Generalized Extreme Value Distribution

Zhangsheng Lai

April 12, 2017

Model Formulation

The model focuses on the statistical behaviour of $M_n = \max\{X_1, \ldots, X_n\}$ 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

$$\begin{aligned}
\mathbb{P}\{M_n \leq z\} &= \mathbb{P}\{X_1 \leq z, \ldots X_n \leq z\} \\
&= \prod_{i=1}^{n} \mathbb{P}\{X_i \leq z\} \\
&= \{F(z)\}^n \qquad (1)
\end{aligned}$$

Model Formulation

Problems encountered:

- not knowing what F is does not allow us to find the distribution of M_n.
- maximum likelihood can be used to get an estimate $\widehat{F}$, but small errors in the estimate of F can lead to substantial errors in F^n.

The approach to the problem is to find an estimate of F^n directly using extreme data only which gives us the extreme value analogue of the central limit theory.

Model Formulation

Observe that for a distribution function F with upper end-point z^+, i.e. z^+ is the smallest value of z such that $F(z^+) = 1$, 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.

Extremal Types Theorem

Theorem (Fisher-Tippett-Gnedenko)

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

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

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

$$\begin{aligned}
\textit{Gumbel}: \quad G(z) &= \exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\}, \quad -\infty < z < \infty \\
\textit{Fréchet}: \quad G(z) &= \begin{cases} 0, & z \leq b \\ \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\}, & z > b \end{cases} \\
\textit{Weibull}: \quad G(z) &= \begin{cases} \exp\left\{-\left(-\frac{z-b}{a}\right)^{\alpha}\right\}, & z \leq b \\ 1, & z > b \end{cases}
\end{aligned}$$

for parameters $a > 0$, b and for families Fréchet and Weibull, $\alpha > 0$.

A much more flexible approach is to reformulate the three models into a single distribution which has the form,

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \tag{2}$$

defined on the set $\{z : 1+\xi(z-\mu)/\sigma > 0\}$ where the parameters satisfy $-\infty < \mu, \xi < \infty$ and $\sigma > 0$. The μ is the location parameter, σ is the scale parameter and ξ is the shape parameter.

Extremal Types Theorem

Theorem (Generalized Extreme Value)

If there exists sequence of constants $\{a_n > 0\}$ *and* $\{b_n\}$ *such that*

$$\mathbb{P}\{(M_n - b_n)/a_n \leq z\} \to G(z) \quad \text{as } n \to \infty \tag{3}$$

for a non-degenerate distribution function G, then G is a member of the GEV family

$$G(x) = \exp\left\{ -\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi} \right\} \tag{4}$$

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

Extremal Types Theorem

By interpreting the limit in Theorem 2 as an approximation for large values of n, suggests the use of the GEV family for modelling the distribution of maxima of long sequences. With this result we can solve for the normalizing constants that are unknown is practice. Given (3) holds

$$\mathbb{P}\{(M_n - b_n)/a_n \leq z\} \approx G(z)$$

for large n. With some manipulation

$$\mathbb{P}\{M_n \leq z\} \approx G((z - b_n)/a_n) = G^*(z)$$

where G^* is another member of the GEV family. This says that we can approximate the distribution of M_n^* by a member of the GEV for large n, the distribution of M_n itself can also be approximated by a different member of the same family.

Definition

A distribution G is said to be **max-stable** if, for every $n = 2, 3, \ldots$ there are constants $\alpha_n > 0$ and β_n such that

$$G^n(\alpha_n z + \beta_n) = G(z)$$

Outline Proof of the Extremal Types Theorem

Theorem

A distribution is max-stable iff it is a generalized extreme value distribution.

Proof.

We shall skip the forward direction proof as it utilizes machinery that are not trivial. To show the converse, let the Fréchet $G(x)$ be given, then

$$\begin{aligned} G^n(z) &= \exp\left\{-n\left(\frac{z-b}{a}\right)^{-\alpha}\right\} \\ &= \exp\left\{-\left(\frac{z-b}{an^{1/\alpha}}\right)^{-\alpha}\right\} \\ &= G(\alpha_n z + \beta_n) \end{aligned}$$

where $\alpha_n = n^{-1/\alpha}$ and $\beta_n = 1 + n^{-1/\alpha}$. In a similar fashion we can prove it for Weibull. $\square$

Proof.

