

# Forecasting probabilities of earthquake in Korea based on seismological data

Seowon Choi<sup>a</sup> · Woncheol Jang<sup>a,1</sup>

<sup>a</sup>Department of Statistics, Seoul National University

(Received August 14, 2017; Revised September 28, 2017; Accepted September 29, 2017)

---

## Abstract

Earthquake concerns have grown after a remarkable earthquake incident on September 12th, 2016 in Gyeongju, Korea. Earthquake forecasting is gaining in importance in order to guarantee infrastructure safety and develop protection policies. In this paper, we adopt a power-law distribution model to fit past earthquake occurrences in Korea with various historical and modern seismological records. We estimated power-law distribution parameters using empirical distributions and calculated the future probabilities for large earthquake events based on our model. We provide the probability that a future event has a larger magnitude than given levels, and the probability that a future event over certain levels will occur in a given period of time. This model contributes to the assessment of latent seismological risk in Korea by estimating future earthquake probabilities.

Keywords: earthquake, forecasting, power-law model, seismological records

---

## 1. 서론

지진은 단층의 갑작스러운 움직임으로 인한 지각의 흔들림 그리고 지진 에너지의 방사로 인한 자연현상을 말한다 (USGS, 2017). 지진은 화산활동, 마그마 활동 혹은 자연적, 인공적으로 지각에 가해지는 에너지에 의해 발생한다. 판 구조론에 따르면 지표의 암석권은 크게 십수 개의 지각 판으로 구성되어있으며 각 지각 판은 접점이 있는 맨틀권 위에 떠있다 (Condie, 1997). 이 지각 판들은 매년 수 센티미터 씩 이동하게 되는데, 이 움직임으로 인해 지각판들이 서로 만나는 경계에서 암석층의 구조적 저항 임계를 초과하는 힘이 축적되고 그 에너지가 방출되면 지진, 화산을 비롯한 다양한 지각활동들이 발생하게 된다 (Condie, 1997).

한반도는 유라시아 판 내부 대륙괴 위에 위치한다. 한반도 북부는 북중국 대륙괴, 남부는 양쯔강 대륙괴에 기반해 있으며, 두 대륙괴의 안정성으로 인해 한반도의 지진 발생 빈도가 적고 발생 규모도 작다 (Reedman과 Um, 1975; Chough, 2000). 계기지진 분석 결과 한반도에서 일어나는 대규모 지진은 대부분 중생대 지각변동에 의해 생성된 NNE-SSW 주향 활성단층이 역단층 매커니즘에 의해 깨어지는 과

---

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2014R1A4A1007895).

<sup>1</sup>Corresponding author: Department of Statistics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea. E-mail: [wjang@snu.ac.kr](mailto:wjang@snu.ac.kr)

정에서 발생하는데, 그 때 활성단층에 가해지는 힘은 유라시아 판이 태평양 판, 필리핀 판과 그리고 인도 판과 충돌할 때 발생하는 E-W 혹은 ENE-WSW 방향의 압력에 기인하는 것으로 생각된다 (Lee, 2010; Jun과 Jeon, 2010; Shimazaki, 1984).

한반도에서의 역사지진 기록은 삼국사기에서 서기 2년부터 시작되었다. 이후 1904년까지 삼국사기, 고려사, 조선왕조실록, 승정원일기 등 역사기록에서 보고된 지진은 최대 2,100여 건으로 파악된다 (KMA, 2012). 현대적 방법론에 입각해 지진계를 이용해 측정한 계기지진 기록은 1905년 조선총독부 관측소에서 시작됐으며 (KMA, 2012), 대한민국에서는 1978년 홍성지진 이후 본격적인 계기 지진 기록이 시작되어 오늘날에 이른다. 20세기부터 역사지진에 대한 연구는 와다(和田)를 비롯한 여러 연구자에 의해 진행되었으며, 북한에서도 조선지진목록을 통해 역사기록에 나타난 지진기록 및 20세기 이후 발생한 계기 지진 기록을 정리 수록하였다 (North Korea Seismology Laboratory, 1986). 한편 기상청에서 2012년 한반도 역사지진 목록 작성 사업을 수행하여 한반도 역사지진 전체 출처와 사서·기록으로부터 결정된 환산진도를 병기하여 공개하였다 (KMA, 2012).

본 논문에서는 역사지진 기록 및 현대 계기지진 기록을 멱법칙 분포(power-law distribution)을 통해 분석한 뒤 이후 한반도 지진 발생확률을 계산한다. 제 2장에서는 분석에 사용할 역사지진 및 계기지진 데이터에 대해 소개하고 데이터를 어떻게 정리했는가에 대해 설명한다. 제 3장에서는 멱법칙 분포를 소개하고, 멱법칙 분포를 설명하는 모수들의 추정 과정을 설명한다. 또한 추정된 모수들로부터 미래의 특정 크기 이상 지진이 일어날 확률을 어떻게 구할 것인가에 대한 방법을 소개한다. 제 4장에서는 앞서 사용한 자료와 방법론을 이용해 실제 한반도 지진 기록 데이터를 분석하고 미래의 지진 발생확률을 구한다. 마지막 제 5절에서는 결론에 대한 분석 및 향후 연구방향에 대한 고찰 및 제언을 적었다.

## 2. 데이터 소개

지진의 크기를 나타내기 위해 규모(magnitude)와 진도(intensity) 두 가지 지표가 사용된다. 규모는 “지진발생시 그 자체의 크기를 정량적으로 나타내는 양으로 지진의 진동에너지에 해당”한다. 규모는 계측관측값을 모아 계산된 절대적 지수이며 지진과의 진폭과 진앙까지의 거리를 이용해 계산된다. 한편 진도는 “특정 장소에 나타난 지진 진동의 세기를 사람의 느낌이나 주변 물체 또는 구조물의 흔들림 정도를 수치로 나타낸 것”으로 설문 혹은 지진계의 가속도 측정을 통해 측정한다. 따라서 진도는 지진이 관측된 지점에 따라 값이 다르며, 지진의 규모, 진앙거리, 진원깊이에 크게 좌우된다 (KISTEC, 2017).

역사시대 지진계를 사용하지 않고 문헌에 기록된 지진기록을 역사지진기록이라고 한다. 역사지진의 경우 그 시대에 지진을 과학적으로 측정할 수 있는 방법이 없고, 지진 크기를 역사 기술에 의존해 추정해야 하는 어려움이 있다. 따라서 역사 기록물로부터 지진에 대한 묘사 혹은 주변 상황 기술에 기반하여 진도를 재구성 후, 감진면적 등 지진과 관련된 다른 요소를 고려해 최대한 사실에 가깝게 지진의 규모를 추정해 내야한다. 이렇게 추정해낸 규모를 역사지진의 환산규모라 한다.

20세기 이후 지진계를 사용해 현대적인 의미에서 측정된 지진 기록을 계기지진기록이라고 한다. 현재 기상청에서는 117개소 디지털 지진관측소를 운영해 지진이 전파되기 시작하는 위치인 진앙, 실제 지진이 발생한 지하 위치인 진원, 지진이 발산한 에너지를 반영하는 지진규모를 계산해낸다 (KISTEC, 2017). 하지만 기상청 관측자료 분량은 불과 30 여년에 불과해 장기간 지진 예보 및 지진 위험도 분석을 하기에는 그 수가 부족하다. 따라서 KMA (2017) “국내 지진목록”을 보완하기 위해 북한 조선지진연구소에서 정리한 “조선지진목록” (North Korea Seismology Laboratory, 1986)을 분석에 활용하겠다.

본 논문에서는 한반도에서 발생한 역사지진 및 계기지진 목록을 데이터로 사용하였다. 역사지진목록으로 “한국 기상기록집 2: 한반도 역사지진기록 (2년-1904년)” (KMA, 2012)을 바탕으로 “조선지진목

록” (North Korea Seismology Laboratory, 1986)과 “한국 지진목록” Kim과 Gao (1995)으로 누락된 자료, 역사지진의 진도결정을 보충하였다. 계기지진목록으로는 “국내 지진목록” (KMA, 2017)을 기본으로 누락된 자료는 “조선지진목록” (North Korea Seismology Laboratory, 1986)으로 보충하였다. 사용된 지진목록에 대한 설명은 다음과 같다.

