Statistics of Extremes and Applications: An Introduction

(2014, October, Grahamstown, South Africa)

M. Ivette Gomes

DEIO e CEAUL, Universidade de Lisboa

Research partially supported by National Funds through FCT — Fundação para a Ciência e a Tecnologia, projects projectos PEst-OE/MAT/UI0006/2014.
Introduction

• In the field of *extreme value theory* (EVT), the ordering of the available sample is of primordial importance.
• Indeed, more generally and in almost all areas of *Statistics*, the ordering of a univariate sample, as the basis for a clear representation of the sample’s content, has been considered highly important for a long time.
• Such an importance enabled us to have access to a broad statistical methodology and associated distributional theory related to ordered samples, as can be seen in the most diverse books on *order statistics* (OS’s) and on extreme OS’s, among which we refer Arnold, B., Balakhrishna, N. & Nagaraja, H.N. (1992; 2008), *A First Course in Order Statistics*, 1st Ed., Wiley; 2nd Ed., SIAM.
There exists, on one side a natural interest by ordering:

- The extreme values are important as an expression of the worst or the best that one can find in a sample (minimal temperatures, high levels of dams, minimal life times in reliability theory).

Alternatively, a set of observations can be deliberately ordered, to facilitate the statistical analysis we want to perform. As an example, we mention:

- Best linear unbiased estimators (BLUE) which are linear combinations of OS’s.
- Quick methods for parameters’ estimation or for significance tests based on systematic statistics, like the range, the inter-quartile and the semi-range.
Statistical issues arising in modeling extremes of a random sample have been successfully used in the most diverse fields.

- **Statistics of univariate extremes (SUE)**, as well as multivariate and spatial extremes, have recently faced a huge development, partially due to the fact that rare events can have catastrophic consequences for human activities, through their impact on natural and constructed environments.

- In the last decades there has been a shift from the area of parametric SUE, based on probabilistic asymptotic results in EVT, towards semi-parametric or even non-parametric approaches.

- This course is essentially based on Beirlant, Caeiro & Gomes (2012) and Gomes & Guillou (2014), two review papers published respectively at REVSTAT and ISR, as well as in a recent book (in Portuguese), by Gomes, Fraga Alves & Neves (2013), SPE & INE eds.
Main topics to be dealt with in the course:

1. *Motivation* for the need of *EVT*. Why *EVT*?
2. A brief reference to *Exact Distributional Theory* of *OS’s*.
3. *Graphical techniques* for a preliminary *extreme value analysis* (*EVA*), such as *QQ*-plots, *PP*-plots, *W*-plots and *ME*-plots.
4. A brief reference to *Asymptotic Distributional Theory* of central, extreme and intermediate *OS’s*, providing some details related to the *non-degenerate limiting behaviour* of
   - the sequence of *maximum* values,
   - other *top OS’s* and
   - *excesses* over high thresholds.
5. *Parametric SUE*: most common parametric models in *SUE*.
6. **Semi-parametric SUE**.

- **Estimation procedures** of the *extreme value index* (EVI) are briefly discussed.
- **Estimation** of the *extremal index* (EI) is also briefly undertaken.
- I also further would like to mention the usefulness of resampling methodologies in the area, particularly the importance of
  - the *Generalized Jackknife* (GJ) and
  - the *Bootstrap*

in the obtention of a reliable semi-parametric estimate of any parameter of *extreme* or even *rare events*, like a *high quantile*, the *expected shortfall*, the *return period* of a high level, the *tail dependence coefficient* or the two primary parameters of extreme events, the *EVI* and the *EI*. 
1. Motivation for the need of EVT

To motivate the interest for this area some examples of high relevance to Society and involving this theory are next provided.

Katrina: An unnatural disaster?

- New Orleans is below the sea level, in between two lakes, at North and East, and the Mississippi river at South.
- According with the authorities’ local information, the flooding due to Katrina hurricane was essentially due to a break of 60 meters in a dam close to Pontchartrain lake.
New Orleans, August 29, 2005, after the hurricane Katrina
We next state, in a free way, part of the chronique published at *New York Times, Sept’05*, entitled ‘*New Orleans After Hurricane Katrina: An Unnatural Disaster?’

- ‘...we need to build an adequate system of dams. For that we need the help of danish engineers, who are able to design such structures ...’
- such a plan will cost billions of dollars, but it is sensible to have learnt the lecture, so that we do **NOT have a repetition within the next 20 years.**

It is obvious that not only this disaster, but also a lot of disasters that have recently happened, as well as **historical floods**, like the one that happened in the **North Sea** in 1953, can serve as a guide.
Indeed, at the early morning of **February 1, 1953**, the level of the water has exceeded 5.6 metres above the sea level, the maritime defenses were destroyed, provoking
- the flooding of areas in **Holland, England, Belgium, Denmark** and **France**,
- with the death of around 2500 people.

As a consequence, the Dutch government formed the so-called ‘**Delta Committee**’.

And it was decided that the dams should be built in such a way that
- **the probability of a flood in a certain year should be 1 out of 10,000**.
The flood at North Sea, February 1, 1953

But data have been observed during a much shorter period!... It is thus necessary to extrapolate beyond observed data!! ...

- And EVT can provide reliable answers on the height of such dam, taking into account the so-called return period of an extreme event.
- Such a return period is merely the mean time interval between consecutive occurrences of a certain extreme event, like Katrina’s hurricane or the flood in the North Sea.
Another example: Extremes in the financial market

- The Basilea Committee on banks control formulates norms and directions for supervision of banks and recommends good practices to all financial institutions.
- Among other risk measures, those norms consider the estimation of the so-called *Value-at-Risk* (VaR). Such a risk measure is merely an extreme quantile of the loss-and-gain distribution.
- How can we estimate the VaR on the basis of a series of daily log-returns $R_t$ (in percentage), defined by

$$R_t := 100 \log \left( \frac{P_t}{P_{t-1}} \right),$$

where $P_t$ is ending price at day $t$? For the PSI20, we present in the following Figure the values of $P_t$ (**left**) and $R_t$ (**right**).
Closing prices (left) and daily log-returns and VaR_{0.01} (right) of PSI20

- The historical data are very poor!
The main points to be taken into account are essentially the following ones:

- There are usually only a few observations in the tail of the distribution.
- Estimates much beyond the observed maximum are often required.
- We thus need to consider models for the tail, usually based on asymptotic results.
- Is it sensible to use those models in all real situations dealing with extreme events?
- We cannot forget, and I am now citing George Box (Box, G.E.P. & Draper, N.R. (1987), Empirical Model-Building and Response Surfaces. Wiley, p. 424), ‘...all models are wrong but some models are useful’.
It is worth mentioning that the areas of application of EVT in the analysis of rare events are as diverse as

- Biology, Environment, Finance, Insurance,
- Quality Control, Social Sciences, Sismology, Sports,
- Structural Engineering and Telecommunications, among others.

**Why EVT? Because the world is not always normal!**

- Most of the real life questions require an estimation related to events for which we have only a few data or even no data — the so-called *extreme or rare events*.
- EVT is just a probabilistic branch supporting *Statistics* that deals with those situations, helping to describe and quantify the so-called *rare events*, *extrapolating beyond the sample*.
In classical data analysis, extremes are sometimes wrongly identified as outliers, being sometimes even ignored in the subsequent statistical study, given the fact that they are a long way from the ‘fitted’ model.

If our objective is to infer on day-to-day events, it can be irrelevant to suppress those tail data. But if our fulcral question is related to those tail events, which do not occur often, it seems sensible to apply the EVT framework, giving particular relevance to those extreme values.

Is there an hidden pattern underlying this type of extreme events?
The Normal versus extreme value (EV) models

• If we measure the heights of several people from a same homogeneous population, and represent them in a simple histogram, we easily discover the same rule, the famous Gauss curve, coming to the conclusion that the Normal model ‘rules’ such heights.

• Surprisingly (or possibly not . . .) most real life data sets follow the Normal distribution and related families of models.

• The reason for this is the central limit theorem (CLT) (for sums of random variables). Indeed, most phenomena are the result of a large number of independent causes that act additively, all with finite variance. Hence the reason for the Normal model, supported by the easiness of related inferential techniques, like the $t$-test for the comparison of mean values.
• However, when we are interested in extreme events, located in the tails of underlying distributions, the use of the Normal model is no longer an irrefutable truth.
• Contrarily to the aforementioned normality condition, it is not difficult to find situations where a unique real observation, staying a long way from the central tendency of data, can have an effect similar to the sum of all other non-dominant observations.
• This is possibly the reason why extreme OS’s have been so strongly used in practice, justifying the constant development of EVT.
• Just as an example, in a financial framework, and in particular regarding possible models for returns, it is usual to find tails heavier than the ones often considered under classical approaches.
• Basically, this means that extreme events, despite of improbable, are more frequent than usual under a gaussian distribution, with light tails, of an exponential type.
Some of the key tools that have led to the way statistical EVT has been exploding in these last decades are:

- the key result obtained by Fréchet (1927), [Ann. Société Polonaise de Math.], on the functional equation of stability for maxima,
- later solved with some restrictions by Fisher & Tippett (1928), [Proceedings Cambridge Philosophical Society], who derived the possible limiting laws of the sample maxima, $X_{n:n}$, of a random sample $(X_1, \ldots, X_n)$, of \textit{independent and identically distributed (IID) random variables (RV’s)},
• **SUE** is essentially based on *asymptotic results* related to the non-degenerate limiting behaviour of
  – the sequence of maximum values,
  – other top OS’s, either individually or jointly, and
  – excesses over high thresholds.
Consequently **SUE** deals essentially with the so-called
  – EV and the *general extreme value (GEV)* models,

\[
G_\xi(x) \equiv \text{GEV}_\xi(x) = \begin{cases} 
  e^{-(1+\xi x)^{-1/\xi}}, & 1 + \xi x \geq 0, \text{ if } \xi \neq 0, \\
  e^{-e^{-x}}, & x \in \mathbb{R}, \text{ if } \xi = 0,
\end{cases}
\]

with \( \xi \) the so-called EVI, the primary parameter in **SUE**.
  – extremal processes (**EP**) and
  – generalized Pareto (**GP=1+ln GEV**) models.
• Indeed, with $GP_\xi(\cdot) = 1 + \ln \text{GEV}_\xi(\cdot)$, the GP cumulative distribution function (CDF), and $F_u$ the distribution of the excesses over a high threshold $u$, conditional to the fact that $X > u$, i.e. the CDF of the RV $Y \mid Y > 0$, with $Y = X - u$,

$$F_u(y) := \mathbb{P}(X - u > y \mid X > u) \approx GP_\xi(y/\sigma_u).$$

• For $\xi = 0$ we get a Gumbel model for maxima and an Exponential model for the excesses.

• The Exponential model, with a tail or survival function,

$$1 - F(x) := \exp(-x/\sigma_u), \ x > 0,$$

plays thus a crucial role in EVT, in the so-called peaks over threshold (POT) approach to SUE.
But we cannot at all be STUCK to these limiting models, essentially because:

- The rate of convergence is sometimes very slow and penultimate approximations, related to BUT often different from GEV and GP laws can appear.

- The scheme is often a randomly stopped scheme, and then a much larger variety of limiting models is achievable. Just as an example, if we consider $M_N := \max(X_1, \ldots, X_N)$, where $N$ is a Geometric($\theta$) RV, the possible limiting laws of $M_N$, linearly normalized, are given by

$$H_\xi(x) = \frac{1}{1 - \ln \text{GEV}_\xi(x)}$$

the so-called max(geo)-stable RV’s (Gnedenko & Korolev, 1996, *Random Summation: Limit Theorems and Applications*).
• The statistical applications of EVT have given emphasis
  – to the relaxation of the independence condition,
  – to the consideration of multivariate and spatial frameworks
    and
  – to a deeper and deeper use of regular variation and point
    process approaches.
• These topics are well-documented in well-known books.
• For an overview of most of the topics in this field I recommend
  the reading of the recent volumes of *Extremes* 11:1 (2008), *Extremes*
  13:2 (2010), essentially dedicated to *Statistics of Extremes in
  Weather and Climate*, and *Revstat* 10:1 (2012), among others.
There are indeed situations where the **EVT** approach is fulcral.

- The aforementioned **GEV** model, associated with the largest observations, is often applied to maxima annual temperatures, just as an example.
- On the other side, the distribution associated with smallest observations, \( \text{GEV}^*_{\eta}(x) := 1 - \text{GEV}_{\eta}(-x) \), is often applied to problems related to material resistance.

In **Statistics**, the **Fisher-Tippett-Gnedenko’s theorem**, also called **extremal types theorem (ETT)**, is a result on the asymptotic distribution of extreme **OS’s**.

The **ETT** has thus a role similar to the **CLT** for averages (sums).
• Basically, the ETT originally established that the sample maxima, after suitable linear normalization, converges to one of 3 possible distributions,
  – Gumbel or
  – Fréchet or
  – Max-Weibull,
also called EV models and particular cases of the aforementioned GEV\(\xi\) model, when \(\xi = 0\), \(\xi > 0\) or \(\xi < 0\), respectively.

• Indeed, independently of the shape of the center of the distribution underlying the data, the *tails always assume very special shapes* when we are sufficiently away from the center of the model.
Pioneering statisticians in the field of extremes

Sir Ronald Alymer **Fisher** (1890-1962),
Leonard Henry Caleb **Tippett** (1902-1985),
Ernst Hjalmar Waloddi **Weibull** (1887-1979),
Emil Julius **Gumbel** (1891-1966),
Maurice René **Fréchet** (1878-1973) and
Richard Edler **von Mises** (1883-1953)
We next refer some of the well-known sentences of Emil Gumbel, one of the pioneering names in the area of Statistics of Extremes:

‘It seems that the rivers know the theory. It only remains to convince the engineers of the validity of this analysis.’

‘Il est impossible que l’improbable n’arrive jamais.’

‘Il y aura toujours une valeur qui dépassera toutes les autres.’

There exist nowadays a large variety of ‘R-Packages for Extreme Values’, such as evd, evdbayes, evir, extRemes, extremevalues, fExtremes, ismev, lmom, lmomRFA, lmomco, POT, SpatialExtremes, and some of them will be used in this course.
Why EVT?

Simple problems related to EVA

- Sample data are typically used to study the properties of the CDF,

\[ F(x) := \mathbb{P}(X \leq x), \]

or of its quantile function

\[ Q(p) := \inf\{x : F(x) \geq p\} =: F^{-}(p), \]

where \( F^{-} \) denotes the generalize inverse function of \( F \).

Scarceness of data in the tails. The following Figure illustrates the possible difficulties associated with a reliable tail estimation. This is due to the fact that most of the data are in the center of the distribution.
The observations in the tail are scarce, being often required an estimation beyond the sample maximum and/or minimum.
**Inadequacy of traditional methodologies.** We could try fitting a probabilistic model, like for instance the Normal model, $\mathcal{N}(\mu, \sigma^2)$, to the whole sample, and to use such a model to estimate tail probabilities, i.e. to consider the estimate

$$\hat{P}(X > x) = 1 - \hat{F}(x) = 1 - \Phi\left(\frac{x - \hat{\mu}}{\hat{\sigma}}\right).$$

The associated problems are essentially the following ones:

- Such a fitting is essentially ruled by the central observations.
- Different models that fit the body of the data can lead to quite different tail extrapolations.
- On another side, if our interest lies on the tails, why compromising the tail fitting forcing simultaneously the fitting of the body of the data?
• In EVT, we often want to estimate the exceedance probability,

\[ p \equiv p_x := P[X > x], \quad \text{where } x(> x_{n:n}) \]

is thus a high value.

• And the empirical CDF, defined by

\[
\hat{F}_n(x) = \begin{cases} 
0 & \text{if } x < x_{1:n}, \\
\frac{i}{n} & \text{if } x_{i:n} \leq x < x_{i+1:n}, \ 1 \leq i < n, \\
1 & \text{if } x \geq x_{n:n},
\end{cases}
\]

where \( x_{i:n} \) is the \( i \)-th smallest sample value, \( 1 \leq i \leq n \), does thus not provide any useful information (it just provides the value ZERO).
• Inversely, similar problems arise when we think on the empirical quantile function,

\[ \hat{Q}_n(p) := \inf\{x : \hat{F}_n(x) \geq p\}, \]

and are interested in extreme empirical quantiles,

\[ \hat{Q}_n(p) \quad \text{and} \quad \hat{Q}_n(1 - p), \quad \text{with} \quad p < 1/n. \]

• We then need to have access to techniques specifically directed to extreme values, usually based on EVT results.
• In the books,
  
  
  
  
  
  it is possible to find several case-studies, in a wide variety of areas of application of the *modeling of rare events*, that fully justify the development of *EVT*. 
2. Exact Distributional Theory of OS’s

• Let us use the notation $X$ for a RV with CDF $F$, possibly dependent on unknown location, $\lambda \in \mathbb{R}$, and scale, $\delta \in \mathbb{R}^+$, respectively. In an univariate set-up, the original random sample $(X_1, \ldots, X_n)$ is immediately ascendingly ordered, with the notation $(X_{1:n} \leq \cdots \leq X_{n:n})$, where

\[
X_{1:n} := \min_{1 \leq i \leq n} X_i \quad \text{and} \quad X_{n:n} := \max_{1 \leq i \leq n} X_i.
\]

• The notation $Z$ is often used for the reduced RV, $Z = (X - \lambda)/\delta$. Without loss of generality, due to the fact that $X_{i:n} = \lambda + \delta Z_{i:n}$, $1 \leq i \leq n$, we work with $(Z_1, \ldots, Z_n)$, from a standard CDF $F$, and the ascending OS’s $(Z_{1:n} \leq \cdots \leq Z_{n:n})$. 


