Confidence Intervals for Extreme Quantile Estimates using Extreme Value Theory

Valentin Erismann

Supervised by
Prof. Alexander J. McNeil

Edinburgh

August 2016

Abstract

In this dissertation, we investigate the theory of order statistics and extreme value theory to construct confidence intervals for high quantile estimates. In an extensive series of simulation experiments, we compare both methods in terms of coverage properties, that is, the average length of computed confidence intervals, and the proportion of them which contain the true quantile.

Contents

1 Introduction

Financial institutions are subject to various requirements by regulatory frameworks such as Basel III in banking or Solvency II in insurance. Their primary concern is to strengthen companies' capital requirements in order to reduce the risk of insolvency. The exact way these requirements are calculated differs from one regulator to the other, and depends on the measure of risk which is used.

There are a variety of ways to define risk such as variance (or standard deviation), semi-variance or expected shortfall. The most widely used measure is the Value-at-Risk (VaR). For instance, under Solvency II, insurers must calculate a solvency capital requirement (SCR). The SCR is the amount of capital an insurer must hold to meet its obligations to policy holders over a one year time horizon with probability larger than 99.5%. Hence, SCR is based on VaR. If L represents the loss over a one year period, then this translates into finding the amount $x_{0.995}$ such that

$$P(L \leq x_{0.995}) = 0.995,$$

which is the definition of $\text{VaR}_{0.995}$. In other words, it ensures that the chance of failure is no more than a 1 in 200 years event.

In probabilistic terms, the VaR is an $\alpha-$quantile of the loss distribution, defined as follows.

Definition 1.0.1. *(Value-at-Risk).* The Value-at-Risk of a distribution function (df) F is given by

$$\text{VaR}_\alpha(F) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\},\ 0 < \alpha < 1.$$

The main disadvantage of using VaR as a risk measure is the uncertainty about the severity of the losses occuring with probability higher than α. It is therefore advised to compute VaR for different levels α: $0.95, 0.975, 0.99, 0.995$.

As with any other statistical quantity, there is uncertainty in the estimation of VaR. We are therefore interested in constructing a confidence interval (CI) for each calculated VaR. Rather than computing a point estimate for $\hat{x}_\alpha$, which does not inform on the uncertainty of the estimate, we want to find $a,\ b \in \mathbb{R}$ such that

$$\mathbb{P}\big(\hat{x}_\alpha \in (a, b)\big) = P,$$

for some confidence level P. We emphasise that in our problem there are two distinct confidence levels. We have

- the confidence level α, which determines x_α, and
- the confidence level P, which determines the confidence interval for $\hat{x}_\alpha$.

Our goal is to construct a confidence interval at level P for an extreme quantile at level α, *e.g.* a 95%CI for $\hat{x}_{0.99}$.

We start with an example to illustrate the Value-at-Risk and motivate the use of order statistics (OS) to estimate it. Let $x_1, ..., x_{100}$ be simulated observations of an insurance

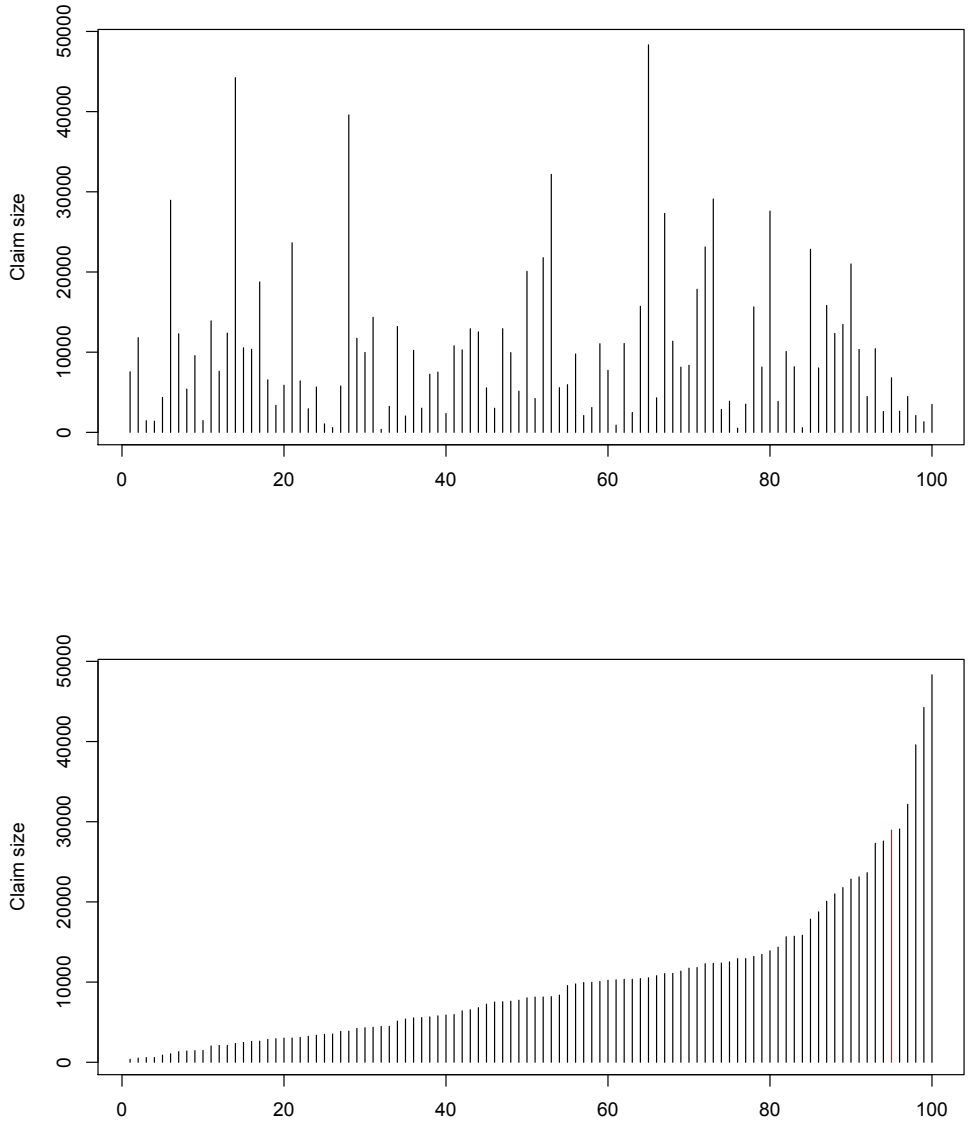

Figure 1: Simulated insurance losses (top) and the corresponding ordered sample (bottom) of a sample of 100 standard exponential random variables. $\hat{x}_{0.95}$ is shown in red.

company's claim sizes. $\text{VaR}_{0.95}$ is the claim size value $x_{0.95}$ such that 95% of the claim sizes are below $x_{0.95}$, that is

$$P(X \leq x_{0.95}) = 0.95.$$

If we rearrange the data in ascending order (see Figure 1), it is easier to identify $\text{VaR}_{0.95}$. In the case of a sample of size 100, $\hat{x}_{0.95}$ can be given by the 95-th smallest (or 6-th largest) value. Indeed, VaR is an order statistic itself. Therefore, by studying the distribution of order statistics, we can construct confidence intervals for extreme quantiles.

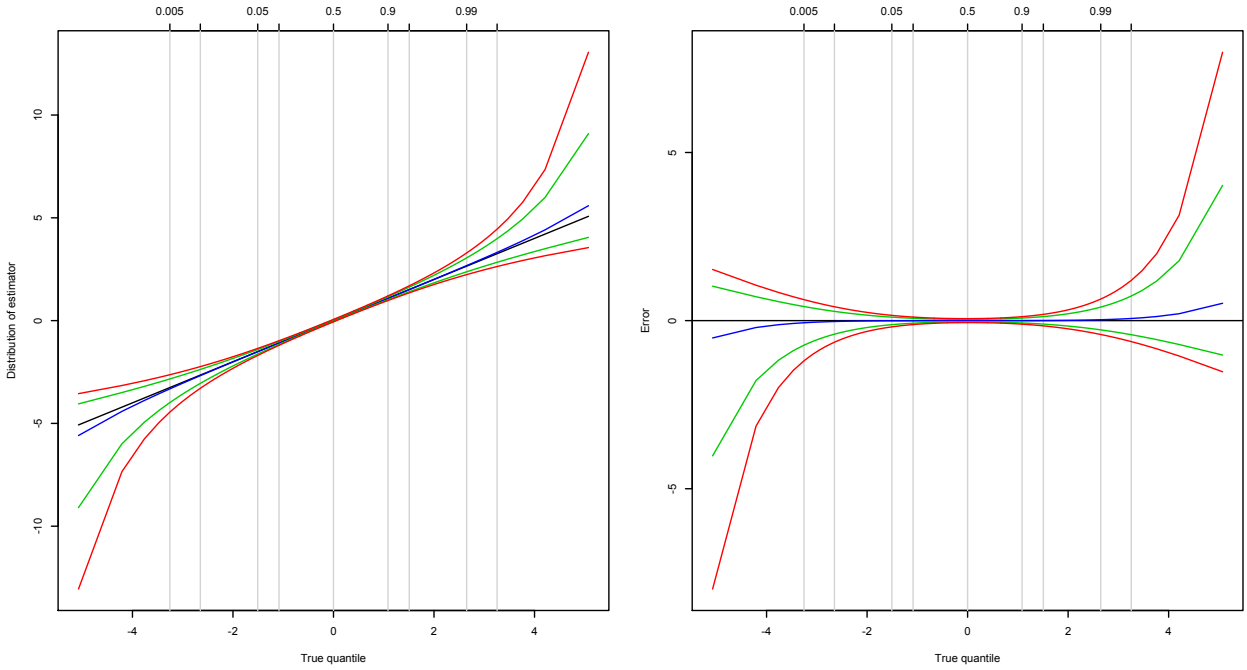

Figure 2: Distribution of order statistics of the Student's t distribution with $\nu = 4$ (left) and its standard error (right). See Appendix C.1 for the code provided by Prof. McNeil.

This is the simplest version of the OS estimation as it is based on a single order statistic. However, if $0.95 \cdot \textit{sample size}$ is not an integer, then it is not straightforward which OS should be used. For example, if the sample size is 272, then $272 \cdot 0.95 = 258.4$. Therefore, the estimator of $\text{VaR}_{0.95}$ could be $x_{(258)}$ or $x_{(259)}$ or a combination of the two.

A more complex estimator is constructed by interpolating between two order statistics. We present two types of estimation based on OS in the next section. However, such estimators have high variability in the tails, as we can see in Figure 2. As we show in the simulation study (Section 4), this becomes problematic for very high quantiles ($\alpha \geq 0.99$). We therefore investigate extreme value theory as an alternative.

This method was developed by hydrologists under the name Peaks Over Threshold (POT), and can be applied to flood estimation. Excesses over high thresholds are modelled with the generalised Pareto distribution (GPD), which is closely related to the behaviour of normalised maxima. We use the maximum likelihood (ML) approach to estimate the parameters of the GPD. The VaR estimator is then computed based on a tail approximation. Finally, we obtain confidence intervals by inverting the likelihood ratio test. This method is presented in Section 3.

In the last section, we compare both methods in an extensive series of simulations. Knowing the value of the true quantile, we can measure the performance of each method. The main research question we are interested in is:

does the EVT method outperform the empirical method in terms of coverage properties?

To answer this, we conduct a simulation study and compare

- the number of CIs that contain the true quantile, and
- the average lengths of the CIs.

We perform this experiment for different classes of distributions and draw conclusions on the performance of each method for various values of the confidence level P, the VaR level α, and the sample size.

2 Confidence intervals using order statistics

After introducing fundamentals on the theory of order statistics, we examine two types of quantile estimation. We then use the distribution of these point estimates to construct confidence intervals. This section mainly builds upon David and Nagaraja [1] and Reiss [12].

2.1 Order statistics

Let $X_1, X_2, X_3, ...$ be a sequence of iid random variables with continuous df F.

Definition 2.1.1. The ordered sample is

$$X_{(1)} \leq ... \leq X_{(n)},$$

with $X_{(1)} = \min(X_1, ..., X_n)$ and $X_{(n)} = \max(X_1, ..., X_n)$. The random variable (rv) $X_{(k)}$ is called the k-th lower order statistic. Similarly we can define upper order statistics as $X_{n,n} \leq ... \leq X_{1,n}$. Note that $X_{k,n} = X_{(n-k+1)}$.

We want to derive an expression for the distribution function of $X_{(m)}$ $(m = 1, ..., n)$, in order to construct a confidence interval for its estimator. First, we introduce the empirical distribution function.