## 2.1. 역사지진 및 계기지진 목록

**2.1.1. 한반도 역사지진기록 (2년–1904년)** “한국 기상기록집 2: 한반도 역사지진기록 (2년–1904년)” (KMA, 2012) (이하 “기상청, 한반도 역사지진기록” 혹은 “Historical Seismological Records in Korea”)은 지질자원연구소에서 수행한 “활성단층지도 및 지진위험 지도 제작” (KIGAM, 2012) 사업에서 작성된 역사지진 목록을 바탕으로 기상청에서 편집, 발간한 최초의 정부 출간 역사지진목록이다 (KMA, 2012; KIGAM, 2012).

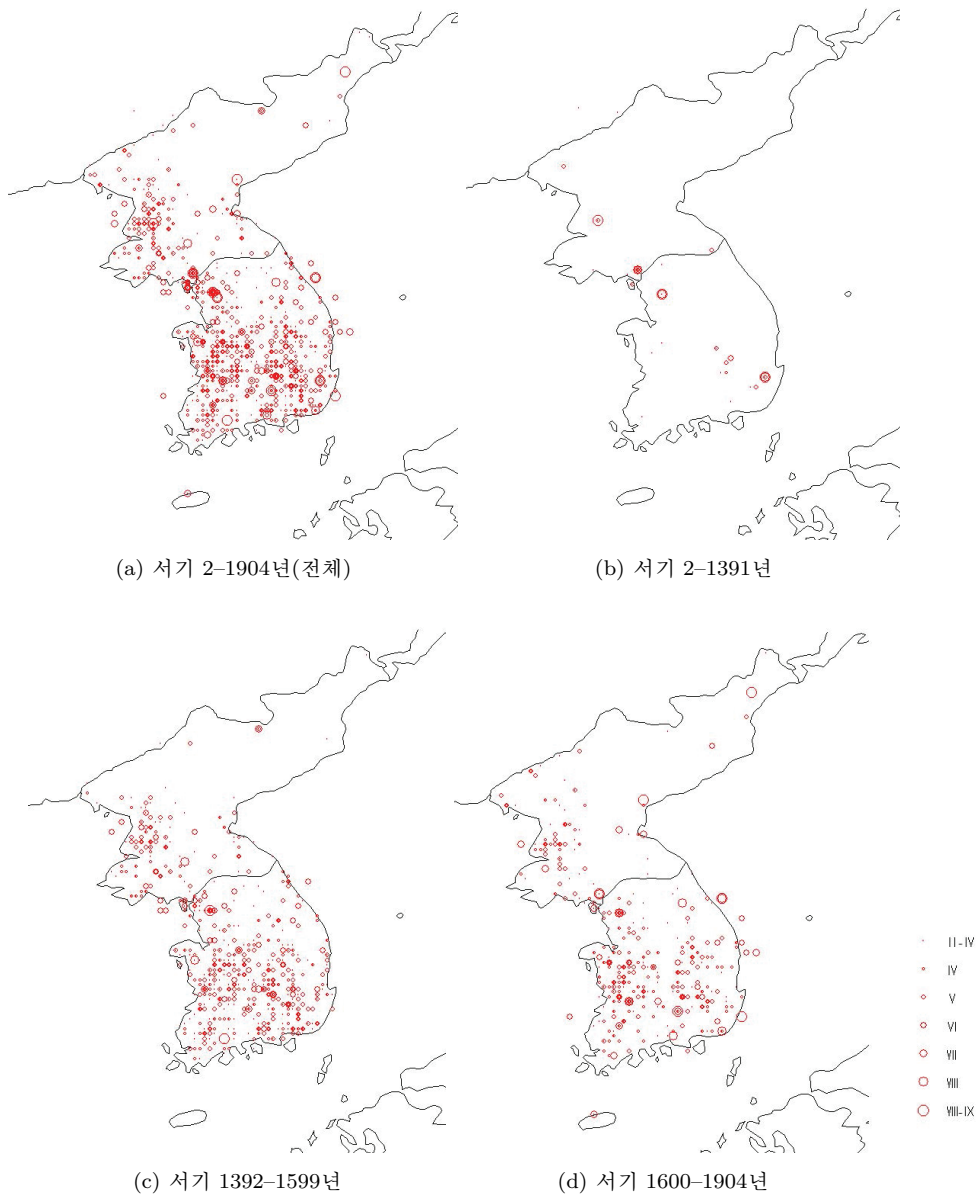
“기상청, 한반도 역사지진기록”은 삼국사기(三國史記), 고려사(高麗史), 조선왕조실록(朝鮮王朝實錄)과 같은 1차 사료와 고려사절요(高麗史節要), 승정원일기(承政院日記), 증보문헌비고(增補文獻備考) 등 2차사료로부터 지진자료를 발췌·해석하였다. 역사지진 평가를 위해 진앙 결정기준과 진도평가기준을 새로 제시했으며, 진도결정은 현재 기상청에서 사용하는 Modified Mercalli Intensity (MMI) 진도 계급을 사용하였다. 사료에 표시된 음력 날짜는 서기 1000년 이후의 경우 양력으로 환산하였으며, 이전의 경우 음력으로 표기하였다 (KMA, 2012; KIGAM, 2012).

“기상청, 한반도 역사지진기록”에는 1,783회의 역사지진 목록이 기재되어있다 (Figure 2.1(a)). 기록된 진앙의 분포도는 한반도 남부 지역에 집중되며, 기록에서 진앙이 특별히 기록되어있지 않는 경향이 많아서 각 시대의 수도에 기록이 집중되는 경향이 있다 (Figure 2.1(b)). Figure 2.1(c), (d)에서 진도 V이상의 지진은 영남, 충청권과 평안남도 평양 부근에서 집중적으로 발생하는 경향을 보여 한반도에서는 판내부 유형 지진이 호발하는 주로 발생하는 사실을 보여주었으며, 이는 현대 계기지진의 경향과도 일치한다 (KMA, 2012; KIGAM, 2012).

**2.1.2. 조선지진목록** “조선지진목록” (North Korea Seismology Laboratory, 1986) (이하 “조선지진목록” 혹은 “Choseon Seismological Records”)은 북한 조선지진연구소에서 작성했으며 서기 2년부터 1889년까지 역사지진목록 1,843건, 1901년부터 1983년까지 계기지진 523건, 총 2,366건의 지진 기록을 제시하고 있다 (Figure 2.2). “조선지진목록”에 기재된 지진 1900년 이전 역사지진 자료에서 진도 4.0 미만, 1900년에서 1977년 자료 중 진도 3.0 미만, 1978년 이후 계기지진 자료에서 진도 2.0 미만 목록은 자료에서 제외되었다 (North Korea Seismology Laboratory, 1986; KIGAM, 2012).

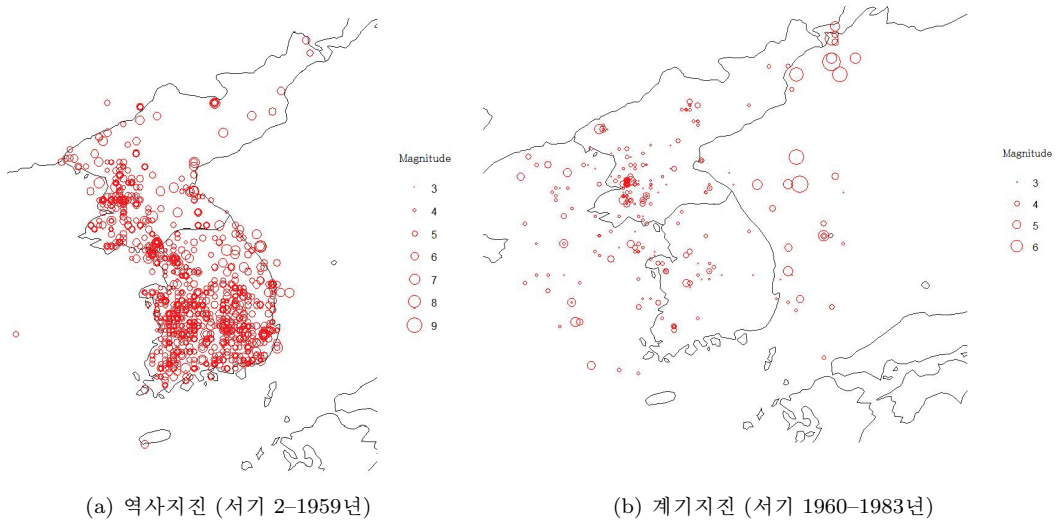
북한은 1960년부터 지진을 관측했으며 “조선지진목록”에 수록된 계기지진의 경우 남한 지역보다 북한 지역에서 발생한 지진이 더 많이 수록되어있다 (Figure 2.2(b)). 특히 북한지역에서 발생한 지진의 경우 “기상청 지진목록”에서는 평양, 평안남도 지역에 지진이 집중되어있는데, “조선지진목록”에 따르면 평양, 평안남도 지역 외에도 동해안, 특히 나진·선봉 지역 해안에 규모 5.0 이상의 지진이 군집되어있는 점이 특이사항이다. 이들 지진은 상대적으로 규모는 크지만 모두 심도 500km 이상 심발지진으로 지진에 의한 진동이 매우 미약해 남한에서 관측이 되지 않은 것으로 생각된다 (North Korea Seismology Laboratory, 1986; KIGAM, 2012).