Theorem 1 (Exact CDF of $Z_{i:n}$). Let $Z_1, \ldots, Z_n$ be IID to $Z$ with CDF $F(\cdot)$. Then, for $1 \leq i \leq n$,

$$F_{i:n}(z) := P[Z_{i:n} \leq z] = P[\text{at least } i \text{ out of } n \text{ RV's } Z_k \leq z] = \sum_{k=i}^{n} \binom{n}{k} F^k(z)[1 - F(z)]^{n-k}.$$ 

If $Z$ is absolutely continuous with PDF $f(z) = F'(z)$, $Z_{i:n}$, $1 \leq i \leq n$, has a PDF

$$f_{i:n}(z) = \frac{1}{B(i, n - i + 1)} F^{i-1}(z)[1 - F(z)]^{n-i} f(z), \quad z \in \mathbb{R},$$

where $B(p, q)$ denotes the complete Beta function.
Relationship to Binomial and Beta models. By definition, $Z_{i:n} \leq z$ if and only if there exist at least $i$ observations, out of $n$, $\leq z$. Consequently, just as seen in Theorem 1, the CDF of $Z_{i:n}$ is, for $F$ either discrete or continuous,

$$F_{i:n}(z) = \sum_{k=i}^{n} \binom{n}{k} F^k(z)(1 - F(z))^{n-k}, \quad 1 \leq i \leq n.$$  

To better understand this Equation, let us think on the relationship between the CDF of an OS and the Binomial model. Consider the following counting RV's:

$$S_z := \sum_{k=1}^{n} I[Z_k > z], \quad S^*_z := \sum_{k=1}^{n} I[Z_k \leq z],$$

where $I_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$ is the indicator function of $A$.  

Obviously, \( S_z + S^*_z = n \), the RV \( S_z \) is Binomial(\( n, 1 - F(z) \)) and the RV \( S^*_z \) is also Binomial(\( n, F(z) \)). Since

\[
Z_{i:n} \leq z \iff S_z < n - i + 1 \iff S^*_z \geq i
\]

we get,

\[
F_{i:n}(z) = \mathbb{P}[\text{Binomial}(n, 1 - F(z)) < n - i + 1]
= \mathbb{P}[\text{Binomial}(n, F(z)) \geq i].
\]

From this second Equation follows the aforementioned CDF of an OS. The first Equation would lead us to

\[
F_{i:n}(z) = \sum_{k=0}^{n-i} \binom{n}{k} (1 - F(z))^k F^{n-k}(z),
\]

which could have been obtained from the first through a change in the summation index, summing in \( k_1 = n - k \).
Further note that we can write the sum as an integral,

\[
F_{i:n}(z) = \frac{1}{B(i, n-i+1)} \int_{0}^{F(z)} t^{i-1}(1-t)^{n-i} dt,
\]

the incomplete Beta function with parameters \(i\) and \(n-i+1\), computed at \(F(z)\). Such a function is usually denoted by \(I_{F(z)}(i, n-i+1)\), and its values can be found in Table 6 of Pearson, E.S. & Hartley, H.O. (1970), *Biometrika Tables for Statisticians*, Cambridge Univ. Press.

Denoting by \(B_{p,q}\) a Beta RV with parameters \(p\) and \(q\), we thus get

\[
F_{i:n}(z) = P\left[ B_{i,n-i+1} \leq F(z) \right].
\]

This is the reason why the CDF of an OS (and often the OS itself) is usually called a transformed Beta. Indeed, for an absolutely continuous \(F\), \(F(Z_{i:n}) \overset{d}{=} B_{i,n-i+1}\). 

39
OS’s in an Uniform model

- Let us denote by $U_i$, $1 \leq i \leq n$, a random sample of size $n$ from an Uniform model in (0,1), $\mathcal{U}(0,1)$, with CDF $F(z) = z$, for $0 \leq z \leq 1$.

- Let $U_{1:n} \leq \cdots \leq U_{n:n}$ denote the sample of associated ascending OS's.

- From the relationship between OS's and the transformed Beta, follows immediately that $U_{i:n} \overset{d}{=} B_{i,n-i+1}$, i.e.

\[
f_{U_{i:n}}(z) = \frac{1}{B(i, n - i + 1)} z^{i-1} (1 - z)^{n-i}, \quad 0 \leq z \leq 1.
\]

We next present a Figure that illustrates the behavior of OS's in an Uniform model.
PDF’s (left) and CDF’s (right) of an \( \mathcal{U}(0, 1) \) RV, \( U \), and of \( U_{1:3}, U_{2:3} \) and \( U_{3:3} \)
Remark 1. Note the symmetric behavior of $U_{2:3}$, with similar ‘light tails’, due to the symmetry and the finite left and right endpoints of the $\mathcal{U}(0, 1)$, with $F(x)$ and $1 - F(x)$ approaching 0 (as $x \to 0$) and 1 (as $x \to 1$), respectively, in a similar way. This is the reason for the same type of asymptotic behaviour for the minimum and maximum.

Role of the probability integral transform in the exact distribution of OS's.

- Given the absolutely continuous RV $Z$, from $F$, $U := F(Z)$ is $\mathcal{U}(0, 1)$. And since $U \overset{d}{=} 1 - U$, $1 - F(Z)$ is also $\mathcal{U}(0, 1)$.
- However, whereas $F$ is non-decreasing, $1 - F$ is non-increasing, i.e. the first transformation preserves the order and the second one reverts it. Formally, we get

$$
F(Z_{i:n}) \overset{d}{=} U_{i:n}, \quad 1 - F(Z_{i:n}) \overset{d}{=} U_{n-i+1:n} \quad 1 \leq i \leq n.
$$
• More generally, for $F$ either discrete or continuous, we can write

$$Z_{i:n} \overset{d}{=} F^{\leftarrow}(U_{i:n}) \overset{d}{=} F^{\leftarrow}(1 - U_{n-i+1:n}), \quad 1 \leq i \leq n,$$

with $F^{\leftarrow}$ denoting again the generalized inverse function of $F$.

• The aforementioned distributional representations play thus a crucial role in
  – the derivation of properties of OS’s in models different from the Uniform
  – on the basis of the behavior of OS’s in an Uniform model.
3. Graphical Methods for a Preliminary EVA

- It is obvious that the linearity of a graph can be easily detected ‘by eye’ whenever we observe a cloud of points, being then quantified on the basis of the correlation coefficient.
- The idea underlying the initial probability paper (PP)-plots, introduced in Hazen (1914; 1930) for a study of floods, or equivalently the quantile vs quantile (QQ)-plots, available in most statistical packages, come from the need to answer the question:
  - Does a specific probability model provide an adequate fit to the distribution underlying a set of data?
- We can then get a visual fast assessment of a certain probabilistic model suggested by the histogram, to data \((x_1, \ldots, x_n)\), and even a preliminary estimation of unknown parameters.
• The PP-plot is usually used when the observed data, \((x_1, \ldots, x_n)\), can be considered as independent observations of an RV \(X\) with CDF of the type \(F((x - \lambda)/\delta)\), with \(\lambda\) a location and \(\delta\) a scale parameter.

• It is a method of linearization of the CDF: given the observed ordered sample, \((x_{1:n} \leq \cdots \leq x_{n:n})\), and \(F(x) = F((x - \lambda)/\delta)\), the points,

\[
(x_{i:n}, y_i := F^{-1}(p_i)), \quad p_i := i/(n + 1), \quad 1 \leq i \leq n,
\]

are plotted, with \(F^{-1}\) the generalized inverse function of the standard \(F\).

• If the plot shows a linear relationship between \(x_{i:n}\) and \(y_i\) we have an informal validation of \(F(\cdot)\).
• Indeed, if $F^{-1}(\cdot)$ exists, and we write

$$p_i = F((x_{i:n} - \lambda)/\delta), \quad 1 \leq i \leq n,$$

we get

$$y_i = F^{-1}(p_i) = x_{i:n}/\delta - \lambda/\delta \iff x_{i:n} = \lambda + \delta y_i, \quad 1 \leq i \leq n,$$

i.e. there exists a linear relationship between $x_{i:n}$ and $y_i = F^{-1}(p_i)$, provided that $p_i$ is a suitable estimate of $F((X_{i:n} - \lambda)/\delta)$.

• If there exists linearity, the estimation of the parameters can be done through a regression procedure, available in any statistical package.

• The intercept and the slope of the fitted line provide thus preliminary estimates of $\lambda$ and $\delta$. 

• The proposed choice is $p_i = i/(n+1)$, $1 \leq i \leq n$, as given above, the values of $\mathbb{E}(F((X_{i:n} - \lambda)/\delta))$, $\forall$ absolutely continuous CDF, $F(\cdot)$. Indeed, again with $B_{p,q}$ denoting a Beta RV with parameters $p$ and $q$, recall that

\[
F\left(\frac{X_{i:n} - \lambda}{\delta}\right) \overset{d}{=} U_{i:n} \overset{d}{=} B_{i,n-i+1},
\]

with mean value $i/(n+1)$, $1 \leq i \leq n$. 
A simple example is given by the Gumbel model (GEV$_0$), quite common in SUE.

**Example 1.** If $F \equiv \Lambda$, the Gumbel CDF,

$$\Lambda(x; \lambda, \delta) = e^{-e^{-(x-\lambda)/\delta}} =: p \implies x = \lambda + \delta(-\log(-\log(p))).$$

Note that the ‘old Gumbel probability paper’ had an arithmetic scale (to plot the ordered observations, $x_{i:n}$, $1 \leq i \leq n$), versus a double-logarithmic scale (to plot the ‘plotting positions’, $p_i = \frac{i}{n+1}$, for $1 \leq i \leq n$).

From a conceptual point of view, we can obviously plot $x_{i:n}$ versus $y_i = -\log(-\log(i/(n+1)))$, or $y_i = -\log(-\log(i/(n+1)))$ versus $x_{i:n}$, $1 \leq i \leq n$, in a common paper, or to use the Gumbel QQ-plot. This last option is nowadays the common situation, due to computational facilities we have, like the ‘R package’, among others.
• **The generation of Gumbel pseudo random numbers is simple.** We have thus generated Gumbe(0,1) observations \([-\ln(-\ln U)]\), placing them in the vector “gumb”.

• **The following Figure (left) is a Normal QQ-plot, that immediately provides indication of the non-normality of data.** The points 

\[
(y_i = -\log(-\log(i/(n + 1))), x_{i:n}), \quad 1 \leq i \leq n,
\]

provide the following Figure (right), and a clear indication of a possible Gumbel underlying model.

• **The least squares method provides the estimates**

\[
\lambda^{**} = -0.102893, \quad \delta^{**} = 0.993790.
\]
Gumbel data in a Normal PP-plot (left) and a Gumbel PP-plot (right)
**Exponential QQ-plot:** Just as mentioned before, in SUE, the Exponential model, $\mathcal{E}(\delta)$, plays a role more important than the one of the Normal model.

- The **right tail-function** (RTF) is for the $\mathcal{E}(\delta)$,

$$F_\delta(x) := 1 - F_\delta(x) = \exp(-x/\delta), \quad x > 0,$$

and

$$1 - F_1(x) := \exp(-x), \quad x > 0,$$

for standard or reduced **Exponential**, $\mathcal{E}(1)$.

- How can we easily assess that the distribution underlying the observations $x_1, \ldots, x_n$ belongs to the family $\mathcal{E}(\delta)$?
• Thinking directly on the *quantile function*,

\[ Q_\delta(p) = F_\delta^\leftarrow(p) = -\delta \log(1 - p), \quad p \in (0, 1), \]

we see that there exists the linear relationship

\[ Q_\delta(p) = \delta Q_1(p) = \delta(- \log(1 - p)), \quad p \in (0, 1). \]

• Given a sample \((x_1, \ldots, x_n)\), it is thus sensible to replace \(Q(p)\) by its empirical counterpart \(\hat{Q}_n(p)\), and to plot the points

\[ (- \log(1 - p), \hat{Q}_n(p)), \quad \text{for values of } p \in (0, 1). \]

More specifically, if the *Exponencial* provides a good fit, the points

\[ \left(y_i = - \log(i/(n + 1)), x_{i:n} \right), \quad 1 \leq i \leq n, \]

are approximately along a line.
• The slope of the line can be identified with $\delta$ and used to obtain a preliminary estimate of $\delta$.
• Note that the value at the origin should be zero ($Q(0) = 0$).
• The line (slope=$a$; intercept=0) fitted to the points by the least squares method, obtained by the minimization of

$$\sum_{i=1}^{n} \left(x_i:n + a \log(1 - p_i)\right)^2,$$

is

$$\hat{a} = \frac{\sum_{i=1}^{n} x_i:n q_i}{\sum_{i=1}^{n} q_i^2}, \quad \text{with } q_i := -\log(1 - p_i), \quad i = 1, \ldots, n.$$
Example 2. For a set of 100 $\mathcal{E}(\delta)$ observations, with $\delta = 1/2$, generated in $\mathbf{R}$, the Exponential QQ-plot is presented in the following Figure. The slope of the line that crosses the origin is $\hat{\delta} = 0.5091$. 

![Exponential QQ-plot](image)

The equation $x = 0.5091(-\log(1-p))$ is plotted on the graph.
Let’s see an interpretation that looks closer to the PP-plot:

- The function that we want to approximate when plotting the points

\[ x_i: n \leftrightarrow -\log(1 - p_i), \ i = 1, \ldots, n, \]

is \( x \mapsto -\log(1 - F(x)) \).

- This is exactly the transformation that converts any RV \( X \), with a continuous CDF \( F \), in an \( \mathcal{E}(1) \) RV. Indeed,

\[
P[-\log(1 - F(X)) \leq x] = P[X \leq Q(1 - \exp(-x))] \\
= F(Q(1 - \exp(-x))) = 1 - \exp(-x)
\]
i.e., \( -\log(1 - F(X)) \sim \mathcal{E}(1) \).
• Let us next think on the observations above a level \( t \), usually the ones considered in the POT method. Indeed, most of the times data are available only above a threshold \( t \). For instance, a Reinsurer receives only information on the compensations above a certain high level \( t \).

• Let \( X \) be \( \mathcal{E}(\delta) \), and let us condition on the event \( \{ X > t \} \),

\[
\mathbb{P}[X > x | X > t] = \frac{\mathbb{P}[X > x]}{\mathbb{P}[X > t]} = \exp(-\frac{x - t}{\delta}), \text{ for } x > t,
\]

i.e. \( X|X > t \) is still Exponential, with a quantile function

\[
Q(p) = t - \delta \log(1 - p), \quad 0 < p < 1.
\]

Then, the QQ-plot has an intercept equal to \( t \).

• How to estimate an upper extremal quantile

\[
q_p := Q(1 - p), \text{ with } p \text{ small?}
\]
• If it is sensible to assume an Exponential model, then
  \[ \hat{q}_p = t - \hat{\delta} \log(p). \]

• Reciprocally, a small exceedance probability
  \[ p \equiv p_x := \mathbb{P}[X > x | X > t] \]
  can be estimated by
  \[ \hat{p}_x = \exp \left( - \frac{x - t}{\hat{\delta}} \right). \]

**Preliminary estimation of \( \delta \):** We can estimate \( \lambda \) through the QQ-plot and the *least squares* method. Alternatively, we can consider the *maximum likelihood* (ML) estimate, \( \hat{\delta} = \bar{x} - t. \)
A first example: the maximum wind-speed in Zaventem

- We next proceed to the application of a few graphical methods to the data in ‘zaventem.txt’,
- There exist 74 speeds above the threshold $t = 82\, \text{Km/h}$.
- The associated histograms of the original maxima and of the exceedances over the threshold $t = 82$ are presented in the following Figure.
Histograms associated with the daily maxima wind speeds in Zaventem (left) and to the sub-sample of wind speeds above 82 Km/h (right)
• Let’s try the fitting of an exponential model to the sub-sample of size $n = 74$.

• The fitted conditional probability density function (PDF) is

$$f(x) = \exp\left(- \frac{x - t}{\hat{\delta}}\right)/\hat{\delta}, \text{ with } t = 82, \quad \hat{\delta} = \bar{x} - t = 12.063.$$  

• If we use the R-package to fit the line

$$\hat{Q}_n(p) = t - \hat{\delta} \log(1 - p),$$

and on the basis of the QQ-plot presented in the following Figure (maxima daily speeds above 82Km/h versus standard exponential quantile), we get the value 80.86 for the $y$-intercept, a value that differs a bit from the true value $t = 82$. 
QQ-plot for the wind speeds above 82 Km/h
• It is thus sensible to impose the exact value $t = 82$ for the ordinate at the origin of the line

$$
\hat{Q}_n(p) = t - \hat{\delta} \log(1 - p).
$$

• The fitted line (slope=$a$; $y$-intercept=$t=82$), obtained through the least squares method, is obtained through the minimization of

$$
\sum_{i=1}^{n} (x_{i:n} - t + a \log(1 - p_i))^2.
$$

• We then get

$$
\hat{a} = \frac{\sum_{i=1}^{n} x_{i:n} q_i - t \sum_{i=1}^{n} q_i}{\sum_{i=1}^{n} q_i^2}, \text{ with } q_i := -\log(1 - p_i), \ i = 1, \ldots, n.
$$
For the data in ‘zaventem.txt’, we get the following QQ-plot, and

\[ \hat{a} = \hat{\delta} = 12.96368. \]
**General QQ-plot.**

In the general case, let $Q_s$ be the quantile function for the standard model of a specific family. The model can be considered adequate as an underlying model if:

1. There exist a linear relationship between the theoretical quantiles $Q(p)$ and $Q_s(p)$.
2. The theoretical unknown quantiles $Q(p)$ should thus be replaced by the empirical quantiles $\hat{Q}_n(p)$.
3. We should thus plot the points
   \[
   \left\{ \left( Q_s \left( \frac{i}{n+1} \right), \hat{Q}_n \left( \frac{i}{n+1} \right) \right) = (Q_s(p_i), x_{i:n}), \quad i = 1, \ldots, n \right\}.
   \]
4. Finally, the linearity of these points should be checked, possibly through a linear regression applied to the QQ-plot.
• **Quantiles and return periods** can thus be estimated after the acceptance of a linear relationship in the QQ-plot, given by \( y = \hat{b} + \hat{a}x \), with

\[
\bar{q} = \frac{1}{n} \sum_{i=1}^{n} Q_s(p_i), \quad \hat{a} = \frac{\sum_{i=1}^{n} (x_i:n - \bar{x}) Q_s(p_i)}{\sum_{i=1}^{n} (Q_s(p_i) - \bar{q})^2} \quad \text{and} \quad \hat{b} = \bar{x} - \hat{a}\bar{q}.
\]

• Denoting by \( F_s \) the standard CDF and \( Q_s = F_s^{-1} \) the generalized inverse function of \( F_s \), \( q_p = Q(1 - p) \) and \( p_x = \mathbb{P}[X > x] \), and consequently:

  – **Extremal quantiles**: \( \hat{q}_p = \hat{b} + \hat{a}Q_s(1 - p) \).

