1 Extreme Value Statistics

1.1 The Gumbel distribution

Let us consider a collection of random variables $\{H_\alpha\}$ chosen *independently* from some PDF $p(H)$, and the extremum

$$X = \max\{H_1, H_2, \ldots, H_K\}\,. \tag{1.1}$$

The *cumulative probability* that $X \leq S$ is the product of probabilities that any selected H_α is less than S, and thus given by

$$\begin{aligned} P_K(X \leq S) &= \text{Prob.}(H_1 \leq S) \times \text{Prob.}(H_2 \leq S) \times \cdots \times \text{Prob.}(H_K \leq S) \\ &= \left[\int_{-\infty}^{S} dH p(H)\right]^K \\ &= \left[1 - \int_{S}^{\infty} dH p(H)\right]^K . \end{aligned} \tag{1.2}$$

For large K, typical values of S are in the tail of $p(H)$, which implies that the integral is small, justifying the approximation

$$P_K(S) \approx \exp\left[-K \int_S^\infty dH p(H)\right]. \tag{1.3}$$

Assume, as we shall demonstrate shortly, that $p(H)$ falls exponentially in its tail, as $ae^{-\lambda H}$. We can then write

$$\int_S^\infty dH p(H) = \int_S^\infty dH a e^{-\lambda H} = \frac{a}{\lambda} e^{-\lambda H}. \tag{1.4}$$

The cumulative probability function $P_K(S)$ is therefore

$$P_K(S) = \exp\left[-\frac{Ka}{\lambda} e^{-\lambda S}\right], \tag{1.5}$$

with a corresponding PDF of

$$p_K(S) = \frac{dP_K(S)}{dS} = Ka \exp\left(-\lambda S - \frac{Ka}{\lambda} e^{-\lambda S}\right). \tag{1.6}$$

The exponent

$$\phi(S) \equiv -\lambda S - \frac{Ka}{\lambda} e^{-\lambda S}, \tag{1.7}$$

has an extremum when

$$\frac{d\phi}{dS} = -\lambda + Kae^{-\lambda S^*} = 0\,, \tag{1.8}$$

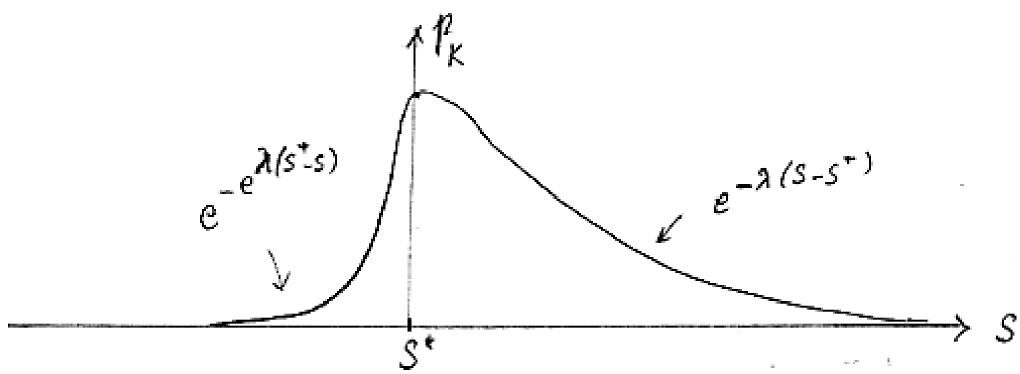
corresponding to

$$S^* = \frac{1}{\lambda} \log \left(\frac{Ka}{\lambda} \right) . \tag{1.9}$$

Equation (1.9) gives the most probable value for the extremum S^*, in terms of which we can re-express the PDF in Eq. (1.6) as

$$p_K(S) = \lambda \exp \left[-\lambda(S - S^*) - e^{-\lambda(S-S^*)} \right] . \tag{1.10}$$

Evidently, once we determine the most likely score S^*, the rest of the probability distribution is determined by the single parameter λ. The PDF in Eq. (1.10) is known as the *Gumbel* or the *Fisher-Tippett* extreme value distribution (EVD). It is characterized by an exponential tail above S^* and a much more rapid decay below S^*. It looks *nothing like* a Gaussian, which is important if we are trying to gauge significances by estimating how many standard deviations a particular value falls above the mean.