**2.1.3. 한국 지진목록** “한국 지진목록” (Kim과 Gao, 1995) (이하 “한국지진목록” 혹은 “Korea Seismological Records”)은 서기 2–1995년까지 총 2,618건의 지진을 제시하고 있다. 서기 2–1977년은

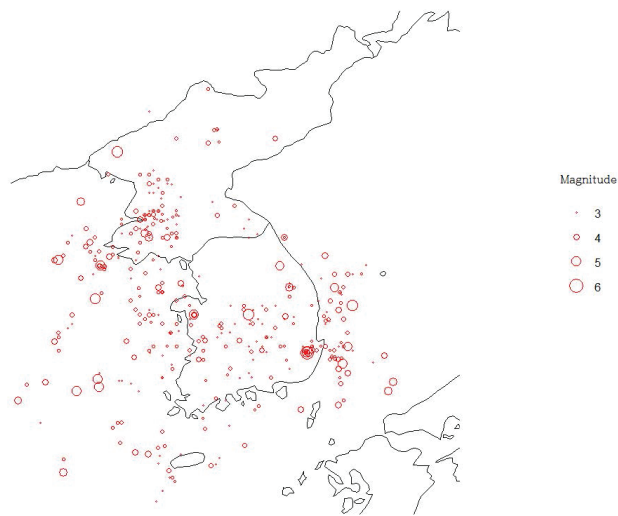


**Figure 2.1.** Distribution of historical earthquake records in Korea by periods listed in “KMA, Historical Seismological Records in Korea”.

“조선지진목록”의 지진목록을 저자가 다시 정리를 한 것이며, 1978년부터 1995년까지 계기지진목록은 “조선지진목록”과 “기상청 지진목록”을 취합하였다. 특히 사항으로 1978년 이후 계기지진 기록에서 북한의 진도기록이 대한민국 기상청의 진도기록보다 평균 0.3 정도 큰 것으로 나왔으며 저자는 “조선지진목록”과 “기상청 지진목록”의 지진기록이 겹치는 경우 “기상청 지진목록”의 진도를 따랐다 (Kim과



**Figure 2.2.** Distribution of historical and modern earthquake records in Korea listed in “Choseon Seismological Records” (A.D. 2–1904).



**Figure 2.3.** Distribution of modern earthquake records in Korea listed in “National Seismological Records” (A.D. 1978–2016).

Gao, 1995).

**2.1.4. 국내지진목록** “국내지진목록” (KMA, 2017) (이하 “기상청 지진목록” 혹은 “National Seismological Records”)은 현재 기상청 홈페이지(<http://www.kma.go.kr>)에 공개되고 있는 계기지진목록이다 (Figure 2.3). 기상청은 1978년 홍성지진 발생을 계기로 장비 현대화를 통해 지진관측을 시작했으며, 이후 지진관측소 확충사업을 통해 초광역 및 광지역 지진계를 도입, 운영하고 있다 (KIGAM,

**Table 2.1.** Seismological records in Korea

연 도 (서기)	지진 목록 이름	데이터 형
2-1904	한반도 역사지진기록	재구 MMI 진도
2-1904	조선지진목록	규모
1905-1977	조선지진목록	규모
1978-1983	한국 지진목록 (국내지진목록) (조선지진목록)	규모 규모 규모
1984-2016	국내지진목록	규모

2012). 1978년 이후 2017년 1월 1일 기준 총 1,472건의 지진이 관측되었으며, 주로 영남, 충청권과 평양 주변 평안남도, 울산 인근 및 경북지역 동해안에서 빈발했다 (KMA, 2017).

## 2.2. 지진 데이터 정리

본 논문에서 통계적 분석에 사용할 데이터는 역사지진의 환산규모와 계기지진의 규모이다 (이하 한반도 지진 기록이라고 함). 한반도 지진 기록의 (환산)규모는 0.1단위로 기록되며 분석시 데이터는 연속형 데이터로 생각하겠다. 다만 “기상청, 한반도 역사지진기록”은 환산규모를 제공하지 않고 재구된 MMI진도만을 제공했는데, 진도는 로마자 정수 단위로 기록되므로 “기상청, 한반도 역사지진기록” 자료를 분석하는 경우 이산형 정수 데이터로 간주하겠다. 2년에서 1904년까지 역사지진은 “기상청, 한반도 역사지진기록”과 “조선지진목록”을 사용했다. “기상청, 한반도 역사지진기록”은 재구성된 MMI 진도를 데이터로 사용해 이산형 정수 데이터로 구성되어있기 때문에, (환산)규모를 사용하는 다른 지진목록 데이터와 데이터의 취합이 어렵다. 따라서 역사지진의 분석에서는 “기상청, 한반도 역사지진기록”과 “조선지진목록”을 따로 분석하기로 하고, 계기지진 목록과 취합은 “조선지진목록”만을 사용하기로 한다. 역사지진의 경우 두 자료 모두 진도, 환산규모가 3 이상인 경우만 기록되었기 때문에 전체를 분석에 사용하기로 한다. 1905년에서 1977년 사이에 발생한 지진은 “조선지진목록”에 기재된 환산규모 데이터를 사용하였다. 1978년에서 1983년 사이의 계기지진은 “한국 지진목록”을 사용해 “조선지진목록”과 “국내지진목록”의 자료 모두 반영되게 하였고 꼬리분포를 사용하기 위해 규모 3 이상 지진 목록만 사용했다. 1984년 이후 계기지진은 “국내지진목록”을 이용하였으며 역시 규모 3 이상 지진 목록만 사용했다. 각 지진 목록에서 사용된 자료의 연도, 데이터의 형태는 Table 2.1에 정리하였다.

## 3. 분석 방법

본 논문에서는 과거 한반도 지진 기록을 바탕으로 앞으로 한반도에서 일어날 지진을 예측하기 위해 Clauset 등 (2009)와 Clauset와 Woodard (2013)을 따라 멱법칙 분포(power-law distribution)을 통한 꼬리분포 구간 확률 추정 모델을 사용한다. 꼬리분포 구간 확률 추정 모델의 목적은 데이터로부터 다음 번 지진이 관측 되었을 때 그 크기가  $x$  혹은 그 이상일 확률과 이제까지 지진을 관측한 전체 기간과 맞먹는 미래의 기간 중 그 크기가  $x$  혹은 그 이상인 지진이 발생할 확률을 구하는 것이다.

### 3.1. 멱법칙 분포의 소개

멱법칙 분포는 큰  $x$ 값에서 확률밀도가  $x$ 의 역으로  $\alpha$ 를 가지는, 즉  $p(x) \propto x^{-\alpha}$ 를 따르는 분포를 말한다. 멱법칙 분포에서 두 모수  $\alpha$ 와  $x_{\min}$ 를 각각 척도모수, 멱법칙 현상의 하한이라 부르며 꼬리 부분에서 굉장히 큰 값을 가지는 데이터들이 어떻게 분포되어 있는가를 설명하는 지표로 사용된다 (Clauset

등, 2009).

