GENERALIZED EXTREME VALUE DISTRIBUTIONS

ZHANGSHENG LAI

Introduction

For a given random variable X with cumulative probability distribution F, we are able to learn the distribution function of the maximum of n independent and identically distributed X_i's i.e. $M_n = \max\{X_1, \ldots, X_n\}$ which is given by,

$$\mathbb{P}(M_n \leq k) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq k) = (F(k))^n$$

Knowing the distribution of X would easily allow us to obtain the distribution function for X_n using the formula above. However, it is very often in real-life problems where we do not know the explicit distributions and can only find suitable models to describe the distribution of our data. With a suitable model in mind, we then depend on statistical methods like method of moments, maximum likelihood estimators and others to estimate the parameters of the hypothesized distribution. With the learnt parameters, we have an estimate of F which might contain small errors when which raised to sufficiently large powers, the errors are not small anymore.

An alternative approach is to not work with F, but instead look for approximate families of F^n which is estimated using extreme data only. Looking at F^n as $n \to \infty$, we see that $F^n(x) \to 0$ as $n \to \infty$ for any $x < x^+$ where x^+ is the upper end-point of F and M_n degenerates to a single point mass at x^+. By doing some linear renormalization of M_n

$$M_n^* = \frac{M_n - b_n}{a_n}$$

the asymptotic of M_n^* will belong to one of the three families: Gumbel, Fréchet or Weibull. These three families can be represented by a single distribution, the generalized extreme value (GEV) distribution,

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-t/\xi}\right\}$$

defined on the set $\{x : 1 + \xi(x-\mu)/\sigma > 0\}$ where the parameters satisfy $-\infty < \mu, \xi < \infty$ and $\sigma > 0$.

Aims

Extreme value analysis is particular useful in the fields of hydrology, where the study of rainfall, extreme floods and others help in the design of hydraulic infrastructure. Underestimation will lead to infrastructure failures and overestimation will lead to high costs that are unnecessary. In the 10th Extreme Value Analysis Conference (EVA 2017[1]), they proposed a challenge to predict spatio-temporal extremes where participants are to estimate high quantiles. From my learnings about extreme value theory, I would then like to apply it for the predictions for the challenge. In my project presentation, I would like to share about my learnings of extreme value theory and how it is applied for the challenge to predict extreme precipitation.

---

[1] http://www.eva2017.nl/Challenge/index.html