코시 군집 과정이라고 한다. 정규 분포보다 코시 분포가 훨씬 더 두꺼운 꼬리를 가지고 있으므로 자손 과정이 분포할 수 있는 범위는 코시 군집 과정의 경우 훨씬 넓어지게 되어 실제 자료 분석에 이점을 갖는다.

위의 세단계를 종합하여 얻어지는 네이만-스콧 과정의 강도 함수는

$$\lambda_c(s) = \alpha \kappa \exp(z(s)^T \beta) \quad (4.1)$$

으로 주어짐이 잘 알려져있다 (Waagepetersen, 2007). 대부분의 문제에서 주된 관심은 β 의 추론에 있으며, κ, α 등은 장애 모수(nuisance parameter)로 처리된다.

본 연구에서는 특히 코시 커널을 사용한 네이만-스콧 과정, 즉 코시 군집 과정을 실제 분석에 사용하고 자 한다. 코시 커널의 함수 형태는

$$k_\sigma(s - c) = \frac{1}{2\pi\sigma^2} \left[1 + \left(\frac{\|s - c\|}{\sigma} \right)^2 \right]^{-\frac{3}{2}}$$

으로 주어진다. 코시 군집과정의 적합은 R의 spatstat 패키지에 kppm 함수로 구현되어 있다.

한편, 코시 군집 과정의 K -함수는 Ghorbani (2012)에 의해 계산되었는데, 그 형식이

$$K_{CCP}(r; \kappa, \sigma) = \pi r^2 + \frac{1 - (1 + r^2/4\sigma^2)^{-\frac{1}{2}}}{\kappa}$$

으로 주어진다. 코시 군집 과정의 K -함수는 비동질적 포아송 과정의 K -함수인 πr^2 뒤에 항상 양수인 항을 가지고 있기 때문에 $K(r) > \pi r^2$ 이 되어 자료에 나타나는 군집 현상을 설명하는 것이 가능하다.

5. 코시 군집 과정의 모수 추정 방법의 비교

일반적으로 군집 과정의 모수를 추정하는 일은 매우 어렵다. 그 이유는 Tanaka 등 (2008)에서 밝히고 있듯이, 부모 과정으로부터 실현된 값이 관측되지 않고, 한 부모로부터 생성된 군집과 다른 부모로부터 생성된 군집이 서로 겹치는 경우가 많기 때문에 가능도 함수를 구성하기 어렵기 때문이다. 이러한 어려움을 극복하기 위한 다양한 방법들이 제안되어 왔는데 이를 간단히 살펴보고, 코시 군집 과정에 가장 적합한 방법을 선택하기 위한 수치 연구를 진행하여 본다.

코시 군집 과정의 계수를 추정하는 방법에는 대표적으로 최소 대비 방법(minimum contrast method), 가중 1차 복합 가능도 방법(weighted first order composite likelihood method), 준가능도 방법(quasi-likelihood method)을 꼽을 수 있다. 먼저 최소 대비 방법이란 두 단계로 구성되는 단순한 방법인데, 첫 번째 단계에서 식 (4.1)에 주어진 강도 함수를 갖는 비동질적 포아송 과정의 가능도 함수로부터 β 를 추론한다. 구해진 β 를 가지고, 두 번째 단계에서 나머지 모수에 대해 이론적인 K -함수와 경험적 K -함수 사이 거리를 최소화하는 α 와 κ 값을 사용하는 방법인데 세부적인 내용은 아래와 같다. 먼저 경험적 K -함수는

$$\hat{K}_{emp}(r) = \sum_i \sum_j \frac{1[0 < \|s_i - s_j\| \leq r]e(s_i, s_j)}{\exp((z(s_i) + z(s_j))^T \beta)}.$$

Table 5.1. Comparison of MSE for the three estimation methods

Estimation method	MSE
Minimum contrast	0.5348
Weigthed first order composite likelihood	0.3752
Quasi-likelihood	0.3541

MSE = mean squared error.