  – **Exceedance probabilities**: \( \hat{p}_x = \overline{F}_s((x - \hat{b})/\hat{a}), \overline{F}_s := 1 - F_s \).
• More generally than for models with location/scale, the PP-plot can be used when $F(x_{i:n}, \theta)$, with $x_{i:n}$ the $i$-th top OS associated with the sample $(x_1, \ldots, x_n)$ and $\theta$ a vector of 2 unknown parameters, can be transformed in a linear relationship, i.e. there exist functions $g_i(\cdot)$, $i = 1, 2, 3, 4$ such that

\[
g_1[F(x_{i:n}, \theta)] = g_2(\theta) + g_3(\theta) g_4(x_{i:n}),
\]


• We then get Figures of the type of the following one.
A possible and general PP/QQ-plot

inclinação: $g_3(\theta)$
Let's next see a few examples.

**Example 3.** If we are interested in another very common model in SUE, the Fréchet model (for maxima) with location $\lambda = 0$ and with CDF,

$$F(x; 0, \delta, \alpha) = \exp(-{(x/\delta)}^{-\alpha}), \quad x \geq 0,$$

we get

$$-\log(-\log(p_i)) = -\alpha \log \delta + \alpha \log x_{i:n}, \quad 1 \leq i \leq n,$$

i.e. an ‘old probability paper’ for this population would have a logarithmic scale (to plot $x_{i:n}$), and the other scale would be double-logarithmic, the Gumbel functional scale.

**Remark 2.** Indeed, the logarithm of a Fréchet RV (for maxima) is a Gumbel RV (for maxima).
Example 4. Something similar happens with a Log-normal $X$, with location $\theta = 0$. Then, $\log X$ is Normal.

An ‘old Lognormal probability paper’ had a logarithmic scale (to plot $x_{i:n}$), and a Normal functional scale, to plot $i/(n+1)$. Nowadays, we merely need to plot

$$\left( \log x_{i:n}, \Phi^{-1}\left(\frac{i}{n+1}\right) \right), \quad 1 \leq i \leq n,$$

or to use, in $\mathbb{R}$, the function `qqnorm(y)` with $y = \log x$. 
**QQ-plot: Table of distributions.** The following table was taken from Beirlant et al. (2004):

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$F(x)$</th>
<th>Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>$\frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{(u-\mu)^2}{2\sigma^2} \right) du$</td>
<td>$(\Phi^{-1}(p_{i,n}), x_{i,n})$</td>
</tr>
<tr>
<td></td>
<td>$x \in \mathbb{R}; \mu \in \mathbb{R}, \sigma &gt; 0$</td>
<td></td>
</tr>
<tr>
<td>Log-normal</td>
<td>$\frac{1}{\sqrt{2\pi}\sigma u} \exp \left( -\frac{(\log u-\mu)^2}{2\sigma^2} \right) du$</td>
<td>$(\Phi^{-1}(p_{i,n}), \log x_{i,n})$</td>
</tr>
<tr>
<td></td>
<td>$x &gt; 0; \mu \in \mathbb{R}, \sigma &gt; 0$</td>
<td></td>
</tr>
<tr>
<td>Exponential</td>
<td>$1 - \exp(-\lambda x)$</td>
<td>$(-\log(1 - p_{i,n}), x_{i,n})$</td>
</tr>
<tr>
<td></td>
<td>$x &gt; 0; \lambda &gt; 0$</td>
<td></td>
</tr>
<tr>
<td>Pareto</td>
<td>$1 - x^{-\alpha}$</td>
<td>$(-\log(1 - p_{i,n}), \log x_{i,n})$</td>
</tr>
<tr>
<td></td>
<td>$x &gt; 1; \alpha &gt; 0$</td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>$1 - \exp(-\lambda x^\tau)$</td>
<td>$(\log(-\log(1 - p_{i,n})), \log x_{i,n})$</td>
</tr>
<tr>
<td></td>
<td>$x &gt; 0; \lambda, \tau &gt; 0$</td>
<td></td>
</tr>
</tbody>
</table>
W-plots: general $F(\cdot | \theta)$

- Just as mentioned above, the PP/QQ-plot is based on the

\[
F_\theta(X_{i:n}) \overset{d}{=} U_{i:n}, \quad i = 1, \ldots, n,
\]

with $U_{i:n}, i = 1, \ldots, n$ the ascending OS's associated with an $\mathcal{U}(0, 1)$ random sample. Then

\[
- \log(1 - F_\theta(X_{i:n})) \overset{d}{=} E_{i:n}, \quad i = 1, \ldots, n,
\]

with $E_{i:n}$ the ascending OS's associated with an $\mathcal{E}(1)$ random sample.

- In a **W-plot**, the points

\[
\left( -\log(1 - p_i), -\log(1 - F_\theta(x_{i:n})) \right) \quad i = 1, \ldots, n,
\]

are plotted and the closeness to the diagonal is checked.
Mean excess function and ME-plot

• In actuarial practice, the conditioning on the event \( \{X > t\} \) is of high importance, particularly in Reinsurance.
• Let us consider an excess-of-loss treaty, with retention \( t \), for any compensation of the contract.
• The reinsurer pays a random amount \( X - t \), the excess above \( t \), but only if \( X > t \).
• For the computation of the premium, the actuary needs to establish a deductible or a threshold \( t \), the so-called \( t \)-threshold of the POT methodology, computing next the expected amount to be paid by the client, given the choice of the threshold \( t \).
• As an example, the actuary computes the mean excess function,

\[
e(t) := \mathbb{E} [X - t|X > t],
\]

assuming that \( \mathbb{E}[X] < \infty \).
ME-plots (mean excess plots).

- In practice, the ME function, \( e(\cdot) \), is replaced by its empirical counterpart, \( \hat{e}_n(\cdot) \), based on the data \( x_1, \ldots, x_n \), and defined by

\[
\hat{e}_n(t) := \frac{\sum_{i=1}^{n} x_i I(t, +\infty)(x_i)}{\sum_{i=1}^{n} I(t, +\infty)(x_i)} - t,
\]

with \( I(t, +\infty)(x_i) := \begin{cases} 1 & x_i > t \\ 0 & x_i \leq t. \end{cases} \)

- Usually, \( \hat{e}_n \) is plotted in the values \( t = x_{n-k:n}, \ k = 1, \ldots, n-1 \).

- We then have \( \sum_{i=1}^{n} x_i I(t, +\infty)(x_i) = \sum_{j=1}^{k} x_{n-j+1:n}, \) with \( k \equiv \# x_i : x_i > t \) and the ME estimates are given by

\[
e_{k,n} := \hat{e}_n(x_{n-k:n}) = \frac{1}{k} \sum_{j=1}^{k} x_{n-j+1:n} - x_{n-k:n}.\]
Patterns of ME functions.

- Let’s see the type of pattern of the ME functions associated with specific models. Note first that we can write

\[
e(t) = \mathbb{E} [X - t | X > t] = \frac{\int_t^{x_F} (1 - F(u))du}{1 - F(t)},
\]

with \( x_F := \sup\{x : F(x) < 1\} \leq \infty \), the right endpoint of \( F \).

Remark 3. Note that the numerator of \( e(t) \), above, is obtained through the change in the order of integration (Fubini’s theorem)

\[
\int_t^{x_F} (x - t)dF(x) = \int_t^{x_F} \left( \int_t^x du \right) dF(x) = \int_t^{x_F} \left( \int_u^{x_F} dF(x) \right) du = \int_t^{x_F} (1 - F(u))du.
\]
Due to the lack of memory of the Exponential model, we get \( \forall t > 0, \)

\[
e(t) := \mathbb{E}[X - t|X > t] = \frac{\int_t^{x_F} (1 - F(u))du}{1 - F(t)} = \frac{\int_t^{+\infty} e^{-u/\delta}du}{e^{-t/\delta}} = \delta,
\]
i.e., the conditioning in \( \{X > t\} \) is irrelevant.

Generally, the shape of \( e(\cdot) \) provides information on the heaviness of the RTF, comparatively to that of the Exponential.

If the RTF of \( X \) is heavier than that of the Exponential, the ME function \( e(t) \) is increasing for high values of \( t \).

In the presence of a lighter RTF, the trend of the aforementioned function is decreasing.
4. Asymptotic Distributional Theory of OS’s

Introduction

- Let us assume that we have access to a sample, \((X_1, \ldots, X_n)\) of IID, or even stationary and weakly dependent RV’s from an underlying model \(F\), and let us denote by \((X_{1:n} \leq \cdots \leq X_{n:n})\) the sample of associated ascending OS’s. Let us also use the notation,

\[
M_n := X_{n:n} = \max_{1 \leq i \leq n} X_i, \quad n \geq 1.
\]

Then, the CDF of \(M_n\), given by

\[
\mathbb{P}\{M_n \leq x\} = F^n(x),
\]

is not too useful since \(F\) is unknown.
• But we are usually interested in the maxima of a large number of RV’s. We can thus use a related asymptotic result.

• What is the possible limiting behavior of $M_n$, as $n \to \infty$? We obviously have

$$F^n(x) \to \begin{cases} 0, & \text{if } F(x) < 1 \\ 1, & \text{if } F(x) = 1 \end{cases}.$$ 

• Consequently, $M_n \xrightarrow{d} x^F := \sup\{x : F(x) < 1\}$, the right endpoint of $F$.

• Due to the degeneracy of the limiting distribution of $M_n$, if we want to get a non-degenerate limit law, we need to normalize the sequence of maximum values, in a way similar to the well-known asymptotic theory for sums.
Asymptotic EVT for IID sequences

We mention the asymptotic behaviour of the 3 types of OS's:

1. **Central OS's**: $X_{k:n}$ where the order $k = k_n \to \infty$, but $k/n \to p$, $0 < p < 1$, as $n \to \infty$. Usually $k = \lfloor np \rfloor + 1$, $0 < \lambda < 1$, and $X_{k:n}$ is the empirical quantile of order $p$. Then, under adequate and not too restrictive regularity conditions on $F$, the asymptotic behaviour of $X_{k:n}$ is going to be Normal.

2. **Extremal OS's**: $X_{k:n}$ or $X_{n-k:n}$ with $k$ a fixed integer. The limiting model is then totally different from the Normal model.

3. **Intermediate OS's**: $X_{k:n}$, with $k = k_n \to \infty$, just as in 1., but with $k/n \to 0$ or $k/n \to 1$, as $n \to \infty$. Under adequate regularity conditions, we again have a Normal asymptotic behaviour.
In this course we shall consider the following topics:

- We first state Gnedenko’s theorem, and the 3 possible non-degenerate limiting distributions of the sequence of maxima, $M_n$, suitably linearly normalized, as $n \to \infty$.
- We next refer again the unified shape of these 3 types.
- We briefly refer to which max-domain of attraction (MDA) belong some of the most common models in applications.
- We further refer, also without proofs, the asymptotic distribution of the $k$-th maximum, for a fixed $k$, and of the $k$ top OS’s associated with a random sample of size $n$, as $n \to \infty$. 
Gnedenko’s theorem

- As mentioned above, the CDF of $M_n$ is $F^n(x)$, that converges to $x^F$, the right endpoint of $F$, as $n \to \infty$. Such a limiting distribution, even when proper, is degenerate and of limited interest.

- It seems thus sensible to put the question:

  - Is it possible to find real constants $\{a_n\}_{n \geq 1}$ ($a_n > 0$) and $\{b_n\}_{n \geq 1}$, and a non-degenerate RV $Y$, with CDF, $G(x)$, such that

    \[
    \frac{M_n - b_n}{a_n} \xrightarrow{d} Y \iff F^n(a_n x + b_n) \xrightarrow{n \to \infty} G(x)?
    \]

**Definition 1.** If there exist such sequences of constants, they are called attraction coefficients of $F$ to $G$, we say the $F$ belongs to the MDA of $G$, and use the notation $F \in \mathcal{D}_M(G)$. 

It is next sensible to place the following 2 problems, stated explicitly and partially solved by Gnedenko (1943):

(i) **Identification** of the possible limiting laws $G$.

(ii) **Characterization** of the MDA’s of the possible limiting laws for maxima, i.e. caracterization of:

\[
\mathcal{D}_M(G) := \left\{ F : \exists \{a_n\}_{n \geq 1} (a_n > 0) \text{ and } \{b_n\}_{n \geq 1} \text{ for which } F^n(a_nx + b_n) \xrightarrow{n \to \infty} G(x), \forall x \in \mathcal{C}(G) \right\},
\]

a topic partially beyond the scope of this course.
We first introduce Khinchine’s concept of type:

**Lemma 1** (Khinchine’s convergence of types). Let $U_1(x)$ and $U_2(x)$ be 2 non-degenerate CDF’s. If for a sequence of CDF’s, $\{F_n\}_{n \geq 1}$, there exist sequences of real numbers, $\{a_n\}_{n \geq 1}$, $\{b_n\}_{n \geq 1}$, $\{a'_n\}_{n \geq 1}$ and $\{b'_n\}_{n \geq 1}$ ($a_n, a'_n > 0$), such that

$$\lim_{n \to \infty} F_n(a_n x + b_n) = U_1(x) \quad \text{and} \quad \lim_{n \to \infty} F_n(a'_n x + b'_n) = U_2(x),$$

then

$$a'_n/a_n \xrightarrow{n \to \infty} A \quad (A > 0), \quad (b'_n - b_n)/a_n \xrightarrow{n \to \infty} B,$$

and

$$U_2(x) = U_1(Ax + B), \quad \forall x \in \mathbb{R}.$$
Definition 2. The CDF’s $U_1(x)$ and $U_2(x)$ are said to be of the same type if there exist real constants $A > 0$ and $B$ such that $U_2(x) = U_1(Ax + B)$, $\forall x \in \mathbb{R}$.

Theorem 2 (ETT: Gnedenko, 1943). Under the aforementioned framework, $F \in \mathcal{D}_M(G)$ if and only if $G$ belongs to one of the three following types of EV distributions (EVD’s):

Type I: $\Lambda(x) = e^{-e^{-x}}$, $x \in \mathbb{R}$ [Gumbel],
Type II: $\Phi_\alpha(x) = e^{-x^{-\alpha}}$, $x > 0$, $\alpha > 0$ [Fréchet],
Type III: $\Psi_\alpha(x) = e^{-(x)^\alpha}$, $x < 0$, $\alpha > 0$ [Max – Weibull].
• **The unified or general EV (GEV) distribution and EVI.**

Von Mises (1936 [Revue Math. Union Interbalcanique]; 1954) and Jenkinson (1955), [Quart. J. Royal Meteorol. Soc.] unified the three families, Gumbel, Fréchet and Max-Weibull, considering the **GEV distribution** (GEVD), with the functional expression,

\[ G_{\xi}(x) = \exp \left\{ - \left[ 1 + \xi x \right]_+^{-1/\xi} \right\}, \quad x_+ := \max(0, x), \]

where \( \xi \) is the **EVI**, the crucial parameter in **Statistics of Extremes**. Such a parameter measures essentially the weight of the **RTF**, \( \overline{F} := 1 - F \).

• We can thus state **Gnedenko’s theorem (Theorem 2)** as:
**Theorem 3 (Unified ETT).** If there exist sequences $a_n > 0$ and $b_n$, such that, as $n \to \infty$

\[
\mathbb{P} \left[ \frac{M_n - b_n}{a_n} \leq x \right] \to G(x) \text{ for some non-degenerate CDF } G,
\]

then $G$ is of the type of a GEVD, $G_\xi \equiv \text{GEV}_\xi$, for some $\xi \in \mathbb{R}$.

We then obviously write $F \in \mathcal{D}_M(\text{GEV}_\xi)$.