Definition 2.1.2. The empirical df is defined as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}, \ x \in \mathbb{R}.$$

Notice that

$$\begin{aligned} X_{(m)} \leq x &\iff \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} \geq m \\ &\iff \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} \geq \frac{m}{n} \\ &\iff F_n(x) \geq \frac{m}{n} \end{aligned}$$

Therefore,

$$P(X_{(m)} \leq x) = P(F_n(x) \geq \frac{m}{n}),$$

which implies that $F_n^{\leftarrow}(t) = X_{(m)}$, for $\frac{m-1}{n} < t \leq \frac{m}{n}$ and $m = 1, ..., n$.

Next we calculate the df of the m-th lower order statistic.

Proposition 2.1.1. *Distribution of the m-th lower order statistic.* For $m = 1, ..., n$ let G_m denote the df of $X_{(m)}$. Then

$$G_m(x) = \sum_{j=m}^{n} \binom{n}{j} [1 - F(x)]^{n-j} [F(x)]^j = F_{m,n-m+1}\big(F(x)\big),$$

where $F_{m,n-m+1}$ denotes the beta df.

Proof. Define $B_n = \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}$. Then B_n is a sum of iid Bernoulli variables with success probability $F(x)$. That is, $B_n \sim \text{Bin}(n, F(x))$.

$$\begin{aligned} G_{(m)}(x) &= P(B_n \geq m) \\ &= \sum_{j=m}^{n} P(B_n = j) \\ &= \sum_{j=m}^{n} \binom{n}{j} [1 - F(x)]^{n-j} [F(x)]^j. \end{aligned}$$

□

As we will see in the next section, some estimators are based on two order statistics, e.g. $(1-\gamma)X_{(m)} + \gamma X_{(m+1)}$, $\gamma \in [0,1]$. Therefore, we are interested in the df of a convex linear combination between two consecutive OS. Since we need their joint distribution to calculate it, we first look at the joint pdf $g_{r,s}$ of $(X_{(r)}, X_{(s)})$, for arbitrary r and s. A detailed proof of the following result can be found in David and Nagaraja [1], p.10.

Proposition 2.1.2. Let $r < s \in \{1, ..., n\}$. The joint pdf of $(X_{(r)}, X_{(s)})$ is given by

$$g_{r,s}(x,y) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} [F(x)]^{r-1} [F(y) - F(x)]^{s-r-1} [1 - F(y)]^{n-s} f(x) f(y),$$

with $x, y \in \mathbb{R}, x \leq y$.

Proof. Let $x \leq y$. The event $x < X_{(r)} < x + \delta x,\ y < X_{(s)} < y + \delta y$. By continuity, there are almost surely no ties and therefore,

$$\begin{aligned} r-1 &\text{ observations are less than } x, \\ 1 &\text{ observation is in } (x, x+\delta x), \\ s-r-1 &\text{ observations are in } (x+\delta x, y), \\ 1 &\text{ observation is in } (y, y+\delta y), \\ n-s &\text{ observations are above } y + \delta y. \end{aligned}$$

□

Corollary 2.1.1. The joint pdf of two consecutive OS $(X_{(k)}, X_{(k+1)})$ is given by

$$g_{k,k+1}(x,y) = \frac{n!}{(m-1)!(n-m-1)!} [F(x)]^{k-1} [1 - F(y)]^{n-k-1} f(x) f(y),\ x < y.$$

Proof. Apply Proposition 2.1.2 to two consecutive OS. □

Before obtaining the density of $(1-\gamma)X_{(m)} + \gamma X_{(m+1)}$, we need to find the joint pdf of $\big((1-\gamma)X_{(m)}, \gamma X_{(m+1)}\big)$.

Corollary 2.1.2. The joint p.d.f of $(X,Y) = ((1-\gamma)X_{(m)}, \gamma X_{(m+1)})$ is given by

$$f_{X,Y}(x,y) = g_{m,m+1}\Big(\frac{x}{1-\gamma}, \frac{y}{\gamma}\Big) \cdot \gamma^{-1}(1-\gamma)^{-1}.$$

Proof. Let X and Y be two dependent rvs. By definition, the joint density can be written as

$$f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x).$$

Assume $X = (1-\gamma)X_{(m)}$ and $Y = \gamma X_{(m+1)}$, then

$$\begin{aligned} f_{X,Y}(x,y) &= f_{Y|X}(y|x) \cdot f_X(x) \\ &= f_{Y|X}\Big(\gamma X_{(m+1)} = y \Big| (1-\gamma)X_{(m)} = x\Big) \cdot \gamma^{-1} \cdot f_{X_{(m)}}\big(x/(1-\gamma)\big) \cdot (1-\gamma)^{-1} \\ &= f_{X_{(m+1)}|X_{(m)}}\Big(X_{(m+1)} = \frac{y}{\gamma} | X_{(m)} = \frac{x}{(1-\gamma)}\Big) \cdot \gamma^{-1} \cdot g_m(x/(1-\gamma)) \cdot (1-\gamma)^{-1} \\ &= g_{m,m+1}\Big(\frac{x}{1-\gamma}, \frac{y}{\gamma}\Big) \cdot \gamma^{-1} \cdot (1-\gamma)^{-1}. \end{aligned}$$

□

Recall that for two independent rvs, the density of their sum is the convolution of their densities. However, OS are not independent, even if the original rvs are. For instance, for $k < j$ and $x > y$ we have that $P(X_{(k)} = x | X_{(j)} = y) = 0$, but $P(X_{(k)} = x) \cdot P(X_{(j)} = y) \neq 0$.

The following result and its proof can be found in Grimmett [6].

Theorem 2.1.1. If X and Y have joint density function f, then $Z = X + Y$ has density function

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z-x)\, dx.$$

The df is

$$F_Z(z) = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{z} f(x, y-x)\, dy\, dx.$$

We now have all the tools we need to find the df of a linear combination of two OS.

Corollary 2.1.3. The df of $(1-\gamma)X_{(m)} + \gamma X_{(m+1)}$ denoted by $G_{\gamma,m}$ is given by

$$G_{\gamma,m}(z) = \Big(\frac{1}{1-\gamma} \cdot \frac{1}{\gamma}\Big) \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{z} g_{m,m+1}\Big(\frac{x}{1-\gamma}, \frac{y-x}{\gamma}\Big) dy dx.$$

Proof. Apply Theorem 2.1.1 and Corollary 2.1.2 to the linear combination. □

Note that we do not obtain an explicit formula for $G_{\gamma,m}$, since this double integral is usually analytically intractable. Furthermore, in order to find lower and upper limits of a confidence interval, one only needs to solve it numerically, that is, find z such that

$$G_{\gamma,m}(z) = \delta,$$

for a given δ. However, we have an explicit formula for the df of any order statistic. The next step is to define the sample quantile and the df will follow from our calculations.

2.2 Sample quantile

As mentioned in the introduction, there are several ways to define the sample quantile. Hyndman and Fan [9] discuss the most widely used sample quantile definitions and their properties. We present two of them. The first is based on a single order statistic, defined as the inverse of the empirical df. Then, we investigate linear interpolation between two order statistics. The second is the default definition in the `R` function `quantile`.

2.2.1 Sample quantile using one order statistic

This is the definition we used in the first example.

Definition 2.2.1. (*Type 1).* Let $X_1, ..., X_n$ be rvs and $X_{(1)}, ..., X_{(n)}$ the ordered sample. Then the $\alpha-$th quantile can be estimated by

$$\hat{Q}_1(\alpha) := \begin{cases} X_{(np)}, & \text{if } [np] = np, \\ X_{([np]+1)}, & \text{if } [np] < np. \end{cases}$$

where $[y]$ denotes the greatest integer no greater than y.

Recall the example in the introduction. Let $x_1, ..., x_{100}$ be iid rvs from an exponential df. We want to estimate $\text{VaR}_{0.95}$. By the latter, we obtain $\hat{Q}_1(0.95) = x_{(95)}$.

The obvious drawback of this definition is the discontinuity of $\hat{Q}_1$.

2.2.2 Sample quantile using linear interpolation between two order statistics

The next definition is the default method used in the software `R`.

Definition 2.2.2. *(Type 2).*

$$\hat{Q}_2(\alpha) = (1-\gamma)X_{(j)} + \gamma X_{(j+1)}$$

with $\gamma = \alpha(n-1) - [\alpha(n-1)]$ and $j = [n\alpha + 1 - \alpha]$.

Example 2.2.1. For $n = 100$ and $\alpha = 0.95$ we obtain $\hat{Q}_2(\alpha) = 0.95X_{(95)} + 0.05X_{(96)}$. For some values, there is quite a difference between the two estimators. For $n = 252$ and $\alpha = 0.99$ for instance, $\hat{Q}_1(\alpha) = X_{(249)}$ and $\hat{Q}_2(\alpha) = 0.51X_{(249)} + 0.49X_{(250)}$.

2.3 Confidence invervals

2.3.1 Introduction

Let x_α be the $\alpha-$th quantile. To construct a confidence interval for $\hat{x}_\alpha$, we first study CIs of the form $[X_{(r)}, X_{(s)}]$, for $r < s$, and calculate the confidence level. First note that the event,

$$\{X_{(r)} \leq x_\alpha\} = \{X_{(r)} \leq x_\alpha\} \bigcap \{X_{(s)} \geq x_\alpha\} \bigcup \{X_{(r)} \leq x_\alpha\} \bigcap \{X_{(s)} < x_\alpha\}.$$

Also,

$$X_{(s)} < x_\alpha \Rightarrow X_{(r)} \leq x_\alpha.$$

Thus, $P(X_{(r)} \leq x_\alpha) = P(X_{(r)} \leq x_\alpha \leq X_{(s)}) + P(X_{(s)} < x_\alpha)$, which implies that

$$\begin{aligned} P(X_{(r)} \leq x_\alpha \leq X_{(s)}) &= P(X_{(r)} \leq x_\alpha) - P(X_{(s)} < x_\alpha) \\ &= \sum_{i=r}^{s-1} \binom{n}{i} \alpha^i (1-\alpha)^{n-i} \end{aligned}$$

However, we do not obtain a CI for a given confidence level. Instead, we use the df of the order statistic, or the linear combination of two, depending on the choice of sample quantile estimator.

2.3.2 Confidence interval of type 1

If the VaR point estimate is of type 1, then the estimator is an exact OS as defined in Definition 2.2.1. Recall that

$$G_m(x) = \sum_{j=m}^{n} \binom{n}{j} [1 - F(x)]^{n-j} [F(x)]^j = F_{m,n-m+1}\big(F(x)\big).$$

Therefore, we can easily find the percentile points and thus construct a confidence interval. Indeed, a quantile is an order statistic and we need to invert the df $G_m(x)$.

Suppose we wish to estimate a $100P\%$CI for an $\alpha-$th quantile. First we find the values of $F(x)$ for which

$$\begin{aligned} G_m(x) &= \frac{1-P}{2}, \text{ and} \\ G_m(x) &= \frac{1+P}{2}. \end{aligned}$$

We then find the lower and upper bounds of the CI by taking the inverse of the underlying df F or the empirical df F_n.

Remark 2.3.1. The exact df is of the form $G_m(x) = F_{m,n-m+1}\big(F(x)\big)$. However, in practice, F is usually unknown. For instance, it could be the loss distribution of an insurance portfolio. Therefore, one should use the empirical df when computing CIs.

Example 2.3.1. Let $x_1, ..., x_{1000}$ be observations from a standard normal distribution. Suppose we want to find a 90%CI for the $\text{VaR}_{0.95}$. Following Definition 2.2.1, point estimation is given by $X_{(950)}$ with distribution G_{950}. We use the incomplete beta distribution to find the 5th and 95th percentile points of the binomial. This gives us values of 0.9371 and 0.9599 for $F(x)$. Now, we use the inverse of the normal distribution function to find the bounds of the CI. $F(x) = 0.9371 \Rightarrow x = F^{-1}(0.9371) = \Phi^{-1}(0.9371) = 1.5312$. Similarly the upper bound is 1.7501.

Remark 2.3.2. Some percentile points of the df of the OS can be found in Table 1. Note that these do not depend on the underlying df F but only on the quantile level, the confidence level and the sample size. Therefore, the values in the table are always the same, no matter what the df F is. We say that this method is distribution-free.

Quantile	Sample size	2.5−th percentile	97.5−th percentile
0.99	100	0.9455	0.9976
	500	0.9768	0.9956
	1000	0.9817	0.9945
0.95	100	0.8872	0.9777
	500	0.9271	0.9658
	1000	0.9346	0.9618
0.975	100	0.9148	0.9890
	500	0.9559	0.9846
	1000	0.9633	0.9829

Table 1: Percentile points of a beta distribution for different quantiles and sample sizes.

2.3.3 Confidence interval of type 2

With definitions of type 2, we do not have an explicit formula for the df. Consequently we cannot use the same approach as before. Indeed, we need to solve a double integral. This is difficult to do analytically. Hence, we use the `R` package `cubature` to solve the double integral numerically. To obtain a $100P\%$CI we then use `uniroot` to solve

$$G_{\gamma,m}(z_1) = \frac{1-P}{2}, \text{ and}$$
$$G_{\gamma,m}(z_2) = \frac{1+P}{2}.$$

The $100P\%$CI is then given by $[z_1, z_2]$.