으로 주어지고, 적절한 상수 a 에 대해

$$\int_0^a \left(\widehat{K}_{emp}(r)^{\frac{1}{4}} - K(r; \kappa, \alpha)^{\frac{1}{4}} \right)^2 dr$$

을 최소화 하는 κ 와 α 를 사용하는 방법이다. 이러한 방식으로 얻어진 추정량의 점근적 성질에 대해서는 Waagepetersen (2007)에서 자세히 논하고 있다.

β 의 추론에 사용되는 포아송 가능도 함수는 사건의 발생 지점 사이의 공간적 의존성에 대한 정보를 무시한다는 단점이 있기 때문에 이를 극복하고자 다양한 시도가 있었는데, Guan과 Shen (2010)에서는 포아송 분포로부터 얻어지는 스코어 함수에 특별한 가중치를 도입하였다. 여기서 가중치는 β 의 추정치가 갖는 점근적 분산이 최소가 되도록 선택하였다.

가장 최근에 제안된 방법인 준가능도 방법은 Guan 등 (2015)에 의해 제안되었는데, 이론적으로 가장 건전한 방법으로, 포아송 가능도로 부터 유도되는 스코어 함수의 형태를 갖는 모든 추정 방정식(estimating equation) 가운데 가장 효율적인(efficient) β 의 추정치를 주는 방법이다. 또한, Guan과 Shen (2010)은 준가능도 방법의 한 단순한 근사에 해당함을 밝혔다.

본 논문에서는 세 종류의 계수 추정 방법의 실제 성능을 시뮬레이션 연구를 통해 진행하였다. 코시 군집 과정으로 부터 자료를 생성하기 위해서 spatstat의 rCauchy 함수를 이용하였는데, 먼저 강도 함수를 지정하는 것이 필요하다. 본 시뮬레이션 연구에서는 $\kappa = 30$, $\sigma = 0.01$, $\log \alpha = 3$ 으로 하였다. 또한, 2차원 공간 $(0, 1) \times (0, 1)$ 위에서 설명변수는 $z_1(s) = \exp(x - 0.5)$ 이 되도록 생성한 후, 식 (4.1)의

$$\exp(z(s)^T \beta) = \exp(\beta_1 z_1(s))$$

에서 $\beta_1 = 3$ 으로 하였다. 위의 코시 군집 과정으로 부터 자료를 생성하였고, 매 생성된 데이터셋에 kppm 함수를 이용하여 최소 대비 방법을, improve.kppm 함수를 이용하여 가중 1차 복합 가능도 방법과 준가능도 방법을 실행하였다. β_1 의 평균제곱오차(mean squared error; MSE)를 추정방법 간 성능을 비교하는 측도로 하였으며, 평균제곱오차의 계산을 위해 500회의 반복된 시뮬레이션 결과를 바탕으로 $\sum_{i=1}^{500} (\hat{\beta}_1 - 3)^2 / 500$ 을 구하였다. 시뮬레이션 결과는 Table 5.1과 같다.

시뮬레이션 결과 준가능도 방법이 가장 작은 평균제곱오차를 가지는 것으로 나타났다. 가중 1차 복합 가능도 방법과의 평균제곱오차의 차이는 그리 크지 않은 것으로 나타났지만, 추가적으로 고려한 다른 시뮬레이션 세팅 하에서도 준가능도 방법이 지속적으로 작은 평균오차를 가짐을 확인할 수 있었다. 이는 코시 군집 과정에서도 이론적으로 예측된 결과가 수치적으로 잘 나타나고 있음을 확인한 것이다. 따라서 다음 절의 실제 사례 분석에서는 강도함수를 가장 정확하게 추정할 것이라 기대되는 준가능도 방법을 적용하기로 하였다.

6. 코시 군집 과정을 이용한 보은 지역 산사태 자료 분석

Figure 6.1은 보은 지역의 산사태 자료에 대해 코시 군집 과정으로 추정된 강도함수의 값과 실제 산사태 발생지점을 같이 표시한 것이다. 그림에서 색이 진할수록 강도함수의 값이 높음을 의미한다.