- If $\xi < 0$, the RTF is short, $x_F < \infty$.
- If $\xi = 0$, the RTF is of exponential-type, $x_F < \infty$ or $x_F = \infty$.
- If $\xi > 0$, the RTF is heavy of a Pareto-type, and $x_F = \infty$. 
**EV PDF**, $g_\xi(x) = dG_\xi(x)/dx$, for $\xi = -0.5$, $\xi = 0$ and $\xi = 2$, together with the normal PDF, $\phi(x) = \exp(-x^2/2)/\sqrt{2\pi}$, $x \in \mathbb{R}$.
Remark 4. If $\xi = 0$, we get the Gumbel CDF,

$$\text{GEV}_0(x) = \exp[- \exp(-x)] = \Lambda(x), \ x \in \mathbb{R}.$$  

- If $\xi > 0$ ($\alpha = 1/\xi$), we get the Fréchet CDF. With the notation, $\text{GEV}_\xi(x; \lambda, \delta) = \text{GEV}((x - \lambda)/\delta)$, $\Phi_\alpha(x) = \text{GEV}_{1/\alpha}(x; 1, 1/\alpha)$, i.e.

$$\text{GEV}_\xi(x) = \begin{cases} 0, & x < -1/\xi, \\ \exp\{-[1 + \xi x]^{-1/\xi}\}, & x \geq -1/\xi. \end{cases}$$

- If $\xi < 0$ ($\alpha = 1/\xi$), we get the Max-Weibull CDF ($\Psi_\alpha(x) = \text{GEV}_{-1/\alpha}(x; -1, -1/\alpha)$), i.e.

$$\text{GEV}_\xi(x) = \begin{cases} \exp\{-[1 + \xi x]^{-1/\xi}\}, & x \leq -1/\xi, \\ 1, & x > -1/\xi. \end{cases}$$
Remark 5. Note that if there exists the aforementioned non-degenerate behavior for maxima,

\[
\lim_{n \to \infty} \mathbb{P}(M_n \leq a_{nk}x + b_{nk}) = \lim_{n \to \infty} F^n(a_{nk}x + b_{nk}) \\
= \lim_{n \to \infty} \left(F^{nk}(a_{nk}x + b_{nk})\right)^{1/k} \\
= G^{1/k}, \quad \forall \ k \geq 1.
\]

Consequently, the validity of the equation above for \( k = 1 \) implies its validity for all \( k > 1 \), i.e. \( G(x) \) and \( G^{1/k}(x) \) are of the same type for all \( k > 1 \), i.e. \( G \) is a max-stable CDF, and

\[
G^k(\alpha_kx + \beta_k) = G(x), \quad \forall x \in \mathbb{R}.
\]

Moreover, if the aforementioned equation holds, then \( G \) is an EVD, of one of three above mentioned types, type I, II or III (equivalently, of a GEV-type).
Remark 6. Note that this stability equation was less formally postulated by Fréchet (1927), observing that the maximum of the $m \times n$ values $X_1, X_2, \ldots, X_{m \times n}$ is also the maximum of the $n$ maxima values of $X_{(i-1)m+1}, \ldots, X_{im}$, $1 \leq i \leq n$.

This stability postulate was also used by Fisher & Tippett (1928), in a pioneering paper where we can find the most important ideas and problems of classical EVT.
Remark 7. Any result for maxima (top OS’s) can be easily reformulated for minima (low OS’s). Indeed,

\[
\min_{1 \leq i \leq n} X_i = - \max_{1 \leq i \leq n} (-X_i),
\]

and consequently, when \( n \rightarrow \infty \),

\[
P[X_{1:n} \leq x] = 1 - (1 - F(x))^n \approx \text{GEV}_\xi^*(\frac{x - \lambda_n'}{\delta_n'}),
\]

with

\[
\text{GEV}_\xi^*(x) = 1 - \text{GEV}_\xi(-x) = 1 - \exp\left\{-(1 - \xi x)^{-1/\xi}\right\},
\]

for \( 1 - \xi x > 0 \) and \((\lambda_n', \delta_n') \in (\mathbb{R}, \mathbb{R}^+)\).
**Example 5** (Models in the Fréchet MDA, $F \in \mathcal{D}(\text{GEV}_\xi), \xi > 0$). Among others, we provide the following examples:

- **Pareto**, $\text{Pa}(\alpha) : F(x) = 1 - x^{-\alpha}, \ x > 1; \ \alpha > 0$; **EVI**: $\xi = 1/\alpha$;

- **Generalized Pareto**, $\text{GP}(\sigma, \xi) : F(x) = 1 - \left(1 + \xi \frac{x}{\sigma}\right)^{-\frac{1}{\xi}}, \ x > 0$; $\sigma, \xi > 0$; **EVI**: $\xi$;

- **Burr**($\eta, \tau, \lambda$) : $F(x) = 1 - \left(\frac{\eta}{\eta + x^\tau}\right)^\lambda, \ x > 0$; $\eta, \tau, \lambda > 0$; **EVI**: $\xi = 1/(\lambda \tau)$;

- **Fréchet**($\alpha$) : $F(x) = \exp(-x^{-\alpha}), \ x > 0; \ \alpha > 0$; **EVI**: $\xi = 1/\alpha$;

- **Student-$t$, with $\nu$ degrees of freedom** **EVI**: $\xi = 1/\nu$;

- **Cauchy**: $F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x, \ x \in \mathbb{R}$; **EVI**: $\xi = 1$;

- **Log-Gama**($\alpha, \lambda$) : $F(x) = \int_1^x \frac{\lambda^\alpha}{\Gamma(\alpha)} (\log t)^{\alpha-1} t^{\lambda-1} dt, \ x \geq 0$; **EVI**: $\xi = 1/\lambda$.  


Example 6 (Distributions in the Max-Weibull MDA, $F \in \mathcal{D}_M(\text{GEV}_\xi)$, $\xi < 0$). Classical examples in this MDA are:

- **Uniforme**, $\mathcal{U}(0,1)$: $F(x) = x$, $0 < x < 1$; **EVI**: $\xi = -1$;

- **Beta**($p, q$): $F(x) = \int_0^x \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} u^{p-1}(1-u)^{q-1} du$, $0 < x < 1$; $p, q > 0$; **EVI**: $\xi = -1/q$;

- **Reversed Burr**($\beta, \tau, \lambda$): $F(x) = 1 - \left(\frac{\beta}{\beta + (-x)^{-\tau}}\right)^\lambda$; $x < 0$; $\beta, \tau, \lambda > 0$; **EVI**: $\xi = -1/(\lambda\tau)$;

- **Max-Weibull**: $F(x) = \exp\left(-(-x)^\alpha\right)$, $x < 0$; $\alpha > 0$; **EVI**: $\xi = -1/\alpha$. 
Example 7 (Models in the Gumbel MDA, $F \in \mathcal{D}_M(\text{GEV}_\xi)$, $\xi = 0$). Among others, we mention the following distributions:

- **Exponential**, $\mathcal{E}(1)$: $F(x) = 1 - \exp(-x)$, $x > 0$;
- **Weibull** (of minima): $F(x) = 1 - \exp(-\lambda x^\tau)$, $x > 0$; $\lambda, \tau > 0$;
- **Logistic**: $F(x) = 1 - \frac{1}{1 + \exp(x)}$, $x \in \mathbb{R}$;
- **Gumbel**: $\Lambda(x) = \exp(-\exp(-x))$, $x \in \mathbb{R}$;
- **Normal**: $\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)dt$, $x \in \mathbb{R}$;
- **Log-normal**: $\log X \sim \text{Normal}$;
- **Gama** ($\alpha, \beta$): $F(x) = \int_0^x \frac{\alpha^\beta}{\Gamma(\alpha)} (\log t)^{\beta - 1} t^{-\alpha - 1} dt$, $x > 0$;
- **Min-Fréchet**: $\Phi_\alpha^*(x) = 1 - \Phi_\alpha(-x) = 1 - \exp(-(-x)^{-\alpha})$, $x < 0$; $\alpha > 0$. 

93
Asymptotic distribution of the $k$-th top OS

**Theorem 4.** Let us consider $M^{(k)}_n = X_{n-k+1:n}$ ($M^{(1)}_n \equiv M_n$), for fixed $k$. Assume that there exist $a_n > 0$ and $b_n$ such that $F \in \mathcal{DM}(G)$. Then, $\forall k \geq 1$, a fixed integer, and for the same $(a_n, b_n)$,

$$
P[M^{(k)}_n \leq a_n x + b_n] \xrightarrow{n \to \infty} G_k(x) = G(x) \sum_{i=0}^{k-1} \frac{[- \log G(x)]^i}{i!}.
$$

**Remark 8.** The corresponding limiting PDF is given by

$$
g_k(x) = \{- \log G(x)\}^{k-1} g(x)/(k - 1)!
$$
Joint limiting behaviour of top OS’s

- Apart from the ETT and the already mentioned GEVD, it is also sensible mentioning the multivariate GEVD (MGEVD), related to the limiting distribution of the $k$ largest OS’s $X_{n-i+1:n}$, $1 \leq i \leq k$, also called the extremal process [Lamperti, 1964, AMS; Dwass, 1964, AMS].

Theorem 5 (Joint limiting behaviour of the $k$ top OS’s). $F \in \mathcal{D}_\mathcal{M}(G)$, with constants $b_n \in \mathbb{R}$ and $a_n > 0$, if and only if, for fixed $k$, and with $G' = g$, the $k$-vector $\left( (X_{n:n-b_n}/a_n, \ldots, (X_{n-k+1:n-b_n})/a_n \right)$ has a non-degenerate limiting behaviour, with an associated PDF,

$$g_{1, \ldots, k}(w_1, \ldots, w_k; G) := G(w_k) \prod_{i=1}^{k} \frac{g(w_i)}{G(w_i)}, \text{ for } w_1 > \cdots > w_k.$$
Limiting structure of the excesses over a high threshold

- Rather than the maxima, we can consider all values larger than a given threshold.
- The differences between these values and a given threshold are called excesses over the threshold.
- The conditional distribution of excesses over a high threshold $u$, i.e. the CDF of the RV $Y|Y > 0$, with $Y = X - u$, is given by

$$F_u(x) := P\left(X - u \leq x \mid X > u\right), \text{ for } 0 \leq x \leq x^F - u.$$

This is the so-called excess-life or residual lifetime in reliability or medical statistics, and excess-of-loss in insurance.
According to Balkema & de Haan (1974), [Ann. Probab.] and Pickands (1975), [Ann. Statist.] the key result in EVT which explains the importance of the GP distribution (GPD) is the following:

**Theorem 6** (Pickands-Balkema-de Haan’s Theorem).

\[
F \in D_M(\text{GEV}_\xi) \iff \lim_{u \to x^F, x \in [0, x^F - u]} \sup_{x \in [0, x^F - u]} |F_u(x) - \text{GP}_\xi(x/\sigma(u))| = 0
\]

for some positive function \( \sigma(u) \).

Thus, the GPD, defined by

\[
\text{GP}_\xi(x) = \begin{cases} 
1 - (1 + \xi x)^{-1/\xi}, & 1 + \xi x > 0, \ x > 0, \text{ if } \xi \neq 0, \\
1 - \exp(-x), & x > 0, \text{ if } \xi = 0.
\end{cases}
\]

is the natural model for the unknown excess distribution above sufficiently high thresholds in the POT approach.
5. PARAMETRIC SUE

Parameters of extreme events

- One of crucial parameters in SUE is the already defined EVI.
- For dependent samples, we also have the extremal index (EI), that we shall briefly define later on.
- Beyond these primordial parameters, we next recall the definitions of extremal quantile and return period of a high level $t$.

**Definition 3** (reciprocal tail quantile function). Let $F$ be a continuous CDF with generalized inverse $F^{-}$. The associated reciprocal tail quantile function (RTQF) is given by

$$U(t) = F^{-}(1 - 1/t) = \inf\{x : F(x) \geq 1 - 1/t\}, \quad t \in [1, \infty].$$
Definition 4 (extremal quantile). An upper extremal quantile is, for small $p$ (usually smaller than $1/n$, with $n$ the sample size), the value 

$$\chi_p := F^{-1}(1 - p) = U(1/p),$$

with $U(\cdot)$ the RTQF defined in Definition 3.

Definition 5 (return level and return period). Again with $U(\cdot)$ the RTQF defined in Definition 3, the value $u := U(t)$ is usually called the return level of the value $t$, and given $u$, the value $T(u)$ such that $u = U(T(u))$ is the so-called return period of the level $u$. We thus get

$$T(u) = 1/(1 - F(u)),$$

the mean number of exceedances of the level $u$ in an IID framework.
In finance, the *Value-at-Risk* (VaR) is a common indicator.

**Definition 6 (Value-at-Risk, VaR).** Let $X$ be a RV associated with the profits and losses (P&L) or to the returns of a financial product at a certain temporal horizon. The VaR is an extremal percentile of the CDF $F$ associated with $X$,

$$\text{VaR}_p := F^{\leftarrow}(1 - p),$$

or a value that is exceeded with a small probability $p$,

$$\text{VaR}_p : \mathbb{P}[X > \text{VaR}_p] = p.$$

**Remark 9.** In practice, the most usual values of $p$ are of the type $p = 0.01, 0.001, 0.0001$. 
• **Statistical inference** about **rare events** is clearly linked to observations which are **extreme** in some sense. Different ways to define such observations lead to different **alternative approaches** to **SUE**.

• One of the difficulties associated with a tail analysis lies on the limited number of available data.

• For instance, in one of the the case-studies to be considered in the illustrations, the yearly maximal discharges of the river Meuse (1911-1995), available at ‘maasmax.txt’, **ONLY** the **yearly maxima** \(m = 85\) **years** are available.

• It looks thus sensible to use the main limiting result in **EVT**, that provides the **max-stable CDF’s** as the unique possible limiting **CDF’s** for the linearly normalized maximum of a random sample.
• Even more generally, assume that we have for instance access to 
$$(x_1, x_2, \ldots, x_n)$$, the average daily discharges of a river in a certain place, **BUT** that we are interested
  – in **floods** ($\max_{1 \leq i \leq n} x_i$)
  – or in **droughts** ($\min_{1 \leq i \leq n} x_i$) of that river in that same place.
• It is then also sensible to work with those **maxima** (or **minima**) of a high number of observations.
• But the exact distributional behaviour of the largest (or lowest) values in a sample is usually difficult. Moreover, we can work with the **maximum** (or **minimum**) of a **large number** of observations.
• It is thus common to use the **asymptotic behavior** of those tail values, with a known **CDF**, up to unknown parameters.
Gumbel’s approach or block maxima method

- When the sample size $n \to \infty$, and due to the limiting result for the normalized sequence of maximum values, i.e. Gnedenko’s ETT, we can write

$$P[X_{n:n} \leq x] = F^n(x) \approx \text{GEV}_\xi((x - \lambda_n)/\delta_n),$$

with $\text{GEV}_\xi(x)$ the GEVD and $(\lambda_n, \delta_n) \in (\mathbb{R}, \mathbb{R}^+) \text{ an unknown vector of location and scale parameters}$, that replaces the attraction coefficients $(b_n, a_n)$ in the normalized sequence of maximum values, $(X_{n:n} - b_n)/a_n$.

- The aforementioned ETT was used by Gumbel in several papers which culminated in his 1958 book, to give approximations of the type of the one provided above but for any of the 3 EVD’s, Fréchet, Gumbel and max-Weibull.
• **Gumbel** has thus suggested the first model in **SUE**, 
  – the *annual maxima* or block maxima (**BM**), or 
  – **EV univariate model**, or merely 
  – Gumbel’s model.

• Observed values of \((X_1, \ldots, X_N), \ N \text{ large, are then splitted in}
  – \(k\) sub-samples (often corresponding to the \(k\) years) of size \(n\),
  that should also be large \((N = nk)\)

• and one of the **EV models**, obviously with extra unknown location
  and scale parameters,
  – **Type I** or **II** or **III**, or
  – **GEV**
  is fitted to the sample of the \(k\) maxima of the \(\text{RV} \ Y = \max(X_1, \ldots, X_n)\).
Associated with the GEV fitting to $Y$, we have a few parameters of rare events, like:

- **Probability of exceedance** of a high level $u$ (for $Y$),

$$1 - \text{GEV}_\xi(u; \lambda, \delta) = \begin{cases} 
1 - \exp \left\{ -\left[ 1 + \xi \left( \frac{u-\lambda}{\delta} \right) \right]^{-1/\xi} \right\}, & \text{if } \xi \neq 0, \\
1 - \exp \left\{ -\exp \left[ -\frac{u-\lambda}{\delta} \right] \right\}, & \text{if } \xi = 0.
\end{cases}$$

- **Return period** of a high level $u$ (again for $Y$), i.e. the mean number of exceedances of the level $u$ in an IID framework.

$$T_u = \frac{1}{1 - \text{GEV}_\xi(u; \lambda, \delta)}.$$
• *T*-years return level (also for $Y$), i.e. the value

$$U(T) = \text{GEV}_{\xi}^{-}(1 - \frac{1}{T}; \lambda, \delta)$$

$$= q_{Y,p} = \begin{cases} 
\lambda + \frac{\delta}{\xi} \left[ (-\log(1 - p))^{-\xi} - 1 \right], & \text{if } \xi \neq 0, \\
\lambda - \delta \log(-\log(1 - p)), & \text{if } \xi = 0,
\end{cases}$$

for $p = 1/T$.

• And if $\xi < 0$, the right endpoint (also for $Y$) is estimated by

$$x^F = q_{Y,0} = \lambda - \delta/\xi.$$
• If we want to infer on $X \sim F$, we can use the max-stability, i.e. the fact that $Y := \max_{1 \leq i \leq n} X_i = X_{n:n}$ has a CDF

$$F_Y \equiv F_{X_{n:n}} = F^n \approx \text{GEV}_\xi.$$  

• Then, for blocks with size $n$, the $(1 - p)$-quantiles for $X$, $q_{X,p}$, values such $F(q_{X,p}) = 1 - p$ are well approximated by

$$F^n(q_{X,p}) = (1 - p)^n \approx \text{GEV}_\xi(q_{X,p}; \lambda, \delta),$$

if $p$ is small, i.e.

$$q_{X,p} = \text{GEV}^\leftarrow_\xi((1 - p)^n; \lambda, \delta) = \begin{cases} \lambda + \delta \left[ (-n \log(1 - p))^{-\xi} - 1 \right] / \xi, & \text{if } \xi \neq 0, \\ \lambda - \delta \log(-n \log(1 - p)), & \text{if } \xi = 0. \end{cases}$$
• All statistical inference is then associated with the aforementioned models, and the estimation of $\lambda$, $\delta$, and $\xi$ (or $\alpha$) is crucial and the basis for the estimation of any other relevant parameter.

• We get to know that the usefulness of a model depends partially on the existence of reliable methods for the estimation of unknown parameters. Since such estimation is (was?) not easy for the GEVD, given a sample of maxima, it is common to try the fitting of one of the 3 CDF’s,
  – Gumbel (the easiest one),
  – Fréchet or
  – Max-Weibull,
performing first a statistical choice test of one of the 3 models. Such a test can be as simple as a QQ-plot.
• Just as mentioned above, if the underlying model is the Gumbel, there exists a linear relationship between the ascending ordered observations, \( x_{i:n} \), and \(-\log(-\log(i/(n + 1)))\), \(1 \leq i \leq n\).

