Statistical aspects of Extreme Value Theory with an application in seismic analysis

貝淵　響　総合研究大学院大学　統計科学専攻　5年一貫制博士課程3年

Introduction

Extreme event **risk** is present in all areas of risk management. Managing risk is vital in our modern society due to the political, economic and environmental impacts of catastrophic disasters. It is known that the statistically justifiable analysis, modelling and prediction of rare events are challenging because the historical data on extreme events are inherently scarce. In order to prevent or prepare for such unfavourable scenarios, the approaches based on the **Extreme Value Theory (EVT)** have been devised.

Main Objectives:

- Model the tail of distribution of the process at risk for risk management.
- Estimate extreme value index (EVI) and extreme quantile.
  Extreme quantile is of particular interest because ability to assess it accurately translates into ability to manage risks effectively.

Why not Central Limit Theory?:

- A Gaussian distribution is often assumed when the conditions of applicability of the Central Limit Theorem are satisfied.
  However, the Central Limit Theorem is not adapted to the very nature of the problem of fitting the large and extreme deviations away from the centre of the distribution because it does not apply to the tails of distributions.

Approach & Challenge:

- We use semi-parametric Hill-based tail estimation methods (i.e. estimators based on a set of upper order statistics and log-transformed data).
- An important challenge in the applications of EVT is the choice of the optimal tail sample fraction. The number of order statistics k can be considered as the effective sample size for the extrapolation outside the range of available observations. Thus, successful practical applications of extreme values heavily depend on the determination of optimal k.

Theoretical foundation of EVT

Assumption and Notation:

- We consider the sample $X_1, X_2, ..., X_n$ of n independent and identically distributed (i.i.d.) random variables with common distribution function (df) F.
- Underlying df F is assumed to be continuous and strictly increasing.
- Maximum is defined as $X_{n,n} = \max(X_1, X_2, ..., X_n)$.

Fisher-Tippett-Gnedenko Theorem:

If there exist a sequence of positive numbers $\{a_n; n \geq 1\}$ and a sequence of numbers $\{b_n; n \geq 1\}$ such that

$$P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) \to G(x) \quad \text{as} \quad n \to \infty,$$

where G is a non-degenerate distribution function, then G must belong to one of the following three distribution functions ($\alpha > 0$)

- *Fréchet*: $\Phi(x) = \begin{cases} 0, & \text{if} \quad x \leq 0, \\ \exp(-x^{-\alpha}), & \text{if} \quad x \geq 0, \end{cases}$
- *Weibull*: $\Psi(x) = \begin{cases} \exp(-|x|^{\alpha}), & \text{if} \quad x \leq 0, \\ 1, & \text{if} \quad x \geq 0, \end{cases}$
- *Gumbel*: $\Lambda(x) = \exp(-\exp(-x)), \quad -\infty < x < \infty.$
- Three types of distributions are the **only possible** limits for the distribution of the normalized maxima, regardless of the underlying df F.

Extreme value distribution & EVI:

Above distributions can be thought of as members of a simple one-parametric family of distribution (*extreme value distribution*), given by

$$G_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma}), \quad \text{for} \quad 1 + \gamma x > 0,$$

where $\gamma = 1/\alpha$ is a new parameter introduced.
The real quantity γ is called the *extreme value index* (EVI). It is a key parameter in the whole of extreme value analysis since it indicates the heaviness of the tail. Furthermore, the sign of EVI is the dominating factor in the description of the tail behaviour of underlying F.

- $\gamma > 0 \Rightarrow$ **Fréchet-Pareto** (heavy-tailed, e.g. earthquakes)
- $\gamma < 0 \Rightarrow$ **Weibull** (short-tailed, e.g. world records in athletics)
- $\gamma = 0 \Rightarrow$ **Gumbel** (light-tailed, e.g. life span)

Tail estimation methods for Pareto-type distribution

EVI estimators:

1. **Hill estimator**
   $$\hat{\gamma}_H = \frac{1}{k}\sum_{j=1}^{k} \log X_{n-j+1,n} - \log X_{n-k,n}, \quad k = 1, ..., n-1.$$
2. **Moment estimator**
   $$\hat{\gamma}_M(k) = \hat{\gamma}_H(k) + 1 - \frac{1}{2}\left(1 - \frac{\hat{\gamma}_H(k)^2}{\hat{\gamma}_H^{(2)}(k)}\right)^{-1} \quad \text{where}$$
   $$\hat{\gamma}_H^{(2)}(k) = \frac{1}{k}\sum_{j=1}^{k}(\log X_{n-j+1,n} - \log X_{n-k,n})^2.$$
3. **Peng's asymptotically unbiased Hill estimator**
   $$\hat{\gamma}_{UH}(k) = \hat{\gamma}_H(k) - \frac{\hat{\gamma}_H^{(2)}(k) - 2(\hat{\gamma}_H(k))^2}{2\hat{\gamma}_H(k)\hat{\rho}}(1-\hat{\rho}) \quad \text{where}$$
   $$\hat{\rho} = \frac{1}{\log 2}\log\frac{\hat{\gamma}_H^{(2)}\left(\frac{n}{2\log n}\right) - 2\left(\hat{\gamma}_H\left(\frac{n}{2\log n}\right)\right)^2}{\hat{\gamma}_H^{(2)}\left(\frac{n}{\log n}\right) - 2\left(\hat{\gamma}_H\left(\frac{n}{\log n}\right)\right)^2}.$$

Extreme quantile estimator:

$$\hat{x}_p = X_{n-k,n} + (X_{n-k,n}\hat{\gamma}_H(k))\left[\left(\frac{k+1}{(n+1)p}\right)^{\hat{\gamma}} - 1\right]\hat{\gamma}^{-1}$$

where $p = \frac{1}{n}$ so that we can extrapolate outside the range of available observations, and for $\hat{\gamma}$ any one of the estimators of EVI γ above can be used.