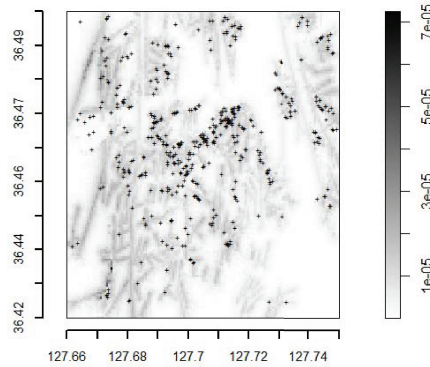


Figure 6.1. The estimated intensities of Cauchy cluster process: Boeun data.

Table 6.1. The regression coefficient estimates of Cauchy cluster process

	$\hat{\beta}$	s.e.($\hat{\beta}$)	Z-value	Z-test
(Intercept)	-13.57611	1.12002	-12.12133	***
Slope	0.08290	0.03262	2.54118	*
Elevation	0.01000	0.00469	2.13186	*
Distance	-0.16907	0.03077	-5.49398	***
Slope: Elevation	-0.00033	0.00014	-2.37997	*
Slope: Distance	0.00270	0.00142	1.89475	.

(Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1).

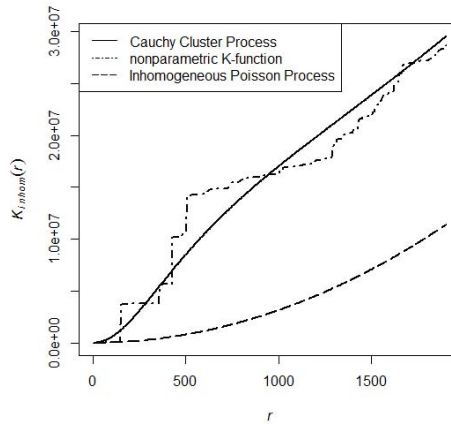


Figure 6.2. Comparison of three K-functions: nonparametric, Cauchy cluster process and inhomogeneous Poisson process.

Table 6.1은 Table 3.2에서 주어진 설명변수를 갖는 코시 군집 과정 모형을 준가능도 방법을 이용하여 적합한 결과이다. 사용된 설명변수 대부분 유의한 결과를 유지하고 있으나, 일부 비동질적 포아송 과정에서 과대 추정된 변수효과가 조정되어 나타났다. 코시 군집 과정에서도 고도에 따라 경사의 효과가 다르게 나타나는 것으로 보이고, 선구조로부터의 거리에 따라 경사가 산사태 발생에 미치는 영향도 확인할 수 있다.

앞서 3.2절에서 K -함수를 이용하여 모형의 적합도를 비교하는 방법에 대해 소개하였다. Figure 6.2는 보은 지역의 산사태 자료를 코시 군집 과정으로 적합한 모형의 K -함수이다. 이 모형의 K -함수는 비모수 K -함수와 매우 유사한 결과를 보여주고 있기 때문에, 코시 군집 과정 모형이 기존의 비동질적 포아송 과정보다 보은 지방의 산사태 자료가 가지는 군집을 잘 설명한다고 할 수 있다.

7. 결론

본 논문에서는 보은 지역의 산사태 자료에 비동질적 포아송 과정을 적합하였을 때의 문제점을 지적하고, 대표적인 군집 과정인 코시 군집 과정을 이용하는 것을 제안하였다. 개선된 정도를 파악하는 방법으로 K -함수를 살펴볼 것을 제안하였고, 코시 군집 과정에 가장 적합한 모수 추정 방법을 수치연구로 살펴본 후 실제 자료에 적용하여 코시 군집 과정의 적합도가 비동질적 포아송 과정의 적합도를 상당히 개선시킬 수 있음을 확인하였다. 따라서 산사태 발생 자료를 분석하는 경우 현재 많은 연구자들이 사용하고 있는 비동질적 포아송 과정을 논의없이 사용하는 것을 지양하는 대신, 산사태의 발생 위치들 간의 의존성, 특히 군집현상을 통계 모형에 적절히 고려하는 단계가 필요하다고 판단된다.