떡법칙 분포는 데이터의 종류에 따라 연속형 떡법칙 분포(continuous power-law distribution)과 이산형 떡법칙 분포(discrete power-law distribution)으로 분류할 수 있다. 연속형 떡법칙 분포의 확률 분포는  $\alpha > 1$ 의 범위에서

$$p(x) = \frac{\alpha - 1}{x_{\min}} \left( \frac{x}{x_{\min}} \right)^{-\alpha}$$

이며, 이산형 떡법칙 분포의 확률 분포는

$$p(x) = \frac{x^{-\alpha}}{\zeta(\alpha, x_{\min})}$$

이다. 여기서  $\zeta(\alpha, x_{\min})$ 는  $\sum_{i=1}^n (n + x_{\min})^{-\alpha}$ 로 일반화 제타 함수이다 (Clauset 등, 2009).

### 3.2. 떡법칙 분포의 모수 추정

**3.2.1. 떡법칙 현상의 하한  $x_{\min}$ 이 주어졌을 때 척도모수  $\alpha$ 의 추정** 연속형 떡법칙 분포에서  $x_{\min}$ 이 주어졌을 때 척도모수  $\alpha$ 의 최대가능도추정량(maximal likelihood estimator)를 구한다.  $n$ 개의 데이터  $x_i > x_{\min}$ 이 주어졌을 때, 이 데이터로부터 척도모수  $\alpha$ 의 최대가능도추정량  $\hat{\alpha}$ 를 로그-가능도함수

$$\mathcal{L}(x; \alpha) = \log p(x; \alpha) = \log \prod_{i=1}^n \frac{\alpha - 1}{x_{\min}} \left( \frac{x_i}{x_{\min}} \right)^{-\alpha}$$

를 이용해 구할 수 있으며, 그 값은  $\partial \mathcal{L} / \partial \alpha = 0$ 의 해인

$$\hat{\alpha} = 1 + n \left[ \sum_{i=1}^n \log \left( \frac{x_i}{x_{\min}} \right) \right]^{-1}$$

이다.

이산형 떡법칙 분포의 로그-가능도함수  $\mathcal{L}$ 은 다음과 같다.

$$\mathcal{L}(x; \alpha) = \log p(x|\alpha) = \log \prod_{i=1}^n \frac{x_i^{-\alpha}}{\zeta(\alpha, x_{\min})}.$$

로그-가능도함수  $\mathcal{L}$ 을 통해  $x_{\min}$ 이 주어졌을 때 척도모수  $\alpha$ 를 구하기 위해  $\mathcal{L}$ 을  $\alpha$ 에 대해 편미분한 뒤 그 식을 0으로 만드는  $\alpha$  값을 구한다.  $\partial \mathcal{L} / \partial \alpha = 0$ 을 정리하면

$$\frac{\zeta'(\hat{\alpha}, x_{\min})}{\zeta(\hat{\alpha}, x_{\min})} = -\frac{1}{n} \sum_{i=1}^n \log x_i$$

이고,  $\hat{\alpha}$ 는 이 식의 해이며 수치적으로 계산한다 (Clauset 등, 2009).

**3.2.2. 떡법칙 현상의 하한  $x_{\min}$ 와 척도모수  $\alpha$ 의 추정의 동시 추정** 현실 세계에서 관측되는 떡법칙 분포 데이터들은 고정된 떡법칙 현상의 하한  $x_{\min}$ 이 알려지지 않은 경우가 많으므로, 한반도 지진 데이터의 떡법칙 현상의 하한  $x_{\min}$ 을 추정하기 위해 본 논문에서는 Clauset 등 (2009)에 따라 Kolmogorov-Smirnov statistics (이하 KS 통계량)을 사용했다.

떡법칙 분포의 KS 통계량은

$$D_{x_{\min}} = \max_x \left| G(x; \alpha_{ML}, x_{\min}) - \hat{G}(x; x_{\min}) \right|$$

로 나타낼 수 있으며,  $D_{x_{\min}}$ 을 최소화(minimization)하는

$$\hat{x}_{\min} = \arg \min_{x_{\min}} D_{x_{\min}} = \arg \min_{x_{\min}} \left[ \max_x \left| G(x; \alpha_{ML}, x_{\min}) - \hat{G}(x; x_{\min}) \right| \right]$$

가 떡법칙 현상 하한  $x_{\min}$ 의 추정량이라 할 수 있다. 여기서  $G(x; \alpha_{ML}, x_{\min})$ 는 척도모수  $\alpha_{ML}$ 와  $x_{\min}$ 이 주어졌을 경우 참고 보누적분포함수,  $\hat{G}(x; x_{\min})$ 는  $x_{\min}$ 이 주어졌을 때 경험적 보누적분포함수이다.  $\alpha_{ML}$ 는  $x \geq x_{\min}$ 인 데이터의 부분집합에서 추정된 척도모수  $\alpha$ 이다. 떡법칙 현상 하한  $x_{\min}$ 의 추정량  $\hat{x}_{\min}$ 을 구한 다음, 그 값이 주어졌을 때 척도모수의 추정량  $\hat{\alpha}$ 을 따라서 구할 수 있다 (Clauset 등, 2009).

### 3.3. 붓스트랩 방법을 이용한 떡법칙 분포 모수 추정

앞 절에서 추정된 떡법칙 분포 모수들은 붓스트랩 방법을 이용해서도 추정할 수 있다. 붓스트랩 방법을 주어진  $x_{\min}$ 에 대한  $\alpha$ 를 추정하는 방법은 다음과 같다. 우선 데이터의 경험적 분포  $E$ 로부터 값이  $x_{\min}$ 보다 큰 꼬리부분의 크기  $p_{\text{tail}} = \#(x_i \geq x_{\min})/n$ 을 구한다. 붓스트랩 표본을 총  $N_B$ 번 만든다고 했을 때, 각각의 붓스트랩 표본은 원 데이터와 같은 크기  $n$ 을 가지는 새로운 표본이다. 새로운 표본을 만들 때 원 데이터에서  $p_{\text{tail}}$ 의 확률로 값이  $x_{\min}$ 보다 큰 꼬리부분을 복원추출하는데, 새로운 분포에서도 전체 데이터에서  $x_{\min}$ 보다 큰 꼬리부분이 차지하는 비중이 같게 된다. 즉, 붓스트랩 방법은 원 데이터에서 복원추출로 새로운 데이터 표본을 만들되 꼬리 구조는 보존하여, 추정시 꼬리부분을 사용하는 떡법칙 분포 모수들의 추정이 원 데이터에서와 붓스트랩 표본에서 같은 조건에서 이루어지게 만드는 것이다. 이 붓스트랩 표본의  $x_{\min}$ 보다 큰 꼬리 부분에서 척도모수  $\alpha$ 의 값을 구한 후, 총  $N_B$ 개의  $\alpha$ 에 대한 추정량의 평균과 표준편차를 구하여  $\alpha$ 의 추정량과 그 오차를 구할 수 있다 (Clauset와 Woodard, 2013).

$x_{\min}$ 과  $\alpha$ 는 몬테카를로 방법을 사용해 동시에 추정할 수 있다. 몬테카를로 방법을 따를 경우 우선 데이터의 경험적 분포  $E$  전체에서 ‘균일분포(uniform distribution)’에 따라 표본을 복원 추출한다. 추출한 표본에서 앞 절의 방법에 따라  $x_{\min}$ 에 대한  $\alpha$ 를 동시에 추정해내고, 총  $N_B$ 개의  $\alpha$ 와  $x_{\min}$ 에 대한 추정량의 평균과 표준편차를 구하면, 데이터의  $\alpha$ 와  $x_{\min}$ 의 추정량과 그 오차를 구할 수 있다 (Clauset와 Woodard, 2013).

### 3.4. 적합도 검정을 통한 $p$ -값 추정

한반도 지진 목록 데이터가 실제로 떡법칙 분포를 따르는가 확인하기 위해 적합도 검정을 사용한다. 적합도 검정에서는 우선 관측 데이터의 경험적 분포의 KS 통계량을 구한 뒤 경험적 분포를 설명하는 모수를 추정하고, 다음 경험적 분포로부터 추정한 모수로 가상의 합성분포(synthetic distribution)를 준모수적 방법(semiparametric approach)으로 추출해 만든다. 준모수적 방법은 붓스트랩 방법과 유사하게 관측 데이터의 경험적 분포  $E$ 에서 값이  $x_{\min}$ 보다 큰 꼬리부분을 선택해 이 꼬리부분으로부터 크기  $n$ 짜리 가상의 합성분포를 만드는 것이다 (Clauset 등, 2009).

