Daniel Matheson
UW ID 20270871
ACTSC 991 – Spring 2019

Chapter 4: The Weibull distribution and its uses

The Weibull distribution is a probability distribution with various uses in Actuarial Science, Electrical and Industrial Engineering, Weather Forecasting, Hydrology and other fields. In the context of this course, it is a particular case of the *Generalized Extreme Value (GEV) distribution.* In this lesson, we will introduce the GEV and explain how it is used, with emphasis on the role of the Weibull distribution.

Suppose we have a sequence of financial losses $(X_i)_{i\in\mathbb{N}}$, and our goal is to find the limiting distribution (as $n \uparrow \infty$) of the *block maximum*, given by

$$M_n := \max\{X_1, \ldots, X_n\}. \tag{1}$$

For this M_n, one can show that $M_n \underset{n\uparrow\infty}{\longrightarrow} \chi_F$, where

$$\chi_F := \sup\{x \in \mathbb{R} \mid F(x) < 1\} = F^{\leftarrow}(1) \leq \infty. \tag{2}$$

To give some intuition before we move on, let us recall the Central Limit Theorem (CLT).

Theorem 4.1: *(CLT)*
Let $(X_k)_{k=1}^{\infty}$ be an independent and identically distributed (iid) sequence of random variables such that $\mu = \mathbb{E}[X_1], \sigma^2 = \mathrm{Var}(X_1) < \infty$ and $S_n = \sum_{k=1}^{n} X_k$. Then

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \underset{n\uparrow\infty}{\xrightarrow{\mathrm{d}}} \mathcal{N}(0,1).$$

Note that for $c_n = \sqrt{n}\sigma, d_n = n\mu$, we can write this convergence as

$$\lim_{n\uparrow\infty} \mathbb{P}\left(\frac{S_n - d_n}{c_n} \leq x\right) = \Phi(x). \tag{3}$$

With this in mind, we consider the following definition.

Definition 4.2: *(Maximum Domain of Attraction)*
For the block maximum M_n as defined in (1), suppose we find *normalizing sequences* of real numbers $(c_n)_{n=1}^{\infty} \subseteq (0,\infty), (d_n)_{n=1}^{\infty} \subseteq \mathbb{R}$ such that

$$\mathbb{P}\left(\frac{M_n - d_n}{c_n} \leq x\right) = \mathbb{P}(M_n \leq c_n x + d_n) = F^n(c_n x + d_n) \underset{n\uparrow\infty}{\longrightarrow} H(x),$$

for some non-degenerate CDF H (i.e not a point-mass). Then F is in the *maximum domain of attraction (MDA)* of H. Which we denote as $F \in \mathrm{MDA}(H)$.

As you might notice, there are some similarities in this definition to the equation (3) from the CLT. As in the CLT, the normalization of the sequence $(M_n)_{n\in\mathbb{N}}$ allows for a non-degenerate limiting distribution. Indeed, we will soon define an analogue to the CLT for block maxima – however, in this case the resulting limiting distribution will not be a normal distribution, but the *Generalized Extreme Value (GEV) distribution*, which is given by

$$H_\xi(x) = \begin{cases} \exp\left(-(1+\xi x)^{-\frac{1}{\xi}}\right), & \text{if } \xi \neq 0 \\ \exp\left(-e^{-x}\right), & \text{if } \xi = 0 \end{cases}, \quad \text{where } (1+\xi x) > 0. \tag{4}$$

If required, one can use the three-parameter GEV family with the location-scale transform

$$H_{\xi,\mu,\sigma}(x) = H_\xi\left(\frac{x-\mu}{\sigma}\right), \mu \in \mathbb{R}, \sigma > 0.$$

Note: The *Weibull distribution* is H_ξ when $\xi < 0$. What distinguishes the Weibull distribution from the other cases of the GEV (Fréchet when $\xi > 0$, and Gumbel when $\xi = 0$) is that it has an upper bound, i.e $\chi_F < \infty$. Therefore it is should only be used if losses $(X_i)_{i\in\mathbb{N}}$ are capped below a certain threshold.

Theorem 4.3: *(Fisher-Tippett-Gnedenko) – "CLT" for block maxima*
If $F \in \mathrm{MDA}(H)$, for a non-degenerate H, then $H = H_{\xi,\mu\,\sigma}$ (GEV) for some $\xi \in \mathbb{R}, \mu \in \mathbb{R}, \sigma > 0$.

What can we take away from this?