References

- Baddeley, A., Moller, J., and Waagepetersen, R. (2000). Non- and semiparametric estimation of interaction in inhomogeneous point patterns, *Statistica Neerlandica*, **54**, 329–350.
- Baddeley, A. (2008). Analyzing spatial point patterns in R, *CSIRO Division of Mathematics and Statistics*, October 2008.
- Ghorbani, M. (2012). Cauchy cluster process, *Metrika*, **76**, 697–706.
- Guan, Y. and Shen, Y. (2010). A weighted estimating equation approach for inhomogeneous spatial point processes, *Biometrika*, **97**, 867–880.
- Guan, Y., Jalilian, A., and Waagepetersen, R. (2015). Quasi-likelihood for Spatial Point processes, *Journal of the Royal Statistical Society, Series B*, **77**, 677–697.
- Park, N. W., Chi, K. H., Chung, C. F., and Kwon, B. D. (2003). GIS-based data-driven geological data integration using fuzzy logic: theory and application, *Economic and Environmental Geology*, **36**, 243–255.
- Park, N. W. (2015). Using maximum entropy modeling for landslide susceptibility mapping with multiple geoenvironmental data sets, *Environmental Earth Sciences*, **73**, 937–949.
- Ripley, B. D. (1976). The second-order analysis of stationary point processes, *Journal of Applied Probability*, **13**, 255–266.
- Tanaka, U., Ogata, Y., and Stoyan, D. (2008). Parameter estimation and model selection for Neyman-Scott point processes, *Biometrical Journal*, **50**, 43–57.
- Tonini, M., Pedrazzini, A., Penna, I., and Jaboyedoff, M. (2013). Spatial pattern of landslides in Swiss Rhone Valley, *Natural Hazards*, **73**, 97–110.
- Waagepetersen, R. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes, *Biometrics*, **63**, 252–258.

코시 군집 과정을 이용한 산사태 자료 분석

이기세^a · 김정환^a · 박노욱^b · 이우주^{a,1}

^a인하대학교 통계학과, ^b인하대학교 공간정보공학과

(2015년 12월 28일 접수, 2016년 1월 26일 수정, 2016년 1월 26일 채택)

요약

산사태 자료에서 환경변수들이 산사태 발생 위험에 어떻게 영향을 주는지 분석하기 위해 현재까지 비동질적 포아송 과정 모형이 주로 사용되어 왔다. 그렇지만, 이 모형은 산사태 자료에서 쉽게 관측되는 산사태 위치의 군집 현상에 대해 설명하지 못한다. 이러한 한계점을 극복하기 위해 우리는 코시 군집 과정을 사용할 것을 제안한다. 그리고, 제안된 방법이 실제 산사태 자료에서 얼마나 모형의 적합도를 개선시키는지 K -함수의 관점에서 살펴보고자 한다. 또한, 코시 군집 과정의 모수 추론을 위해 제안된 다양한 추정 방법의 성능을 비교하기 위해 시뮬레이션 연구를 진행하였다.

주요용어: 산사태, 점 패턴 자료, 포아송 과정, 코시 군집 과정

본 논문은 정부(국민안전처)의 재원으로 자연재해저감기술개발사업단의 지원을 받아 수행된 연구입니다 (MPSS-자연-2015-79). 박노욱의 참여는 2015년도 정부(미래창조과학부)의 재원으로 한국연구재단 기초연구사업의 지원을 받아 수행되었음 (NRF-2015R1A1A1A05000966).

¹교신저자: 인천광역시 남구 인하로 100, 인하대학교 통계학과. E-mail: lwj221@gmail.com

처음의 두 저자는 논문에 대한 공헌도가 동등합니다.