• Then, \((\lambda, \delta)\) can be estimated through the least squares method.

• Moreover, if the points \((x_{i:n}, -\log(-\log(i/(n + 1))))\) are concave (⌢), we get an informal validation of the Fréchet model.

• Convex (⌣) plots, \((x_{i:n}, -\log(-\log(i/(n + 1))))\), \(1 \leq i \leq n\), provide an informal validation of a Max-Weibull model.

• We can also go on with a preliminary estimation of \(\xi\) as the value that maximizes the correlation in the QQ-Plot.

• But it is nowadays more common to use
  – the ML method, or
  – the probability weighted moments (PWM) method.

• These methods are implemented in several R-packages.
Data in ‘maasmax.txt’ and the BM method

• We have here only access to the yearly maxima discharges, $Y_1, \ldots, Y_m$, of Meuse’s river (Borgharen, NL), in $m^3/s$, 1911–1995, in a total of $m = 85$ years (Beirlant et al., 2004; http://lstat.kuleuven.be/Wiley/).

• These data are replicates of the RV $Y \equiv M_n$, with $M_n :=$ yearly maximum $= \max(X_1, \ldots, X_n), \ n = 365$ (years).

• These data are presented in the following Figure, together with the associated box-and-whiskers plot.
Yearly maxima of Meuse’s river in 1911-1995 (*left*) and associated box-and-whiskers plot (*right*)

- Let’s see what would have happened should we have decided for a ‘traditional statistical analysis’, with the fitting of a Normal, $\mathcal{N}(\mu, \sigma^2)$ model to the *yearly maxima*. 
• In the following Figure we present the associated Normal QQ-plot.

Normal QQ-plot: $\{(\Phi^{-1}(i/(n + 1)), y_{i:n}) : i = 1, \ldots, n\}$
The fitting of a least squares line leads to the estimates $\hat{\mu} = 1495.962$ and $\hat{\sigma} = 551.0057$. The corresponding correlation coefficient is $r_Q = 0.9788504$. The superposition of the histogram (yearly maxima) and the estimate Normal,

$$Y \sim \mathcal{N}(\hat{\mu} = 1495.962, \hat{\sigma} = 551.0057),$$

leads to the following Figure.
• Given the nature of the data, the importance of Gumbel’s model in EVT, and the right asymmetry of the histogram, let us try the fitting of a Gumbel’s model to the yearly maxima (BM methodology).

• In the following Figure, we present the associated QQ-plot.

• The fitting of a least squares line, leads to the estimates \( \hat{\lambda} = 1247.363 \) and \( \hat{\delta} = 445.6884 \), respectively for location and scale unknown parameters, and to an empirical correlation coefficient \( r_Q = 0.9924606 \) (higher than the one found for the Normal fitting).
Gumbel QQ-plot: \( \{ (\Lambda^{-1}(i/(m + 1)), y_{i:m}) : i = 1, \ldots, m \} \)
Surely the Gumbel model provides a better fit than the Normal.
- Let us now think on the estimation of the \( T \)-years return level, given the Normal fitting. We have

\[
U(T) = F_Y^{-1}(1 - 1/T),
\]

and the average level of discharges overpassed by yearly maxima every \( T = 100 \) years would be estimated by

\[
\hat{U}(100) = \hat{\mu} + \hat{\sigma} \Phi^{-1}(1 - 1/100) \\
= 1495.962 + 551.0057 \Phi^{-1}(0.99) = 2777.793.
\]

- Regarding the return period of the level \( y_T = 3175 = y_{85:85} \), the maximal discharge along the 85 years, we get

\[
T = \frac{1}{1 - F_Y(y_T)}, \quad \text{and} \quad \hat{T} = \frac{1}{1 - \Phi\left(\frac{3175 - \hat{\mu}}{\hat{\sigma}}\right)} \approx 866 \text{ years.}
\]
Let us next consider the estimated Gumbel,

\[ Y \sim \Lambda(\cdot; \hat{b} = 1247.363, \hat{a} = 445.6884). \]

Let us think again on the \( T = 100 \)-years return level, \( U(T) = F_Y^{-\left(1 - 1/100\right)} \). Such a return level is estimated by

\[ \hat{U}(100) = \hat{b} + \hat{a} \Lambda^{-\left(1 - 1/100\right)} = 3297.596, \]

above the estimated value associated with the Normal fitting (2777.793), i.e. normal under-estimates such a return level.

Regarding the return period of the level \( y_T = 3175 \), we get

\[ \hat{T} = \frac{1}{1 - \Lambda\left(\frac{3175 - \hat{b}}{\hat{a}}\right)} \approx 76 \text{ years}, \]

much below the one associated to a normal fitting (866).
• More generally, let us consider the GEVD as a candidate to model the yearly maximum $Y$.

• Let us consider the estimate of $\xi$ that maximizes the sample correlation in the associated QQ-plots, for $\xi \in [-0.5, 0.5]$.

• Such a correlation, as a function of $\xi$, is reproduced in the following Figure.

Sample correlation in the QQ-plots associated with different values of $\xi \in [-0.5, 0.5]$.
• We then get $\hat{\xi} = -0.03419188$, and the **GEV QQ-plot** associated with this value of $\xi$ is represented in the following Figure.

• The least squares method leads to the location/scale parameters’ estimates, $\hat{\lambda} = 1251.887 / \hat{\delta} = 461.6413$, and to a correlation coefficient $r_Q = 0.9928135$, slightly above the one found for the Gumbel fitting, as expected.

![GEV QQ-plot](image)

**GEV ($\hat{\xi}$) QQ-plot:** $\left\{ (G_{\hat{\xi}}(i/(m+1)), y_{i:m}) : i = 1, \ldots, m \right\}$
• The $T$-years return level, $U(T) = F_Y^{<} (1 - 1/T)$, is estimated by

$$
\hat{U}(100) = \hat{b} + \hat{a} G_{\hat{\xi}}^{<} (1 - 1/100) \\
= 1251.887 + 461.6413 G_{\hat{\xi}}^{<} (0.99) = 3216.919,
$$

with $\hat{\xi} = -0.03419188$.

• As expected, since $\hat{\xi} < 0$, the Gumbel model slightly overestimates the $(T = 100)$-years return level.

• The return period of the level $y_T$ is estimated by

$$
\hat{T} = \frac{1}{1 - G_{\hat{\xi}} \left( \frac{3175 - \hat{b}}{\hat{a}} \right)} \simeq 90 \text{ years},
$$

a value slightly above the one found for the Gumbel fitting, but much below the one found for the normal fitting.
• We have fitted a $\text{GEV}(\xi; \lambda, \delta)$ to the RV $Y := \max(X_1, \ldots, X_n)$, based on a sample of $m$ yearly maxima $Y_1, Y_2, \ldots, Y_m$, and we have conducted a preliminary estimation, by choosing the value of $\xi$ that maximizes the correlation in the QQ-plot. Next, the location and scale parameters, $\lambda$ and $\delta$, were estimated through the least squares method.

• We shall next estimate the unknown parameters through the ML methodology, using some of the aforementioned R packages for the fitting to these data of max-stable models.
**ML-estimation**

“library(ismev)” (gum.fit(y))

- For a **Gumbel** fitting, we got the ML-estimates $\hat{\lambda} = 1243.567$ and $\hat{\delta} = 456.454$ (*recall that the preliminary estimates associated with a Gumbel fitting were $\hat{\lambda} = 1247.363$ and $\hat{\delta} = 445.688$*).
- Other possible alternatives could be “library(fitdistrplus)”, “library(evir)”, “library(evd)”, “library(fExtremes)”.
- Obtained results are summarized in the following Table.

<table>
<thead>
<tr>
<th>“library”</th>
<th>$\hat{\lambda}$ (location)</th>
<th>$\hat{\delta}$ (scale)</th>
<th>100-years return level</th>
</tr>
</thead>
<tbody>
<tr>
<td>“ismev”</td>
<td>1243.6</td>
<td>456.5</td>
<td>3343.32</td>
</tr>
<tr>
<td>“fitdistrplus”</td>
<td>1243.9</td>
<td>456.6</td>
<td>3344.36</td>
</tr>
<tr>
<td>“evir”</td>
<td>1243.6</td>
<td>456.5</td>
<td>3343.32</td>
</tr>
<tr>
<td>“evd”</td>
<td>1240.3</td>
<td>455.2</td>
<td>3334.29</td>
</tr>
<tr>
<td>“fExtremes”</td>
<td>1243.6</td>
<td>456.5</td>
<td>3343.47</td>
</tr>
</tbody>
</table>

**Gumbel fitting**: estimation of location, scale and 100-years return level
The use of the aforementioned *packages* in a GEV fitting, and taking as initial values in the ML iterative procedure the estimates obtained through least squares estimation in the QQ-plot, \( \hat{\lambda} = 1251.887, \hat{\delta} = 461.6413, \hat{\xi} = -0.03419188 \), leads to the estimates provided in the following Table.

<table>
<thead>
<tr>
<th>“library”</th>
<th>( \hat{\xi} )</th>
<th>( \hat{\lambda} ) (location)</th>
<th>( \hat{\delta} ) (scale)</th>
<th>100-years return level</th>
</tr>
</thead>
<tbody>
<tr>
<td>“ismev”</td>
<td>-0.09243</td>
<td>1267.22750</td>
<td>466.79293</td>
<td>3016.41</td>
</tr>
<tr>
<td>“fitdistrplus”</td>
<td>-0.09220</td>
<td>1266.97383</td>
<td>466.43461</td>
<td>3015.67</td>
</tr>
<tr>
<td>“evir”</td>
<td>-0.09243</td>
<td>1267.22750</td>
<td>466.79293</td>
<td>3016.41</td>
</tr>
<tr>
<td>“evd”</td>
<td>-0.08944</td>
<td>1259.08675</td>
<td>466.12376</td>
<td>3016.97</td>
</tr>
<tr>
<td>“fExtremes”</td>
<td>-0.09232</td>
<td>1266.88569</td>
<td>466.45138</td>
<td>3015.17</td>
</tr>
</tbody>
</table>

**GEV fitting**: estimation of shape, location, scale and 100-years return level
Remark 10. As expected, note the smaller values of the estimates of the 100-years return level whenever we try a GEV fitting, comparatively to the Gumbel fitting.

- We have further got the observed values of several goodness-of-fit tests, provided by Kolmogorov-Smirnov, Cramer-von Mises and Anderson Darling statistics, through the use of “library(fitdistrplus)”

- As expected, in the case of a GEV fitting, the goodness-of-fit statistics provide values slightly below the corresponding ones provided for Gumbel fitting.

- But we have never been led to the rejection of either the GEV or the Gumbel model.
Alternatively, we can use “library(fExtremes)”

- In the case of a negative EVI ($\xi < 0$), the right endpoint of the GEVD is finite and can be estimated by
  \[
  \hat{x}^F = \hat{\lambda} - \hat{\delta}/\hat{\xi}.
  \]

- In this case-study, and for the yearly maximum $Y$ we obtain a right endpoint estimate around 6000.

- Assume now that we want to estimate the $T = 100$ months return level. We can identify the yearly maximum $Y$ with the maximum of $m = 12$ (approximately IID) monthly maxima, $X$, i.e. $Y = X_{m:m} = \max(X_1, \ldots, X_{12})$. For these blocks of size $m = 12$, the $(1 - p) = 1 - 1/100$-quantile of $X$ is estimated by
  \[
  \hat{q}_{X,p}^* = \text{GEV}_{\hat{\xi}}^{-1}\left((1 - 1/100)^{12}; \hat{\lambda}, \hat{\delta}\right).
  \]
Multivariate EV (MEV) and multi-dimensional EV (MDEV) approaches: the method of largest observations (LOB)

- Although Gumbel’s statistical procedure has proved to be fruitful in the most diverse situations, several criticisms have been made on Gumbel’s technique:
  - One of them is the fact that we are wasting information when using only observed maxima and not further OS’s, if available, because they certainly contain useful information about the RTF underlying the data.
  - On the other hand, in most areas of application there is no natural seasonality of the data, and in such a framework the method of sub-samples may look a bit subjective and artificial.
• To infer on the **RTF weight** of the underlying model, it seems sensible to think on a small number $k$ of **top OS's** from the original data.

• Indeed, if we have daily data, some years may have several values among those **top OS's** (that are for sure relevant to make inference upon the **RTF**), and other years may contain none of those top values.

• We can thus say that such an approach provides **additional information**, that has been **disregarded** in the traditional Gumbel's methodology.

• This approach surely depends on the **joint limiting** behaviour of those $k$ top **OS's**.

• When the sample size $n$ is large and for fixed $k$, it is sensible to consider the **PDF** $h_\xi(\cdot,\ldots,\cdot) := g_{1,\ldots,k}(\cdot,\ldots,\cdot; \text{GEV}_\xi)$, already defined.
A possible parametric approach to SUE is thus the modeling of the top observations through the joint behaviour in Theorem 5. As $N \to \infty$, with fixed $k$,

$$\mathbb{P}[X_{N:N} \leq x_1, \ldots, X_{N-k+1:N} \leq x_k] \approx H_\xi\left(\frac{x_1 - \lambda}{\delta}, \ldots, \frac{x_k - \lambda}{\delta}\right)$$

where, with $\text{GEV}_\xi$ the univariate GEVD, and $\text{gEV}_\xi$ the associated PDF,

$$h_\xi(x_1, \ldots, x_k) = \frac{\partial^k}{\partial x_1 \ldots \partial x_k} H_\xi(x_1, \ldots, x_k)$$

$$= \frac{1}{\delta^k} \text{gEV}_\xi\left(\frac{x_k - \lambda}{\delta}\right)^{k-1} \prod_{j=1}^{k} \frac{\text{gEV}_\xi\left(\frac{x_j - \lambda}{\delta}\right)}{\text{GEV}_\xi\left(\frac{x_j - \lambda}{\tilde{\delta}}\right)},$$

the so-called MGEV distribution (MGEVD).
• This approach to **SUE** is the so-called **MGEV model** or **LOB method** or **extremal process**.

• Under this approach it is easier to increase the number $k$ of observations, contrarily to what happens in **Gumbel's approach**, where a larger number $N$ of original observations is usually needed.

• Such an approach has been introduced first, in a slightly different context, by **Pickands (1975)** and was used by **Weissman (1978)** and **Gomes (1978, 1981)**.
**Remark 11.** Note finally that it is easy to **combine both approaches.** In each of the sub-samples associated with **Gumbel's classical approach,** we can collect a few **top OS’s** modelled through a **MGEV model,** and next consider the **MDGEV model,** based on the **multivariate sample,**

$$(X_1, X_2, \ldots, X_k),$$

where

$$X_j = (X_{1j}, \ldots, X_{ij}), \quad 1 \leq j \leq k,$$

are **MGEV vectors.**
The POT approach

- Another approach to SUE, in a certain sense parallel to the MGEV model, is the one in which we restrict our attention only to the observations that exceed a certain high threshold $u$, fitting the appropriate statistical model to the excesses over such a high level $u$. On the basis of Theorem 6, we get the GPD approximation,

$$P[X - u \leq x | X > u] \approx GP_\xi(x/\sigma(u)).$$

- We are then led to consider a deterministic high level $u$, work with the excesses and the GPD model.

- Such a model is the so-called Paretian excesses model or POT model, and was introduced in Smith (1987), [Ann. Statist.]. Here, all statistical inference is related to the GPD.
• Indeed, in most data basis, as happens for instance with data in ‘soa.txt’ from the Group Medical Insurance, and related to Large Indemnizations, we have access to the $k = 75789$ exceedances above the threshold $u = 25000$ USD, in 1991.

• We can then work with those $k$ exceedances, or, as also usual in SUE, with the excesses above a certain extreme level $u$, $X_i - u$, using the so-called POT methodology.

• The MGEV method takes into account the $k$ largest observations in the sample. This corresponds to take a random level, $X_{n-k:n}$, working next with the $k$ observations above that level, $X_{n-i+1:n}$, $1 \leq i \leq k$, or with the excesses $X_i - X_{n-k:n} | X_i > X_{n-k+1:n}$.

• This last approach is nowadays called peaks over random threshold (PORT) methodology, an acronymous coined in Araújo Santos, P., Fraga Alves, M.I. & Gomes, M.I. (2006), Revstat.
Let us consider the original sample, $X_1, \ldots, X_N$. Let $u$ be a high level. Let us denote by $N_u$, the number of exceedances (observations that exceed $u$). Let

$$Y_j := X_i - u | X_i > u, \quad j = 1, 2, \ldots, N_u \text{ (excesses)}.$$ 

We now need to estimate $\xi$ and $\sigma$ from the sample of excesses. It is again usual to consider the ML or the PWM methodologues.

**Remark 12.** The level $u$ is usually one of the top observations, i.e. $u = X_{n-k:n}$. In this case the ordered excesses are

$$Y_{j:k} := X_{n-k+j:n} - X_{n-k:n}, \quad j = 1, \ldots, k.$$ 

Choice of the level \( u \) (or of \( k \)).

