BASICS OF EXTREME VALUE THEORY FOR ACTUARIES

NOTES OF OLIVIER WINTENBERGER

1. Motivation for Actuaries

Extreme Value Theory (EVT) is a critical tool for actuaries, particularly in the assessment of rare but high-impact events. Its applications span risk management, capital requirement estimation, and solvency assessment.

1.1. **Solvency Assessment.**

Solvency II is a regulatory framework for insurance and reinsurance companies in the European Union. It requires firms to hold sufficient capital to withstand extreme financial stress. The Solvency Capital Requirement (SCR) is the amount of capital required to ensure that the (re)insurance company can meet its obligations over the next 12 months with a probability of at least 99.5%. Formally, this is expressed as:

$$\mathbb{P}\left(\min_{0\leqslant t\leqslant 1} R_t \leqslant -c_0\right) = 0.005,$$

where:

- $(R_t)_{0\leqslant t\leqslant 1}$ represents the company's reserve evolution over the next year,
- c_0 is the capital requirement to be estimated to cover potential losses,
- 0.005 is the probability level associated with the ruin event.

This risk measure is complex and often requires closed-form solutions under specific conditions, such as the standard formula (which assumes a Gaussian setting) or ruin probability calculus (e.g., Lindeberg's model under the Cramér assumption). The Cramér assumption requires the existence of a constant $c > 0$ such that:

$$\mathbb{E}[\exp(c|R_t|)] < \infty, \quad 0 < t < 1.$$

This assumption is restrictive, as it implies that the cost process $-R_t$ has moments of all orders. However, in practice, costs $-R_t$ can exhibit extremely large values, raising the question:

What happens in more realistic scenarios where the costs $-R_t$ can have infinite moments of order $k \geqslant 1$?

EVT provides tools to address this question by modeling the tails of distributions, allowing actuaries to estimate the probability and impact of extreme events even when traditional gaussian or exponential moment-based methods fail [3, 6]. In cases where no closed-form solution exists, stress testing and scenario analysis are commonly used in practice.

Date: October 9, 2025.

1.2. **Insurability.** EVT is especially relevant for assessing operational risk and catastrophic losses in insurance. The fundamental principle of insurability is rooted in the cost/frequency formula, which assumes well-behaved claim costs and frequencies. Specifically, consider a portfolio where:

- Claim costs (C_i) are independent and identically distributed (iid),
- The number of claims N is a random variable, independent of (C_i).

The total risk I is given by the aggregate claim amount:

$$I = \sum_{i=1}^{N} C_i.$$

Under these assumptions, the expected total risk is:

$$\mathbb{E}[I] = \mathbb{E}\Big[\sum_{i=1}^{N} C_i\Big] = \mathbb{E}[N] \cdot \mathbb{E}[C].$$

The total premium π must at least cover the expected risk, i.e., $\pi \geq \mathbb{E}[I]$. This premium is then spread across policyholders, reflecting the principle of risk mutualization.

However, this basic principle only holds under the following assumptions:

- Independence between claim costs and claim frequency,
- Finite first moment of claim costs, i.e., $\mathbb{E}[C] < \infty$.

Insurability issues arise when these assumptions are violated, particularly when $\mathbb{E}[I] = \infty$. In such cases, the expected value $\mathbb{E}[I]$ is no longer a meaningful measure of risk. Instead, the focus shifts to the tail distribution of I, specifically the probability $\mathbb{P}(I > \pi)$, where π is the total premium.

When the catastrophic event $\{I > \pi\}$ (for large π) is primarily driven by an individual claim $M = \max_{1\leqslant i\leqslant N} C_i$ with an extremely large cost, the mutualization principle breaks down. We will see that this may occur when the claim distribution is so heavy-tailed that $\mathbb{E}[C^k] = \infty$ for some $k \geqslant 1$.

In such scenarios, reinsurance becomes essential to mitigate the risk of catastrophic losses and ensure the insurability of the portfolio. For example, natural catastrophes often trigger specialized reinsurance mechanisms, sometimes activated or backed by governments to manage systemic risks.

1.3. **Reinsurance.** Reinsurance is a risk management tool used by insurance companies to transfer a portion of their risk portfolio to another party, typically a reinsurer. This practice helps insurers stabilize their financial results, protect against catastrophic losses, and maintain solvency.

The primary insurer cedes a part of its risk exposure to the reinsurer, reducing its liability for large or unexpected claims. Common forms include:

- Proportional reinsurance, where risks and premiums are shared proportionally between the primary insurer and the reinsurer.
- Non-proportional reinsurance, where the reinsurer covers claims exceeding a certain threshold $u > 0$, e.g., excess-of-loss reinsurance.

In non-proportional reinsurance, the risk is formalized as follows:

- The primary insurer's risk is $C_i \wedge u$ for each claim C_i.
- The reinsurer's risk is $(C_i - u) \cdot \mathbf{1}(C_i > u)$.

The expected risk for the primary insurer is then:

$$\mathbb{E}\Big[\sum_{i=1}^{N} C_i \wedge u\Big] = \mathbb{E}[N] \cdot \mathbb{E}[C \wedge u].$$

For large π, the event $\{X > \pi\}$ is no longer equivalent to $\{M > \pi\}$, as the reinsurer's risk may be driven by a single catastrophic claim M satisfying $M > u$. Mutualization is not possible for the reinsurer, as their risk exposure is concentrated on extreme events. Therefore, reinsurers must design contracts with primary insurers that focus on specific potential catastrophic events.

EVT is useful for the primary insurer to leverage reinsurance, allowing insurance companies to underwrite policies that would otherwise be uninsurable due to the potential for catastrophic losses. For the reinsurer, EVT is even more critical, as it informs the design and pricing of reinsurance contracts or treaties, particularly for extreme and unpredictable events such as natural disasters.

2. First Principles in Extreme Value Theory

2.1. **Order Statistics.** Most actuarial problems involve estimating a quantile q_p of level $p \in (0,1)$, defined as:

$$q_p = F^{\leftarrow}(p) := \inf\{x \colon F(x) \geq p\},$$

where $F^{\leftarrow}$ is the generalized inverse of the cumulative distribution function (CDF) $F(x) = \mathbb{P}(X \leq x)$, which is right-continuous (cdlg). The Value-at-Risk (VaR) serves as the actuarial analog of a quantile and is the most widely used risk measure in quantitative risk management:

Definition 2.1 (VaR)**.** *The Value-at-Risk of order $p \in (0,1)$ is defined as:*

$$\mathrm{VaR}_p = \inf\{x \in \mathbb{R} \colon \mathbb{P}(X \leq x) \geq p\}.$$

Estimating VaR relies primarily on the asymptotic behavior of order statistics:

Definition 2.2 (Order Statistics)**.** *Let $(X_i)_{1 \leq i \leq n}$ be an iid sample with distribution F. The ordered sample is denoted by $(X_{(i)})_{1 \leq i \leq n}$, such that:*

$$X_{(n)} \leq \cdots \leq X_{(2)} \leq X_{(1)} \quad \textit{almost surely (a.s.)}.$$

Here, $M_n := X_{(1)}$ is the sample maximum, $m_n := X_{(n)}$ is the sample minimum, and $X_{(i)}$ is the i-th largest order statistic for $i = 1, \ldots, n$.