1.2 Other extreme value distributions

The Gumbel distribution describes the extreme of any large number of independent random variables, as long as the tail of the distribution vanishes more rapidly than a power law. There are in fact three classes of extremal distributions depending on the behavior of the tail of $p_1(r)$ for a single variable, as follows:

1. If at large x, $\overline{P}_1(x) \simeq c \exp[-(x/a)^p]$, then $P_N(x)$ again assumes the Gumbel form as shown below.

$$p_1(x) = -\frac{d\overline{P}_1(x)}{dx} = \frac{cpx^{p-1}}{a^p} \exp\left(-(x/a)^p\right) \tag{1.11}$$

$$p_1'(x) = \frac{dp_1(x)}{dx} = \left\{ \frac{cp(p-1)x^{p-2}}{a^p} - \frac{cp^2x^{2p-2}}{a^{2p}} \right\} \exp\left(-(x/a)^p\right) \tag{1.12}$$

The most likely value of x^* is now obtained as:

$$
\begin{aligned}
0 &= p_1'(x^*) + Np_1^2(x^*) \\
0 &= \frac{cpx^{*p-2}}{a^p} \exp\left(-\left(x^*/a\right)^p\right) \left\{ p - 1 - p\frac{x^{*p}}{a^p} + Ncp\frac{x^{*p}}{a^p} \exp\left(-\left(x^*/a\right)^p\right) \right\}
\end{aligned}
$$

To the leading order in N we get:

$$N c \exp\left(-\left(x^*/a\right)^p\right) = 1 = N\overline{P}_1(x^*) \tag{1.13}$$

$$x^* = a\left\{\ln(Nc)\right\}^{1/p} \tag{1.14}$$

Cumulative distribution is:

$$
\begin{aligned}
P_N(x) &\approx \exp\left(-N\overline{P}_1(x)\right) \\
P_N(x) &= \exp\left(-\exp\left(\frac{x^{*p} - x^p}{a^p}\right)\right) \\
P_N(x) &\approx \exp\left(-\exp\left(\frac{p\ x^{*p-1}}{a^p}(x^* - x)\right)\right) \qquad (1.15)
\end{aligned}
$$

Again we have obtained cumulative Gumbel distribution with parameter $\lambda = p\, x^{*p-1}/a^p$.

2. The *Frechet distribution* is obtained when at large x, $P_1(x) \simeq cx^{-p}$.

$$
\begin{aligned}
p_1(x) &= -\frac{d\overline{P}_1(x)}{dx} = cp\ x^{-p-1} \qquad (1.16) \\
p_1'(x) &= \frac{dp_1(x)}{dx} = -cp(p+1)\ x^{-p-2} \qquad (1.17)
\end{aligned}
$$

Now we can use equation (**??**) to find x^*:

$$
\begin{aligned}
0 &= p_1'(x^*) + Np_1^2(x^*) \\
0 &= cp\ x^{*(-p-2)} \left(-(p+1) + Ncp\ x^{*(-p)}\right)
\end{aligned}
$$

To the leading order in N we get:

$$Nc\ x^{*(-p)} = \frac{p+1}{p} = N\overline{P}_1(x^*) \tag{1.18}$$

$$x^* = \left(\frac{Ncp}{p+1}\right)^{1/p} \tag{1.19}$$

We can see that this time $N\overline{P}_1(x^*) \neq 1$. Cumulative distribution is:

$$\begin{aligned} P_N(x) &\approx \exp\left(-N\overline{P}_1(x)\right) \\ P_N(x) &= \exp\left(-\left(\frac{p+1}{p}\right)\left[\frac{x}{x^*}\right]^{-p}\right) \\ P_N(x) &= \exp\left(-\left(\frac{p+1}{p}\right)\left[1+\frac{x-x^*}{x^*}\right]^{-p}\right) \end{aligned} \tag{1.20}$$

3. Finally, the *Weibull distribution* is obtained when the random variable takes only values $r \leq a$, and with $P_1(x) \simeq c(a-x)^p$ as $x \to a$.

$$p_1(x) = -\frac{d\overline{P}_1(x)}{dx} = cp(a-x)^{p-1} \tag{1.21}$$

$$p_1'(x) = \frac{dp_1(x)}{dx} = -cp(p-1)(a-x)^{p-2} \tag{1.22}$$

Now we can use equation (**??**) to find x^*:

$$\begin{aligned} 0 &= p_1'(x^*) + Np_1^2(x^*) \\ 0 &= cp(a-x^*)^{p-2}\left[-(p-1)+Ncp(a-x^*)^p\right] \end{aligned}$$

To the leading order in N we get:

$$Nc(a-x^*)^p = \frac{p-1}{p} = N\overline{P}_1(x^*) \tag{1.23}$$

$$x^* = \left(\frac{Ncp}{p+1}\right)^{1/p} \tag{1.24}$$

We can see that again $N\overline{P}_1(x^*) \neq 1$. Cumulative distribution is:

$$\begin{aligned} P_N(x) &\approx \exp\left(-N\overline{P}_1(x)\right) \\ P_N(x) &= \exp\left(-\left(\frac{p-1}{p}\right)\left[\frac{a-x}{a-x^*}\right]^{p}\right) \\ P_N(x) &= \exp\left(-\left(\frac{p-1}{p}\right)\left[1-\frac{x-x^*}{a-x^*}\right]^{p}\right) \end{aligned} \tag{1.25}$$