- The choice of the level \( u \) is still an interesting topic of research.
- It is similar to the choice of \( k \), whenever we consider the random level \( X_{n-k:n} \), and the excesses over that level (the so-called PORT methodology).
- Such a choice deals with a trade-off between high values of \( u \), where the estimators’ bias is smaller, and small values of \( u \), where the variance is smaller.
- Davison & Smith (1990), *J. Royal Statist. Soc. B*, propose a choice of \( u \) based on the mean excess function, given by

\[
e(t) := \mathbb{E}[X - t|X > t] = \mathbb{E}[Y|Y > 0] = \frac{\sigma + \xi t}{1 - \xi}, \quad \text{if} \quad \xi < 1,
\]

in the GP model.
• In practice, we merely need to assess the *linearity* to the right of \( u \) for the plot of \( \hat{e}_n(t) := \sum_{j=1}^{N_t} y_j / N_t \), given the observed excesses \( y_1, \ldots, y_{N_t} \), i.e. we can consider

\[
\hat{e}_n(t) := \frac{\sum_{i=1}^{n} x_i I(t, +\infty)(x_i)}{\sum_{i=1}^{n} I(t, +\infty)(x_i)} - t.
\]

Alternatively, we can consider

\[
\hat{e}_n(x_{n-k:n}) = \frac{1}{k} \sum_{j=1}^{k} x_{n-j+1:n} - x_{n-k:n},
\]

for the choice of \( k \).
Estimation of other parameters of rare events under the POT approach

- For the GPD, extreme quantiles with an exceedance probability \( p \), i.e. \( U_{\text{GP}_\xi}(1/p) = \text{GP}_\xi^- (1 - p) \), are given by

\[
U_{\text{GP}_\xi}(1/p) = \begin{cases} 
\sigma(p^{-\xi} - 1)/\xi, & \text{if } \xi \neq 0, \\
-\sigma \log p, & \text{if } \xi = 0.
\end{cases}
\]

- For \( \xi < 0 \), the right endpoint of \( \text{GP}_\xi \) is finite and given by

\[
U_{\text{GP}_\xi}(\infty) = \text{GP}_\xi^-(1) = -\sigma/\xi.
\]

- We next merely need to replace the vector \((\xi, \sigma)\) of the unknown parameters by suitable estimates.
• If we want to infer on \( X \sim F \), we use the fact that the conditional distribution of the excesses above \( u \), given by

\[
F_u(y) = \mathbb{P}[X - u \leq y | X > u] = \frac{F(u + y) - F(u)}{1 - F(u)},
\]

can be approximated by the GPD, i.e. \( F_u(y) \approx \text{GP}_\xi(y/\sigma) \).

• Then, with \( x = u + y \), we get

\[
1 - F(x) = \{1 - F(u)\}[1 - F_u(x - u)],
\]

i.e.

\[
\overline{F}(x) \approx \overline{F}(u)[1 - \text{GP}_\xi((x - u)/\sigma)].
\]
To estimate the *probability of exceedance* of a high \( x \), \( F(x) \), we should consider the approximation

\[
F(x) \approx F(u) \left( 1 + \xi (x - u)/\sigma \right)^{-1/\xi}.
\]

With \( F(u) \) estimated by the relative frequency, \( N_u/n \), we get

\[
\hat{F}(x) = \frac{N_u}{n} \left( 1 + \hat{\xi} (x - u)/\hat{\sigma} \right)^{-1/\hat{\xi}}.
\]

An estimator of a *high quantile* of \( F \), \( U(1/p) = F^\leftarrow(1 - p) \), is

\[
\hat{U}(1/p) = u + \frac{\hat{\sigma}}{\hat{\xi}} \left( \left( \frac{np}{N_u} \right)^{-\hat{\xi}} - 1 \right)
\]

For \( \xi < 0 \), an estimator of the *right endpoint* of \( F \) is

\[
\hat{x}^F = \hat{U}(\infty) = u - \hat{\sigma}/\hat{\xi}.
\]
Data in ‘soa.txt’ and the POT method

• In this data set one can find large compensations, in a total amount of $N_u = 75789$ exceedances above the threshold $u = 25000$ USD (Beirlant et al., 2004).

• The histogram and box-and-whiskers plot of the log-data are next presented, showing a high right asymmetry.

soa.txt: Histogram (left) and box-and-whiskers plot (right) of log-data in ‘soa.txt’.
• On another side, the increasing behaviour of the ME-plot, presented in the following Figure, strongly suggests a RTF heavier than the Exponential RTF.
• Moreover, the convex shape of the Exponential QQ-plot associated with the excesses, \( y_i = x_i - u \), for \( u = 25000 \) USD, represented at the same Figure, right, provides the same information.

soa.txt: ME-plot (left) and Exponential QQ-plot (right) associated with ‘soa’ data
Despite of the ME-plot, we can work with another higher threshold, like $u = 400000$ USD. We then have $N_u = 397$ exceedances, just as illustrates in the following Figure.

\[\text{soa.txt: whole data basis and exceedances of } u = 400000 \text{ USD}\]
We next summarize the usual procedure associated with the POT method:

- Compute the observed values of the *excesses* \((X - u) | X > u\), i.e.
  \[
  \{y_i = x_i - u\}_{i=1}^{N_u} = \{y_i = x_i - 400000\}_{i=1}^{397}.
  \]

- Fit a GPD, \(GP_\xi(y/\sigma)\), \(\sigma = \sigma_u\), to the RV \((X - u) | X > u\)
  \[
  GP_\xi(y/\sigma) := 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi}.
  \]

- Obtain the QQ-plot associated with a standard GPD, \(GP_\xi(y)\), i.e. plot the points,
  \[
  \{(GP_\xi^{-1}(p_i), y_i: N_u) : i = 1, \ldots, N_u\}, \quad p_i := \frac{i}{N_u + 1}.
  \]
• Go on with a **preliminary estimation** of $\xi$, like the value of $\xi \in (-0.5, 0.5)$ that **maximizes the empirical correlation** between the observations $\{y_i\}_{i=1}^{Nu}$ and the line fitted to the **QQ-plot**. We have got $\hat{\xi} = 0.453717$.

• Obtain an estimate of $\delta$ on the basis of the **line fitted to the QQ-plot**, i.e. the line with slope $= \delta$ fitted to the points

$$\left\{(\text{GP}_{\xi}^- (p_i) = ((1 - p_i)^{-\xi} - 1)/\xi, \ y_i: Nu) : i = 1, \ldots, Nu\right\}$$

through the **least squares method**, obtained thus through the minimization of

$$h(\delta) := \sum_{i=1}^{Nu} \left(y_i: Nu - \delta q_i\right)^2, \ \text{with} \ q_i := ((1 - p_i)^{-\xi} - 1)/\xi.$$
• We then get

\[ \hat{\delta} = \frac{\sum_{i=1}^{N_u} y_i n q_i}{\sum_{i=1}^{N_u} q_i^2}. \]

• In the following Figure, we present the QQ-plot associated to the model \( GP_{\hat{\xi} = 0.453717} \), i.e. the points

\[ \{(GP_{\hat{\xi}}(p_i := i/(N_u + 1)), y_i; N_u), i = 1, \ldots, N_u\}. \]

• The fitting of the least squares line, provides the estimate \( \hat{\delta} = 126788 \), to which corresponds a sample correlation \( r_Q = 0.9932243 \).
If we want to estimate the 'the probability that a future compensations (exceeding the 25000 USD) exceeds the observed maximum', the reasoning is the following:
Recalling that given $X \sim F$, the conditional distribution of the excesses above $u$ is

$$F_u(y) = \mathbb{P}[X - u \leq y|X > u] = \frac{F(u + y) - F(u)}{1 - F(u)}, \quad 0 \leq y \leq x^F - u,$$

and equivalently,

$$1 - F(u + y) = \{1 - F(u)\}[1 - F_u(y)],$$

being $y = x_{n:n} - u = y_{Nu:Nu} = 4518420 - 400000 = 4118420$, the observed maximum of the excesses above 400000 USD,

$$\mathbb{P}[X > x_{n:n}] = 1 - F(x_{n:n}) = 1 - F(u + y_{Nu:Nu}) = \{1 - F(u)\}[1 - F_u(y_{Nu:Nu})].$$
Using the GP approximation to the underlying conditional distribution of excesses, \( F_u(\cdot) \approx \text{GP} \hat{\xi}(\cdot/\hat{\delta}) \),

\[
F_u(y_{Nu:Nu}) \simeq \text{GP} \hat{\xi}(y_{Nu:Nu}/\hat{\delta}) = \text{GP}0.454(y397:397, 126788) = 0.998.
\]

On another side,

\[
1 - \overline{F}(u) = \frac{N_u}{n} = 397/75789 = 0.005,
\]

and consequently,

\[
\mathbb{P}[X > x_n:n] \simeq \frac{N_u}{n} \left[ 1 - \text{GP} \hat{\xi}(y_{Nu:Nu}, \hat{\delta}) \right] = 1.205e - 05.
\]
Let us now consider the **ML** estimation of the parameters associated with the **POT** methodology applied to the $k = 75789$ values above $u = 25000$ USD (Beirlant et al., 2004).

**ML estimation:**
We have used the "**library(ismev)**" (gpd.fit(x,threshold=u)) and the "**library(evir)**" (gpd(x,u, method = "ml"))

"**library(ismev)**": $\hat{\delta} = 142558$, $\hat{\xi} = 0.3819$ and $L^*(\hat{\xi}, \hat{\sigma}) = -5260$.

"**library(evir)**": $\hat{\delta} = 142147$, $\hat{\xi} = 0.3847$ and $L^*(\hat{\xi}, \hat{\sigma}) = -5260$. 
We have further built the **W-plot**, using these later **ML** estimates, \( \hat{\delta} = 142147 \) and \( \hat{\xi} = 0.3847 \), representing the points,

\[
\left( -\log(1 - p_i), -\log(1 - \text{GP}_{\hat{\xi}}(y_i; N_u, \hat{\delta})) \right) \quad i = 1, 2, \ldots, N_u.
\]

Just as can be seen in the following Figure such points are reasonably close to the **diagonal**.
Regarding the **EVI-estimation** the following Figure illustrates the behaviour of the **ML** and **PWM EVI-estimates** versus $k < 500$, with $k$ the number of **OS’s** above $u$.

soa.txt: **ML** and **PWM EVI-estimates** versus $k$, the number of **OS’s** above $u$
soa.txt: **ML** estimation of the **EVI** vs $k$ and 95% confidence intervals (CI’s)
- Under the **POT** framework, another relevant problem is the estimation of *extreme quantiles*.
- The following Figure illustrates the sample path of the estimates of $U(100,000)$ versus $k < 500$, and under a **POT** framework.

**soa.txt**: Estimates of $U(100,000)$ versus $k < 500$
• For the estimation of $U(100\,000)$ we have used the ML estimates in the Equation of $\hat{U}(1/p)$ for $p = 1/100\,000$, and compute them for $k < 500$.
• For those values of $k$ and in the region $(200, 500)$, there is a reasonable stability around the value $4\,000\,000$.

Remark 13. Alternatively, the ML estimates of high quantiles, $\hat{U}(1/p)$, can be directly obtained through a re-parameterization of the GP log-likelihood as a function of $U(1/p)$. For instance, we can consider that

$$\sigma = \frac{\xi (U(1/p) - u)}{(np/Nu)^{-\xi} - 1}.$$
Summary of parametric approaches and a link to semi-parametric frameworks

• More recently, the LOB and the POT methodologies have been considered under a semi-parametric framework.

• There is then NO fitting of a specific parametric model, dependent upon a location parameter $\lambda$, a scale parameter $\delta$ and a ‘shape’ parameter $\xi$.

• It is merely assumed that $F \in \mathcal{D}_M(\text{GEV}_\xi)$, with $\text{GEV}_\xi$ the EVD, $\xi$ being the unique primary parameter of extreme events to be initially estimated, on the basis of a few top observations, and according to adequate methodology, to be dealt with later on.

• We now summarise the most relevant approaches to SUE:
1. **Parametric approaches:**

   **I** The *univariate EV model* (for the $k$ maximum values of sub-samples of size $n$, $N = n \times k$) (*Gumbel's classical approach* or BM method).

   **II** The *MEV model* or *LOB method* (for the $k$ top OS's associated with the original sample of size $N$).

   **III** The *multi-dimensional EV model* (*MEV model for the* $i_j$ top observations, $j = 1, 2, \ldots, m$, in sub-samples of size $n$, $m \times n = N$).

   $m = k$ and $i_j = 1$ for $1 \leq j \leq k$ originates **I**;

   $m = 1$ ($n = N$) and $i_1 = k$ originates **II**.

   **IV** The *Paretian model* for the excesses, $X_j - u$, given $X_j > u$, $1 \leq j \leq k$, of a high deterministic threshold $u$, suitably chosen [POT approach].

   **V** *Bayesian* approaches.
2. Semi-parametric approaches:

VI Under these approaches we work with

- the $k$ top OS's associated with all $N$ observations
- or with the excesses over a high deterministic or random threshold,

assuming only that the model $F$ underlying the data is in $\mathcal{D}_M(\text{GEV}_\xi)$ or in specific sub-domains of $\mathcal{D}_M(\text{GEV}_\xi)$, with $\text{GEV}_\xi(\cdot)$ the EVD. This case will be briefly considered in the following Section, but can be seen with all detail in de Haan, L. & Ferreira, A. (2006), *Extreme Value Theory: an Introduction*, Springer Science+Business Media, LLC, New York.
Estimation under parametric frameworks

- We have now several ‘R-Packages for Extreme Values’, such as evd, evdbayes, evir, evt0, extRemes, extremevalues, fExtremes, ismev, lmom, lmomRFA, lmomco, POT and SpatialExtremes, among others, that can help us in most of the inferential procedures related to:
  
  - Gumbel's approach or BM method
  - MEV and MDEV approaches
  - The POT approach

- The BM method has been recently revisited in Dombry (2013), [arXiv:1301.5611] and Ferreira & de Haan (2013), [arxiv.org/pdf/1310.3222].
6. SEMI-PARAMETRIC SUE

- As mentioned above, in a semi-parametric framework we do not consider any parametric model underlying the available sample, \((X_1, \ldots, X_n)\), with unknown underlying \(F(.)\).
- We merely assume that
  1. \(F \in \mathcal{D}_M(\text{GEV}_\xi)\), for some \(\xi \in \mathbb{R}\),
  2. and consider the \(k\) top observations

\[
X_{n:n} \geq X_{n-1:n} \geq \cdots \geq X_{n-k+1:n},
\]

above the random threshold \(X_{n-k:n}\), which needs to be an upper intermediate OS, i.e. \(k \equiv k_n\) is such that

\[
k \to \infty \quad \text{and} \quad k/n \to 0, \quad \text{as} \quad n \to \infty.
\]
First and second-order frameworks

- We further use the notations
  - \( F^{-}(y) := \inf \{ x : F(x) \geq y \} \), \( y \in (0,1) \), for the generalized inverse function of \( F \),
  - \( U(t) := F^{-}(1 - 1/t) \), \( t > 1 \), for the RTQF, already given in **Definition** 3, and
  - \( R_{a} \) for the class of *regularly varying* functions at infinity with an index of regular variation \( a \), i.e. positive Borel measurable functions \( g(\cdot) \) such that \( g(tx)/g(t) \to x^{a} \), as \( t \to \infty \), for all \( x > 0 \) (in *Bingham et al.*, 1987, *Regular Variation*. Cambridge Univ. Press, we find details on regular variation).
• A necessary and sufficient condition for \( F \in \mathcal{D}_M(\text{GEV}_\xi) \), provided in de Haan (1984), is the so-called first-order condition:

\[
F \in \mathcal{D}_M(\text{GEV}_\xi) \iff \lim_{t \to \infty} \frac{U(tx) - U(t)}{a(t)} = \varphi_\xi(x)
\]

for all \( x > 0 \), where \( a \) is a positive measurable function, and

\[
\varphi_\xi(x) =: \begin{cases} 
\frac{x^{\xi-1}}{\xi}, & \text{if } \xi \neq 0, \\
\ln x, & \text{if } \xi = 0.
\end{cases}
\]

• For instance, in the Fréchet MDA, we can choose \( a(t) = \xi U(t) \) and thus \( F \in \mathcal{D}_M(\Phi_{\xi>0}) \) iff \( \lim_{t \to \infty} U(tx)/U(t) = x^\xi \), \( x > 0 \), i.e.

\[
F \in \mathcal{D}_M(\Phi_{\xi>0}) \iff U \in \mathcal{R}_\xi \iff \overline{F} \in \mathcal{R}_{-1/\xi}.
\]
But a first-order condition is in general not sufficient to study non-degenerate properties of tail parameters' estimators. In that case a second-order condition is required specifying the rate of convergence in the first-order condition. Different types of conditions exist, some expressed in terms of either $\overline{F}$ or $U$ or $\ln U$, like

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} - \varphi_\xi(x) = \begin{cases} 
\frac{1}{\rho} \left( \frac{x^\rho - 1}{\rho} - \ln x \right), & \text{if } \xi = 0, \rho \neq 0, \\
\frac{1}{\xi} \left( x^\xi \ln x - \frac{x^\xi - 1}{\xi} \right), & \text{if } \xi \neq 0, \rho = 0, \\
\ln^2 \frac{x}{2}, & \text{if } \xi = \rho = 0, \\
\frac{1}{\rho} \left( \frac{x^{\xi + \rho} - 1}{\xi + \rho} - \frac{x^\xi - 1}{\xi} \right), & \text{otherwise,}
\end{cases}$$

for all $x > 0$, where $\rho \leq 0$ is a second-order parameter, $A$ is a function possibly not changing in sign and tending to 0 as $t \rightarrow \infty$, such that $|A| \in \mathcal{R}_\rho$ [de Haan & Stadtmüller, 1996, JAMS-A].
Semi-parametric EVI-estimation

Classical Estimators

Hill (H) estimator.

- The most famous estimator of $\xi > 0$ is the Hill (H) (Hill, 1975, [Ann. Statist.]) EVI-estimator,

$$\hat{\xi}_{H,k,n} = \frac{1}{k} \sum_{i=1}^{k} \log X_{n-i+1:n} - \log X_{n-k:n}.$$  

- One of the interesting facts concerning the H EVI-estimator is that various asymptotically equivalent versions of $\hat{\xi}_{H,k,n}$ can be derived through essentially different methods (such as the ML method or the mean excess function approach), showing that the Hill estimator is very natural.
Kernel estimator.

- A general class of kernel estimators, also restricted to the case $\xi > 0$, was given in Csörgő, Deheuvels and Mason (1985), \[Ann. Statist.\],

$$\hat{\xi}_{k,n}^K = \frac{\sum_{j=1}^{k} \frac{1}{k} K\left(\frac{j}{k}\right) \left(\log X_{n-j+1:n} - \log X_{n-j:n}\right)}{\frac{1}{k} \sum_{j=1}^{k} K\left(\frac{j}{k}\right)}.$$

- The H EVI-estimator is a member of this class, since it can be obtained by taking a uniform kernel $K(x)$.