먼저 값이  $x_{\min}$ 보다 큰 꼬리부분 자료갯수를  $n_{\text{tail}}$ 이라고 했을 때  $n_{\text{tail}}/n$ 의 확률로 앞서 원 데이터로부터 추정한 모수를 가지는 떡법칙 분포에서 생성되는 난수를 만들어(random number generation) 새로 만들 가상의 합성분포 데이터에 추가한다. 또  $1 - n_{\text{tail}}/n$ 의 확률로 관측 데이터에서 균일분포로 자료를



추출해 가상의 합성분포 자료에 추가한다. 이 과정을  $n$ 번 거치게 되면 관측 데이터의 경험적 분포와 같은 성질을 가지리라 생각되는 멱법칙 분포를 합성한 새로운 가상의 합성분포를 가지게 된다. 이 가상의 합성분포에서 다시 KS 통계량을 구한다. 가상의 합성분포를  $N_B$ 개 합성했다고 했을 때  $N_B$ 개 가상의 합성분포는 각각의 KS 통계량을 가지게 된다. 이때  $N_B$ 개의 KS 통계량 중 맨 처음 구한 KS 통계량보다 큰 갯수를  $N_B$ 로 나눈 것을  $p$ 라고 하자. 이  $p$ 가 우리가 구하고자 하는  $p$ -값이다 (Clauset 등, 2009). 실제 데이터 분석에서는 총 1,000개의 가상의 합성분포를 만들며,  $p$ -값이 0.1보다 큰 경우 관측 데이터가 멱법칙 분포에 적합하다고 간주하였다.

### 3.5. 한반도 지진 발생확률의 추정

앞서 구한 멱법칙 분포 모수를 이용해 미래의 한반도 지진 발생확률을 추정한다. 우리가 추정하고자 하는 한반도 지진 발생확률은 두 가지이다. 첫째는 앞으로 한반도에서 지진이 한 번 관측되는 경우 그 크기가  $x$  혹은 그보다 클 확률(per-event probability; 이하 사건확률)이다. 이 확률은 지진의 발생이 연속-시간 포아송 과정(continuous-time Poisson process)을 따른다고 했을 때, 발율(rate)이 멱법칙 분포함수의 보누적분포함수임을 이용해 구할 수 있다. 둘째는 이제까지 데이터를 관측한 전체 기간에 맞먹는 미래의 기간 중 그 크기가  $x$  혹은 그 이상인 지진이 관측될 확률(per-period probability; 이하 기간확률)을 구하는 것이다. 여기서 “전체 기간에 맞먹는 미래의 기간”이라는 말의 뜻은 “자료가 관측된 전체 기간 동안 발생한 총 지진수만큼 미래에 지진이 발생했을 때”를 의미한다. 다시 말해 자료에 지난 100년간 일어난 1,000건의 지진이 있고 이 자료를 통해 기간확률을 구했다면, 이는 앞으로 미래에 1,000여건의 지진이 발생했을 때 주어진 크기 이상의 지진이 한번이라도 발생할 확률이 기간확률이라는 뜻이다. 기간확률은 앞서 사건확률을 구할 때 사용한 연속-시간 포아송 과정을 근사해 구할 수 있다 (Clauset와 Woodard, 2013).

우선 한반도에서의 지진 발생은 그 발생율이  $q(x)$ 인 연속-시간 포아송 과정을 따르고, 그 규모 또는 정도의 크기가 서로 독립적이며 같은 분포를 가진다(independent and identically distributed)라고 가정하자. 이 가정으로부터  $q(x)$ 는 포아송 과정에서 한번 사건이 관측 되었을 때 그 크기가  $x$  이상일 확률, 즉 앞에서 언급한 첫째 확률인 한반도에서 미래에 지진이 한 번 관측 되었을 때 크기가  $x$  이상일 사건확률로 사용될 수 있다. 한편  $q(x)$ 는 그 정의 상 멱법칙 분포함수의 보누적분포함수일 뿐이므로

$$q(x) = \Pr(X \geq x | X \geq x_{\min}) \Pr(X \geq x_{\min}) = \Pr(X \geq x) = \left( \frac{x}{x_{\min}} \right)^{-\alpha+1}$$

로 계산할 수 있다 (Clauset와 Woodard, 2013).

한편 앞에서 언급한 두 번째 확률인 전체 기간에 맞먹는 미래의 기간 중 지진이 관측될 기간확률  $\hat{p}$ 는 연속-시간 포아송 과정에서 사건의 크기  $x$ 가 연속이라 가정하여 근사한 뒤,

$$\hat{p} = 1 - [1 - q(x)]^n \approx 1 - e^{-nq(x)}$$

이다 (Clauset와 Woodard, 2013).

한반도 지진 발생확률 또한 붓스트랩 방법을 이용해 추정하고 그 오차를 계산할 수 있다. 한반도 지진 발생확률을 붓스트랩 방법으로 추정할 때 앞절에서 소개한 방법과 마찬가지로 주어진  $x_{\min}$ 에 대한 척도 모수  $\alpha$ 를 추정하는 방법과  $x_{\min}$ 과  $\alpha$ 를 동시에 추정하는 방법 두 가지를 사용해 모수  $\alpha$ 와  $x_{\min}$ 을 구하고, 그에 따른 지진 발생확률을 위의  $q(x)$ 와  $\hat{p}$ 를 구하는 공식에 대입해 계산해낼 수 있다. 각각 방법으로 붓스트랩 표본에서 도출해낸 한반도 지진 발생확률  $q(x)$ 와  $\hat{p}$ 의 평균과 표준편차를 구하여, 주어진 데이터의  $q(x)$ 와  $\hat{p}$ 의 추정량과 오차를 구할 수 있다 (Clauset와 Woodard, 2013).

**Table 4.1.** List of seismological data

분석 자료	구성
기상청 역사	기상청, 한반도 역사지진 기록 서기 4-1904년
조선지진 역사	조선지진목록 서기 4-1904년
계기지진1	국내지진목록 서기 1978-2016년
계기지진2	국내지진목록 서기 1978-2016년, 한국지진목록 서기 1960-2016년
계기지진3	국내지진목록 서기 1978-2016년, 한국지진목록 서기 1960-2016년, 조선지진목록 서기 1905-1959
전체목록	조선지진목록 국내지진목록 서기 4-2016년

#### 4. 분석 결과

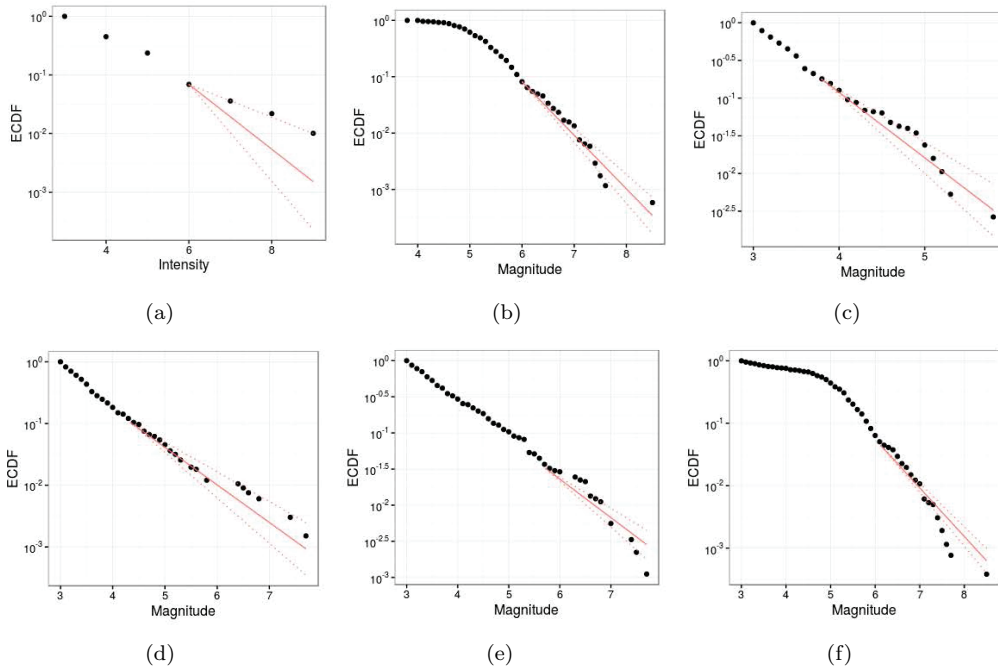
앞서 제시한 지진목록 중 역사지진 분석을 위해 “기상청, 한반도 역사지진기록” (이하 “기상청 역사” 혹은 “KMA Historical”)과 “조선지진목록” 중 서기 4-1904년 부분 (이하 “조선지진 역사” 혹은 “Choseon Historical”)을 사용한다. 계기지진 분석은 3가지 목록을 사용하겠다, 첫 번째는 “국내지진목록”만 수록된 목록으로 1978년 이후 기상청 자료만을 분석했다 (이하 “계기목록1” 혹은 “Modern1”). 두 번째 자료는 1978년 이후 “국내지진목록”과 1960년 이후 “한국 지진목록”을 이용해 현대적 의미에서 남북한 모두 지진계 사용이 확실하다고 근거가 밝혀진 지진 자료를 분석 대상으로 삼았다 (이하 “계기목록2” 혹은 “Modern2”). 마지막으로 “국내지진목록”, 1960년 이후 “한국 지진목록, 1905-1959년 “조선지진목록”을 사용해 계기지진으로 알려진 모든 지진 자료를 분석 대상으로 삼았다 (이하 “계기목록3” 혹은 “Modern3”). 2017년을 기준으로 “계기목록1”은 39년, “계기목록2”는 57년, “계기목록3”은 112년 간의 자료를 수록하고 있다. 마지막으로 “조선지진목록”, “국내지진목록”의 모든 데이터를 사용해 2013년 간 한반도 지진기록 전체를 분석한 결과도 제시하였다 (이하 “전체목록” 혹은 “Whole records”).