2.4 Concluding notes

We presented two types of sample quantiles and how to construct their corresponding confidence intervals. In the simulation study (Section 4), we use the definition of type 1. One way to deal with the case where $n\alpha$ is not an integer is to use a bisection algorithm, which is suggested by Dowd [2]. Consider the df of an order statistic, that is,

$$G_m(x) = \sum_{j=m}^{n} \binom{n}{j} [1-F(x)]^{n-j} [F(x)]^j.$$

We use the bisection algorithm in Appendix C.2 to determine the percentile points of $G_m(x)$. As we will see in Section 4, the confidence intervals obtained with the algorithm are more accurate than with the beta df, for the case where $n\alpha$ is not an integer. However, for very high quantiles, we obtain better results with the EVT method, which we introduce in the next section.

3 Confidence intervals using extreme value theory

Classical extreme value theory (EVT) is interested in the fluctuations of the maximum of a sample of rvs. In this section, we summarise the main results from EVT. We then explore more precisely the generalised Pareto sistribution (GPD), which is used to model excesses over a high threshold. We fit this distribution to Peaks Over Threshold (POT) and use the profile likelihood to obtain point estimates as well as confidence intervals for high quantiles. This sections follows mainly Embrechts, Klüpperlberg and Mikosch (EKM) [4] and McNeil, Frey and Embrechts (MFE) [10].

3.1 Theoretical background

We begin by presenting the limit laws for normalised maxima of iid rvs. The crucial result is that the non-degenerate limits can only be of three types. We then provide a one-parameter representation known as the Generalised Extreme Value distribution (GEV), which leads to the GPD.

3.1.1 The generalised extreme value distribution

Let $X_1, ..., X_n$ be a sequence of iid non-degenerate rvs with df F. We are interested in the behaviour of the rv $M_n = \max(X_1, ..., X_n)$ or $X_{(n)}$ using the OS notation. The exact distribution of the maximum is given by

$$\begin{aligned} P(M_n < x) &= P(X_1 < x, ..., X_n < x) \\ &= P(X_1 < x) \cdot ... \cdot P(X_n < x), \text{ by independence} \\ &= F^n(x). \end{aligned}$$

F is a df and therefore, as $n \to \infty$, $F_n(x) \to \{0, 1\}$. The tipping point is called the *right endpoint* of F and is defined as

$$x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}.$$

Hence, $\forall x < x_F$,

$$F^n(x) \to 0, \ n \to \infty$$

and if x_F is finite,

$$F^n(x) = 1, \text{ for } x > x_F.$$

By definition (see Appendix A) $M_n \xrightarrow{P} x_F$.

Example 3.1.1. The Weibull distribution has a finite right endpoint whereas $x_F = \infty$ for the Fréchet and the Gumbel distributions (see Figure 3).

The main result in classical EVT is similar to the Central Limit Theorem (CLT) for the sum of rvs. Suppose $X_1, ..., X_n \sim F$ and $S_n := \sum_1^n X_i$. The General Central Limit problem consists in finding $a_n > 0, b_n \in \mathbb{R}$ such that

$$\frac{S_n - a_n}{b_n}$$

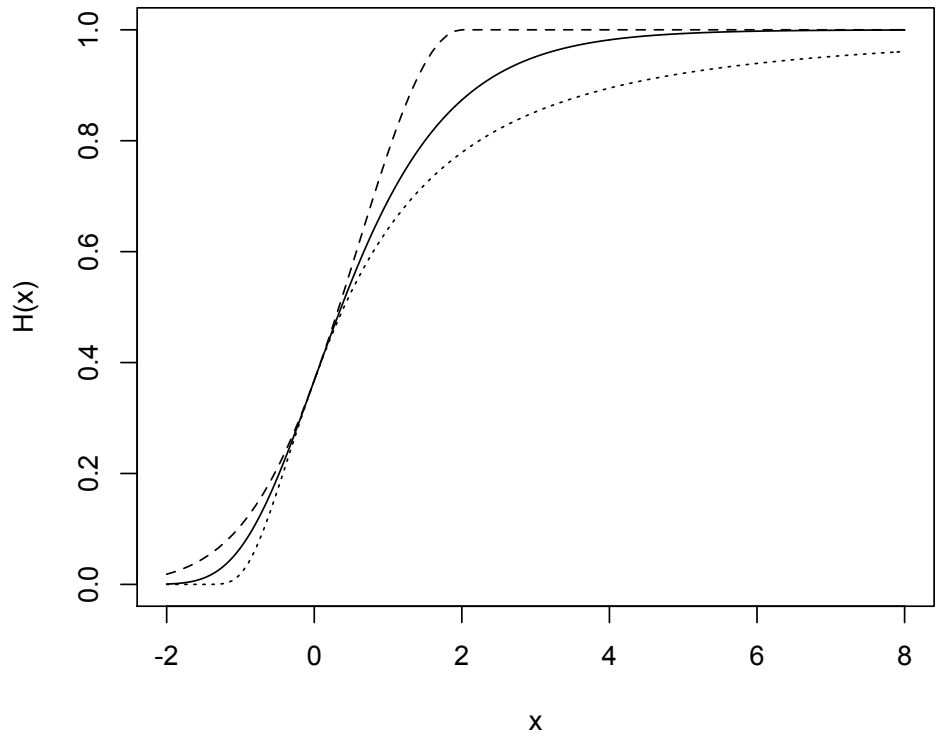

Figure 3: Distribution function of the GEV for different parameters ξ and $\mu = 0, \sigma = 1$. The solid line represents the Gumbel ($\xi = 0$); the dotted line represents the Fréchet ($\xi = 0.5$); the dashed line represents the Weibull ($\xi = -0.5$).

converges to some non-degenerate rv Z. The CLT tells us that if F has a finite second moment, then $Z \sim \mathcal{N}(0, 1)$.

Similarly, we want to find $c_n > 0, d_n \in \mathbb{R}$ such that

$$\frac{M_n - c_n}{d_n}$$

converges to some non-degenerate rv H. The following result is the foundation of EVT. A detailed proof can be found in Resnick [13].

Theorem 3.1.1. *(The Fisher-Tippett theorem).* If there exist $c_n > 0, d_n \in \mathbb{R}$ and some non-degenerate df H such that

$$\frac{M_n - c_n}{d_n} \xrightarrow{d} H, \tag{1}$$

then H is one of the three following dfs:

$$\text{Fréchet:}\quad \phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp\{-x^{-\alpha}\}, & x > 0, \end{cases} \quad \alpha > 0,$$

$$\text{Weibull:}\quad \psi_\alpha(x) = \begin{cases} \exp\{-x^{-\alpha}\}, & x \leq 0 \\ 1, & x > 0 \end{cases} \quad \alpha > 0.$$

$$\text{Gumbel:}\quad \Lambda(x) = \exp\{-e^{-x}\},\ x \in \mathbb{R}.$$

Definition 3.1.1. We say that a rv X belongs to the maximum domain of attraction of the distribution H ($X \in \text{MDA}(H)$) if Equation (1) holds.

A very practical representation of the three previous dfs is given by the Jenkinson-von Mises representation. It combines all three dfs in a one-parameter definition.

Definition 3.1.2. *(Jenkinson-von Mises representation).* The generalised extreme value distribution is defined as

$$H_\xi(x) = \begin{cases} \exp\{-(1+\xi x)^{-1/\xi}\} \text{ if } \xi \neq 0, \\ \exp\{-\exp\{-x\}\} \text{ if } \xi = 0, \end{cases}$$

where $1 + \xi x > 0$.

H_ξ gives either the Fréchet, Weibull or Gumbel df depending on the value of the parameter ξ:

$$\xi > 0 \text{ represents the Fréchet,}$$
$$\xi = 0 \text{ represents the Gumbel,}$$
$$\xi < 0 \text{ represents the Weibull.}$$

The quantity $1/\xi$ is known as the tail index. We say that a df F is heavy-tailed if its (right) tail is not exponentially bounded, that is,

$$\lim_{x\to\infty} e^{\lambda x}\bar{F}(x) = \infty,$$

for all $\lambda > 0$, and $\bar{F}(x) = 1 - F(x)$. Gnedenko [5] showed that if this is the case, then $F \in MDA(H_\xi)$ for $\xi > 0$. This class of distributions includes the Pareto, Cauchy and Student's-t distributions. The maximum domain of attraction of H_0 includes the lognormal and the normal distributions. These are called medium-tailed distributions. These two classes are the center of interest of the simulation study. Distributions in $\text{MDA}(H_\xi)$ with $\xi < 0$ are called short-tailed distributions. The larger the parameter ξ, the heavier the tail.

Definition 3.1.3. *(Excess distribution function).* Let X be a rv with df F. Then, the excess df over the threshold $u < x_F$ is given by

$$F_u(x) = P(X - u \leq x | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)},\ x \geq 0.$$

ξ	Distribution
$\xi > 0$	Pareto
$\xi = 0$	Exponential
$\xi < 0$	Pareto Type II

Table 2: Special cases of the generalised Pareto distributions given the paramater ξ.

The mean excess function of X is

$$e(u) = E(X - u | X > u).$$

Remark 3.1.1. F_u is referred to as the excess-of-loss df in the context of reinsurance.

The relation in Definition 3.1.3 can also be written as

$$\bar{F}(u)\bar{F}_u(x) = \bar{F}(u + x) \tag{2}$$

Theorem 3.1.2. *(Characterisation of MDA(H_ξ)).* The following are equivalent:

$$F \in MDA(H_\xi) \iff$$

$$\lim_{u \to x_F} \frac{\bar{F}\big(u + xa(u)\big)}{\bar{F}(u)} = \begin{cases} (1 + \xi x)^{-1/\xi}, & \text{if } \xi \neq 0, \\ e^{-x}, & \text{if } \xi = 0. \end{cases} \tag{3}$$

Proof. A sketch of the proof can be found in Embrechts, Kluppelberg and Mikosch [4], Theorem 3.4.5. □

This theorem is crucial for the modelling of excesses over a threshold. We can rewrite the LHS of Equation 3 as

$$\lim_{u \to x_F} P\Big(\frac{X - u}{a(u)} \geq u \Big| X > u\Big).$$

Therefore, we obtain a limit of excess distribution over a threshold. This is the motivation for the generalised Pareto distribution, which we introduce in the next section.

3.1.2 The generalised Pareto gistribution

This definition is prompted by the limiting behaviour of excesses over a threshold (see RHS of Equation 3).

Definition 3.1.4. *(GPD).* The df of the generalised Pareto distribution is given by

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1 - \exp\left(-x/\beta\right), & \text{if } \xi = 0, \end{cases}$$

where

$$\begin{cases} x \geq 0 & \text{if } \xi \geq 0, \text{ and} \\ 0 \leq x \leq -\beta/\xi & \text{if } \xi < 0. \end{cases}$$

and $\beta > 0$.

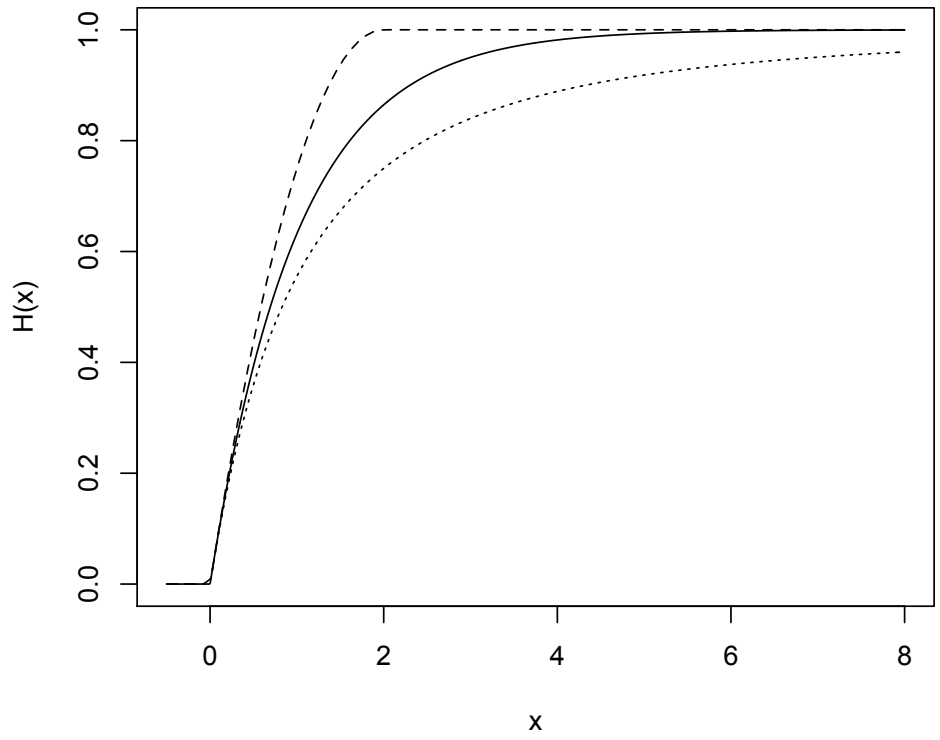

Figure 4: Distribution function of the GPD for different parameters ξ and $\beta = 1$. The solid line is the exponential ($\xi = 0$); the dotted line is the Pareto ($\xi = 0.5$); the dashed line is the Pareto type II ($\xi = -0.5$).

