GENERALIZED EXTREME VALUE DISTRIBUTIONS

ZHANGSHENG LAI

1 Asymptotic Models

1.1 Model Formulation

The model focuses on the statistical behaviour of $M_n = \max\{X_1, \dots, X_n\}$ where all the X_i's are independent and identically distributed with distribution function F. The distribution of M_n can be derived exactly for all values of n

$$\begin{aligned} \mathbb{P}(M_n \le z) &= \mathbb{P}(X_1 \le z, \dots X_n \le z) \\ &= \prod_{i=1}^{n} \mathbb{P}(X_i \le z) \\ &= \{F(z)\}^n \end{aligned} \tag{1.1}$$

Although we know how to find the distribution of M_n the maximum of iid random variables, not knowing what F makes knowing the above not helpful. We could utilise standard statistical techniques like maximum likelihood to get an estimate $\widehat{F}$ from the observed data the substitute into (1.1). However small errors in the estimate of F can lead to substantial errors in F^n.

The approach we are going to look at here is to accept that F is unknown, instead of estimating F to estimate F^n, we find an estimate of F^n directly, which can be estimated using extreme data only. The idea is similar to the usual method of approximating the distribution of sample means by the normal distribution. So essentially we are doing the extreme value analogue of the central limit theory.

Observe that for a distribution function F with upper end-point z^+, i.e. $F(z^+) = 1$ for any $z < z^+$, $F^n(z) \to 0$ as $n \to \infty$, thus M_n degenerates to a point mass on z^+. To avoid this problem, we do a linear renormalization of the variable M_n:

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

for a sequence of constants $\{a_n > 0\}$ and $\{b_n\}$. By choosing appropriate $\{a_n\}$ and $\{b_n\}$ it stabilizes the location and scale of M_n^* as n grows avoiding problems of degeneracy. Thus we seek limit distributions of M_n^* instead of M_n with appropriate choices of $\{a_n\}$ and $\{b_n\}$.

1.2 Extremal Types Theorem

Theorem 1.1. If there exists sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$\mathbb{P}(M_n - b_n/a_n \le z) \to G(z) \quad \text{as } n \to \infty$$

where G is a non-degenerate distribution function, then G belongs to one of the following families:

$$\begin{aligned} \mathbb{I}: \quad G(z) &= \exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\}, \quad -\infty < z < \infty \\ \mathbb{II}: \quad G(z) &= \begin{cases} 0, & z \le b \\ \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\}, & z > b \end{cases} \\ \mathbb{III}: \quad G(z) &= \begin{cases} \exp\left\{-\left[-\left(\frac{z-b}{a}\right)^{-\alpha}\right]\right\}, & z \le b \\ 1, & z > b \end{cases} \end{aligned}$$

for parameters $a > 0, b$ and for families $\mathbb{II}, \mathbb{III}$, $\alpha > 0$.

These three classes of distribution are called **extreme value distributions** with the types $\mathbb{I}, \mathbb{II}$ and $\mathbb{III}$ widely known as the **Gumbel**, **Fréchet** and **Weibull** families respectively. Theorem 1.1 implies that when M_n can be stabilized with suitable sequences $\{a_n > 0\}$ and b_n the corresponding normalized M_n^* has a limiting distribution that must be one of the three extreme distributions. It is in this sense that the theorem provides an extreme value analog of central limit theorem.

1.3 The Generalized Extreme Value Distribution

The three types of limits have different characteristic, corresponding to the different kind of tail behaviour for the distribution function F of the X_i's. We have the Gumbel to be unbounded, Fréchet bounded below and the Weibull bounded above. The density of Gumbel decays exponentially whereas it is polynomially for the Fréchet.

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 $\sigma > 0$.

2 Hill's Estimator

The motivation of this paper came from considering a random sample $Z_1, \ldots, Z_k$ from a distribution of F from a unit interval with $F(x) \sim Cx^\alpha$ as $x \to 0$. We would also like to be able to draw inference about α without making assumptions about the form of F elsewhere.

Suppose that a sample $Y_1, \ldots, Y_k$ is drawn from the population with distributino G and let

$$Y^{(1)} \geq Y^{(2)} \geq \ldots \geq Y^{(k)}$$

be the order statistics. On the basis of theoretical arguments or previous data, it is believed, or at least the hypothesis is tentatively entertained, that G has a known functional form, say $G(y) = w(y; \theta)$, for y *sufficiently large* and θ be a vector of parameters.