The empirical distribution function connects order statistics to the quantiles of the empirical distribution:

Definition 2.3 (Empirical Distribution)**.** *Given a sample $X_1, \ldots, X_n$ with distribution F, the empirical distribution function is defined as:*

$$\hat{F}_n(x) = \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}(X_t \leq x), \quad x \in \mathbb{R}.$$

The empirical distribution function $\hat{F}_n(x) : \mathbb{R} \to [0,1]$ is right-continuous and monotonically increasing, with jumps at the order statistics of the sample.

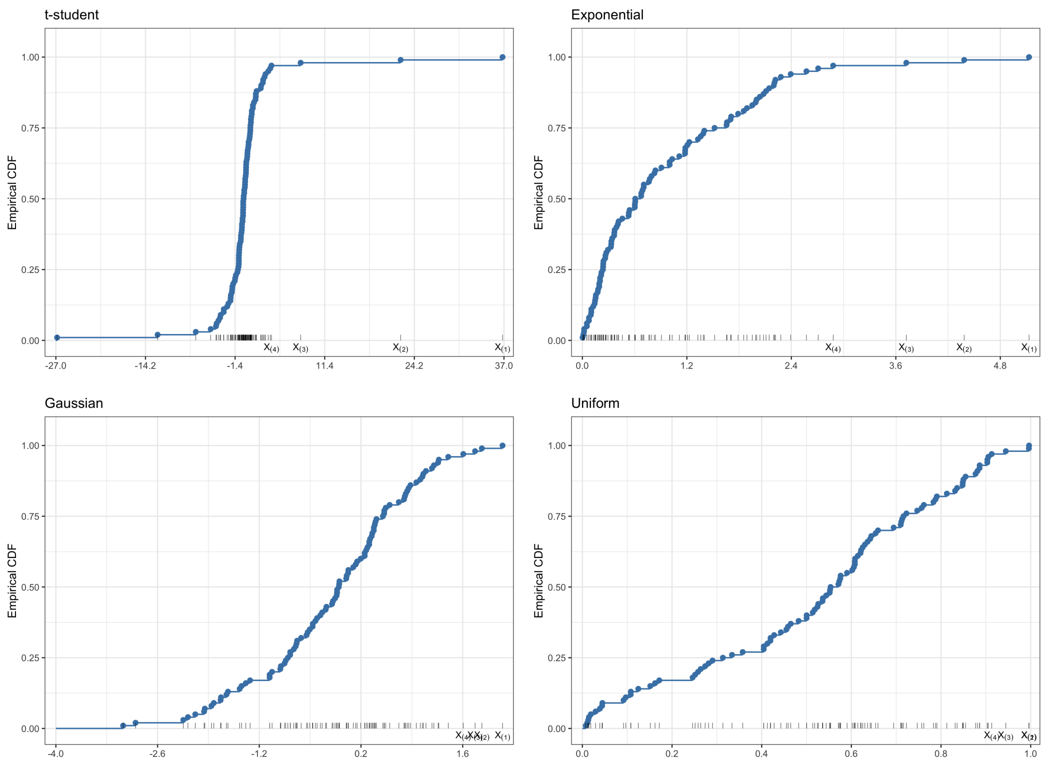

FIGURE 1. Empirical CDFs for t-Student, Exponential, Gaussian, and Uniform distributions.

Note that $\hat{F}_n(X_{(1)}) = 1$, $\hat{F}_n(X_{(2)}) = 1 - \frac{1}{n}$, and in general, $\hat{F}_n(X_{(k+1)}) = 1 - \frac{k}{n}$ for $k = 1, \ldots, n-1$. This justifies using order statistics as estimators of the quantile function. An estimate $\hat{q}_{1-k/(n+1)}$ of the $(1 - \frac{k}{n+1})$-th quantile is given by:

$$\hat{q}_{1-k/(n+1)} = F_n^{\leftarrow}\left(1 - \frac{k}{n+1}\right) = \inf\{x \in \mathbb{R} : \hat{F}_n(x) > 1 - \frac{k}{n+1}\}.$$

This relation holds if:

$$\frac{1}{n}\sum_{t=1}^{n} \mathbf{1}(X_t > \hat{q}_{1-k/(n+1)}) < \frac{k}{n+1} = \frac{k}{n}\left(1 - \frac{1}{n+1}\right),$$

which implies $\hat{q}_{1-k/(n+1)} = X_{(k)}$. These are the empirical quantiles of the sample. To assess the quality of this approximation, we use a standard tool called the QQ-plot.

Definition 2.4 (QQ-plot)**.** *The QQ-plot is the scatterplot of points:*

$$\left(F^{-1}\left(1 - \frac{k}{n+1}\right), X_{(k)}\right),$$

where F is the true distribution function of the sample.

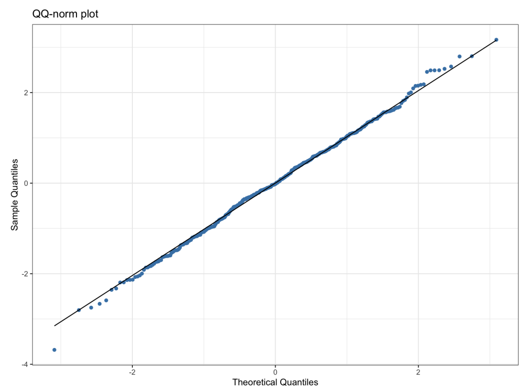

FIGURE 2. QQ-plot for a sample from a normal distribution.

The following theorem establishes the asymptotic normality of order statistics as empirical quantiles:

Theorem 2.5. *Let $X_1, X_2, \ldots, X_n$ be iid with distribution F. For $0 < p < 1$, let $q_p = F^{-1}(p)$. Assume F is differentiable at q_p with density f, where $f(q_p) > 0$, and let $k = k(n) = \lceil (1-p)n \rceil$. Then,*

$$\sqrt{n}(X_{(k)} - q_p) \xrightarrow{d} \mathcal{N}\left(0, \frac{p(1-p)}{f(q_p)^2}\right), \quad \textit{as } n \to \infty.$$

2.2. **Asymptotics for Maxima.**

2.2.1. *Extreme Order Statistics.* We first examine the asymptotic properties of $X_{(1)} = \max_{1 \leq i \leq n} X_i = M_n$ (where $k_n = 1$ satisfies $k_n/n \to 0$ as $n \to \infty$). For every $x \in \mathbb{R}$, we have:

$$\mathbb{P}(M_n \leq x) = \mathbb{P}\left(\max_{1 \leq i \leq n} X_i \leq x\right) = F(x)^n.$$

In practice, this is not directly useful because F is unknown, and estimating F does not resolve the challenge of determining the extreme quantile $q_{1-1/n}$ satisfying $F(q_{1-1/n}) = 1 - 1/n$. Instead, we focus on the tail of Fspecifically, on $\overline{F} = 1 - F$ when its values are close to 0–or directly on the distribution of M_n.