##### 4.1. 한반도 지진데이터 경험적 보누적분포함수 그림

한반도 지진 목록의 멱법칙 분포 보누적분포함수를 로그-로그 도표에 그린 결과는 Figure 4.1과 같다. (a)에서는 “기상청, 한반도 역사지진기록”, (b)에는 “조선지진목록” 중 서기 2년-1904년 역사지진 부분, (c)에는 “계기지진1”, (d)에는 “계기지진2”, (e)에는 “계기지진3”, (f)에는 “전체목록”의 보누적분포함수를 그렸다.

(b) “조선지진목록” 데이터의 분포는 일직선이 아니라 오목한(concave) 모양을 그리는데, 이는 “조선지진목록” 저자가 환산규모를 산정함에 있어 주관적으로 규모 4.0에서 7.5 사이에 환산치를 집중시킨 결과로 생각된다. “조선지진목록” 데이터가 다수 반영된 (f) “전체목록”에서도 비슷한 경향을 가진다. 이후 분석에서 “조선지진목록” 역사지진 부분과 “전체목록”에서는 환산규모 5.0 이상 자료만을 분석에 사용해 최대한 멱법칙 분포에 부합하는 부분만을 분석에 반영하도록 하겠다.

(c) “계기지진1”, (d) “계기지진2”, (e) “계기지진3”은 대체적으로 선형성을 잘 띄는 것을 볼 수 있다. (d) “계기지진2”와 (e) “계기지진3”에서는 규모 5.5 이후 데이터의 분포가 상방으로 이탈하는 모습을 보이는데 이는 규모 4-5크기의 지진이 평양 인근에서 빈발하기 때문으로 생각된다. 다만 (e) “계기지진3”에서 다시 규모 6.5 이후 데이터 분포가 하방으로 꺾이는 모습을 볼 수 있는데 이는 1940년대 초반 서해와 동해 먼바다에서 발생한 규모 6에서 7 사이 지진들이 도표에 반영된 것으로 보인다.



**Figure 4.1.** Complementary cumulative distribution functions of seismological data in Korea and fittings of scale parameters.

**Table 4.2.** Estimations of power-law distribution parameters (without bootstrapping)

	주어진 $x_{\min}, \hat{\alpha}$		$\hat{\alpha}, \hat{x}_{\min}$ 동시		$p$ -값
	$\hat{\alpha}$	주어진 $x_{\min}$	$\hat{\alpha}$	$\hat{x}_{\min}$	
기상청 역사	1.552	5.0	1.010	6.0	0.894
조선지진 역사	1.836	3.0	2.551	7.4	0.088
계기지진1	2.004	3.0	3.369	5.1	0.115
계기지진2	1.782	3.0	1.704	4.9	0.409
계기지진3	1.558	3.0	4.257	7.4	0.085
전체목록	1.814	5.0	2.670	7.5	0.068

**4.2. 한반도 지진데이터 척도모수 추정량 계산**

척도모수  $\alpha$ 의 추정과 그 신뢰구간 계산을 한반도 지진 목록에 적용시켜 그 결과를 표로 제시했다 (Table 4.2, Table 4.3). 또 Figure 4.1에서는 척도모수 추정량  $\hat{\alpha}$ 과 그 신뢰구간을 사용해 그린 적합 직선(fitting line)을 로그-로그 도표에 표시했다.

떡법칙 분포 모수 추정 결과는 붓스트랩을 사용하지 않은 경우와 붓스트랩을 사용한 경우 나누어 표로 작성했다 (Table 4.2, Table 4.3). 붓스트랩을 사용하지 않은 경우 추정된 척도모수  $\hat{\alpha}$ 는 대개 1.5에서 2 사이의 값을 가지는 것으로 나타나 데이터의 분포가 전체적으로 안정적으로 떡법칙 분포를 따른다고 해석할 수 있다 (Clauset와 Woodard, 2013). 다음  $\alpha$ 와  $x_{\min}$ 를 동시에 추정한 경우  $\hat{x}_{\min}$ 가 5와 8 사이의 값을 가져 분포 안에서도 상당히 우측에 자리잡고 있음을 알 수 있다. 로그-로그 도표에서 데이터의 분포를 살펴보면 큰 규모쪽 데이터가 적합 직선을 벗어나는 것을 볼 수 있다 (Figure 4.1). 이런 불안정한

**Table 4.3.** Estimations of power-law distribution parameters (with bootstrapping)

	주어진 $x_{\min}, \hat{\alpha}$		$\hat{\alpha}, \hat{x}_{\min}$ 동시				$p$ -값
	$\hat{\alpha}$	표준편차	$\hat{\alpha}$	표준편차	$\hat{x}_{\min}$	표준편차	
기상청 역사	1.553	0.019	1.068	0.116	6.457	1.167	*
조선지진 역사	1.836	0.024	2.240	0.214	6.000	0.641	0.639
계기지진1	2.008	0.059	2.317	0.518	3.793	0.721	0.794
계기지진2	1.781	0.036	1.870	0.317	4.339	0.996	0.871
계기지진3	1.557	0.020	2.253	1.170	5.720	0.939	0.190
전체목록	1.815	0.022	2.236	0.335	6.122	0.748	0.092

꼬리분포를 KS 통계량으로 추정하게 되면, 꼬리머리에서 이상값들 중 조금이라도 선형성을 보이는 일부 구간을 멱법칙 분포로 적합을 해 전체 데이터를 설명할 수 있는 충분히 작은  $x_{\min}$ 을 추정할 수 없게 되는 현상이 발생하였다.

따라서 붓스트랩 방법을 사용해  $x_{\min}$ 을 추정하게 되면 불안정한 꼬리분포에 있는 이상값들의 영향이 선형성을 잘 보이는 작은 규모 데이터들과 희석이 되는데, 이는 모수 추정시 원 데이터로부터 무작위 추출된 새로운 분포로부터 모수를 추정하고 그 값을 평균해 결과를 얻기 때문이다. 꼬리부분의 불안정성이 선형성을 잘 보이는 작은 규모의 데이터들과 희석이 되면서, 꼬리 부분의 데이터들의 선형성을 추정할 수 있는 영역이 왼쪽으로 커지기 때문에 붓스트랩 방법을 사용하지 않은 경우보다 더 작은  $x_{\min}$ 값이 계산할 수 있다 (Figure 4.1). 따라서 붓스트랩 방법으로 계산된 모수 추정량들이 분포의 전체적인 관점에서 선형성을 더 잘 설명하는 결과를 얻을 수 있다. 실제로 로그-로그 도표에 그린 적합직선과 신뢰구간 직선을 보면 꼬리부분의 대부분 데이터들이 붓스트랩 방법으로 구한  $\hat{\alpha}$ 의 90% 신뢰구간 안에 들어옴을 눈으로 볼 수 있다 (Figure 4.1). 이는 추정된 모수들이 불안정한 꼬리부분 만을 설명하는 것이 아니라 더 작은  $x_{\min}$ 에서부터 전체적인 데이터의 선형성, 즉 데이터가 멱법칙 분포로 잘 설명이 되는지를 반영할 수 있게 된 것이라고 할 수 있다.