This df is *generalised* because, like the GEV distribution, it contains special cases (see Table 2 and Figure 4).

If $\xi < 1$, the expectation of $X \sim GPD_{\xi,\beta}$ is given by

$$E[X] = \frac{\beta}{1-\xi}.$$

For $\xi > 0$, $E[X^r] = \infty$, for $r \geq 1/\xi$.

The following theorem by Pickands, Balkema and de Haan essentially states that the GPD appears as the natural distribution for modelling excess losses over high thresholds.

Theorem 3.1.3. $F \in \mathrm{MDA}(H_\xi), \xi \in \mathbb{R}$ if and only if

$$\lim_{u \to x_F} \sup_{0<x<x_F-u} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0$$

for some positive and measurable function $\beta(\cdot)$.

Proof. Recall Theorem 3.1.2. Together with Equation (2), the theorem becomes

$$\lim_{u \to x_F} |\bar{F}_u(x) - \bar{G}_{\xi,\beta(u)(x)}| = \lim_{u \to x_F} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0,$$

with $a(u) = \beta(u)$. The GPD is continuous and therefore, the convergence is uniform. □

Proposition 3.1.1. If X follows a GPD with $\xi < 1$. Then, for $u < x_F$,

$$e(u) = E[X - u | X > u] = \frac{\beta + \xi u}{1 - \xi}, \ \beta + u\xi > 0.$$

Proof. Using the definitions of the mean excess function and the expectation of a rv, we have,

$$\begin{aligned} e(u) &= \frac{1}{\bar{F}(u)} \int_u^{x_F} (x - u) dF(x) \\ &= \frac{1}{\bar{F}(u)} \int_u^{x_F} \bar{F}(x) dx, \ 0 < u < x_F. \end{aligned}$$

□

The last proposition implies that the mean excess function is linear in u, the threshold. This is a very useful property. The choice of threshold is an important problem in fitting a GPD. A threshold too high (with few exceedances) results in a high variance. Conversely a threshold too low produces a biased estimator. One way to choose an suitable threshold is to look at the Mean-Excess plot (ME plot) and identify one (or various) threshold for which $e(x)$ is approximately linear for $x > u$. This is presented in the next section when we fit the GPD.

3.2 Modelling excesses over a threshold

In the previous section, we showed that the GPD appears as the natural distribution for excesses over a high threshold. We now show how to fit this distribution to a data set and how to derive a confidence interval for VaR using the maximum likelihood approach.

3.2.1 Fitting the generalised Pareto distribution

Suppose we have a sequence of iid rvs $X_1, ..., X_n$ from an unknown df F. We want to fit a GPD to the data to model excesses over a high threshold. To make the theoretical calculations, we assume that for some threshold u we have $F_u(x) = G_{\xi,\beta}(x)$ for $0 \leq x < x_F - u$ and some $\xi \in \mathbb{R}$ and $\beta > 0$.

First, we need to find an appropriate threshold u. Denote by $N_u := \#\{i | X_i > u\}$, the number of exceedances of the sample above u (see Figure 5).

Recall that the excess df is given by

$$F_u(y) = P(X - u \leq y | X > u), \ y \geq 0.$$

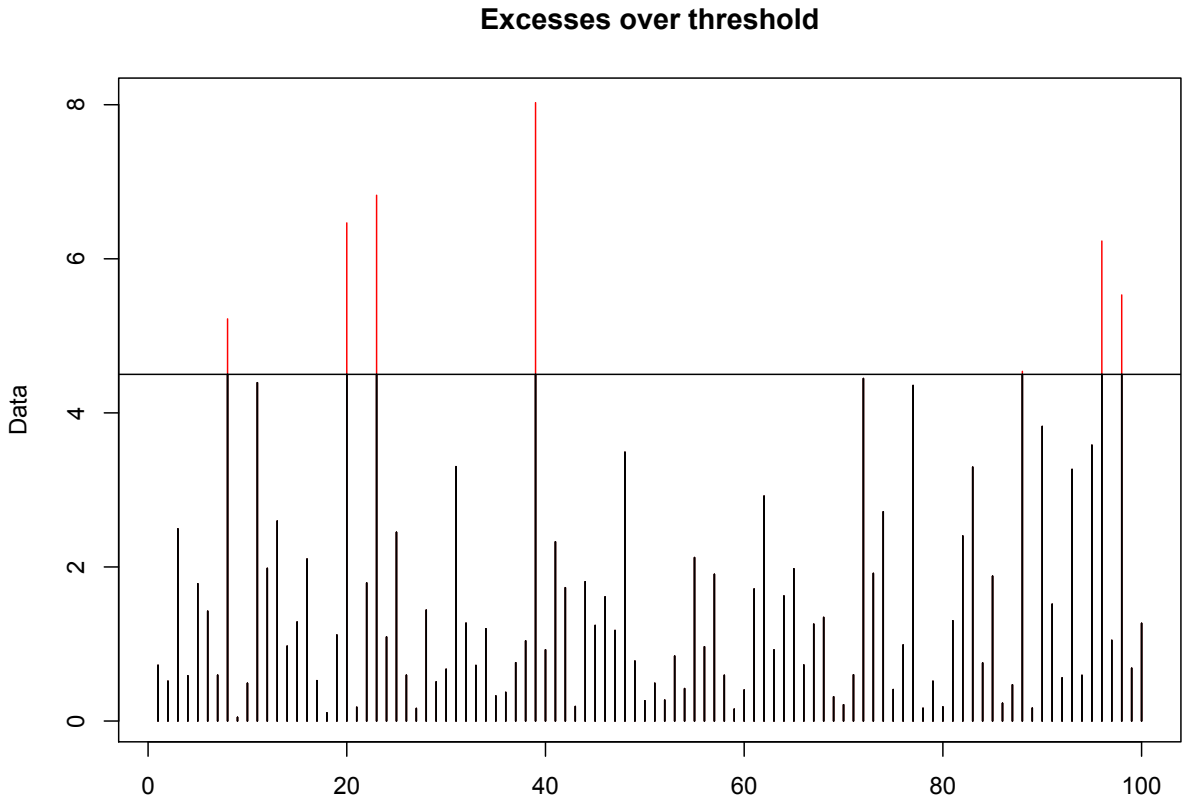

Figure 5: Data $X_1, ..., X_{100}$ and the excesses over the threshold $u = 4.5$.

First, we can get a formula for tail probabilities, that is, for $x \geq u$,

$$\begin{aligned}\bar{F}(x) &= P(X > u)P(X > x|X > u) \\ &= \bar{F}(u)P(X - u > x - u|X > u) \\ &= \bar{F}(u)\bar{F}_u(x - u) \\ &= \bar{F}(u)\big(1 + \xi\frac{x-u}{\beta}\big)^{-1/\xi}.\end{aligned}$$

We then invert this formula to obtain a high quantile of F,

$$x_\alpha = u + \frac{\beta}{\xi}\left(\left(\frac{1-\alpha}{\bar{F}(u)}\right)^{-\xi} - 1\right), \tag{4}$$

and replace the unknown quantities by their estimates. For $\bar{F}(u)$ we take $\frac{N_u}{n}$. Finally, we get the following estimator, first proposed by Smith [16],

$$\hat{\bar{F}}(u) = \frac{N_u}{n}\left(1 + \hat{\xi}\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\xi}}, \tag{5}$$

for $x \geq u$.

In Proposition 3.1.1 we have shown that the mean-excess function is asymptotically linear in u. Therefore, we plot the mean-excess function and graphically identify a threshold such that the function is approximately linear (see Figure 6).

3.2.2 Maximum likelihood estimation

Assume $\boldsymbol{X} = (X_1, ..., X_n)$ is the data and that $G_{\xi,\beta}$ is a GPD with parameters ξ and β. The density $g_{\xi,\beta}$, is given by

$$g_{\xi,\beta}(x) = \frac{\xi}{\beta}\big(1 + \xi\frac{x}{\beta}\big)^{-1-\frac{1}{\xi}},\ x \in D(\xi,\beta).$$

where

$$D(\xi,\beta) = \begin{cases} [0,\infty) & \text{if } \xi \geq 0, \text{ and} \\ (0, -\beta/\xi) & \text{if } \xi < 0. \end{cases}$$

Therefore, the log-likelihood is

$$l(\xi,\beta;\boldsymbol{x}) = -n\ln\beta - (\frac{1}{\xi}+1)\sum_{i=1}^{n}\ln(1+\frac{\xi}{\beta}x_i).$$

We can solve this numerically and obtain the MLE $\hat{\xi}$ and $\hat{\beta}$. Smith [14] shows that this method works well for $\xi > -1/2$. In fact,

$$\sqrt{n}\Big(\hat{\xi} - \xi, \frac{\hat{\beta}}{\beta} - 1\Big) \xrightarrow{d} \mathcal{N}(0, M^{-1}),$$

where

$$M^{-1} = (1+\xi)\begin{pmatrix} 1+\xi & 1 \\ 1 & 2 \end{pmatrix},$$

and $\mathcal{N}(\mu,\Sigma)$ stands for the bivariate normal distribution with mean vector μ and covariance matrix Σ. The standard properties of consistency and asymptotic efficiency hold. However, Smith [15] states that for $\xi \leq -1/2$, *"the problem is nonregular and special procedures are needed."*

We wish to construct a confidence interval for a high quantile, or VaR. However, we need to reparametrise the GPD in terms of the VaR (x_α), *i.e.*

$$(\xi,\beta) \to (\xi, x_\alpha).$$

Recall Equation 4,

$$x_\alpha = u + \frac{\beta}{\xi}\bigg(\Big(\frac{1-p}{\bar{F}(u)}\Big)^{-\xi} - 1\bigg),$$

which allows us to write β in terms of the parameters (ξ, x_α):

$$\beta(\xi, x_\alpha) = \xi(x_\alpha - u)\Big(\big(\frac{1-p}{N_u/n}\big)^{-\xi} - 1\Big)^{-1}$$

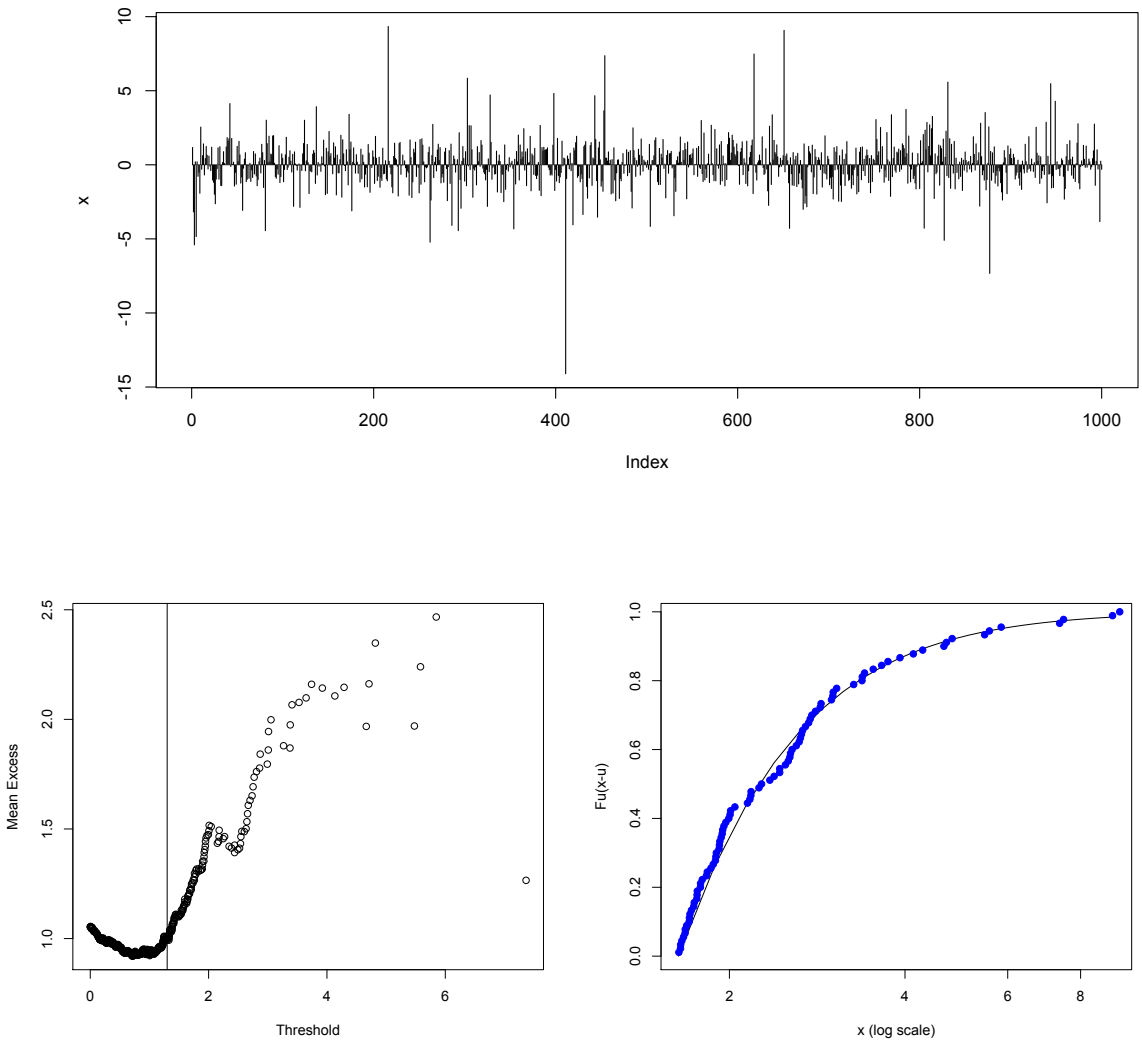

Figure 6: 1,000 generated data points from a Student's t distribution with 4 degrees of freedom (top); the corresponding ME plot with a potential choice of threshold $u_1 = 1.3$ (bottom left); the empirical distribution of excesses and the fitted GPD with estimated parameters based on ML approach $\hat{\xi} = 0.35$ and $\hat{\beta} = 0.78$ (bottom right).