Note that:

$$F^n(x) \xrightarrow[n \to \infty]{} \begin{cases} 0 & \text{if } x < x_F, \\ 1 & \text{if } x \geq x_F, \end{cases}$$

where $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$ is the right endpoint of F. Thus, M_n converges in distribution to $\delta_{\{x_F\}}$ (a degenerate distribution).

2.2.2. *The Extreme Value Theorem.* Assume there exist sequences (a_n) and (b_n) such that:

$$a_n^{-1}(M_n - b_n) \xrightarrow{d} MS, \quad \text{as } n \to \infty.$$

Then,

$$\mathbb{P}(a_n^{-1}(M_n - b_n) \leq x) = F(a_n x + b_n)^n \to \mathbb{P}(MS \leq x).$$

For simplicity, consider the case where $b_n = 0$ for all $n \geq 1$. Then,

$$\mathbb{P}(a_{2n} X_{(1)} \leq x) = F(x/a_{2n})^{2n} \to \mathbb{P}(MS \leq x),$$

but also, if $a_n/a_{2n} \to c$ as $n \to \infty$,

$$\mathbb{P}(a_{2n}^{-1} M_n \leq x) = F(a_n(a_{2n}/a_n)x)^{2n} \sim (F(a_n(cx))^n)^2 \to \mathbb{P}(MS \leq cx)^2.$$

This leads to a stability equation for $g(x) = \log(\mathbb{P}(MS \leq cx))$, satisfying:

$$\text{There exists } c > 0 \text{ such that } g(x) = 2g(cx), \quad x > 0.$$

The solution g is a power function. Being non-decreasing and negative, it admits the form $-x^{-1/\gamma}$, $x > 0$, $\gamma > 0$. More generally,

Theorem 2.6 ([4, 5])**.** *If there exist sequences $a_n > 0$ and b_n, and a non-degenerate distribution G, such that:*

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n) \xrightarrow[n\to\infty]{d} G(x),$$

then G belongs to the Generalized Extreme Value (GEV) family:

$$G_{\mu,\sigma,\gamma}(x) = \exp\left(-\left(1 + \gamma\frac{x-\mu}{\sigma}\right)_+^{-1/\gamma}\right), \quad x \in \mathbb{R}.$$

For $\gamma = 0$, the right-hand side extends by continuity to $\exp(-\exp(-(x-\mu)/\sigma))$*.*

The GEV family $(G_{\mu,\sigma,\gamma})_{\mu,\sigma,\gamma}$ is called the Generalized Extreme Value distribution, where:

- $\mu \in \mathbb{R}$ is the location parameter,
- $\sigma > 0$ is the scale parameter,
- $\gamma \in \mathbb{R}$ is the shape parameter, reflecting tail heaviness.

The theorem suggests modeling the maximum of ”long sequences” using the GEV family.

2.2.3. *Max-Stable Laws and Fréchet Distribution.* A random variable Y is max-stable if, for any $n \in \mathbb{N}$, there exist $a_n > 0$ and $b_n \in \mathbb{R}$ such that:

$$M_n = \max\{Y_1, \ldots, Y_n\} \stackrel{d}{=} a_n Y + b_n,$$

where Y_i are iid copies of Y. When $b_n = 0$, the logarithm of the pdf satisfies the stability equation:

$$\text{For every } n \geq 2, \text{ there exists } a_n > 0 \text{ such that } g(a_n x) = n g(x), \quad x > 0.$$

Thus, Y follows a GEV distribution, and $a_n^{-1} M_n - b_n$ converges to Y.

Max-stability extends to $\mathbb{R}^d$ for $d \geq 1$. For $d = 1$, GEV distributions can be reparametrized into three max-stable families based on the sign of γ:

- Fréchet: $\Phi_\alpha(x) = \exp(-x^{-\alpha})$, $\alpha = 1/\gamma > 0$, $x > 0$,

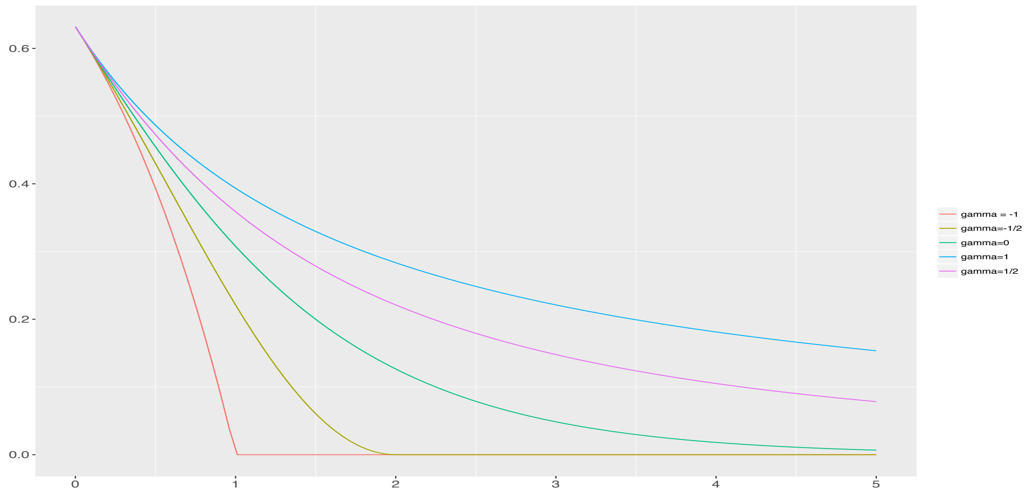

FIGURE 3. Survival functions of the GEV distribution.

- Weibull: $\Psi_\alpha(x) = \exp(-(-x)^\alpha)$, $\alpha = -1/\gamma > 0$, $x < 0$,
- Gumbel: $\Lambda(x) = \exp(-e^{-x})$, $x \in \mathbb{R}$.

In the Fréchet case, Y is max-stable with $\mathbb{P}(Y \leq x) = \exp(-x^{-\alpha})$, and:

$$\mathbb{P}(M_n \leq x) = \exp(-(a_n x + b_n)^{-\alpha}).$$

Choosing $a_n = n^{1/\alpha} = n^\gamma$ and $b_n = 0$ yields:

$$\mathbb{P}(M_n \leq x) = \exp(-x^{-\alpha}),$$

demonstrating max-stability. This case is critical in actuarial science because $x_F = +\infty$, $M_n \to \infty$ a.s., and the maximum M_n is stochastically of order $a_n \to \infty$. The Fréchet distribution is thus useful for modeling maxima in risk settings.

For the Weibull case, $a_n = n^{-1/\alpha} = n^\gamma$ ensures $M_n < 0$ converges to $x_F = 0$. This case is typically excluded in actuarial science as it does not involve risk beyond $x_F < \infty$.

