Estimating extreme quantiles of Weibull tail-distributions

Laurent Gardes, Stéphane Girard

Estimating extreme quantiles of Weibull tail-distributions

Laurent Gardes & Stéphane Girard

SMS/LMC, Université Grenoble 1, BP 53,
38041 Grenoble cedex 9, France.
{Laurent.Gardes,Stephane.Girard}@imag.fr

Abstract: We present a new estimator of extreme quantiles dedicated to Weibull tail-distributions. This estimate is based on a consistent estimator of the Weibull tail-coefficient. This parameter is defined as the regular variation coefficient of the inverse cumulative hazard function. We give conditions in order to obtain the weak consistency and the asymptotic distribution of the extreme quantiles estimator. Its asymptotic as well as its finite sample performances are compared to classical ones.

1 Introduction

Let $X_1, X_2, \dots, X_n$ be a sequence of independent and identically distributed random variables with cumulative distribution function F. We address the problem of estimating extreme quantiles of Weibull tail-distributions. In such a case, the tail of such distributions satisfies

$$\textbf{(A.1)}: 1 - F(x) = \exp(-H(x)),\ V(t) = H^{\leftarrow}(t) = \inf\{x,\ H(x) \geq t\} = t^{\theta}\ell(t),$$

where $\theta > 0$ is called the Weibull tail-coefficient and where ℓ is a slowly varying function *i.e.*

$$\ell(\lambda x)/\ell(x) \to 1 \text{ as } x \to \infty \text{ for all } \lambda > 0.$$

The inverse cumulative hazard function V is said to be regularly varying at infinity with index θ and this property is denoted by $V \in \mathcal{R}_\theta$, see [4] for more details on this topic. When ℓ is a constant function, this problems reduces to estimating extreme quantiles of a Weibull distribution. Distributions with non-constant slowly varying functions include for instance normal, gamma and extended Weibull distributions (see Section 3 for their definitions). Such distributions have been used to model large claims in non-life insurance [3]. They can be be classified as sub-exponential distributions ($\theta < 1$), exponential-like distributions ($\theta = 1$) and super-exponential distributions ($\theta > 1$).
An extreme quantile x_{p_n} of order p_n is defined by the equation

$$1 - F(x_{p_n}) = p_n, \text{ with } 0 < p_n < 1/n. \tag{1}$$

The condition $p_n < 1/n$ is very important in this context. It usually implies that x_{p_n} is larger than the maximum observation of the sample. This necessity to extrapolate sample results to areas where no data are observed occurs in reliability [10], hydrology [24], finance [13],... Dedicated estimation methods have been proposed to address the problem of estimating extreme quantiles. Most of these methods rely on two steps (see Section 5 for some examples). First, a classical quantile u_n of order c_n defined by

$$1 - F(u_n) = c_n, \text{ with } 1/n \leq c_n < 1, \tag{2}$$

is estimated by the corresponding order statistic $X_{n-k_n+1,n}$, where $k_n = nc_n$. The condition $c_n \to 0$ as $n \to \infty$ is required so that u_n remains in the distribution tail. Second, an extrapolation of the inverse distribution tail is computed from c_n to p_n. Our estimate belongs to this family. It is defined by

$$\hat{x}_{p_n} = X_{n-k_n+1,n} \left(\frac{\log(1/p_n)}{\log(1/c_n)} \right)^{\hat{\theta}_n} \tag{3}$$

where $\hat{\theta}_n$ is an estimator of the Weibull tail-coefficient θ. Some conditions will be imposed to $\hat{\theta}_n$ later, and some examples will be given in Section 4. The estimator (3) is motivated by the remark that, under **(A.1)**, we have

$$\frac{F^{\leftarrow}(1-x)}{F^{\leftarrow}(1-y)} = \frac{V(\log(1/x))}{V(\log(1/y))} \simeq \left(\frac{\log(1/x)}{\log(1/y)} \right)^{\theta},$$

where x and y are close to zero. In Section 2 we state the main asymptotic properties of our estimate. These results are illustrated in Section 3 on some examples of distributions. Some examples of Weibull tail-estimates are compared in Section 4 and some estimates of extreme quantiles are listed in Section 5. Finally, these estimates are compared in Section 6 on finite sample situations.

2 Asymptotic results

In this section, we state our main results on the limiting behavior of the estimate $\hat{x}_{p_n}$. Its weak consistency is established in Theorem 1 and its asymptotic normality is studied in Theorem 2 and Theorem 3. The proofs are postponed to the Appendix. For the sake of simplicity, in the following, we shall use the notation

$$\tau_n = \frac{\log(1/p_n)}{\log(1/c_n)},$$

leading to $\hat{x}_{p_n} = X_{n-k_n+1,n}(\tau_n)^{\hat{\theta}_n}$. Let us note that (1) and (2) imply $\tau_n > 1$. The weak consistency of $\hat{x}_{p_n}$ requires the weak consistency of $\hat{\theta}_n$:

Theorem 1 *Suppose* **(A.1)** *holds. If* $k_n \to \infty$, $k_n/n \to 0$ *and*

$$1 \leq \liminf \tau_n \leq \limsup \tau_n < \infty \tag{4}$$

then $\hat{\theta}_n \xrightarrow{P} \theta$ *implies* $\hat{x}_{p_n}/x_{p_n} \xrightarrow{P} 1$ *as* $n \to \infty$.