and therefore obtain the log-likelihood

$$\begin{aligned}\ln L(\xi, \beta(\xi, x_\alpha); \boldsymbol{y}) &= \sum_{i=1}^{N_u} \ln g_{\xi,\beta(\xi,x_\alpha)}(y_i) \\ &= -N_u \ln(\beta(\xi, x_\alpha)) - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^{N_u} \ln\left(1 + \xi \frac{y_i}{\beta(\xi, x_\alpha)}\right).\end{aligned}$$

This is what is done in `ML.HO.GPD` (see appendix C).

In the simulation study, we use only the ML approach. Hosking and Wallis [8] present

two other methods of parameters estimation and compare them to the ML method. They establish that the method of moments gives a reliable alternative except for $\xi \geq 0.2$. Furthermore, they advocate the use of the method of probability-weighted-moments (PWM) in the case $\xi \geq 0$. However, the CIs we construct in the next section are based on the profile log-likelihood function, which is why we focus on the ML approach.

3.3 Confidence intervals

In this section, we show how we can use the generalised Pareto distribution to construct confidence intervals for high quantiles. The CIs are based on the Likelihood Ratio Test (see appendix D). In the previous section, we showed how we can fit a GPD to a data set. For simplicity, assume we want to construct a 95%CI for 99%VaR.

First, we compute the partial log-likelihood in terms of the parameters (ξ, x_α). We obtain the limits of the confidence intervals by finding the points of intersection between the partial log-likelihood curve and the horizontal line $h = 0.95$ (see Figure 7).

We use the `R` function `ML.H0.GPD` to maximise the negative partial log-likelihood for values of ξ. We then use the function `uniroot` to find the roots of the equation given by the likelihood ratio test (Equation 6 in Appendix D)

$$\ln L(x_\alpha^0, \hat{\xi}_0; \boldsymbol{y}) - \ln L(\hat{\alpha}_p, \hat{\xi}, \boldsymbol{y}) + 0.5 c_{1,0.95} = 0,$$

where $c_{1,0.95}$ is the 95th quantile of the χ_1^2 distribution, $\ln L(\hat{\alpha}_p, \hat{\xi}, \boldsymbol{y})$ is the global maximum and $\ln L(x_\alpha^0, \hat{\xi}_0; \boldsymbol{y})$ is the partial log-likelihood function.

The solutions of the previous equation define the lower and upper bounds of the confidence interval for $\hat{x}_\alpha$.

Remark 3.3.1. We need to specify a search interval for `uniroot`. Therefore, one might not always find the roots of the LRT equation either because the interval is too wide (and contains both roots) or too narrow (and contains none). To minimise this issue, we restrict the search intervals to $(0.25\hat{\alpha}_p, \hat{\alpha}_p)$ and $(\hat{\alpha}_p, 4\hat{\alpha}_p)$. Another solution is to use `ExtendInt` within the function `uniroot`, which extends the search interval if no solution is found. However, this increases the computing time considerably.

Remark 3.3.2. Note that in general, the CI is asymmetric, that is, the distance between the MLE and the lower and upper bounds are different.

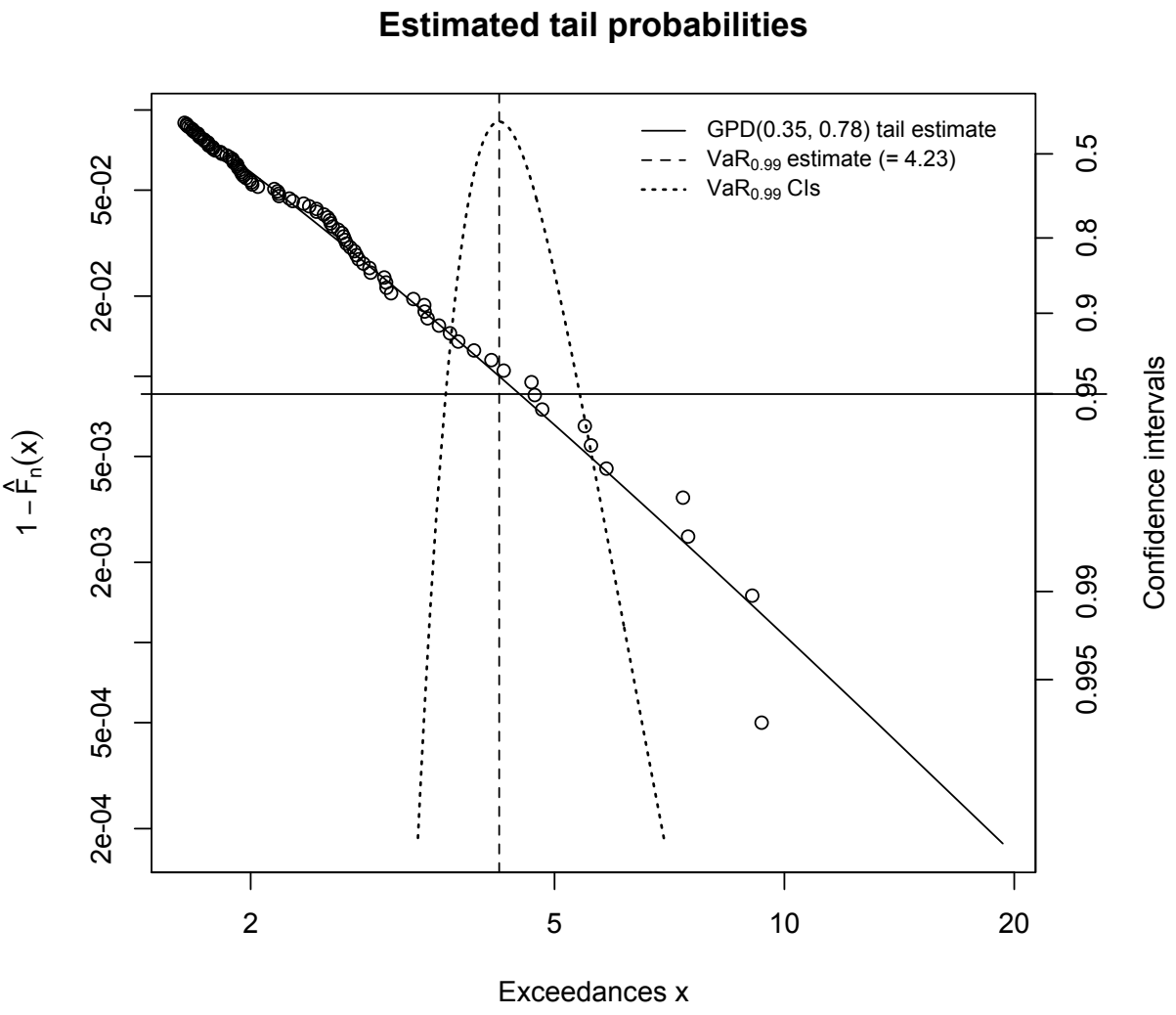

Figure 7: The smooth curve through the points shows the fitted GPD for 1,000 Student's t rvs with 4 degrees of freedom at a threshold u=1.3. The intersections between the dotted curve and the straight horizontal line give the lower and upper limits of a 95%CI for $\text{VaR}_{0.99}$.

4 Simulation study

We now present the methodology, and analyse the results of the simulation study. Our goal is to compare the empirical method, based on order statistics and the EVT method, based on the generalised Pareto distribution. The performance of each method is measured by the percentage of CIs that contain the true quantile, and the average length of the CIs. We also record the percentage of simulation runs producing errors.

4.1 Simulation procedure

Data is generated from a known distribution function F. Therefore, we deduce the values of the true quantiles. Confidence intervals are then constructed for each method using the function `doOne` (see Appendix C.5). The function returns two values: `lengthCI`, the length of each CI, and `inCI`, a logical value verifying if the true quantile is in the CI. These results are stored in the variable `res`, along with error and warning indicators as well as the simulation time in *ms*. We compare the methods for different

- distributions `F`,
- sample sizes `n`,
- VaR levels `alpha`, and
- confidence levels `P`.

We summarise the variables and their type within Table 3 (see Hofert and Maechler [7] for more details on the types and the package `simsalapar`). The variable of type N defines the number of simulation replications (here, N=10000). The variables of type grid produce a data frame. Each row contains a unique combination of all those variables. The simulation iterates N times over all rows.

Variable	Expression	Type	Value
`n.sim`	N	N	10000
`n`	n	grid	100, 500, 1000
`P`	P	grid	0.90, 0.95, 0.99
`alpha`	α	grid	0.950, 0.990, 0.995
`method`	Method	grid	GPD, OS
`Df`	Df	grid	Logormal, Normal, Pareto, Student's t

Table 3: Variables which determine the simulation study.

As summarised in Table 3, we are interested in several confidence levels for different VaRs. But for a given data set this can be computed *simultaneously*. This is done using the `R` package `simsalapar` (see Appendix C for the code).

In the simulation study, we consider the distribution functions with different indexes, that is, medium and heavy-tailed distributions, as suggested by McNeil and Saladin [11] (see

MDA(H_0)	MDA($H_{0.5}$)	MDA(H_1)
Standard Lognormal	Pareto ($\alpha = 2, x_m = 1$)	Pareto ($\alpha = 1, x_m = 1$)
Standard Normal	Student's t ($\nu = 2$)	Student's t ($\nu = 1$)

Table 4: Summary of distributions used in simulation study.

Table 4). The standard normal and lognormal dfs are in the maximum domain of attraction of the Gumbel (MDA(H_0). The Pareto and the Student's-t distributions are in the maximum domain of attraction of the Fréchet (MDA($H_{0.5}$) or MDA(H_1) depending on the parameters). These are popular choices for loss distributions among insurers.

In an individual analysis, one would plot the mean-excess function in order to choose a suitable threshold. But for a simulation study, this cannot be done manually every time. Therefore, we fix the number of exceedances as $N_u = \frac{\text{sample size}}{4}$. We show in Table 7 the success rates of the simulation runs for $N_u = \frac{\text{sample size}}{10}$. The rates are higher for the normal, but this is a special case, as the lognormal (and the other dfs) has lower success rates than for a greater value of N_u.

It is important to mention that the underlying distribution function F is treated as unknown in the construction of CIs. We only use that information to compute the true quantile. The aim is to assess how well the methods perform with limited information on the data. The additional assumption we make is that the generated rvs are iid, which is standard practice in insurance.

4.2 Error handling

The package `simsalapar` allows to record the percentage of failure in a code. Clearly, the OS method never fails but the EVT one is likely to fail for distributions such as the standard normal because convergence of the GPD is very slow, as we will see in Section 4.3.1. It is therefore important to report the percentage of success.

As we see in Table 5 (which only displays statistics for successful attempts) the EVT method performs well for a normal distribution. Although, the GPD converges in so little occurrences that the results are not relevant to the study (see Table 6). One could argue that the normal distribution is a poor choice for an insurance loss distribution. Nevertheless, this information needs to be taken into account when comparing methods.