The Gumbel case is intermediate: $x_F = \infty$, $a_n = 1$, and $b_n = \log n$, so the stochastic order of the maximum is not significantly higher than that of an individual observation. While the Gumbel distribution requires caution in quantitative risk management, it is common in environmental sciences for bounded measurements where some risk extrapolation is desired.

2.2.4. *Max-Domain of Attraction.* If F satisfies the assumptions of the extreme value theorem (i.e., there exist $a_n > 0$ and b_n such that $a_n^{-1}(M_n - b_n)$ converges to G_γ), then F is said to belong to the max-domain of attraction of G_γ, denoted $F \in \mathrm{MDA}(G_\gamma)$.

To fully describe the MDA, it is convenient to use quantile functions. We compute:

$$G_\gamma^{\leftarrow}(p) = \frac{(-\log p)^{-\gamma} - 1}{\gamma}.$$

As $n \to \infty$,

$$F(a_n x + b_n)^n \to G_\gamma(x) \iff \frac{F^{\leftarrow}(p^{1/n}) - b_n}{a_n} \to \frac{(-\log p)^{-\gamma} - 1}{\gamma}.$$

Definition 2.7 (Extreme Quantile Function)**.** *For $t > 1$, the extreme quantile function is:*

$$U(t) = \left(\frac{1}{1-F}\right)^{\leftarrow}(t) = F^{\leftarrow}\left(1 - \frac{1}{t}\right).$$

The following result characterizes the MDA in terms of the asymptotic behavior of U:

Theorem 2.8 ([1, 7])**.** *For $\gamma \in \mathbb{R}$, the following are equivalent:*

(1) There exist sequences $a_n > 0$ and b_n such that:

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(x).$$

(2) There exists a positive function a such that for all $x > 0$:

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = D_\gamma(x) = \frac{x^\gamma - 1}{\gamma}.$$

For $\gamma = 0$, the right-hand side equals $\log x$.

Statement (1) holds with $b_n = U(n)$ and $a_n = a(n) = \gamma(U(ne) - U(n))/(e^\gamma - 1)$.

2.2.5. *Regular Variation.* For $\gamma > 0$, we have:

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{U(te) - U(t)} = \frac{x^\gamma - 1}{e^\gamma - 1},$$

or equivalently, $\lim_{t\to\infty} U(tx)/U(t) = x^\gamma$ for all $x > 0$. Thus, U generalizes a power function.

Definition 2.9. *A function ℓ is slowly varying if $\ell(tx)/\ell(t) \to 1$ as $t \to \infty$.*

Definition 2.10. *A function f is regularly varying of index $\gamma > 0$ if $f(x) = x^\gamma \ell(x)$ for some slowly varying ℓ.*

Regularly varying functions of index γ are the only functions satisfying $\lim_{t\to\infty} f(tx)/f(t) = x^\gamma$. If f is invertible, its inverse $f^{\leftarrow}$ is regularly varying of index $1/\gamma$.

From the above, $F \in \mathrm{MDA}(G_\gamma)$ for $\gamma > 0$ if and only if U is regularly varying of index γ, which in turn holds if and only if $U^{\leftarrow}$ is regularly varying of index $1/\gamma$. Since:

$$U^{\leftarrow}(t) = F^{\leftarrow}\left(1 - \frac{1}{t}\right) = \left(\frac{1}{1-F}\right)^{\leftarrow}(t),$$

we conclude:

Proposition 2.11. *A random variable $X \in MDA(G_\gamma)$ for $\gamma > 0$ if and only if $\mathbb{P}(X > x) = x^{-1/\gamma}\ell(x)$ for some slowly varying ℓ.*

For $\gamma > 0$, we find:

$$a_n = a(n) = \gamma \frac{U(ne) - U(n)}{e^\gamma - 1} \sim \gamma U(n).$$

Thus, $a_n \propto b_n$ when $\gamma > 0$, simplifying estimation. This relation no longer holds for $\gamma = 0$.

2.2.6. *Von Mises Condition.* A tractable sufficient condition for $F \in \mathrm{MDA}(G_0)$ is more complex. Under regularity assumptions:

Theorem 2.12 ([8])**.** *If F'' exists, F' is positive, and:*

$$\lim_{t \to x_F} \left(\frac{1-F}{F'}(t) \right)' = \gamma,$$

then $F \in MDA(G_\gamma)$.

For $\gamma = 0$, this provides a sufficient condition based solely on F and its derivatives:

$$\lim_{x \to \infty} \left(\frac{\mathbb{P}(X > x)}{f(x)} \right)' = 0.$$

This condition can be verified for classical distributions such as the exponential or Gaussian.

2.3. **Asymptotics for Exceedances.**

2.3.1. *Exceedance Distribution ($\gamma > 0$).* In EVT, we focus on high values such as maxima. Another approach is to fix a high threshold $u > 0$ and analyze exceedances:

Definition 2.13 (Exceedance)**.** *An exceedance above level $u > 0$ is $X - u$ given $X > u$.*

Exceedances are particularly relevant in excess-of-loss reinsurance contracts. For a sample $X_1, \dots, X_n$ and threshold $u > 0$, the exceedances $X_i - u$ for $X_i > u$ follow the distribution:

$$\mathbb{P}(X - u \le x \mid X > u) = 1 - \frac{\mathbb{P}(X > x + u)}{\mathbb{P}(X > u)}.$$

Thus, $X \in \mathrm{MDA}(G_\gamma)$ for $\gamma > 0$ if and only if $\mathbb{P}(X > x) = x^{-1/\gamma}\ell(x)$, which implies:

$$\mathbb{P}(X - u \le (\gamma u)x \mid X > u) \to 1 - (1 + \gamma x)^{-1/\gamma}, \quad u \to \infty.$$

This means $X \in \mathrm{MDA}(G_\gamma)$ for $\gamma > 0$ if and only if its exceedances asymptotically follow a power-law tail.

Definition 2.14. *A random variable Y is Pareto-distributed with tail index $\alpha > 0$ if $\mathbb{P}(Y > x) = x^{-\alpha}$ for $x > 1$.*

A Pareto random variable has finite moments of order $k < \alpha$ and infinite moments of order $k \ge \alpha$. We can summarize our findings as:

$$\exists (a_n), (b_n), \mathcal{L}(a_n^{-1}(M_n - b_n)) \to G_\gamma, \gamma > 0 \iff \mathcal{L}(X/u \mid X > u) \to \mathcal{L}(Y), \quad u \to \infty,$$

where Y is Pareto-distributed with tail index $\alpha = 1/\gamma$.

2.3.2. *Balkema–de Haan–Pickands Theorem.* Theorem 2.8 has an equivalent form involving exceedances, which is useful in statistics:

Theorem 2.15 ([1, 7])**.** *For $\gamma \in \mathbb{R}$, the following are equivalent:*

(1) There exist sequences $a_n > 0$ and b_n such that:

$$\lim_{n \to \infty} F^n(a_n x + b_n) = G_\gamma(x).$$

(2) There exists a scaling function $a' > 0$ such that for all $x > 0$:

$$\lim_{u\to\infty} \mathbb{P}(X - u \le a'(u)x \mid X > u) \to H_{\sigma,\gamma}(x) = 1 - \Big(1 + \gamma\frac{x}{\sigma}\Big)_+^{-1/\gamma},$$

with $\sigma > 0$. One can choose $a'(u) = a(1/\mathbb{P}(X > u))$.