$x_{\min}$ 이 주어졌을 때  $\alpha$ 의 추정량은 붓스트랩 방법을 사용하는 경우와 붓스트랩 방법을 사용하지 않은 경우 거의 같은 결과값을 얻었다. 이는 데이터를 적합하는 시작 점이 같아 붓스트랩 방법의 사용여부가 적합직선의 기울기를 계산하는데 큰 영향을 주지 않기 때문이다.

#### 4.3. 멱법칙 분포 적합 검정

앞서 3.4에서 소개한 멱법칙 분포 적합 검정을 시행하여 한반도 지진 목록 각각이 실제 멱법칙 분포를 따르는지 검정하였다. 검정기준은  $p$ -값이 0.1보다 크게 나오는 경우 데이터가 멱법칙 분포를 따르는 것으로 해석하였다.

우선 붓스트랩 방법을 사용하지 않고 모수를 추정한 경우  $p$ -값들이 0.05에서 0.10 사이에 분포하는 경우가 많았다. 이는 붓스트랩을 사용하지 않고  $\alpha$ 와  $x_{\min}$ 를 동시에 추정하는 경우 추정된 모수로 합성한 실제 멱법칙 분포와 원데이터가 가지는 경험적 분포의  $\hat{x}_{\min}$  이후 꼬리부분과 형태상 차이가 어느 정도 있어, 원데이터의 꼬리부분이 정말로 멱법칙 분포를 따른다고 말하기에 석연찮은 부분이 있다고 해석할 수 있다. 이렇게  $p$ -값이 낮게 나오는 이유는 대부분 지진목록에서  $\hat{x}_{\min}$  꼬리 부분의 데이터 분포가 적합 직선을 따르지 않고 밖으로 벗어나는 불안정한 경향을 보이고 있기 때문에, 추정 모수로 만들어낸 멱법칙 분포와 일치하는 정도가 낮아  $p$ -값이 작게 나온 것으로 생각된다 (Table 4.2).

반면 붓스트랩 방법을 사용하여  $x_{\min}$ 이 더 낮게 추정된 경우, 선형성을 잘 보이는 낮은 규모의 지진 데이터도 실제 멱법칙 분포와의 비교에 반영되므로  $\hat{x}_{\min}$  이후 꼬리부분이 전체적으로 적합직선에 잘 따라

**Table 4.4.** Estimations of earthquake forecasting probabilities in Korea

		발생확률 추정 결과					
목록	모수	6.5		7		7.5	
		$x$	표준편차	$x$	표준편차	$x$	표준편차
계기지진1 (39년간)	$q(x)$	0.0003311	0.0001559	0.0001066	5.81E-05	3.30E-05	2.04E-05
	$\hat{p}$	0.1158660	0.0504876	0.0391746	2.07E-02	1.23E-02	7.54E-03
	$\hat{\alpha}$	2.0079417	0.0597501	2.0084631	5.92E-02	2.01E+00	5.82E-02
	$q(x)$	8.30E-04	8.74E-04	0.0002165	0.0002383	5.98E-05	6.16E-05
	$\hat{p}$	2.39E-01	1.69E-01	0.0750448	0.0715109	2.20E-02	2.22E-02
	$\hat{\alpha}$	2.33E+00	5.25E-01	2.3067092	0.5289257	2.28E+00	4.95E-01
	$\hat{x}_{\min}$	3.7538000	0.6922896	3.7637	0.7114059	3.73E+00	6.83E-01
계기지진2 (51년간)	$q(x)$	0.0018643	0.0005070	0.0007595	0.0002539	0.0003226	0.0001157
	$\hat{p}$	0.6935289	0.0970797	0.3874275	0.0964663	0.1902009	0.0604328
	$\hat{\alpha}$	1.7845014	0.0343956	1.7857235	0.0358599	1.7821148	0.0353779
	$q(x)$	0.0937535	0.2331135	0.0260432	0.0514859	0.0083070	0.0136876
	$\hat{p}$	0.9395838	0.0894375	0.8231553	0.2167116	0.6548360	0.3349746
	$\hat{\alpha}$	1.8695750	0.2924579	1.8619616	0.2982523	1.8618303	0.2810232
	$\hat{x}_{\min}$	4.3023000	0.9899822	4.3493000	0.9803676	4.3131000	1.0061763
계기지진3 (112년간)	$q(x)$	0.0112502	0.0018133	0.0058849	0.0010807	0.0031162	0.0006555
	$\hat{p}$	0.9998794	0.0002150	0.9921652	0.0078614	0.928052	0.0407848
	$\hat{\alpha}$	1.5584024	0.0198824	1.5594122	0.0201717	1.5591316	0.0206031
	$q(x)$	0.2095711	2.66E-01	0.1292275	0.144964	0.0670619	0.1247493
	$\hat{p}$	1.00E+00	1.83E-05	0.9997477	0.0018143	0.9972713	0.0157977
	$\hat{\alpha}$	2.21E+00	9.79E-01	2.2395702	1.2092162	2.2853894	1.1648968
	$\hat{x}_{\min}$	5.67E+00	9.96E-01	5.6683	1.0048457	5.733	0.9934777
전체목록 (2014년간)	$q(x)$	0.0600401	0.0047036	0.0235625	0.0023826	0.0092151	1.18E-03
	$\hat{p}$	1	0	1	0	1	1.95E-08
	$\hat{\alpha}$	1.8152647	0.0227873	1.8150068	0.0221098	1.8156362	2.24E-02
	$q(x)$	0.1527976	0.1260682	0.0507806	0.0775254	0.1589676	2.93E-01
	$\hat{p}$	1	0	1	0	1	5.01E-08
	$\hat{\alpha}$	2.2416040	0.3094273	2.2493133	0.3439827	2.2381571	3.23E-01
	$\hat{x}_{\min}$	6.1201000	0.7417926	6.1412000	0.7619863	6.1267000	6.1267000

계 되는 모습을 확인할 수 있다. 이는 로그-로그 도표에서도  $\hat{x}_{\min}$  이후 꼬리 부분 데이터가 대부분  $\hat{\alpha}$ 의 신뢰구간에 들어는 것으로 직접 확인할 수 있다. 붓스트랩 방법으로 추정된 모수를 사용해 멱법칙 분포를 만드는 경우,  $\hat{x}_{\min}$  이후 꼬리 부분과 분포상 거리가 작으므로 멱법칙 분포가 꼬리부분 데이터를 더 잘 설명하고 있다고 말할 수 있다. 따라서 붓스트랩 방법을 사용한 경우, 사용하지 않은 경우보다 더 높은  $p$ -값을 보이게 되고, 우리가 계산한  $p$ -값도 더 큰 값을 보여주고 있다. 다만 “계기지진3”과 “전체목록”처럼 꼬리부분의 이상값이 두드러지는 경우에는  $p$ -값이 여전히 낮게 나왔다 (Table 4.3). 이를 통해 우리가 사용한 지진목록들이 대체적으로 멱법칙 분포를 잘 따르게 되어 멱법칙 분포 사용의 정당성이 보장되었다고 말할 수 있다.

#### 4.4. 한반도 지진 발생확률의 추정

한반도 지진 발생확률 추정 결과를 Table 4.4에 정리했다. 사용된 데이터는 “계기지진1”, “계기지진2”, “계기지진3”, “전체목록”이며 각각 377, 663, 896, 2,800건의 지진 기록을 기재했다.

비교적 최근에 관측된 지진기록인 “계기지진1”과 “계기지진2”에서 대규모 지진이 발생할 사건확률은 대부분 1/1000 이하로 매우 낮은 수치를 보였다. 하지만 동자료의 기간확률은 그에 비해 상당히 높은 값을 보였다. 다만, 한반도 남부 지역의 기록이 집중된 “계기지진1”에서 사건확률과 기간확률이 한반도 북부 지역에 집중된 “계기지진2”보다 훨씬 낮은 값을 보이고 있는데, 이는 평양 인근, 함경북도 해안지역 혹은 동해 먼바다에서 발생한 대규모 지진 기록이 반영되었기 때문이라 생각한다.