4.3 Results

We want to compare the lengths of the confidence intervals because for a given confidence level, a wider CI means a loss in precision. Additionally, we are interested in the percentage of CIs that contain the true quantile. If we are constructing a 95%CI, say, then in theory, 95% of the CIs should contain the true quantile. But recall that we are considering the underlying df F as unknown. We evaluate the performance of both methods under that assumption. Overall, no method is perfect, that is, no method produces CIs that contain the true quantile systematically 100P% of the time. We say that the method performs well if the proportion is close to the desired level P. This also means that a proportion higher

than P is not satisfactory. Indeed, it would suggest that the CI is larger than what is needed to reach the wanted outcome.

The results of the simulation study for distributions in MDA(H_0), MDA($H_{0.5}$) and MDA(H_1) are shown in Tables 5, 8 and 10, respectively.

For both method, the size of the CIs increases with the value of α and decreases as the sample size grows. This holds regardless of the type of distribution or confidence level P.

The EVT method is more affected by a change in the sample size than the OS method. For a given confidence level, the size can be reduced by a factor of three between a large and small sample, whereas the empirical method remains stable. As a result, for a large sample, we obtain smaller CIs irrespective of the underlying distribution or the VaR level.

4.3.1 Distributions in MDA(H_0)

This class of distributions includes the normal and the lognormal dfs. They both have medium-sized tails but the lognormal has a longer tail than that of a normal distribution. As a result, the EVT method fails very often for the latter (around 95% of the time). This is mostly due to the convergence of the GPD. With a tail like that of a normal, the MLE is likely to produce a negative estimate of the parameter ξ. As discussed in Section 3.2.2, for these values of ξ, the convergence is likely to fail. One way to deal with this problem is to increase the threshold (see Table 7), but this reduces the success rates of the lognormal. Nevertheless, this is an isolated issue, as it is not the case for heavier tails. Finally, when the method does not fail, it performs well, both in terms of CI length and of proportion of CIs containing the true quantile.

Because of the longer tail, the lognormal distribution provides good results with the EVT method. The success rate is over 90%. For very extreme quantiles such as $\text{VaR}_{0.99}$, the empirical method gives smaller CIs but isn't as accurate as the EVT method.

Therefore, for this class of distributions, we would suggest the use of the EVT method (if it works) over the OS method. The conclusion is not as obvious for heavy-tailed distributions, which we examine in the final two sections.

4.3.2 Distributions in MDA($H_{0.5}$)

In this section we consider the Pareto ($\alpha = 2, x_m = 1$) and the Student's-t distribution with 2 degrees of freedom.

For small samples ($n = 100$), the EVT method is more accurate. Even though the OS method produces smaller CIs, the precision is unsatisfactory. Indeed, less than 62% of CIs contain the true quantile. We recommend to use the former, especially for high quantiles such as $\alpha = 0.99$. For these quantiles the OS method should not be considered, regardless of the sample size, unless the GPD does not converge. The success rate of the EVT is less than 10% for $n = 100$ and $\alpha = 0.99$ (see Table 9)

However, for $\alpha = 0.95, 0.975$ and $n = 500, 1000$, no method stands out as the results are very close. The empirical method seems to produce more accurate CIs whereas the EVT method generates smaller CIs, but both methods perform well overall. Therefore, the choice should be made according to the particular analysis one is trying to achieve.

4.3.3 Distributions in MDA(H_1)

Finally, we investigate the Pareto ($\alpha = 1, x_m = 1$) and the Cauchy distributions. These have heavier tails than the previous distributions, which explains why we obtain such large CIs. With a low index, we need a large sample to obtain precise results, which is confirmed by the study.

Indeed, for large samples ($n = 500$ and $n = 1000$), we obtain very satisfactory results with both methods. A similar trend to the previous distributions can be observed, that is, the EVT method gets more accurate as α increases and conversely for the OS method. But the latter produces larger CIs, apart for some rare occurrences ($\alpha = 0.99, n = 500, P = 0.90$ and 0.95, for the Pareto distribution).

For small samples, notice that the success rate of the GPD is quite low (see Table 11) except for $\alpha = 0.95$. Thus, the results obtained by that method are not relevant. For the Pareto distribution, the CIs obtained by the OS method are substantial (more than 10 times the size we obtain with a larger sample), it does not seem pertinent to use them. For the Cauchy however, we recommend to use the empirical method.

4.4 Concluding notes

We have conducted this experiment for different classes of dfs and various values of the variables. Both methods can be improved on an individual basis. The choice of threshold plays an important role for the EVT method. A different definition of the sample quantile might yield better results for the empirical method. Furthermore, we can see in Table 12 that we obtain better results when using a bisection algorithm in the case where $n\alpha$ is not an integer. Note that when $n\alpha$ is an integer, the results are the same. Therefore, an "expert judgement" is needed to evaluate the relevance of the results.

		Method	GPD			OS		
Df	Sample size	$P \mid \alpha$	0.950	0.975	0.990	0.950	0.975	0.990
Standard lognormal	100	0.90	3.090 86.54%	7.134 90.41%	18.309 93.36%	3.094 80.74%	5.241 79.43%	6.710 54.95%
		0.95	3.983 92.62%	9.740 95.48%	20.193 97.88%	3.693 85.76%	6.432 84.92%	7.588 58.20%
		0.99	6.576 98.57%	14.375 99.39%	20.354 100.00%	5.022 93.59%	8.213 89.23%	8.832 61.58%
	500	0.90	1.193 84.02%	2.203 87.35%	5.040 89.69%	1.554 88.01%	2.701 87.79%	4.838 80.62%
		0.95	1.435 90.92%	2.693 93.22%	6.286 95.05%	1.859 93.18%	3.226 92.48%	5.774 85.91%
		0.99	1.930 97.51%	3.776 98.42%	9.255 98.86%	2.463 97.97%	4.308 97.77%	7.821 93.73%
	1000	0.90	0.838 84.80%	1.519 87.07%	3.365 89.58%	1.120 88.70%	1.917 87.90%	3.673 84.61%
		0.95	1.003 90.84%	1.832 93.12%	4.099 94.51%	1.338 94.01%	2.289 93.02%	4.415 90.46%
		0.99	1.332 97.40%	2.483 98.34%	5.692 98.87%	1.764 98.53%	3.031 97.95%	5.893 96.25%
Standard normal	100	0.90	0.656 91.51%	1.110 96.82%	2.327 95.59%	0.612 80.39%	0.710 77.86%	0.749 53.25%
		0.95	0.821 96.88%	1.503 98.71%	3.309 98.38%	0.725 85.64%	0.845 83.81%	0.860 56.89%
		0.99	1.269 99.69%	2.734 99.86%	5.650 99.67%	0.951 93.44%	1.068 88.90%	1.041 61.23%
	500	0.90	0.271 84.58%	0.371 92.00%	0.572 86.69%	0.301 88.00%	0.377 87.77%	0.481 80.20%
		0.95	0.323 94.27%	0.449 96.75%	0.706 92.40%	0.360 93.18%	0.448 92.48%	0.570 85.78%
		0.99	0.429 99.12%	0.615 100.00%	1.012 97.91%	0.473 97.97%	0.590 97.75%	0.749 93.58%
	1000	0.90	0.190 84.62%	0.254 94.29%	0.381 82.71%	0.217 88.70%	0.271 87.89%	0.362 84.57%
		0.95	0.227 92.31%	0.305 94.29%	0.462 89.47%	0.258 94.01%	0.323 93.02%	0.432 90.38%
		0.99	0.300 100.00%	0.408 100.00%	0.635 93.98%	0.339 98.53%	0.424 97.95%	0.568 96.24%

Table 5: Simulation results for distributions in MDA(H_0). The table displays the average length of the confidence intervals (black) and the proportion of CIs (blue) that contain the true quantile out of 10,000 simulations.

		Method	GPD			OS		
Df	n	$P \mid \alpha$	0.950	0.975	0.990	0.950	0.975	0.990
Lognormal	100	0.90	0.833	0.837	0.723	1	1	1
		0.95	0.833	0.826	0.539	1	1	1
		0.99	0.830	0.621	0.108	1	1	1
	500	0.90	0.998	0.999	0.993	1	1	1
		0.95	0.998	0.999	0.993	1	1	1
		0.99	0.998	0.999	0.993	1	1	1
	1000	0.90	1	1	0.999	1	1	1
		0.95	1	1	0.999	1	1	1
		0.99	1	1	0.999	1	1	1
Normal	100	0.90	0.128	0.148	0.138	1	1	1
		0.95	0.128	0.148	0.136	1	1	1
		0.99	0.128	0.146	0.091	1	1	1
	500	0.90	0.023	0.040	0.053	1	1	1
		0.95	0.023	0.040	0.053	1	1	1
		0.99	0.023	0.040	0.053	1	1	1
	1000	0.90	0.003	0.007	0.013	1	1	1
		0.95	0.003	0.007	0.013	1	1	1
		0.99	0.003	0.007	0.013	1	1	1

Table 6: Success rates for distributions in MDA(H_0) with $N_u = \frac{n}{4}$. Rates below 0.60 are shown in orange; rates below 0.20 are shown in red.

		Method	GPD			OS		
Df	n	$P \mid \alpha$	0.950	0.975	0.990	0.950	0.975	0.990
Lognormal	100	0.90	0.437	0.540	0.368	1	1	1
		0.95	0.448	0.515	0.174	1	1	1
		0.99	0.354	0.260	0.001	1	1	1
	500	0.90	0.939	0.917	0.947	1	1	1
		0.95	0.939	0.917	0.947	1	1	1
		0.99	0.939	0.917	0.947	1	1	1
	1000	0.90	0.992	0.983	0.992	1	1	1
		0.95	0.992	0.983	0.992	1	1	1
		0.99	0.992	0.983	0.992	1	1	1
Normal	100	0.90	0.119	0.179	0.153	1	1	1
		0.95	0.129	0.178	0.116	1	1	1
		0.99	0.072	0.143	0.006	1	1	1
	500	0.90	0.126	0.226	0.114	1	1	1
		0.95	0.126	0.226	0.114	1	1	1
		0.99	0.126	0.226	0.114	1	1	1
	1000	0.90	0.066	0.200	0.062	1	1	1
		0.95	0.066	0.200	0.062	1	1	1
		0.99	0.066	0.200	0.062	1	1	1

Table 7: Success rates for distributions in MDA(H_0) with $N_u = \frac{n}{10}$. Rates below 0.60 are shown in orange; rates below 0.20 are shown in red.

Df	Sample size	Method $P \mid \alpha$	GPD 0.950	GPD 0.975	GPD 0.990	OS 0.950	OS 0.975	OS 0.990
Pareto	100	0.90	2.966 85.34%	8.148 88.67%	17.978 92.11%	2.893 81.69%	6.639 80.50%	10.208 55.71%
($\alpha = 2, x_m = 1$)		0.95	3.909 92.08%	10.364 95.26%	17.158 96.35%	3.506 86.34%	8.662 86.45%	11.548 58.75%
		0.99	6.531 98.16%	12.557 99.47%	16.413 99.51%	5.006 93.84%	11.585 90.09%	13.320 61.45%
	500	0.90	1.094 84.63%	2.419 87.42%	6.778 88.93%	1.400 88.30%	2.891 88.01%	6.481 81.16%
		0.95	1.321 91.00%	2.974 93.10%	8.526 94.10%	1.679 93.35%	3.481 92.77%	7.823 86.49%
		0.99	1.796 97.35%	4.228 98.26%	12.834 98.68%	2.242 98.05%	4.742 97.81%	11.193 94.15%
	1000	0.90	0.762 85.33%	1.645 87.75%	4.436 88.95%	1.001 88.83%	2.004 88.05%	4.849 84.10%
		0.95	0.913 91.25%	1.989 93.39%	5.425 94.18%	1.198 94.29%	2.401 93.36%	5.871 90.40%
		0.99	1.219 97.76%	2.715 98.31%	7.608 98.71%	1.588 98.60%	3.217 98.20%	7.988 96.61%
Student's t	100	0.90	2.332 85.79%	5.975 88.02%	14.001 94.35%	2.209 81.20%	4.842 80.23%	7.365 56.20%
($\nu = 2$)		0.95	3.025 92.35%	7.673 94.29%	13.988 98.67%	2.668 86.49%	6.308 85.83%	8.343 58.95%
		0.99	4.895 98.18%	9.619 99.31%	13.904 100.00%	3.790 94.06%	8.433 89.67%	9.647 61.69%
	500	0.90	0.876 84.34%	1.748 84.84%	4.322 85.40%	1.068 87.45%	2.117 87.54%	4.685 80.55%
		0.95	1.054 90.65%	2.139 91.06%	5.398 91.21%	1.280 93.06%	2.544 92.53%	5.668 85.69%
		0.99	1.424 97.16%	3.007 97.40%	7.983 97.83%	1.711 98.18%	3.453 97.71%	8.080 93.06%
	1000	0.90	0.613 82.58%	1.198 83.83%	2.859 85.60%	0.767 88.46%	1.471 86.76%	3.482 84.08%
		0.95	0.734 89.89%	1.445 90.76%	3.485 91.92%	0.917 94.00%	1.764 92.66%	4.217 90.16%
		0.99	0.977 97.04%	1.961 97.31%	4.846 97.61%	1.213 98.47%	2.356 98.02%	5.733 96.24%

Table 8: Simulation results for distributions in MDA($H_{0.5}$). The table displays the average length of the confidence intervals (black) and the proportion of CIs (blue) that contain the true quantile out of 10,000 simulations.