For $\gamma = 0$, the limit reads $1 - \exp(-x/\sigma)$.

If F satisfies these assumptions, we say F belongs to the domain of attraction of the excesses of $H_{\sigma,\gamma}$. For $\gamma > 0$, we have $a'(u) \sim \gamma u$, consistent with $a'(u) = a(1/\mathbb{P}(X > u))$ and $a \sim \gamma U$.

In summary, belonging to the domain of attraction of a max-stable distribution is equivalent to two properties:

(1) The standardized sequences of maxima converge to a GEV distribution.
(2) The standardized sequences of excesses converge to a Generalized Pareto Distribution (GPD). We denote this as $F \in \mathrm{MDA}(H_\gamma)$, $\gamma \in \mathbb{R}$.

These equivalent properties underpin statistical methods for extremes.

3. Statistical Methods for Extreme Value Theory

Let $X_1, \dots, X_n$ be an iid sample from F. Assume $F \in \mathrm{MDA}(G_\gamma) = \mathrm{MDA}(H_\gamma)$.

3.1. **Parametric Models and Estimation.**

3.1.1. *Block Maxima Method.* Since $F \in \mathrm{MDA}(G_\gamma)$, there exist sequences (a_n) and (b_n), with $a_n > 0$, such that:

$$a_n^{-1}(\max(X_1, \dots, X_n) - b_n) \xrightarrow{d} G_{0,1,\gamma}(x), \quad n \to \infty.$$

The theorem suggests modeling block maxima using the GEV family. In practice, estimating (a_n) and (b_n) is challenging. For large n, we approximate:

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \le x\right) \approx G_{0,1,\gamma}(x) \iff \mathbb{P}(M_n \le x) \approx G_{b_n,a_n,\gamma}(x).$$

Thus, (a_n) and (b_n) can be interpreted as the first two parameters of the GEV model applied to the maximum.

The block maxima method constructs an iid sequence of maxima to fit a GEV distribution. Given iid random variables $X_1, \dots, X_n$, assume $n = kr$ for simplicity. Partition these into k blocks of equal size r:

$$\underbrace{X_1, \dots, X_r}_{\text{Block 1}} \mid \underbrace{X_{r+1}, \dots, X_{2r}}_{\text{Block 2}} \mid \cdots \mid \underbrace{X_{(k-1)r+1}, \dots, X_{kr}}_{\text{Block } k}.$$

For each block j, compute the maximum:

$$M_{r,j} = \max\{X_{(j-1)r+1}, \dots, X_{jr}\},$$

yielding an iid sequence of block maxima $M_{r,1}, \dots, M_{r,k}$. A GEV distribution is then fitted to these maxima. In practice, blocks often correspond to annual periods, where r is the number of observations per year and $M_{r,j}$ represents annual maxima.

The most common statistical approach is Maximum Likelihood Estimation (MLE), though this is technically a quasi-likelihood since we do not assume $M_{r,1}$ is exactly GEV-distributed (true only if F is max-stable). The approach may thus be biased due to model misspecification.

The MLE requires the GEV density $g_{\mu,\sigma,\gamma}$ on its support $1+\gamma(x-\mu)/\sigma > 0$:

$$g_{\mu,\sigma,\gamma}(x) = \frac{1}{\sigma}\left(1+\gamma\frac{x-\mu}{\sigma}\right)^{-(\gamma+1)/\gamma} \exp\left(-\left(1+\gamma\frac{x-\mu}{\sigma}\right)^{-1/\gamma}\right).$$

Let $\theta = (\mu,\sigma,\gamma) \in \Theta = \mathbb{R}\times(0,\infty)\times\mathbb{R}$. The pseudo-likelihood is:

$$L_k(\theta) = \prod_{i=1}^{k} g_\theta(M_{r,i}).$$

The MLE is defined as:

$$\hat{\theta}_k \in \arg\max_{\Theta} L_k(\theta),$$

provided the optimization is well-posed. Here, $\hat{\mu}_k$ approximates b_r and $\hat{\sigma}_k$ approximates a_r, though they may not converge.

Under model regularity and if $M_r \sim G_\gamma$, the MLE satisfies:

$$\sqrt{k}(\hat{\gamma}_k - \gamma) \xrightarrow{d} \mathcal{N}(0,\sigma_\gamma^2), \quad k\to\infty,$$

where σ_γ^2 is the asymptotic variance. Key remarks: - For $\gamma > -0.5$, the MLE is regular, and asymptotic normality is plausible. - For $-1 \le \gamma \le -0.5$, the MLE exists but may not satisfy asymptotic properties. - For $\gamma < -1$, the optimization may lack a solution, precluding a valid MLE.

In practice, the assumption $M_r \sim G_\gamma$ is replaced by $X \in \mathrm{MDA}(G_\gamma)$. Thus, asymptotic normality:

$$\sqrt{k}(\hat{\gamma}_k - \gamma) \xrightarrow{d} \mathcal{N}(0,\sigma_\gamma^2), \quad k\to\infty,$$

may only hold if $r\to\infty$, ensuring the model becomes well-specified as $n\to\infty$. Both $k\to\infty$ and $r\to\infty$ are thus required. Selecting the block size r involves a bias-variance trade-off: small blocks may render the approximation unreasonable (high bias), while large blocks reduce the number $k=n/r$ of maxima (high variance). Practical constraints (e.g., only annual maxima available) often dictate yearly blocks. Even with finer data (e.g., daily), annual maxima analyses tend to be more robust, as shorter blocks (e.g., seasonal) may violate iid assumptions. For example, daily temperatures exhibit seasonal patterns, making summer maxima systematically larger than winter maxima.

3.1.2. *Peaks Over Threshold (POT) Method.* The Peaks-over-Threshold (POT) method analyzes extreme events in an iid dataset $X_1,\ldots,X_n$ where $X \sim \mathrm{MDA}(H_\gamma)$, $\gamma > 0$. An extreme event is any observation exceeding a threshold u, i.e., $X_j > u$ for $j=1,\ldots,k$. The corresponding exceedances are:

$$Z_u = X_j - u \quad \text{for } X_j > u,$$

forming an iid sequence of random length k. We condition on k, achieved by setting $u = X_{(k+1)}$, the $(k+1)$-th largest order statistic. This choice is sometimes called POT with a random threshold (and deterministic exceedances). Dependence in $Z_{u,1},\ldots,Z_{u,k}$ due to $u = X_{(k+1)}$ is neglected by conditioning on u.