- If the location-scale transformed maxima of iid random variables (M_n) converges in distribution to a non-degenerate limit as $n \uparrow \infty$, then the limiting distribution must be of GEV type.
- All common continuous distributions are in the MDA of some GEV distribution. e.g Normal, log-normal, Gamma, Pareto, student-t etc.
- Note that if the sequences $(c_n)_{n=1}^\infty$ and $(d_n)_{n=1}^\infty$ can be found (see Example 4.4), they can be modified so that the resulting $H_{\xi,\mu,\sigma}$ appears in standard form (i.e with $\mu = 0, \sigma = 1$.)

Example 4.4: (Exponential Distribution)
For $(X_i)_{i\in\mathbb{N}} \overset{\text{iid}}{\sim} \mathrm{Exp}(\lambda)$, setting $c_n = \frac{1}{\lambda}, d_n = \frac{\log(n)}{\lambda}$ gives

$$F^n(c_n x + d_n) = \left(1 - \exp\left(-\lambda\left(\left(\frac{1}{\lambda}\right)x + \frac{\log(n)}{\lambda}\right)\right)\right)^n = \left(1 - \frac{\exp(-x)}{n}\right)^n \underset{n\uparrow\infty}{=} \exp\left(-e^{-x}\right) = H_0(x)$$

Key Question: How do we know when to use the Weibull distribution?

Definition 4.5: *(Slowly/regularly varying functions)*

(1) A positive function L on $(0,\infty)$ is *slowly varying* at ∞ if $\lim_{x\uparrow\infty} \frac{L(tx)}{L(x)} = 1, t > 0$. The class of all such functions is denoted $\mathcal{R}_0$. e.g $\log \in \mathcal{R}_0$.
(2) A positive function h on $(0,\infty)$ is *regularly varying at* ∞ *with index* $\alpha \in \mathbb{R}$ if $\lim_{x\uparrow\infty} \frac{h(tx)}{h(x)} = t^\alpha, t > 0$. The class of all such functions is denoted by $\mathcal{R}_\alpha$. e.g $x^\alpha L(x) \in R_\alpha \ \forall\ L \in \mathcal{R}_0$.

Detecting whether F is in Weibull MDA analytically
Theorem 4.6 *(Weibull MDA)*
For $\xi < 0$, $F \in \mathrm{MDA}(H_\xi)$ if and only if $\chi_F < \infty$ and $\overline{F}\left(\chi_F - \frac{1}{x}\right) = 1 - F\left(\chi_F - \frac{1}{x}\right) = x^{\frac{1}{\xi}} L(x)$ for some $L \in \mathcal{R}_0$. The normalizing sequences can then be chosen as $c_n = \chi_F - F^{\leftarrow}\left(1 - \frac{1}{n}\right)$ and $d_n = \chi_F, n \in \mathbb{N}$.

Detecting whether F is in Weibull MDA in practice: via Block Maxima Method

(1) Suppose $(x_i)_{i\in\mathbb{N}}$ are realizations of losses $(X_i)_{i\in\mathbb{N}} \overset{\text{iid}}{\sim} F \in MDA(H_\xi), \xi \in \mathbb{R}$, where F is not known. By Theorem 4.3,

$$\mathbb{P}\left(\frac{M_n - d_n}{c_n} \le \frac{x - d_n}{c_n}\right) \simeq H_{\xi,\mu=d_n,\sigma=c_n}(x).$$

Therefore, if we estimate this $H_{\xi,\mu,\sigma}$, we can model the distribution of the maximal losses, without even knowing the original distribution.

(2) To fit $\boldsymbol{\theta} = (\xi,\mu,\sigma)$, divide the realizations into m blocks of size n, denoted $M_{n1},\ldots,M_{nm}$ (e.g daily returns $\longrightarrow$ monthly maxima). Finding the right balance between n and m is difficult, more on this next time. From these m blocks, we use maximum likelihood estimation to find

$$\hat{\boldsymbol{\theta}} = \left(\hat{\xi},\hat{\mu},\hat{\sigma}\right) = \mathrm{argmax}_{\boldsymbol{\theta}}\ \ell\left(\boldsymbol{\theta}; M_{n1},\ldots,M_{nm}\right) = \sum_{i=1}^m \log\left(\frac{1}{\sigma} h_\xi\left(\frac{M_{ni}-\mu}{\sigma}\right)\right),$$

which is easily done in R with packages such as 'qrmtools'. If $\hat{\xi} < 0$, we are fitting the block maxima with the Weibull distribution.