		Method	GPD			OS		
Df	Sample size	$P \mid \alpha$	0.950	0.975	0.990	0.950	0.975	0.990
Pareto	100	0.90	0.923	0.925	0.703	1	1	1
($\alpha = 2, x_m = 1$)		0.95	0.923	0.871	0.438	1	1	1
		0.99	0.902	0.505	0.081	1	1	1
	500	0.90	1	1	1	1	1	1
		0.95	1	1	1	1	1	1
		0.99	1	1	1	1	1	1
	1000	0.90	1	1	1	1	1	1
		0.95	1	1	1	1	1	1
		0.99	1	1	1	1	1	1
Student's t	100	0.90	0.860	0.827	0.685	1	1	1
($\nu = 2$)		0.95	0.860	0.793	0.452	1	1	1
		0.99	0.841	0.481	0.046	1	1	1
	500	0.90	0.999	0.997	0.997	1	1	1
		0.95	0.999	0.997	0.997	1	1	1
		0.99	0.999	0.997	0.997	1	1	1
	1000	0.90	1	1	1	1	1	1
		0.95	1	1	1	1	1	1
		0.99	1	1	1	1	1	1

Table 9: Success rates for distributions in MDA($H_{0.5}$). Rates below 0.60 are shown in orange; rates below 0.20 are shown in red.

Df	Sample size	Method $P \mid \alpha$	GPD 0.950	GPD 0.975	GPD 0.990	OS 0.950	OS 0.975	OS 0.990
Pareto	100	0.90	31.609 84.23%	79.019 87.22%	110.192 64.96%	30.481 82.09%	315.183 82.14%	877.983 57.05%
($\alpha = 1, x_m = 1$)		0.95	38.963 91.01%	70.806 90.34%	91.028 53.95%	38.669 86.49%	560.599 87.12%	1019.467 59.50%
		0.99	42.909 96.72%	60.441 90.46%	74.306 37.04%	67.194 94.07%	907.872 90.44%	1193.761 61.84%
	500	0.90	10.189 84.58%	33.645 87.91%	153.375 89.42%	12.847 88.32%	39.459 88.09%	149.957 81.50%
		0.95	12.437 91.15%	42.049 93.33%	198.641 94.22%	15.562 93.35%	48.433 92.80%	188.276 86.54%
		0.99	17.379 97.37%	62.300 98.25%	262.180 98.99%	21.277 98.06%	69.781 97.83%	326.069 94.28%
	1000	0.90	6.938 85.20%	22.030 87.92%	93.692 89.01%	9.065 88.83%	26.008 88.05%	103.439 84.15%
		0.95	8.363 91.41%	26.859 93.75%	116.190 94.28%	10.902 94.29%	31.433 93.36%	128.603 90.47%
		0.99	11.314 97.65%	37.369 98.25%	168.608 98.74%	14.629 98.60%	43.161 98.21%	186.063 96.63%
Student's t	100	0.90	10.111 83.84%	25.568 86.78%	35.838 66.68%	9.568 80.86%	67.798 81.73%	169.086 56.97%
($\nu = 1$)		0.95	12.445 90.49%	23.184 89.91%	29.214 53.08%	12.014 86.00%	112.612 86.30%	195.207 59.32%
		0.99	14.143 96.76%	19.800 91.79%	25.993 44.12%	20.681 93.97%	176.022 89.62%	227.488 61.59%
	500	0.90	3.312 84.71%	10.392 86.48%	44.480 87.12%	4.120 87.41%	12.527 87.36%	48.516 81.05%
		0.95	4.036 91.04%	12.952 92.41%	57.353 93.08%	4.986 92.91%	15.331 92.65%	61.083 86.01%
		0.99	5.618 97.35%	19.056 98.11%	79.869 98.80%	6.817 97.94%	22.086 97.77%	104.698 94.14%
	1000	0.90	2.265 84.09%	6.832 86.53%	27.283 87.29%	2.908 88.57%	8.297 87.81%	33.065 84.50%
		0.95	2.728 90.76%	8.318 92.57%	33.746 93.12%	3.490 93.96%	10.027 93.20%	41.241 90.68%
		0.99	3.683 97.17%	11.533 97.97%	48.657 98.44%	4.680 98.40%	13.710 98.07%	59.927 96.73%

Table 10: Simulation results for distributions in MDA(H_1). The table displays the average length of the confidence intervals (black) and the proportion of CIs (blue) that contain the true quantile out of 10,000 simulations.

		Method	GPD			OS		
Df	Sample size	$P \mid \alpha$	0.950	0.975	0.990	0.950	0.975	0.990
Pareto	100	0.90	0.980	0.707	0.211	1	1	1
($\alpha = 1, x_m = 1$)		0.95	0.919	0.412	0.073	1	1	1
		0.99	0.503	0.079	0.003	1	1	1
	500	0.90	1	1	1	1	1	1
		0.95	1	1	1	1	1	1
		0.99	1	1	0.860	1	1	1
	1000	0.90	1	1	1	1	1	1
		0.95	1	1	1	1	1	1
		0.99	1	1	1	1	1	1
Student's t	100	0.90	0.972	0.760	0.265	1	1	1
($\nu = 2$)		0.95	0.921	0.471	0.099	1	1	1
		0.99	0.540	0.101	0.003	1	1	1
	500	0.90	1	1	1	1	1	1
		0.95	1	1	1	1	1	1
		0.99	1	1	0.926	1	1	1
	1000	0.90	1	1	1	1	1	1
		0.95	1	1	1	1	1	1
		0.99	1	1	1	1	1	1

Table 11: Success rates for distributions in MDA(H_1). Rates below 0.60 are shown in orange; rates below 0.20 are shown in red.

Df	Sample size	$P \mid \alpha$	0.950	0.975	0.990
Lognormal	100	0.90	3.094 80.74%	4.197 64.82%	6.710 54.95%
		0.95	3.693 85.76%	4.919 70.10%	7.588 58.20%
		0.99	5.022 93.59%	7.001 84.13%	8.832 61.58%
	500	0.90	1.554 88.01%	2.550 83.96%	4.838 80.62%
		0.95	1.859 93.18%	3.053 89.68%	5.774 85.91%
		0.99	2.463 97.97%	4.059 95.97%	7.821 93.73%
	1000	0.90	1.120 88.70%	1.917 87.90%	3.673 84.61%
		0.95	1.338 94.01%	2.289 93.02%	4.415 90.46%
		0.99	1.764 98.53%	3.031 97.95%	5.893 96.25%
Normal	100	0.90	0.612 80.39%	0.669 64.07%	0.749 53.25%
		0.95	0.725 85.64%	0.784 69.98%	0.860 56.89%
		0.99	0.951 93.44%	1.036 82.93%	1.041 61.23%
	500	0.90	0.301 88.00%	0.368 83.91%	0.481 80.20%
		0.95	0.360 93.18%	0.438 89.64%	0.570 85.78%
		0.99	0.473 97.97%	0.575 95.95%	0.749 93.58%
	1000	0.90	0.217 88.70%	0.271 87.89%	0.362 84.57%
		0.95	0.258 94.01%	0.323 93.02%	0.432 90.38%
		0.99	0.339 98.53%	0.424 97.95%	0.568 96.24%

Table 12: Simulation results of the OS method for distributions in MDA(H_0) using a bisection algorithm to find percentile points of G_m.

5 Conclusion

The estimation of VaR has many applications in finance and insurance. A point estimate might be enough, but if we wish to take into account the uncertainty in the estimation, we can construct confidence intervals. Several methods can achieve this and as we have discussed in the simulation study, there is no method that stands out over all criteria.

In sections 2 and 3, we established the necessary theory for the construction of confidence intervals using the theory of order statistics and extreme value theory, respectively. The empirical method produces CIs based on the df of a single OS or a convex linear combination between two OS. We then defined the generalised Pareto distribution, which is the natural distribution for excesses over a high threshold. The parameters are computed by the ML approach, which gives us a formula for a high quantile. The last step consists in inverting the likelihood ratio test to obtain a CI for the estimator of a high VaR.

Finally, in Section 4, we presented the results of the simulation study and provided the scope in which each method performs better. The experiment distinguished medium and heavy-tailed distributions. The empirical method is straightforward, fast, and easy to implement. For very high quantiles, however, the results are unsatisfactory, as mentioned in the introduction. We could improve upon it by using an alternative definition of the sample quantile as described in Section 2.2. The EVT method works well for certain distribution functions, but the GPD does not always converge, especially for the normal distribution. For very high quantiles ($\alpha \geq 0.99$), the EVT method seems to outperform the empirical one. In the context of insurance, this is a convenient choice, since under Solvency II, one needs to compute $\text{VaR}_{0.995}$.

The main challenge with the POT method is the choice of a suitable threshold. For an individual study, we have given a graphical method using the Mean-Excess plot. Clearly this is not possible in this experiment with 10,000 replications. Consequently, we can expect improved results when the threshold is chosen manually for a specific data set.

In our tail approximation (Equation (5)), we have taken the ML approach to estimate the parameters ξ and β. Yet, we used an empirical estimator to evaluate the probability of exceeding the threshold. McNeil, Frey and Embrechts [10] state that *"it is quite easy to give confidence intervals that take into account the uncertainty in $\hat{\xi}$ and $\hat{\beta}$, but neglect the uncertainty in N_u/n as an estimator of $\bar{F}(u)$."* We therefore assume our sample is large enough to estimate $\bar{F}(u)$ accurately. Advancements can be made on this estimator by allowing for this uncertainty. Further research could focus on measuring how far the true quantile is from the closest bound of the CI when the latter does not contain the true value.

The EVT method provides a good estimation of the uncertainty in risk measures evaluation. For small samples, we might obtain larger CIs, but they are more likely to contain the true quantile. Furthermore, distributions such as the lognormal, the Pareto or the Student's-t distributions are popular choices for loss distributions.

A Modes of convergence

The following definitions can be found in Durrett [3]. Let $X, X_1, X_2, ...$, be random variables.
i) X_n converges to X almost surely, denoted $X_n \to X$ *a.s.*, if

$$P(\lim_{n\to\infty} X_n = X) = 1.$$

ii) X_n converges to X in mean $L^r,\ r \geq 1$, denoted $X_n \xrightarrow{L^r} X$, if

$$E[|X_n^r|] < \infty,\ \forall n \text{ and } \lim_{n\to\infty} E[|X_n - X|^r] = 0.$$

iii) X_n converges to X in probability, denoted $X_n \xrightarrow{P} X$, if

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0,\ \forall \epsilon > 0.$$

iv) X_n converges to X in distribution, denoted $X_n \xrightarrow{d} X$, if

$$\lim_{n\to\infty} F_n(x) = F(x),\ \forall F\text{-continuity points x},$$

where F_n and F are the distribution functions of X_n and X respectively.

B Probability distributions

B.1 Exponential

The exponential distribution F_λ has probability density function

$$f_\lambda(x) = \lambda x^{-\lambda x},$$

where $\lambda > 0$ is the rate and $x > 0$.

B.2 Beta

The beta distribution $F_{\alpha,\beta}$ has probability density function

$$f_{\alpha,\beta}(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1},$$

with $x \in [0,1], \alpha, \beta > 0$, and B is the beta function

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt.$$

B.3 Student's-t

The Student's-t distribution F_ν has probability density function

$$f_\nu(x) \propto \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

where ν is known as the degrees of freedom.

B.4 Pareto

The Pareto distribution F_{α,x_m} has cumulative distribution function

$$F_{\alpha,x_m} = 1 - \left(\frac{x_m}{x}\right)^\alpha,$$

where $x_m, \alpha > 0$ are the shape and scale parameters, respectively, and $x \geq x_m$.

B.5 Pareto type II

The Pareto type II distribution $F_{\alpha,x_m,\mu}(x)$ has cumulative distribution function

$$F_{\alpha,x_m,\mu}(x) = 1 - \left(1 + \frac{x-\mu}{x_m}\right)^{-\alpha},$$

where $x > \mu, \mu \in \mathbb{R}$.

B.6 Lognormal

If the rv X is lognormally distributed, then $Y = \ln(X)$ has a normal distribution. Just like for the normal distribution, the parameters are the location $\mu \in \mathbb{R}$ and the scale $\sigma > 0$.