Theorem 2.15 suggests modeling exceedances above a sufficiently high threshold using the Generalized Pareto Distribution (GPD) family. In practice, we estimate functions a and b satisfying:

$$\mathbb{P}(Z_u \leq a'(u)x) \approx H_{1,\gamma}(x) \iff \mathbb{P}(Z_u \leq x) \approx H_{a'(u),\gamma}(x).$$

Let $\theta = (\sigma, \gamma) \in \Theta = (0, \infty) \times \mathbb{R}$. The GPD density on its support $1 + \gamma x/\sigma > 0$ is:

$$h_{\sigma,\gamma}(x) = \frac{1}{\sigma}\left(1 + \gamma\frac{x}{\sigma}\right)^{-(\gamma+1)/\gamma}.$$

The MLE is defined as:

$$\hat{\theta}_k \in \arg\max_{\Theta} L_k(\theta),$$

where L_k is the pseudo-likelihood:

$$L_k(\theta) = \prod_{i=1}^{k} h_\theta(Z_{u,i}).$$

Here, $\hat{\sigma}_k$ approximates $a'(u)$ but may not converge. If $Z_u \sim G_\gamma$ and $\gamma > -0.5$, the model is regular, and:

$$\sqrt{k}(\hat{\gamma}_k - \gamma) \xrightarrow{d} \mathcal{N}(0, \sigma_\gamma^2), \quad k \to \infty.$$

In practice, $Z_u \sim G_\gamma$ only for sufficiently large u; otherwise, the approach is biased due to model misspecification.

Choosing the threshold u involves a bias-variance trade-off similar to block size selection. A threshold set too low may violate asymptotic assumptions, while an excessively high threshold yields too few exceedances, increasing estimation variance. One approach is to select the lowest u for which the GPD approximation remains valid. The Mean Residual Life (MRL) plot assists in threshold selection:

Definition 3.1 (Mean Residual Life (MRL))**.** *The MRL at time u is the expected remaining lifetime given survival up to u:*

$$e(u) = \mathbb{E}[X - u \mid X > u] = \frac{\int_u^\infty \mathbb{P}(X > t)\,dt}{\mathbb{P}(X > u)}.$$

We have the following result:

Proposition 3.2. *If $F \in MDA(H_\gamma)$ with $\gamma < 1$ and $\gamma \neq 0$, then $e(u)$ behaves as $\frac{\gamma}{1-\gamma}u$ for large u.*

The MRL can be estimated non-parametrically by replacing $\mathbb{P}(X > u)$ with $\overline{F}_n(u)$, where F_n is the empirical CDF:

$$\hat{e}(u) = \frac{\sum_{i=1}^n (X_i - u)_+}{\sum_{i=1}^n \mathbf{1}(X_i > u)} = \frac{1}{k}\sum_{i=1}^{k}(X_{(i)} - u),$$

where $X_{(1)} \geq \cdots \geq X_{(n)}$ are the order statistics and k is the number of exceedances. Plotting $\hat{e}(u)$ against u helps identify intervals where the MRL is approximately linear. The threshold u^* is chosen at the start of the linear region to maximize k and minimize variance.

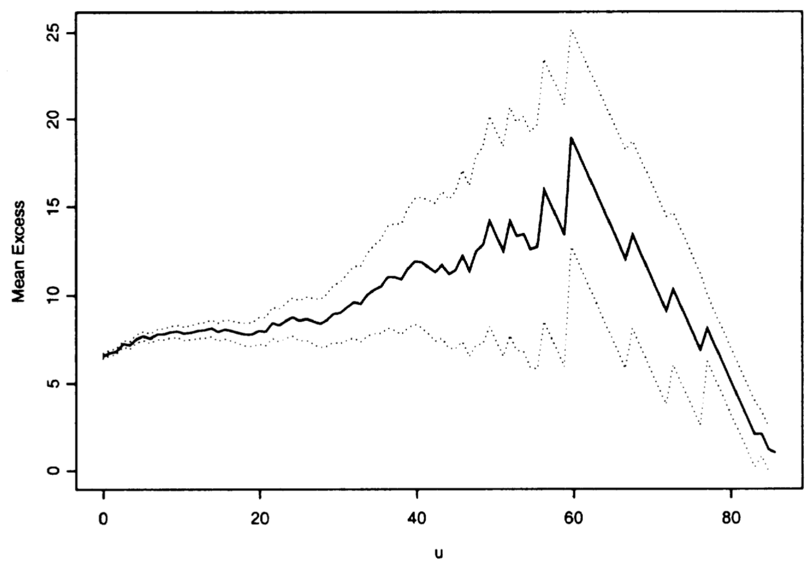

FIGURE 4. Example MRL plot. Threshold $u^* \approx 30$ is chosen at the beginning of the linear increase. Source: [2].

The MRL is related to the Expected Shortfall (ES) or Conditional Value-at-Risk (CVaR), defined as:

$$\mathrm{ES}_p(X) = \mathbb{E}[X \mid X > q_p],$$

for a quantile q_p of level p. The relation "MRL + VaR = CVaR" holds.

3.1.3. *The Hill Estimator for Positive Tail Index ($\gamma > 0$).* The asymptotic normality of block maxima (BM) and POT estimators of γ is complex because $\hat{\gamma}_k$ lacks a closed form. For an iid sample (X_i) with $F \in \mathrm{MDA}(G_\gamma)$ and $\gamma > 0$, we use:

$$\lim_{u\to\infty} \mathbb{P}(X/u \leq y \mid X > u) \to H_{1,\gamma}(y) = 1 - y^{-1/\gamma}, \quad y > 0.$$

The normalized exceedances $Z'_{u,i} = X_i/u$ given $X_i > u$ are approximately Pareto-distributed. The MLE for γ is:

$$\hat{\gamma}_k = \arg\max_{\gamma>0} L'_k(\gamma) = \frac{1}{k}\sum_{i=1}^{k} \ln Z'_{u,i}.$$

The Hill estimator formalizes this heuristic by choosing a random threshold $u = X_{(k+1)}$ to fix the number of exceedances to k:

Definition 3.3 (Hill Estimator)**.** *The Hill estimator is defined as:*

$$\widehat{\gamma}_H(k) = \frac{1}{k}\sum_{i=1}^{k} \ln \frac{X_{(i)}}{X_{(k+1)}}.$$

The following asymptotic normality result holds:

Theorem 3.4 (Asymptotic Normality)**.** *For $k_n \to \infty$ sufficiently slowly as $n \to \infty$:*

$$\sqrt{k_n}\,(\widehat{\gamma}_H(k_n) - \gamma) \xrightarrow{d} \mathcal{N}(0, \gamma^2).$$