For Gumbel, we have

$$\begin{aligned} G^n(z) &= \exp\left\{-n \exp\left[-\left(\frac{z-b}{a}\right)\right]\right\} \\ &= \exp\left\{-\exp\left(\log n\right) \cdot \exp\left[-\left(\frac{z-b}{a}\right)\right]\right\} \\ &= \exp\left\{-\exp\left[\log n - \left(\frac{z-b}{a}\right)\right]\right\} \\ &= G(\alpha_n z + \beta_n) \end{aligned}$$

where $\alpha_n = 1$ and $\beta_n = -a \log n$. $\square$

Outline Proof of the Extremal Types Theorem

Theorem 4 is used directly in the proof of extremal types theorems. We start be considering M_{nk}, the maximum random variable in a sequence of $n \times k$ variables for some large value of n. This can be regarded as the maximum of a single sequence of length $n \times k$ or as the maximum of k maxima, each of which is a maximum of n observations. Let's assume that the limit distribution of $(M_n - b_n)/a_n$ is G, thus for sufficiently large n

$$\mathbb{P}\{(M_n - b_n)/a_n\} \approx G(z)$$

Outline Proof of the Extremal Types Theorem

Hence for any integer k we also have

$$\mathbb{P}\{(M_{nk} - b_{nk})/a_{nk}\} \approx G(z)$$

but now we recall the definitions of M_n and M_{nk}, we have

$$\begin{aligned} G\left(\frac{z - b_{nk}}{a_{nk}}\right) &\approx \mathbb{P}\{(M_{nk} - b_{nk})/a_{nk} \leq z\} \\ &= [\mathbb{P}\{(M_n - b_n)/a_n \leq z\}]^k \approx G^k\left(\frac{z - b_n}{a_n}\right) \end{aligned}$$

Therefore G and G^k are identically apart except for location and scale coefficients. Thus G is max-stable and a member of GEV by Theorem 4

Example

Example

If $X_1, X_2, \ldots$ is a sequence of independent and identically distributed Exp(1) random variables, $F(x) = 1 - e^{-x}$ for $x > 0$, choosing $a_n = 1$ and $b_n = \log n$

$$\begin{aligned}
\mathbb{P}\{(M_n - b_n)/a_n\} &= F^n(z + \log n) \\
&= \left[1 - e^{-(z+\log n)}\right]^n \\
&= \left[1 - n^{-1}e^{-z}\right]^n \to \exp(-e^{-z}) \text{ as } n \to \infty
\end{aligned}$$

here we use the fact that $\lim_{n\to\infty}\left(1 - \frac{1}{n}\right)^n = \frac{1}{e}$. Hence with the chosen a_n and b_n, the limit of M_n converges to the Gumbel distribution as $n \to \infty$. This corresponds to $\xi = 0$ in the GEV family. ▲

Example

If $X_1, X_2, \ldots$ is a sequence of independent Fréchet variables, $F(x) = \exp(-1/x)$ for $x > 0$. Letting $a_n = n$ and $b_n = 0$

$$\begin{aligned}
\mathbb{P}\{(M_n - b_n)/a_n \leq z\} &= F^n(nz) \\
&= \Big[\exp\{-1/nz\}\Big]^n \\
&= \exp(-1/z)
\end{aligned}$$

as $n \to \infty$, for each $z > 0$. Hence the limit in this case which is an exact result for all n since the Fréchet is max-stable is also the Fréchet distribution. This corresponds to $\xi = 1$ in the GEV family. ▲

Example

If $X_1, X_2, \ldots$ are a sequence of independent uniform $\mathrm{U}(0,1)$ variables, $F(x) = x$ for $0 \le x \le 1$. For fixed $z < 0$, suppose $n > -z$ and let $a_n = 1/n$ and $b_n = 1$. Then

$$\begin{aligned} \mathbb{P}\{(M_n - b_n)/a_n \le z\} &= F^n(n^{-1}z + 1) \\ &= \left(1 + \frac{z}{n}\right)^n \\ &\to e^z \text{ as } n \to \infty \end{aligned}$$

Hence the distribution is Weibull type with $\xi = -1$ in the GEV family. ▲