- The kernel estimators can be interpreted as weighted least squares regression estimators of the slope of the Pareto quantile plot in case one considers regression lines passing through a fixed anchor point.

- Kernel EVI-estimators for a real $\xi$ have been studied by Groeneboom, Lopuhaä and de Wolf, (2003), \[Ann. Statist.\].
• **Pickands estimator.** A simple estimator for the general case, $\xi \in \mathbb{R}$, is Pickands’ estimator (Pickands III, J., 1975)

$$\hat{\xi}_{k,n}^P = \left( \frac{1}{\log 2} \right) \frac{X_{n-\lfloor k/4 \rfloor:n} - X_{n-\lfloor k/2 \rfloor:n}}{X_{n-\lfloor k/2 \rfloor:n} - X_{n-k:n}},$$

where $\lfloor x \rfloor$ denotes the integer part of $x$.

• The weak consistency of this **EVI**-estimator was first proved in Pickands (1975), for any $\xi \in \mathbb{R}$. A more detailed study of this estimator, as an estimator based on intermediate and extreme statistics, can be seen in Dekkers, A.L.M. & de Haan, L. (1989), *Ann. Statist.*
• **Moment (M) estimator.** Let us define

\[
M_{k,n}^{(\ell)} := \frac{1}{k} \sum_{i=1}^{k} (\log X_{n-i+1:n} - \log X_{n-k:n})^\ell, \quad \ell = 1, 2,
\]

and

\[
\hat{\xi}^+_{k,n} = M_{k,n}^{(1)} = \xi^H_{k,n}, \quad \hat{\xi}^-_{k,n} = 1 - \frac{1}{2} \left\{ 1 - \frac{(M_{k,n}^{(1)})^2}{M_{k,n}^{(2)}} \right\}^{-1}.
\]

For any real \( \xi = \xi_+ + \xi_- \), with \( \xi_- := \min(0, \xi) \), \( \xi_+ := \max(0, \xi) \), the moment **EVI**-estimator (Dekkers, A.L.M., Einmahl, J.H.J. & de Haan, L., 1989, *Ann. Statist.* ) is given by

\[
\hat{\xi}^M_{k,n} = \hat{\xi}^-_{k,n} + \hat{\xi}^+_{k,n} \equiv M_{k,n}^{(1)} + \frac{1}{2} \left\{ 1 - \left( M_{k,n}^{(2)}/[M_{k,n}^{(1)}]^2 - 1 \right)^{-1} \right\}.
\]
Generalized Hill estimator.

- For $\xi \in \mathbb{R}$, Beirlant, Vynckier & Teugels (1996), $[\text{Bernoulli}]$, proposed an EVI-estimator based on the slope of the generalized quantile plot,

\[
\left( \log \frac{n+1}{j}, \log UH_{j,n} \right), \quad j = 1, \ldots, n, \quad UH_{j,n} = X_{n-j:n} \hat{\xi}_{j,n}.
\]

This leads to the generalized Hill (GH) estimator defined as

\[
\hat{\xi}_{GH}^{k,n} = \frac{1}{k} \sum_{j=1}^{k} \log UH_{j,n} - \log UH_{k+1,n},
\]

further studied in Beirlant Dierckx & Guillou, (2005), $[\text{Bernoulli}]$, and asymptotically equivalent to

\[
\hat{\xi}_{GH}^{*} := H_{k,n} + \frac{1}{k} \sum_{i=1}^{k} \left\{ \ln H_{i,n} - \ln H_{k,n} \right\}.
\]
ML estimator.

- Conditionally on the OS, \( X_{n-k:n} \), \( k \) intermediate, the excesses, \( X_{n-i+1:n} - X_{n-k:n}, 1 \leq i \leq k \), are approximately the \( k \) top OS’s associated with a sample of size \( k \) from GP\(_\xi(\alpha x/\xi)\), \( \xi, \alpha \in \mathbb{R} \), with GP\(_\xi(x)\) the GPD.

- The solution of the ML equations associated with the above mentioned parameterization (Davison, 1984) gives rise to an explicit EVI-estimator, usually called the ML EVI-estimator, that can be named PORT-ML, after Araújo Santos et al. (2006).

- Such an EVI-estimator is given by

\[
\hat{\xi}_{n,k}^{\text{ML}} := \frac{1}{k} \sum_{i=1}^{k} \ln(1 + \hat{\alpha}[X_{n-i+1:n} - X_{n-k:n}]),
\]

where \( \hat{\alpha} \) is the implicit ML estimator of the unknown ‘scale’ parameter \( \alpha \).
Probability weighted moment (PWM) estimators.

- The parametric PWM method, introduced by Landwehr, Matalas & Wallis (1979), [Water Resources Research] and Greenwood, Landwehr, Matalas & Wallis (1979), [Water Resources Research], is an interesting alternative to the ML approach, for small samples.
- The main idea of this method is to match the moments $E[X^p(F(X))^r(1 - F(X))^s]$, with $p, r$ and $s$ real numbers, with their empirical versions, similarly to the classical method of moments.
- For the GEVD, Hosking, Wallis & Wood (1985), [Technometrics], show that $E[X(F'(X))^r]$ can be explicitly computed, which leads to the PWM estimation of the parameters under play.
For the GPD, the PWM EVI-estimators were studied by Hosking & Wallis (1987), [Technometrics], and are quite common in the most diverse applied fields.

The parametric PWM method can also be devised under a semi-parametric framework.

De Haan & Ferreira (2006), Extreme Value Theory: an Introduction, considered the semi-parametric generalized Pareto PWM (GP-PWM) EVI-estimator,

\[
\hat{\xi}_{GPPWM}^{k,n} := \frac{1 - 2\hat{a}_1^*(k)}{\left(\hat{a}_0^*(k) - 2\hat{a}_1^*(k)\right)},
\]

\[
\hat{a}_\ell^*(k) := \frac{1}{k} \sum_{i=1}^{k} \left(\frac{i-1}{k-1}\right)^\ell (X_{n-i+1:n} - X_{n-k:n}), \quad \ell = 0, 1.
\]

See also Diebolt, Guillou, Naveau & Ribereau (2008), [Revstat].
In Caeiro & Gomes (2011), [J. Statist. Planning & Infer.], were introduced and studied Pareto PWM (PPWM) EVI-estimators,

\[ \hat{\xi}_{k,n}^{PPWM} := 1 - \hat{a}_1(k)/(\hat{a}_0(k) - \hat{a}_1(k)), \]

with

\[ \hat{a}_\ell(k) := \sum_{i=1}^{k} \left( \frac{i - 1}{k - 1} \right)^\ell X_{n-i+1:n}/k, \quad \ell = 0, 1. \]
**Mixed-Moment (MM) estimator.**

- We further refer the so-called MM EVI-estimator (Fraga Alves, Gomes, de Haan & Neves, 2009, *Extremes*), valid for all $\xi \in \mathbb{R}$, and given by

\[
\hat{\xi}_{n}^{\text{MM}}(k) := \frac{\hat{\varphi}_{k,n} - 1}{1 + 2 \min\left(\hat{\varphi}_{k,n} - 1, 0\right)}, \quad \text{with} \quad \hat{\varphi}_{k,n} := \frac{\hat{\xi}_{k,n}^{H} - L_{k,n}^{(1)}}{L_{k,n}^{(1)}},
\]

where $\hat{\xi}_{k,n}^{H}$ is the H EVI-estimator, and

\[
L_{k,n}^{(1)} := 1 - \frac{1}{k} \sum_{i=1}^{k} \frac{X_{n-k:n}}{X_{n-i+1:n}}.
\]

- The MM EVI-estimator has a very simple form and is asymptotically very close to the ML EVI-estimator for a wide class of heavy-tailed models.
The mean-of-order-$p$ (MOP) EVI-estimator.

- A competitive generalization of the Hill estimator has been recently introduced in the literature. Note that we can write

$$\hat{\xi}_{k,n}^H = \sum_{i=1}^{k} \ln \left( \frac{X_{n-i+1:n}}{X_{n-k:n}} \right)^{1/k} = \ln \left( \prod_{i=1}^{k} \frac{X_{n-i+1:n}}{X_{n-k:n}} \right)^{1/k}, \ 1 \leq k < n.$$

The H EVI-estimator is thus the logarithm of the geometric mean (or mean-of-order-0) of $U := \{U_{ik}\}$, with $U_{ik} := X_{n-i+1:n}/X_{n-k:n}$, $1 \leq i \leq k < n$.

- More generally, Brilhante, Gomes & Pestana (2013), [Comp. Statist. & Data Anal.], considered as basic statistics the MOP of $U$, with $p \geq 0$, i.e., the class of statistics
\[ A_p(k) = \begin{cases} \left( \frac{1}{k} \sum_{i=1}^{k} U_{ik}^p \right)^{1/p}, & \text{if } p > 0, \\ \left( \prod_{i=1}^{k} U_{ik} \right)^{1/k}, & \text{if } p = 0. \end{cases} \]

- The class of **MOP EVI-estimators** has then the functional form,

\[
\hat{\xi}^H_{k,n} := \begin{cases} 
\frac{1 - A_p^{-p}(k)}{p}, & \text{if } 0 < p < 1/\xi, \\
\ln A_0(k), & \text{if } p = 0,
\end{cases}
\]

with \( \hat{\xi}^H_{k,n} \equiv \hat{\xi}^H_{k,n} \), the **H EVI-estimator**.

The PORT EVI-estimators.

- Apart from Pickands and ML’s estimators, all aforementioned EVI-estimators are scale invariant but not location-invariant.
- And particularly the Hill estimator can suffer drastic changes when we induce a shift in the data, given rise to the so-called ‘Hill horror-plots’, a terminology used in Resnick (1997), [Ann. Statist].
- This led Araújo Santos et al. (2006), [Revstat], to introduce the already referred PORT methodology. The estimators are then functionals of a sample of excesses over a random level $X_{nq:n}$, $n_q := \lceil nq \rceil + 1$, i.e. functionals of the sample,

$$X_{n}^{(q)} := (X_{n:n} - X_{nq:n}, \ldots, X_{nq+1:n} - X_{nq:n}).$$
• Generally, we can have $0 < q < 1$, for any $F \in \mathcal{D}_M(\text{GEV}_\xi)$ (the random level is an empirical quantile). If the underlying model $F$ has a finite left endpoint, $x_F := \inf \{x : F(x) \geq 0\}$, we can also use $q = 0$ (the random level can then be the minimum).

• If we think on H EVI-estimators, the new classes of PORT-H EVI-estimators, theoretically studied in Araújo Santos et al. (2006), and for finite samples in Gomes et al. (2008b), are given by

$$\hat{\xi}_{k,n}(q) := \frac{1}{k} \sum_{i=1}^{k} \left\{ \ln \frac{X_{n-i+1:n} - X_{nq:n}}{X_{n-k:n} - X_{nq:n}} \right\}, \quad 0 \leq q < 1.$$ 

• And a similar dependence on this extra tuning parameter $q$ can be conceived for any other EVI-estimator. For further details, see Gomes, Henriques-Rodrigues & Miranda (2011), [CSSC] and Gomes, Henriques-Rodrigues, Fraga Alves & Manjunath (2013a), [JSSC].
Main asymptotic properties

• In order to obtain weak consistency of all these estimators, we only need to assume that $F \in \mathcal{D}_M(EV_{\xi})$, for the adequate $\xi$-range, and that $k$ is an intermediate sequence.

• However, if we want to obtain asymptotic normality, it is sensible to strengthen this condition into a second-order one, just as given. Then, assuming that $\sqrt{k} A(n/k) \to \lambda < \infty$, we can expand each estimator $\hat{\xi}^*_{k,n}$ as follows

$$\hat{\xi}^*_{k,n} - \xi \overset{d}{=} \frac{\sigma^* N(0, 1)}{\sqrt{k}} + b^* A\left(\frac{n}{k}\right) + o_P\left(A\left(\frac{n}{k}\right)\right),$$

where $N(\mu, \sigma^2)$ denotes a normal RV with mean value $\mu$ and variance $\sigma^2$, and $(\sigma^*, b^*) \in \mathbb{R}^+ \times \mathbb{R}$. Thus,

$$\sqrt{k}(\hat{\xi}^*_{k,n} - \xi) \overset{d}{\to} N(\lambda b^*, \sigma^{*2}).$$
Data in ‘soa.txt’ and semi-parametric SUE

- In the following Figure we present the Hill (left), the moment (center) and the Pickands (right) EVI-estimates, again as a function of $k$.

Remark 14. **The largest variance is clearly associated with Pickands’ estimator.**
A few comments

• It is possible to see from the previous figures that when we take 200 excesses in the POT methodology, the results are similar to the ones we get in most of the semi-parametric approaches.
• It is worth mentioning that the difficulties in the choice of a high threshold in the POT methodology or the number $k$ of largest observations to be taken either in the LOB method or in the semi-parametric approach are quite similar.
• The two approaches, parametric and semi-parametric, need to be considered as complementary for any inference related to parameters of extreme events.
• The choice of a high threshold is not trivial, and despite of the high number of theoretical and heuristic contributions to the topic, a lot still needs to be developed.
Reduced-bias estimators

- The classical EVI-estimators are asymptotically unbiased only for adequate values $k = k_n$.
- Generally,
  - for small values of $k$, or equivalently high thresholds, these EVI-estimators have a high variance,
  - and for large values of $k$, or equivalently low thresholds, the bias is quite high.
- The adequate accommodation of this bias has been extensively addressed since the late nineties.

Second-order reduced-bias (SORB) EVI-estimators

- We mention the pioneering papers by Peng (1998), Beirlant et al. (1999), Feuerverger and Hall (1999), and Gomes et al. (2000), for Pareto-type distributions ($\xi > 0$).
• In these papers, authors are led to propose **SORB EVI-estimators**, with asymptotic variances larger than or equal to 
\( (\xi (1 - \rho)/\rho)^2 \), where \( \rho(\cdot < 0) \) is the aforementioned ‘shape’ second-order parameter.

• Similar results in the general **MDA**, \( D_M(\text{EV}_\xi) \), can be found in Beirlant et al. (2005), and more recently in Cai et al. (2013), who introduced a **SORB EVI-estimator** for \( \xi \) around zero, based on the **PWM** methodology.

• For those estimators, the **asymptotic mean** is **zero**, instead of the value \( \lambda b^* \), whatever the value of \( \lambda \).

• **However**, the **asymptotic variance increases** compared to \( \sigma^*2 \).
Minimum-variance reduced-bias (MVRB) EVI-estimators

- Caeiro et al. (2005), Gomes et al. (2007; 2008) and Caeiro et al. (2009) have been able
  - to reduce the asymptotic bias
  - without increasing the asymptotic variance, kept at $\xi^2$, as happens with the H EVI-estimators.

- These MVRB EVI-estimators, are all based on an adequate ‘external’ and a bit more than consistent estimation of the pair of second-order parameters, $(\beta, \rho) \in (\mathbb{R}, \mathbb{R}^-)$ in $A(t) = \xi \beta t^\rho$, done through consistent estimators, denoted by $({\hat{\beta}, \hat{\rho}})$, such that $\hat{\rho} - \rho = o_p(1/\ln{n})$, and outperform the classical estimators for all $k$. 
• Different algorithms for the estimation of \((\beta, \rho)\) can be found in Gomes & Pestana (2007) [J. Amer. Statist. Assoc.], among others.
• Among the most common **MVRB EVI-estimators**, we just mention the simplest class in Caeiro et al. (2005),

\[
\hat{\xi}_{k,n}^H(\hat{\beta}, \hat{\rho}) := \hat{\xi}_{k,n}^H \left( 1 - \hat{\beta} \left( \frac{n}{k} \right) \frac{\hat{\rho}}{1 - \hat{\rho}} \right),
\]

with \(\hat{\xi}_{k,n}^H\) the **H EVI-estimator**, and where \((\hat{\beta}, \hat{\rho})\) is an adequate consistent estimator of the second-order parameters \((\beta, \rho)\).
• Note that this **CH-MVRB EVI-estimator** is easily justified by the fact that \(b^H = 1/(1 - \rho)\) and \(A(t) = \xi \beta t^\rho\).
• For recent overviews on reduced-bias **EVI-estimation**, see Chapter 6 in Reiss & Thomas (2007), Gomes, Canto e Castro, Fraga Alves & Pestana (2008), [Extremes], Beirlant, Caeiro & Gomes (2012), [Revstat] and Gomes and Guillou (2014), [Intern. Statist. Review].
Semi-parametric estimation of other parameters

Classical semi-parametric estimation

- Despite the fact that the estimation of
  - quantiles,
  - return periods of high levels,
  - exceedance probabilities and
  - coefficient of tail dependence,
among other parameters of extreme events, are at least as important in applications as the estimation of the EVI, we shall only briefly refer the estimation of high quantiles, the so-called Value-at-Risk (VaR) in the areas of insurance and finance.
High quantile or VaR-estimation.

• In a semi-parametric framework, the most usual estimators of a quantile $\text{VaR}_p = \chi_{1-p} := U(1/p)$, with $p$ small, can be easily derived through the approximation

$$U(tx) \approx U(t) + a(t)(x^\xi - 1)/\xi.$$  

The fact that $X_{n-k+1:n} \stackrel{p}{\sim} U(n/k)$ enables us to estimate $\chi_{1-p}$ on the basis of this approximation and adequate estimates of $\xi$ and $a(n/k)$.

• For the simpler case of heavy tails, the approximation is

$$U(tx) \approx U(t)x^\xi,$$

and, with $\hat{\xi}_{k,n}$ is any consistent semi-parametric EVI-estimator, we get
\[ \hat{\chi}_{1-p,k,n} := X_{n-k:n} \left( k/(np) \right) \hat{\xi}_{k,n}, \]

an estimator of the type of the one introduced by Weissman (1978).