Proof. We consider Renyi's representation of extremes order statistics: $X_i = F^{\leftarrow}(U_{(i)})$ where $U_{(1)} > \cdots > U_{(n)}$ is the ordered sample of (U_i) uniformly distributed. Then

$$U_{(i)} = 1 - \frac{E_1 + \cdots + E_i}{E_1 + \cdots + E_{n+1}}\,, \qquad 1 \leqslant i \leqslant n\,,$$

where (E_i) is an iid $\sim Exp(1)$. Recall the function $U(t) = F^{\leftarrow}(1 - 1/t)$ that is regularly varying with index $\gamma > 0$: $U(t) = t^{\gamma}\ell(t)$ for a slowly varying function ℓ. We rewrite the Hill estimator as

$$\frac{1}{k}\sum_{i=1}^{k} \ln \frac{U(\frac{E_1+\cdots+E_{n+1}}{E_1+\cdots+E_i})}{U(\frac{E_1+\cdots+E_{n+1}}{E_1+\cdots+E_{k+1}})} = \frac{\gamma}{k}\sum_{i=1}^{k} \ln \frac{E_1 + \cdots + E_{k+1}}{E_1 + \cdots + E_i} + \frac{1}{k}\sum_{i=1}^{k} \ln \frac{\ell(\frac{E_1+\cdots+E_{n+1}}{E_1+\cdots+E_i})}{\ell(\frac{E_1+\cdots+E_{n+1}}{E_1+\cdots+E_{k+1}})}\,.$$

Applying again Reny's representation on the first summand, we obtain

$$\frac{\gamma}{k}\sum_{i=1}^{k} \ln \frac{E_1 + \cdots + E_{k+1}}{E_1 + \cdots + E_i} = -\frac{\gamma}{k}\sum_{i=1}^{k} \ln(1 - U_{(i)}) = \frac{\gamma}{k}\sum_{i=1}^{k} - \ln U_i = \frac{\gamma}{k}\sum_{i=1}^{k} E_i\,.$$

We prove that the second summand is negligible for k fixed. Denoting $t_{n,k} = \frac{E_1+\cdots+E_{n+1}}{E_1+\cdots+E_{k+1}}$ and using again Renyi's representation we have

$$\frac{1}{k}\sum_{i=1}^{k} \ln \frac{\ell(\frac{E_1+\cdots+E_{n+1}}{E_1+\cdots+E_i})}{\ell(\frac{E_1+\cdots+E_{n+1}}{E_1+\cdots+E_{k+1}})} = \frac{1}{k}\sum_{i=1}^{k} \ln \frac{\ell(t_{n,k}/U_i)}{\ell(t_{n,k})}\,.$$

By the SLLN we have $t_{n,k} \sim n/(E_1 + \cdots + E_{k+1}) \to \infty$ a.s. Therefore $\ell(t_{n,k}/U_i)/\ell(t_n) \to 1$ as $n \to \infty$ by the slow variation property of ℓ, the sum becomes negligible and we obtain

$$\widehat{\gamma}_H(k) \sim \frac{\gamma}{k}\sum_{i=1}^{k} E_i\,, \qquad n \to \infty\,,\ k \text{ fixed.}$$

When (k_n) tends to infinity sufficiently slowly, we can still apply the CLT to the iid sequence E_i with expectation 1 and variance 1:

$$\sqrt{k}\Big(\frac{\gamma}{k}\sum_{i=1}^{k} E_i - \gamma\Big) \xrightarrow{d} \mathcal{N}\left(0, \gamma^2\right)\,, \qquad k \to \infty\,.$$

The desired result follows. □

Lemma 3.5 (Renyi's representation)**.** *Let $U_{(1)} > U_{(2)} > \cdots > U_{(k)}$ be the order statistics of iid $U_i \sim \mathcal{U}(0,1)$. Let $E_i \sim Exp(1)$ be iid, and define the partial sums:*

$$\Gamma_j = \sum_{i=1}^{j} E_i, \quad j = 1, \ldots, k+1.$$

Then, the order statistics of the uniform sample admit the following representation:

$$(U_{(1)}, U_{(2)}, \dots, U_{(k)}) \stackrel{d}{=} \left(\frac{\Gamma_1}{\Gamma_{k+1}}, \frac{\Gamma_2}{\Gamma_{k+1}}, \dots, \frac{\Gamma_k}{\Gamma_{k+1}} \right).$$

Proof. The joint density of $(\Gamma_1, \Gamma_2, \dots, \Gamma_{k+1})$ is:

$$f_{\Gamma_1,\dots,\Gamma_{k+1}}(t_1, \dots, t_{k+1}) = e^{-t_{k+1}} \mathbf{1}(0 \le t_1 \le t_2 \le \dots \le t_{k+1}).$$

This follows from the fact that the increments $E_i = \Gamma_i - \Gamma_{i-1}$ (with $\Gamma_0 = 0$) are iid Exp(1), and their joint density is the product of exponential densities.Define the ratios and the total time as:

$$V_i = \frac{\Gamma_i}{\Gamma_{k+1}} \quad \text{for } i = 1, \dots, k, \quad \text{and} \quad W = \Gamma_{k+1}.$$

The transformation $(t_1, \dots, t_{k+1}) \mapsto (v_1, \dots, v_k, w)$ has Jacobian determinant:

$$J = w^k,$$

since $t_i = v_i w$ for $i = 1, \dots, k$ and $t_{k+1} = w$. The joint density of $(V_1, \dots, V_k, W)$ is therefore:

$$f_{V_1,\dots,V_k,W}(v_1, \dots, v_k, w) = e^{-w} w^k \mathbf{1}(0 \le v_1 \le \dots \le v_k \le 1).$$

Integrate the joint density over $w > 0$ to obtain the marginal density of $(V_1, \dots, V_k)$:

$$f_{V_1,\dots,V_k}(v_1, \dots, v_k) = \int_0^\infty e^{-w} w^k \, dw \cdot \mathbf{1}(0 \le v_1 \le \dots \le v_k \le 1).$$

The integral evaluates to the Gamma function at $k + 1$:

$$\int_0^\infty e^{-w} w^k \, dw = \Gamma(k + 1) = k!.$$

Thus, the density of the ratios is:

$$f_{V_1,\dots,V_k}(v_1, \dots, v_k) = k! \cdot \mathbf{1}(0 \le v_1 \le \dots \le v_k \le 1).$$

This is precisely the joint density of the order statistics $(U_{(1)}, \dots, U_{(k)})$ of k iid uniform random variables on $[0, 1]$. □

The estimator's performance depends critically on k, as illustrated by the "Horror Hill" plot (Figure 5), where the trade-off between variance (large k) and bias (small k) complicates practical application.

3.2. **Quantile Approximation and Extrapolation.** Extreme quantile estimation is fundamental in quantitative actuarial risk analysis, coinciding with VaR estimation required by regulators.

3.2.1. *Extrapolation.* For a sequence of quantiles $q(p_n)$, the probability that the sample maximum M_n does not exceed $q(p_n)$ is asymptotically:

$$\mathbb{P}(M_n \le q(p_n)) \underset{n\to\infty}{=} \exp(-np_n(1 + o(1))).$$

Two cases arise based on np_n:

(1) If $np_n \to \infty$, $\mathbb{P}(M_n \le q(p_n)) \to 0$. The quantile $q(p_n)$ lies within the sample with high probability, and the empirical quantile $X_{(\lfloor np_n \rfloor)}$ is a natural estimator, asymptotically normal when $p_n = p$, $n \ge 1$.