Seismic extreme value analysis

Background:

- Our data set of earthquakes (scalar seismic moment distribution) in Japan is obtained from the Global Centroid Moment Tensor database, known as the Harvard CMT catalog.
- Data constraints $\to$ 01/01/1976 $\sim$ 31/12/2016.
- Location constraints $\to 30 \leq \text{lat} \leq 45$ and $130 \leq \text{lon} \leq 145$.
- We have collected 2364 observations for last 40 years because extreme value analysis requires a long observation time due to extremely low probability of rare events.
- We assumed that the earthquake data consists of a sample of i.i.d observations.
  (**In many real problems, assumption of the independence of extremal events is not realistic**).

Results:

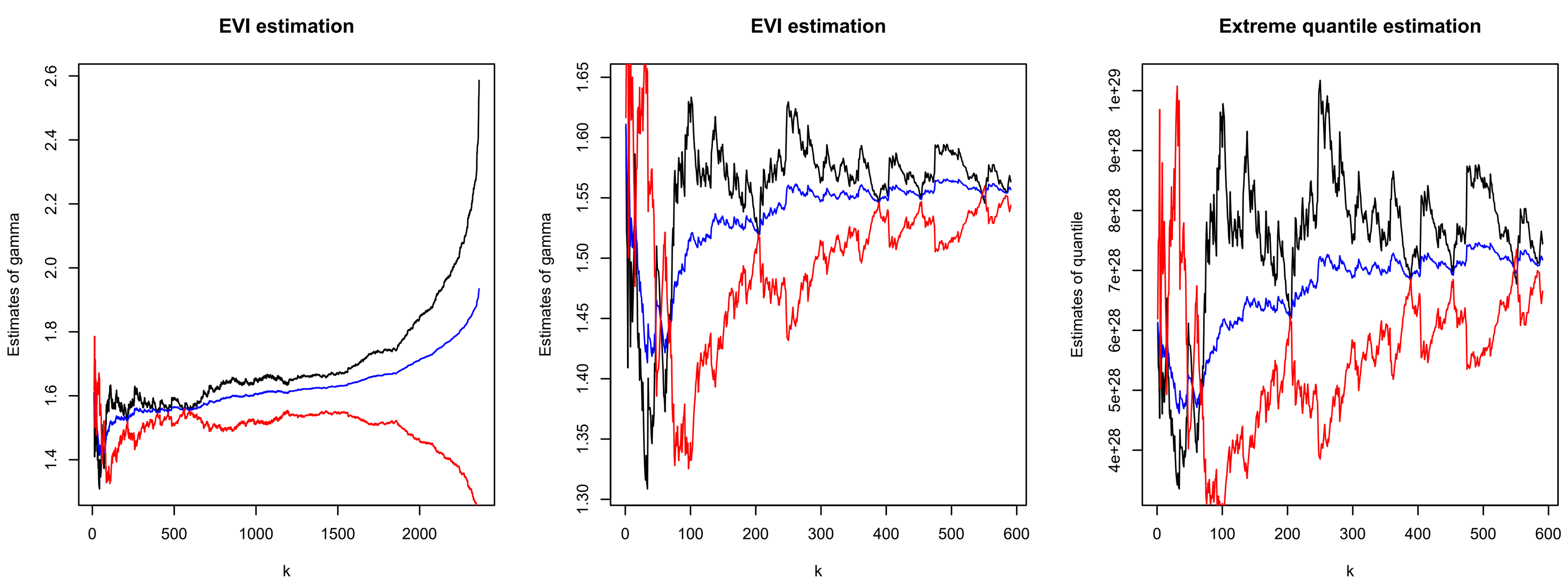

Figure 1: EVI (left, middle) and extreme quantile (right) estimations of scalar seismic moment distribution of earthquakes in Japan between 1976 and 2016, by $\hat{\gamma}_H$ (black), $\hat{\gamma}_M$ (blue) and $\hat{\gamma}_{UH}$ (red) with unrestricted and restricted ranges of k.

Conclusions:

- We found that $\hat{\gamma} = 1.55$, which is in line with the EVI estimations from Pisarenko *et al.* [2] ($1.19 \leq \hat{\gamma} \leq 1.81$) and Goegebeur *et al.* [1] ($1.3 \leq \hat{\gamma} \leq 1.6$).
- Our main finding is that when an earthquake occurs in Japan, probability of an earthquake with magnitude 8.5 or above is $p = 0.0004$ approximately.

References

[1] Goegebeur, Y., Guillou A. and Osmann, M. (2013). *A local moment type estimator for the extreme value index in regression with random covariates.* Canadian Journal of Statistics **42(3)**: 487-507.

[2] Pisarenko, V.F. and Sornette, D. (2003). *Characterization of the frequency of Extreme Earthquake Events by the GPD.* Pure and Applied Geophysics, 160, 2343-2364.