- Details on semi-parametric estimation of extremely high quantiles for \( \xi \in \mathbb{R} \), can be found in Dekkers & de Haan (1989), [Ann. Statist.], de Haan & Rootzén (1993), [J. Statist. Planning & Infer.], Ferreira et al. (2003), [Statistics], and more recently in de Haan & Ferreira (2006).

- Fraga Alves et al. (2009) also provide, jointly with the MM-estimator, accompanying shift and scale estimators that make high quantile estimation almost straightforward.

- Other approaches to high quantile estimation can be found in Matthys & Beirlant (2003), [Statist. Sinica].
• However, none of the aforementioned quantile estimators reacts adequately to a shift of the data, enjoying the empirical counterpart of the theoretical linearity of any quantile $\chi_\alpha$, $\chi_\alpha(\delta X + \lambda) = \delta \chi_\alpha(X) + \lambda$, for any real $\lambda$ and positive $\delta$.

• Araújo Santos et al. (2006) provide a class of semi-parametric $\text{VaR}_p$ estimators which enjoy such a feature. Such class of estimators is based on the PORT methodology, providing exact properties for risk measures in finance: translation-equivariance and positive homogeneity.

• For Pareto-type models and a H EVI-estimation, one has

$$\hat{\chi}_{1-p,k,n}^{(q)} := (X_{n-k:n} - X_{nq:n})(k/(np))^{\hat{\xi}_{k,n}^{(q)}} + X_{nq:n},$$

with $\hat{\xi}_{k,n}^{(q)}$ the PORT-H EVI-estimator.
**SORB semi-parametric quantile estimation**

- **Reduced-bias quantile estimators** have been studied in Matthys, Delafosse, Guillou & Beirlant (2004), [Insurance Math. Econom.] and Gomes & Figueiredo (2006), [Test], who consider the classical SORB EVI-estimators.


- References also need to be done to Diebolt, Gardes, Girard & Guillou (2008), [J. Statist. Planning & Infer.], Beirlant, Joossens & Segers (2009), [J. Statist. Planning & Infer.], Caeiro & Gomes (2009), [Test] and Li, Peng & Qi (2011), [Test].
Dependent frameworks—a brief reference to the EI-estimation

- The same GEVD appears as the limiting CDF of the maximum for a large class of stationary sequences, like, for instance, the ones for which the mixing condition $D$, introduced in Leadbetter, Lindgren & Rootzén (1983), *Extremes and Related Properties of Random Sequences and Processes*, holds.

- Let us assume we have data from a stationary process, $\{X_n\}_{n \geq 1}$, with an underlying CDF, $F$, and let $\{Y_n\}_{n \geq 1}$ be the associated IID sequence (from the same model $F$).

- Under adequate local dependence conditions, the limiting CDF of the maximum $X_{n:n}$ can be directly related to the maximum, $Y_{n:n}$, of the IID associated sequence, through a new parameter, the so-called extremal index (EI).
• More specifically, the stationary sequence \(\{X_n\}_{n \geq 1}\) has an EI, \(\theta\) \((0 < \theta \leq 1)\), if, for every \(\tau > 0\), we may find a sequence of levels \(u_n = u_n(\tau)\) such that

\[
\mathbb{P}(Y_n: n \leq u_n) \xrightarrow{n \to \infty} e^{-\tau} \quad \text{and} \quad \mathbb{P}(X_n: n \leq u_n) \xrightarrow{n \to \infty} e^{-\theta \tau}.
\]

• The extremal index \(\theta\) may thus be informally defined by the approximation

\[
\mathbb{P}[X_n: n \leq x] \approx F^n \theta(x) \approx \text{GEV}_\xi \left( \frac{x - b'_n}{a'_n} \right), \quad \begin{cases} 
a'_n = a_n \theta^\xi 
b'_n = b_n + a_n \frac{\theta^\xi - 1}{\xi} \end{cases},
\]

where \(F(\cdot)\) is the marginal CDF of a strictly stationary sequence \(\{X_n\}_{n \geq 1}\), satisfying adequate local and asymptotic dependence conditions.
• One of the **local dependence conditions** which enable us to guarantee the existence of an **EI** is the \( D'' \) condition, introduced by Leadbetter & Nandagopalan (1989).

• Under the validity of such a condition, the **EI** can also be defined as the **reciprocal** of the ‘**mean time of duration of extreme events**’, being directly related to the **exceedances of high levels**.

• Indeed, we have

\[
\theta = \frac{1}{\text{limiting mean size of clusters}} = \lim_{n \to \infty} \mathbb{P}(X_2 \leq u_n | X_1 > u_n),
\]

where \( u_n \) is a sequence of values such that

\[
F(u_n) = 1 - \frac{\tau}{n} + o(1/n), \quad \text{as } n \to \infty.
\]
• The *auto-regressive for maxima* (ARMAX) processes will be used here for illustration. Such processes are based on an IID sequence of innovations \( \{Z_i\}_{i \geq 1} \), with CDF, \( H \), and are defined through the relation,

\[
X_i = \beta \max(X_{i-1}, Z_i), \quad i \geq 1, \quad 0 < \beta < 1.
\]

• The ARMAX sequence has a stationary distribution \( F \), dependent on \( H \) through the relation \( F(\beta x)/F(x) = H(x) \) [Alpuim, 1989, *J. Appl. Probab*].

• Conditions D and D” hold for these sequences and stationary ARMAX sequences may possess an extremal index \( \theta < 1 \).

• For illustration, we shall consider ARMAX processes with Fréchet innovations. If \( H(x) = \Phi_{\xi}^{-1/\xi-1}(x) \), \( F(x) = \Phi_{\xi}(x) = \exp\left(-x^{-1/\xi}\right), \ x \geq 0 \), and \( \theta = 1 - \beta^{1/\xi} \).
Notice the richness of these processes, regarding clustering of exceedances. Note also that there is a “shrinkage” of maximum values, together with the exhibition of larger and larger clusters’ of exceedances of high values, as $\theta$ decreases.
**Classical EI-estimation**

- Then, given a sample \((X_1, \ldots, X_n)\), an obvious non-parametric estimator of \(\theta\) is immediately suggested: once a suitable threshold \(u\) is chosen, put

\[
\theta_n^{N} = \theta_{n,u}^N := \frac{\sum_{j=1}^{n-1} I\{X_j > u, X_{j+1} \leq u\}}{\sum_{j=1}^{n} I\{X_j > u\}} = \frac{\sum_{j=1}^{n-1} I\{X_j \leq u < X_{j+1}\}}{\sum_{j=1}^{n} I\{X_j > u\}}.
\]

- In order to have consistency of this estimator the high level \(u = u_n\) must be such that

\[
n(1 - F(u_n)) = c_n \tau = \tau_n, \quad \tau_n \to \infty \quad \text{and} \quad \tau_n/n \to 0
\]

To make the semi-parametric **EI-estimation** closer to the semi-parametric **EVI-estimation**, we consider, just as in [Gomes, Hall & Miranda, 2008, *Comp. Statist. Data Anal.*], \( u \in [X_{n-k:n}, X_{n-k+1:n}] \) and the **upcrossing (UC) EI-estimator**

\[
\hat{\theta}_{n}^{\text{UC}}(k) \equiv \hat{\theta}_{k,n}^{\text{UC}} := \frac{1}{k} \sum_{j=1}^{n-1} I[X_j \leq X_{n-k:n} < X_{j+1}].
\]

- Again, **consistency** is attained only if \( k \) is intermediate.
- **Reduced-bias** versions of the aforementioned **UC EI-estimator** have been obtained in the above paper by Gomes, Hall & Miranda (2008), and will be briefly mentioned later on.
- **Limit theorems** for empirical processes of cluster functionals have been obtained by Drees & Rootzén (2010), [Ann. Statist].
• For ARMAX processes, and also for IID data ($\theta = 1$), we get

\[
\mathbb{E} \left[ \hat{\theta}_n^N (k) \right] = \theta - \left( \frac{\theta (\theta + 1)}{2} \left( \frac{k}{n} \right) - \frac{3 - 2 \theta}{2} \right) \left( 1 + o(1) \right).
\]

• We shall next assume that, as $n \to \infty$, and for intermediate $k$,

\[
\text{Bias} \left[ \hat{\theta}_n^N (k) \right] = \varphi_1(\theta) \left( \frac{k}{n} \right) + \varphi_2(\theta) \left( \frac{1}{k} \right) + o \left( \frac{1}{k} \right) + o \left( \frac{k}{n} \right).
\]

• In the semi-parametric EI-estimation we have thus to cope with problems similar to the ones appearing in the EVI-estimation: increasing bias, as the threshold decreases and a high variance for high thresholds.

Is it possible to improve the performance of estimators through the use of computer intensive methods?
The use of resampling methodologies [Efron, 1979, *Ann. Statist.*] has revealed to be promising in the estimation of the nuisance parameter \( k \), and in the reduction of bias of any estimator of a parameter of extreme events.

If we ask how to choose the tuning parameter \( k \) in the estimation, through \( T(k) \), of a parameter of extreme events, \( \eta \), we usually consider the estimation of \( k_0^T := \arg \min_k \text{MSE}(T(k)) \).

To obtain estimates of \( k_0^T \) one can then use a double-bootstrap method applied to an adequate auxiliary statistic like \( A(k) := T(k) - T(\lfloor k/2 \rfloor) \), which tends to zero and has an asymptotic behaviour similar to the one of \( T(k) \) (Gomes and Oliveira, 2001, *Extremes*, among others). We shall not sketch such a double-bootstrap algorithm.
• At such optimal levels, we have a non-null asymptotic bias.

• If we still want to remove such a bias, we can then make use of the generalized jackknife (GJ) methodology.

• The main objectives of the Jackknife methodology are:
  1. Bias and variance estimation of a certain estimator, only through manipulation of observed data \( x \).
  2. The building of estimators with Bias and MSE smaller than those of an initial set of estimators.

• The Jackknife or the GJ are resampling methodologies, which usually give a positive answer to the question: “May the combination of information improve the quality of estimators of a certain parameter or functional?”.
• For the **GJ EVI**-estimation, it is then enough to consider an adequate pair of estimators, possibly also $T(k)$ and $T([k/2])$, of the parameter of extreme events under consideration, and to built a *reduced-bias affine combination* of them. In Gomes, Martins & Neves (2000), [Extremes], also among others, we can find an application of this technique to the Hill estimator.

• In order to illustrate the use of these methodologies in **EVT** , we shall essentially consider, just as performed in Gomes, Martins & Neves (2013), *Comm. Statist.–Theory & Methods*, the aforementioned **MVRB EVI**-estimators $\overline{H}(k) \equiv \hat{\xi}_{k,n}$ in Caeiro et al. (2005), and **UC EI**-estimators.
The jackknife methodology and bias reduction

- The pioneering EVI reduced-bias estimators are all, in a certain sense, GJ estimators, i.e., affine combinations of well-known estimators of $\xi$.
- The generalized jackknife statistic was introduced by Gray and Shucany (1972): Let $T_n^{(1)}$ and $T_n^{(2)}$ be two biased estimators of $\xi$, with similar bias properties, i.e.,

$$\text{Bias}(T_n^{(i)}) = \xi + \phi(\xi)d_i(n), \quad i = 1, 2.$$ 

Then, if $q = q_n = d_1(n)/d_2(n) \neq 1$, the affine combination

$$T_n^G := \frac{T_n^{(1)} - qT_n^{(2)}}{1 - q}$$

is an unbiased estimator of $\xi$. 
• For the **GJ EI**-estimation, we use the fact that given 3 estimators $T_n^{(1)}$, $T_n^{(2)}$ and $T_n^{(3)}$ of $\theta$ such that

$$E\left[T_n^{(i)} - \theta\right] = d_1(\theta) \varphi_1^{(i)}(n) + d_2(\theta) \varphi_2^{(i)}(n), \; i = 1, 2, 3,$$

the **GJ** statistic (of order 2), given by

$$T_n^{GJ} := \begin{vmatrix} T_n^{(1)} & T_n^{(2)} & T_n^{(3)} \\ \varphi_1^{(1)} & \varphi_1^{(2)} & \varphi_1^{(3)} \\ \varphi_2^{(1)} & \varphi_2^{(2)} & \varphi_2^{(3)} \end{vmatrix} / \begin{vmatrix} 1 & 1 & 1 \\ \varphi_1^{(1)} & \varphi_1^{(2)} & \varphi_1^{(3)} \\ \varphi_2^{(1)} & \varphi_2^{(2)} & \varphi_2^{(3)} \end{vmatrix},$$

with $||A||$ denoting, as usual, the determinant of the matrix $A$, is an **unbiased estimator** of $\theta$. 
A GJ corrected-bias EVI-estimator

- Given $\bar{H}$, the most natural GJ RV is the one associated with the random pair $(\bar{H}(k), \bar{H}([\theta k]))$, $0 < \theta < 1$, is

$$\bar{H}^{GJ}(q, \theta)(k) := \frac{\bar{H}(k) - q \bar{H}([\theta k])}{1 - q}, \quad 0 < \theta < 1,$$

with

$$q = q_n = \frac{\text{Bias}_\infty[\bar{H}(k)]}{\text{Bias}_\infty[\bar{H}([\theta k])]} = \frac{A^2(n/k)}{A^2(n/\lfloor \theta k \rfloor)} \xrightarrow{n/k \to \infty} \theta^{2\rho}.$$

It is thus sensible to consider $q = \theta^{2\rho}$, $\theta = 1/2$, and, with $\hat{\rho}$ a consistent estimator of $\rho$, the GJ EVI-estimator,

$$\bar{H}^{GJ}(k) := \frac{2^{2\hat{\rho}} \bar{H}(k) - \bar{H}([k/2])}{2^{2\hat{\rho}} - 1}.$$
A GJ corrected-bias EI-estimator

- Since the bias term of the aforementioned classical EI-estimator reveals 2 main components of \( \neq \) orders, we need to use an affine combination of 3 EI-estimators and a order-2 GJ-statistic.

- The information on the bias of the EI-estimator \( \hat{\theta}_{nC}^{UC}(k) \) led us to consider the levels \( k, \lfloor \delta k \rfloor + 1 \) and \( \lfloor \delta^2 k \rfloor + 1 \), dependent of a tuning parameter \( \delta \), \( 0 < \delta < 1 \), and the class of estimators,

\[
\hat{\theta}_{n}^{GJ(\delta)}(k) := \frac{(\delta^2 + 1) \hat{\theta}_{nC}^{NC}(\lfloor \delta k \rfloor + 1) - \delta \left( \hat{\theta}_{nC}^{NC}(\lfloor \delta^2 k \rfloor + 1) + \hat{\theta}_{nC}^{NC}(k) \right)}{(1 - \delta)^2}
\]

- For a stationary Fréchet(1) ARMAX sample of size \( n = 5000 \), with \( \theta = 0.5 \), we next present
  - the expected values of such an estimator, associated to \( \delta = 0.1, 0.2, 0.4 \text{ e } 0.5 \) are next presented.
• Note the reasonably high stability around the target value \( \theta = 0.5 \), of the sample path and mean value of the GJ EI-estimator for a wide range of \( k \)-values, comparatively to that of Nandagopalan’s (or UC) EI-estimator.
Remark 15. The mean value stability around the target value $\theta$, for a wide range of $k$-levels, is true for all $\theta$ and for all simulated models.

But the GJ EI-estimator, $\hat{\theta}_{n}^{GJ}$, may not overpass, for $n = 1000$ (and small $\theta$), the UC estimator, $\hat{\theta}_{n}^{UC}$, regarding MSE at optimal levels. Extra investment is thus needed on the “optimal” choice of the 3 levels to be used in the building of a GJ EI-estimator or on the use of extra resampling or sub-sampling techniques, as performed in Gomes, Hall & Miranda (2008), who have used simple subsampling techniques, in order to attain a smaller MSE at optimal levels.
Other related topics and open problems

- We think that SUE is still a quite lively topic of research.
- Important developments have appeared recently in the area of *spatial extremes*, where *parametric models* became again relevant.
- And now, that we have access to highly sophisticated computational techniques, a great variety of *parametric models* can further be considered, like, for instance, the penultimate EV parametric model

\[
\text{PEV}_{\xi}(x; r) = \exp \left(- (1 + \xi x)^{-1/\xi} \left(1 + r(1 + \xi x)^{-1/\xi}\right)\right)
\]

or the associated penultimate GP CDF,

\[
\text{PGP}_{\xi}(x; r) = 1 - (1 + \xi x)^{-1/\xi} \left(1 + r(1 + \xi x)^{-1/\xi}\right).
\]

• Recent parametric models, like the extreme value Birnbaum-Saunders model in Ferreira, Gomes & Leiva (2012), [Revstat], can also become relevant in the area of SUE.

• And in a semi-parametric framework, topics like robustness and extremes, threshold selection, and the PORT methodology are still quite challenging.
• **Change-points detection in the tail behaviour** is also a challenging topic of research.

• **Testing** whether $F \in \mathcal{D}_M(\text{EV}_\xi)$, for a certain $\xi$, is a crucial topic, already dealt with in several articles. And what about testing second-order and even third-order conditions?

• And **SUE** for weakly dependent data, with all problems related to clustering of extreme values still deserves further research.

• **SUE** for randomly censored data and endpoint estimation are also still relevant topics in **SUE**.

• Moreover, the estimation of second and higher-order parameters still deserves further attention, particularly due to the importance of such estimation in **SORB** and **MVRB** estimators of parameters of extreme events.
- We could have said much more on the role of EVT in the modelling of rare events.
- The case of non-independence was only slightly touched.
- In this introduction to the study of extreme values, we have considered only the univariate case, but EVT is of high relevance both in the multivariate or in the spacial set-up, whenever dealing with the modelling of extreme or even rare events.
Epilogue

We hope to have increased your interest by this field, relatively recent from an historical point of view, but with a lot of applications, so many as we can conceive.

### END and THANKS ###
References