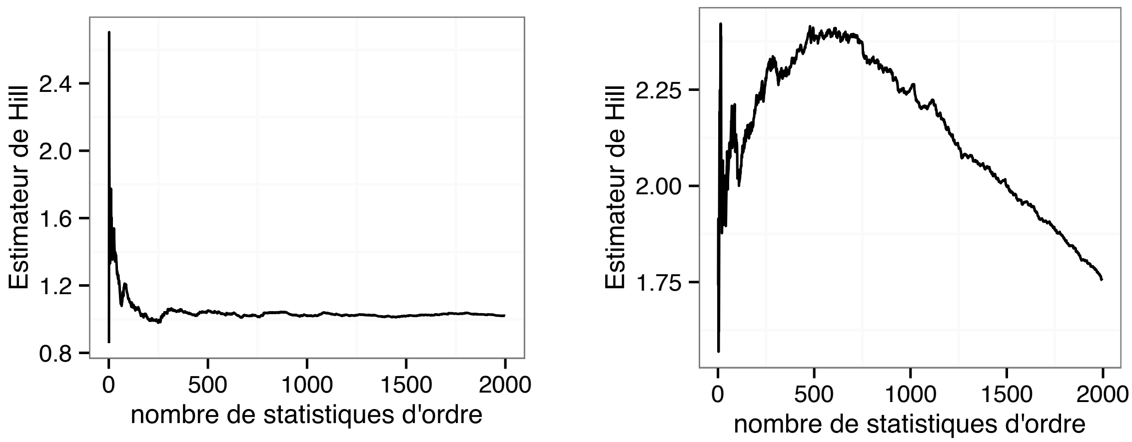

FIGURE 5. Hill plot showing sensitivity to k with the "Horror Hill" phenomenon. The horizontal axis shows the number of upper order statistics k, and the vertical axis displays $\widehat{\gamma}_H(k)$. Source: [2].

(2) If $np_n \to 0$, $\mathbb{P}(M_n \leq q(p_n)) \to 1$. The quantile $q(p_n)$ lies outside the sample with high probability, necessitating out-of-sample extrapolation via EVT methods (e.g., block maxima or POT).

3.2.2. *Quantiles of a GEV Distribution.* The quantiles of a GEV distribution are obtained by inverting its CDF. For $G_{\mu,\sigma,\gamma}$, the quantile q_p corresponding to $1-p$ is:

$$q_p = \begin{cases} \mu - \frac{\sigma}{\gamma}\left(1 - (-\log(1-p))^{-\gamma}\right) & \text{if } \gamma \neq 0, \\ \mu - \sigma \log(-\log(1-p)) & \text{if } \gamma = 0. \end{cases}$$

In practice, $\theta = (\mu, \sigma, \gamma)$ is estimated via the block maxima method with blocks of length $r = n/k$.

The return level plot visualizes the relationship between the return level q_p and return period $T = 1/p$. Defining $y_p = -\log(1-p)$, the return level is:

$$q_p = \begin{cases} \mu - \frac{\sigma}{\gamma}\left(1 - y_p^{-\gamma}\right) & \text{if } \gamma \neq 0, \\ \mu - \sigma \log(y_p) & \text{if } \gamma = 0. \end{cases}$$

Plotting q_p against $-\log y_p \sim \log T$ reveals the tail shape: convex for $\gamma > 0$, linear for $\gamma = 0$, and concave for $\gamma < 0$.

A return level q_p based on block maxima over blocks of length $r = n/k$ corresponds to a return period of $rT = r/p$.

3.2.3. *Quantiles of a GPD.* Extreme quantiles can also be interpreted as GPD quantiles. The GPD CDF is:

$$H_{\sigma,\gamma}(x) = 1 - \left(1 + \gamma \frac{x}{\sigma}\right)^{-1/\gamma}, \quad x > 0.$$

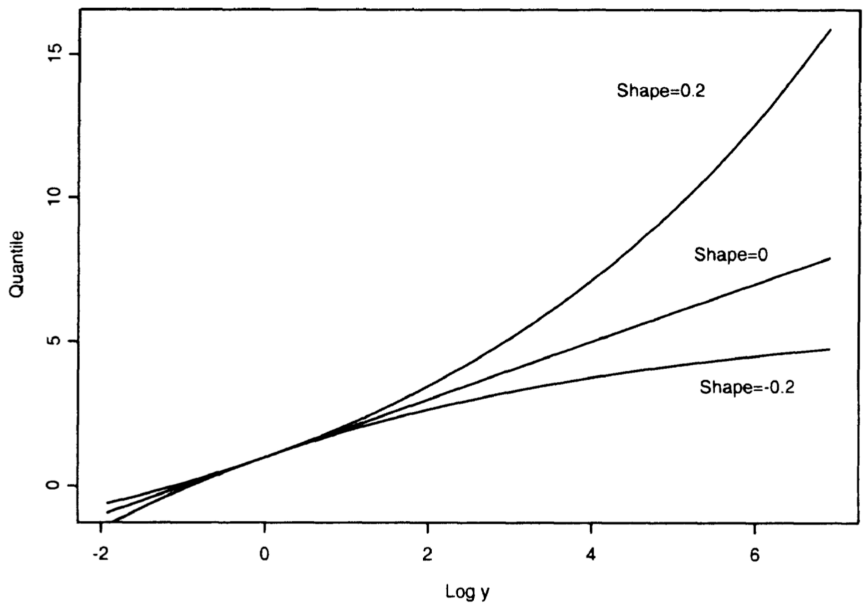

FIGURE 6. Example return level plot. Source: [2].

The quantile x_m for $\mathbb{P}(X > x_m \mid X > u) = 1 - 1/m$ is:

$$x_m = \begin{cases} u + \frac{\sigma}{\gamma}(m^\gamma - 1) & \text{if } \gamma \neq 0, \\ u + \sigma \log(m) & \text{if } \gamma = 0. \end{cases}$$

The parameter $\theta = (\sigma, \gamma)$ is estimated via the POT method for exceedances above u. The value x_m represents the threshold exceeded on average every m exceedances. With k exceedances, $\mathbb{P}(X > u) \approx k/n$ and $\mathbb{P}(X > x_m) \approx k/(mn)$, so x_m estimates the quantile of order $k/(mn)$.

References

[1] A. A. Balkema and L. De Haan. Residual life time at great age. *The Annals of probability*, pages 792–804, 1974.

[2] S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, 2001.

[3] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events: for Insurance and Finance*. Springer, 1997.

[4] R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical proceedings of the Cambridge philosophical society*, volume 24, pages 180–190. Cambridge University Press, 1928.

[5] B. Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, 44(3):423–453, 1943.

[6] A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, 2015.

[7] J. Pickands III. Statistical inference using extreme order statistics. *the Annals of Statistics*, pages 119–131, 1975.

[8] R. Von Mises. *Wahrscheinlichkeit Statistik und Wahrheit: Einführung in die neue Wahrscheinlichkeitslehre und ihre Anwendung*. Springer, 1936.