C R functions

C.1 Distribution of order statistics

```
plot(true.t, q50.t, type="n", ylim=range(q975.t, q025.t), xlab="True quantile", yla
axis(3, at=qst(qs, df=nu, scale=TRUE), labels=qs)
abline(v=qst(qs, df=nu, scale=TRUE), col="lightgray")
lines(true.t, true.t)
lines(true.t, q50.t, col=4)
lines(true.t, q10.t, col=3)
lines(true.t, q90.t, col=3)
lines(true.t, q025.t, col=2)
lines(true.t, q975.t, col=2)
```

```
plot(true.t,q50.t-true.t,type="n",ylim=range(q975.t-true.t,q025.t-true.t),xla
axis(3,at=qst(qs,df=nu,scale=TRUE),labels=qs)
abline(v=qst(qs,df=nu,scale=TRUE),col="lightgray")
abline(h=0)
lines(true.t,q50.t-true.t,col=4)
lines(true.t,q10.t-true.t,col=3)
lines(true.t,q90.t-true.t,col=3)
lines(true.t,q025.t-true.t,col=2)
lines(true.t,q975.t-true.t,col=2)
```

C.2 Bisection algorithm

```
## Bisection Algorithm returning a percentile point of binomial
BiAlgo <- function(n,percentile,alpha){
  # n is the sample size, percentile, alpha is the quantile
  a=0
  b=1
  ga=-1
  gb=+1
  r=n*alpha
  eps=.Machine$double.eps

  while (b-a > eps*b){
    x=(a+b)/2
    gx = 1-percentile-pbinom(r-1,n,x)
    if (sign(gx)==sign(ga)){
      a=x
      ga = gx
    }
    else {
      b=x
      gb=gx
    }
  }
  return(x)
}
```

C.3 Confidence intervals based on order statistics

```
## Percentile points of a beta distribution
beta.PP <- function(x,p,P){
  # x is the sample; p is the quantile; P is the confidence level
  n <- length(x)
  m <- ceiling(n*p)
  return(c(qbeta((1-P)/2,m,n-m+1),qbeta((1+P)/2,m,n-m+1)))
```

```
}

## PE and CI of a quantile using one OS
q.OS <- function(x,p,P){
  # x is the sample; p is the quantile; P is the confidence level
  n <- length(x)
  m <- ceiling(n*p)
  pp <- beta.PP(x,p,P)
  CI <- as.numeric(c(quantile(x,pp[1]),quantile(x,pp[2])))
  return(c(CI[1],sort(x)[m],CI[2]))
}
```

C.4 Confidence intervals based on extreme value theory

```
ML.H0.GPD <- function(xVaR,data,u,k,xi){
  parloglik <- function(xi,xVaR,u,k,data){
    beta <- xi * (xVaR-u)/(k^-xi-1)
    if(beta<=0)
      out <- 1e+19
    else
      out <- -sum(dGPD(data-u,xi,beta,log=TRUE))
    out
  }
  optimfit <- optim(xi, fn=parloglik,xVaR,u,k,data,method="BFGS")
  xi.est <- optimfit$par
  llmax <- -parloglik(xi.est,xVaR,u,k,data)
  list(xi=xi.est,VaR=xVaR,llmax=llmax,conv=optimfit$convergence)
}

rootfunc <- function(xVaR,exceedances,u,k,xi,global.max,P){
  ML.H0.GPD(xVaR,exceedances,u,k,xi)$llmax - global.max
   + qchisq(1-P,df=1)/2
}

q.GPD <- function(x,VaR,exceedances,u,k,xi,global.max,P){
  lower <- uniroot(f=rootfunc,interval=c(VaR/4,VaR),
        exceedances,u,k,xi,global.max,P)
  VaR.lower <- lower$root
  upper <- uniroot(f=rootfunc,interval=c(VaR,4*VaR),
        exceedances,u,k,xi,global.max,P)
  VaR.upper <- upper$root
  return(c(VaR.lower,VaR,VaR.upper))
}
```

C.5 Simulation study

```
## Variables which determine the simulation study
varList <-
  varlist(
    # Replications
    n.sim = list(type="N",expr=quote(N),value=10000),
    # Sample size
    n = list(type="grid", value = c(100,500,1000)),
    # Confidence levels
    P = list(type="grid", value = c(0.90,0.95,0.99)),
    # Quantile levels
    alpha = list(type="grid", value = c(0.95,0.975,0.99)),
    # Method
    method = list(type="grid",expr=quote(Method),value=c("GPD","OS")),
    # Distribution
    Df = list(type="grid",expr=quote(Df),value=c("Pareto11","t1",
                                "Pareto21", "t2",
                                "Lognormal","Normal")))

doOne <- function(Df,n,alpha,P,method){
  data <- switch(Df,
                 Normal = rnorm(n),
                 Lognormal = rlnorm(n),
                 Pareto21 = rpareto(n,1,2),
                 Pareto11 = rpareto(n,1,1),
                 t1 = rt(n,df=1),
                 t2 = rt(n,df=2))
  trueq <- switch(Df,
                  Normal = qnorm(alpha),
                  Lognormal = qlnorm(alpha),
                  Pareto21 = qpareto(alpha,1,2),
                  Pareto11 = qpareto(alpha,1,1),
                  t1 = qt(alpha,df=1),
                  t2 = qt(alpha,df=2))
  #method.func <- eval(parse(text=method))
  if(method=="GPD"){
    Nu <- n/4 #Number of exceendances
    u <- findthreshold(data,Nu) #Threshold
    mod1 <- fit.GPD(data,threshold=u) #GPD fit
    global.max <- mod1$ll.max
    xi <- as.numeric(mod1$par.ests[1])
    beta <- as.numeric(mod1$par.ests[2])
    k <- (1-alpha)/(Nu/n)
    VaR <- u+ beta*(k^(-xi)-1)/xi
    exceedances <- mod1$data
    CI <- q.GPD(data,VaR,exceedances,u,k,xi,global.max,1-P)
```

```
  }
  else {CI <- q.OS(data, alpha, P)}
  inCI <- as.numeric((trueq>CI[1])&(trueq<CI[3]))
  lengthCI <- CI[3]-CI[1]
  return(c(lengthCI, inCI))
}
```

D Profile likelihood method and the likelihood ratio test

Let $\mathbf{X} = (X_1, ..., X_n)$ be a random vector with joint pdf $f_{\mathbf{X}}(\boldsymbol{x}; \boldsymbol{\theta})$ with $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)$. For realisations $\boldsymbol{x}$, we can use the maximum likelihood method to estimate an unknown parameter θ.

The likelihood function for $\boldsymbol{\theta}$ is

$$L(\boldsymbol{\theta}; \boldsymbol{X}) = f_{\mathbf{X}}(\boldsymbol{X}; \boldsymbol{\theta}).$$

The maximum likelihood estimator (MLE), $\boldsymbol{\theta}_{MLE}$, is the value of $\boldsymbol{\theta}$ which maximises the likelihood function.

In practice, some elements of $\boldsymbol{\theta}$ might be more relevant to the study in question than others. Therefore we divide $\boldsymbol{\theta}$ into $(\boldsymbol{\psi}, \boldsymbol{\lambda})$. We say that $\boldsymbol{\psi}$ is the vector of parameters of interest and that $\boldsymbol{\lambda}$ is the vector of nuisance parameters. For instance, we might want to construct a confidence interval for $\boldsymbol{\psi}$. Even though the value of $\boldsymbol{\lambda}$ is irrelevant, we cannot exclude it from the model. But by profiling out the parameter $\boldsymbol{\lambda}$, one can avoid constructing a confidence interval for $\boldsymbol{\lambda}$ too.

Two models are said to be nested if one model reduces to the other when certain parameter values are constrained. We use the likelihood ratio test (LRT) to compare two nested models.

Assume we want to test the hypothesis

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \text{ v } H_1 : \boldsymbol{\theta} \in \Theta_0^c,$$

with $\Theta_0 \subset \Theta$. The Likelihood Ratio Test statistic is

$$\Lambda(X) = \frac{\sup_{\theta \in \Theta_0} L(\boldsymbol{\theta}; \boldsymbol{X})}{\sup_{\theta \in \Theta} L(\boldsymbol{\theta}; \boldsymbol{X})}.$$

Under the null hypothesis, $-2 \ln \Lambda(\boldsymbol{X}) \sim \chi^2_\nu$, for $n \to \infty$, where the degrees of freedom are given by $\nu = \#\{\text{free parameters specified by } \Theta\} - \#\{\text{free parameters specified by } \Theta_0\}$. Suppose $\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\lambda})$ and we want to test $H_0 : \boldsymbol{\psi} = \boldsymbol{\psi}_0$ against $H_1 : \boldsymbol{\psi} \neq \boldsymbol{\psi}_0$. Then, the likelihood ratio test statistic satisfies

$$-2 \ln \Lambda(\boldsymbol{X}) = -2\Big(\ln L(\boldsymbol{\psi}_0, \hat{\boldsymbol{\lambda}}_0; \boldsymbol{X}) - \ln L(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\lambda}}; \boldsymbol{X}) \Big) \sim \chi^2_q$$

where $\hat{\boldsymbol{\lambda}}_0$ is the MLE of $\boldsymbol{\lambda}$ under H_0, and $\hat{\boldsymbol{\psi}}$, $\hat{\boldsymbol{\lambda}}$ are the unconstrained MLEs. We would reject H_0 if

$$-2 \ln \Lambda(\boldsymbol{X}) > c_{q,1-\alpha},$$

where $c_{q,1-\alpha}$ is the $(1-\alpha)$-quantile of the χ^2_q distribution.

Now, suppose we want to construct a confidence interval for $\boldsymbol{\psi}$ (for instance the parameter ξ of a GPD $G_{\xi,\beta}$). We need to invert the likelihood ratio test, and therefore, a $100(1-\alpha)\%$ confidence interval is given by the values for which H_0 is **not** rejected, *i.e.*

$$\begin{aligned} &\{\boldsymbol{\psi}_0 \mid -2\ln\Lambda(\boldsymbol{X}) \geq c_{q,1-\alpha}\}, \\ \iff &\{\boldsymbol{\psi}_0 \mid \ln L(\boldsymbol{\psi}_0, \hat{\boldsymbol{\lambda}}_0; \boldsymbol{X}) \geq \ln L(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\lambda}}; \boldsymbol{X}) - 0.5 c_{q,1-\alpha}\}. \end{aligned}$$

In the case of the GPD, we have $q = 1$ and we test $H_0 : \xi = \xi_0$ against $H_1 : \xi \neq \xi_0$. A 95%CI is therefore given by

$$\{\xi_0 : \ln L(\xi_0, \hat{\beta}_0; \boldsymbol{x}) \geq \ln L(\hat{\xi}, \hat{\beta}, \boldsymbol{x}) - 0.5 \underbrace{c_{1,0.95}}_{=3.84}\}. \tag{6}$$

The curve $(\xi_0, \ln L(\xi_0, \hat{\beta}_0; \boldsymbol{x})$ is referred to as the profile log-likelihood curve.

References

[1] Herbert Aron David and Haikady Navada Nagaraja. *Order statistics*. Wiley online library, 1981.

[2] Kevin Dowd. Using order statistics to estimate confidence intervals for probabilistic risk measures. *The Journal of Derivatives*, 14(2):77–81, 2006.

[3] Rick Durrett. *Probability: theory and examples*, volume 3. Cambridge university press, 2010.

[4] Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media, 2013.

[5] Boris Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, pages 423–453, 1943.

[6] Geoffrey Grimmett and David Stirzaker. *Probability and random processes*. Oxford university press, 2001.

[7] Marius Hofert, Martin Maechler, and Maintainer Marius Hofert. Package 'simsalapar'. 2016.

[8] Jonathan RM Hosking and James R Wallis. Parameter and quantile estimation for the generalized pareto distribution. *Technometrics*, 29(3):339–349, 1987.

[9] Rob J Hyndman and Yanan Fan. Sample quantiles in statistical packages. *The American Statistician*, 50(4):361–365, 1996.

[10] Alexander J McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative risk management: concepts, techniques and tools*. Princeton university press, 2015.

[11] Alexander J McNeil and Thomas Saladin. The peaks over thresholds method for estimating high quantiles of loss distributions. In *Proceedings of 28th International ASTIN Colloquium*, pages 23–43, 1997.

[12] Rolf-Dieter Reiss. *Approximate distributions of order statistics: with applications to nonparametric statistics*. Springer science & business media, 2012.

[13] Sidney I Resnick. *Extreme values, regular variation and point processes*. Springer, 1987.

[14] Richard L Smith. Threshold methods for sample extremes. In *Statistical extremes and applications*, pages 621–638. Springer, 1984.

[15] Richard L Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90, 1985.

[16] Richard L Smith. Estimating tails of probability distributions. *The annals of Statistics*, pages 1174–1207, 1987.