## 5. 결론 및 제언

본 논문에서는 한반도 지진 기록을 통계적으로 설명하고 그것을 바탕으로 앞으로 한반도 지진 발생확률을 예보하기 위해 멱법칙 분포 모형을 도입하였다. 우리가 사용한 한반도 지진 기록은 “기상청, 한반도 역사지진기록”, “조선지진목록”, “한국 지진목록”, “국내지진목록”이었으며, 지진기록 데이터를 분석하기 위해 멱법칙 분포의 척도모수와 멱법칙 현상의 하한을 추정했다. 추정에는 최대가능도통계량추정법과 붓스트랩 방법을 사용되었으며 추정량을 바탕으로 지진기록 데이터가 실제 멱법칙 분포를 따르는지 적합도검정을 통해 검정했고, 마지막으로 한반도 지진 발생확률을 계산해냈다.

우리가 분석한 데이터는 대부분 붓스트랩으로 계산한  $p$ -값이 높게 나와 멱법칙 분포를 대체적으로 잘 따른다고 할 수 있으나, 데이터 수집멱법칙 분포 모수 추정 및 한반도 지진 발생확률의 정확한 계산이 지장이 있었다. 이렇게 데이터 분포에 왜곡이 생기는 이유는 한반도 지진 발생 자체의 내재적인 특성으로 인해 발생하는 부분이 분명히 있겠지만, 관측 기술의 한계로 인한 데이터 생성 자체가 가지는 불완전성, 규모 판정에 있어 연구자의 주관성 등도 원인으로 작용하였다. 이 점과 관련해 KIGAM (2012)는 역사지진의 완결성 부족을 지적하였는데, “기상청, 한반도 역사지진기록”의 MMI 재구 진도의 구텐베르크-리히터 계수(Gutenberg-Richter coefficient)가 전세계적인 추세인 1이 아닌 0.7 정도로 계산이 되어 자료의 완결성이 의심된다는 견해를 보였다. 실제로 “기상청, 한반도 역사지진기록”과 “조선지진목록”의 구텐베르크-리히터 계수를 계산해본 결과 0.723과 0.448이 나와 자료의 신뢰성에 한계가 있음을 알 수 있다. 특히 “조선지진목록”의 환산규모는 “기상청, 한반도 역사지진기록”의 진도와 크기 경향이 불일치되는 기록이 많았는데 “조선지진목록”에서 구체적인 환산규모 추정법을 밝히지 않아 아쉬운 점이 있었다.

데이터 분석 중 모수추정에 있어 척도모수  $\alpha$ 는 최대가능도통계량추정법으로 구하고 멱법칙 현상의 하한  $x_{\min}$ 은 KS 통계량을 이용해 준모수적인 방법으로 추정값을 구했다. 결론부분에서 지적하였듯 데이터의 분포가 불안정한 경우, 특히 꼬리부분에서 불안정한 경우 KS 통계량이 충분히 낮은 멱법칙 현상의 하한  $x_{\min}$ 을 추정하지 못하는 현상이 발생했다. 이는 붓스트랩 방법을 사용하여 꼬리부분의 낮은 빈도를 가지는 이상값의 영향력을 희석시켜 상당히 해소되었으나, 그 결과가 실제  $x_{\min}$ 에 얼마나 접근했는지 확인할 수는 없어 계산값을 완전히 신뢰하기엔 제한이 있었다. Clauset와 Woodard (2013)에선 꼬리 가장 끝부분에 있는 이상값을 제거한 상태에서 모수를 추정하고 그 모수를 바탕으로 이상값이 발생할 확률을 계산하는 방법을 사용했으나, 어떤 데이터가 이상치인지 명확한 기준을 제시하기 못하고 이상치가 여러 개인 경우 이상치들 자체가 가지는 분포의 경향을 제대로 반영하지 못해 분석방법에 있어 한계를 보였다 이는 이후 연구에서 KS 통계량 이외에도 꼬리부분의 왜곡에 더 강건(robust)하게 분포 간 거리를 반영할 수 있는 통계량을 이용해야할 필요성을 나타낸다.

## References

- Chough, S. K. (2000). Tectonic and sedimentary evolution of the Korean peninsula: a review and new view, *Earth-Science Reviews*, **52**, 175–235.

- Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009). Power-law distributions in empirical data, *SIAM Reviews*, **51**, 661–703.
- Clauset, A. and Woodard, R. (2013). Estimating the historical and future probabilities of large terrorist events, *The Annals of Applied Statistics*, **7**, 1838–1865.
- Condie, K. C. (1997). *Plate Tectonics and Crustal Evolution* (4th ed.), Butterworth-Heinemann, **282**.
- Jun M. S. and Jeon, J. S. (2010). Earthquake mechanism characteristics of the Korea Peninsula, *Jigu-Mulli-wa-Mulli-Tamsa*, **13**, 198–202.
- Korea Meteorological Administration (KMA) (2012). *Historical Earthquake Records in Korea (2-1904 Year)*, KMA.
- Korea Meteorological Administration (KMA) (2017). National Seismologic Records, accessed in Jan. 2017: <http://www.kma.go.kr>
- Korea Institute of Geoscience and Mineral Resources (KIGAM) (2012). *Active Fault Map and Seismic Hazard Map*, KIGAM, 2012.
- Kim, S. G and Gao, F. (1995). *Seismic Gaps in Korean Peninsula and Hazard Possibility*, The Seismological Institute, Hanyang University
- Korea Infrastructure Safety and Technology Companion (KISTEC) (2017). Glossary in Seismology - Intensity, Magnitude, accessed in Jan. 2017: <https://www.kistec.or.kr/kistec/earth/earth0303.asp>.
- Lee, K. H. (2010). Comments on seismicity and crustal structure of the Korean Peninsula *Jigu-Mulli-wa-Mulli-Tamsa*, **13**, 256–267.
- North Korea Seismology Laboratory (1986). *The List of Earthquakes in Chosun*, North Korea Seismology Laboratory, 1986.
- Reedman, A. J., and Um, S. H. (1975). *Geology of Korea*, Geological and Mineral Institute of Korea, Seoul, 139.
- Shimazaki, K. (1984). *Mid-plate, plate-margin, and plate-boundary earthquakes and stress transmission in far east*. in A Collection of Papers of the International Symposium on Continental Seismicity and Earthquake Prediction (ISCSEP), The Organizing Committee of ISCSEP, Seismological Press, Beijing.
- USGS (U.S. Geological Survey) (2017). Earthquake Glossary - Earthquake. <https://earthquake.usgs.gov/learn/glossary/?term=earthquake> [accessed in Jan. 2017]

# 지진 관측자료를 기반으로 한 한반도 지진 발생 확률 예측

최서원<sup>a</sup> · 장원철<sup>a,1</sup>

<sup>a</sup>서울대학교 통계학과

(2017년 8월 14일 접수, 2017년 9월 28일 수정, 2017년 9월 29일 채택)

---

## 요약

2016년 9월 경주 지진 이후 원자력발전소, 고층 빌딩, 주택, 교량 등 우리 사회의 설비 자산들이 과연 지진으로부터 얼마나 안전한가, 앞으로 안정성을 담보하기 위해 어떤 정책을 펼쳐야 하는가에 대한 관심이 높아졌다. 본 논문에서는 한반도에서 발생한 역사지진 및 계기지진 목록을 데이터로 사용하여, 맥법칙 분포를 통해 한반도 지진 기록의 크기 분포를 설명하는 모수를 추정한다. 또한 추정한 모수를 바탕으로 미래에 한반도에서 일정 규모 이상의 지진이 발생할 확률을 계산한다. 한반도 미래 지진 발생 확률 계산 모형을 통해 지진 위험도를 파악하고자 하는 것이 본 논문의 목적이다.

주요용어: 한반도 지진 기록, 역사지진, 계기지진, 맥법칙 분포, 한반도 미래 지진 발생확률 예측

---

---

본 논문은 2014년도 정부(미래창조과학부)의 재원으로 한국연구재단 지원을 받아 수행된 기초연구사업임 (No. 2014R1A4A1007895).

<sup>1</sup>교신저자: (08826) 서울특별시 관악구 관악로1, 서울대학교 통계학과. E-mail: wcjang@snu.ac.kr