Condition (4) can be interpreted as a limitation on the order of the extreme quantiles that can be consistently estimated. The order $\log p_n$ of such extreme quantiles should be asymptotically close to the order $\log c_n$ of the classical quantile u_n. Let us now introduce the deterministic approximation of the extreme quantile

$$\tilde{x}_{p_n} = u_n \left(\frac{\log(1/p_n)}{\log(1/c_n)} \right)^{\theta} = u_n \tau_n^{\theta}, \tag{5}$$

where u_n is defined in (2): $u_n = V(\log(1/c_n))$.
In the following, we examine the asymptotic distribution of $\hat{x}_{p_n}/\tilde{x}_{p_n}$ and $\hat{x}_{p_n}/x_{p_n}$. Two main situations appear: Either $\hat{\theta}_n$ converges to θ sufficiently rapidly and then the asymptotic normality of the two previous ratios can be established. This situation is denoted by **(S.1)** and can be summarized by:

(S.1) There exists a sequence (β_n) such that: $\log(\tau_n)\log(1/c_n)k_n^{1/2}(\hat{\theta}_n - \beta_n - \theta) \xrightarrow{P} 0$.

Or, in the converse situation, the limit distribution of $\hat{x}_{p_n}/\tilde{x}_{p_n}$ and $\hat{x}_{p_n}/x_{p_n}$ is driven by $\hat{\theta}_n$. This situation is described by the following conditions:

(S.2) There exist two sequences (α_n) and (β_n) and a distribution $\mathcal{L}$ such that: $\alpha_n(\hat{\theta}_n - \beta_n - \theta) \xrightarrow{d} \mathcal{L}$ with $\log(\tau_n) = o(\alpha_n)$ and $\alpha_n = o(\log(\tau_n)\log(1/c_n)k_n^{1/2})$.

In both situations, β_n represents the asymptotic bias of the Weibull tail-coefficient estimate. In situation **(S.2)**, α_n controls the asymptotic rate of convergence of the estimate. Of course, $\hat{x}_{p_n}/\tilde{x}_{p_n}$ inherits the bias term from $\hat{\theta}_n$:

Theorem 2 *Suppose* **(A.1)** *holds,* $k_n \to \infty$*, and* $k_n/n \to 0$ *as* $n \to \infty$*.*

- *Under* **(S.1)***:*

$$\log(1/c_n) k_n^{1/2} \tau_n^{-\beta_n} \left(\frac{\hat{x}_{p_n}}{\tilde{x}_{p_n}} - \tau_n^{\beta_n} \right) \xrightarrow{d} \mathcal{N}(0, \theta^2).$$

- *Under* **(S.2)***:*

$$\frac{\alpha_n}{\log(\tau_n)} \tau_n^{-\beta_n} \left(\frac{\hat{x}_{p_n}}{\tilde{x}_{p_n}} - \tau_n^{\beta_n} \right) \xrightarrow{d} \mathcal{L}.$$

The study of the limit distribution of $\hat{x}_{p_n}/x_{p_n}$ requires a second order condition on ℓ: There exist $\rho \le 0$ and $b(x) \to 0$ such that uniformly locally on $\lambda \ge 1$ when $x \to \infty$,

(A.2): $\log\left(\dfrac{\ell(\lambda x)}{\ell(x)}\right) \sim b(x) K_\rho(\lambda)$,

with $K_\rho(\lambda) = \int_1^\lambda u^{\rho-1} du$. It can be shown [15] that necessarily $|b| \in \mathcal{R}_\rho$. The second order parameter $\rho \le 0$ tunes the rate of convergence of $\ell(\lambda x)/\ell(x)$ to 1. The closer ρ is to 0, the slower is the convergence. Condition **(A.2)** is the cornerstone in all proofs of asymptotic normality for extreme value estimates. It is used in [19] to prove the asymptotic normality of the Hill estimate and in [1] for one of its refinements. Our result is the following:

Theorem 3 *Suppose* **(A.1)** *and* **(A.2)** *hold,* $k_n \to \infty$*,* $k_n/n \to 0$ *and* $\tau_n \to 1$ *as* $n \to \infty$*.*

- *Under* **(S.1)***, if moreover*

$$k_n^{1/2} \log(\tau_n) \log(1/c_n) b(\log(1/c_n)) \to 0 \tag{6}$$

  then

$$\log(1/c_n) k_n^{1/2} \tau_n^{-\beta_n} \left(\frac{\hat{x}_{p_n}}{x_{p_n}} - \tau_n^{\beta_n} \right) \xrightarrow{d} \mathcal{N}(0, \theta^2).$$

- *Under* **(S.2)***, if moreover*

$$\alpha_n b(\log(1/c_n)) \to 0 \tag{7}$$

  then

$$\frac{\alpha_n}{\log(\tau_n)} \tau_n^{-\beta_n} \left(\frac{\hat{x}_{p_n}}{x_{p_n}} - \tau_n^{\beta_n} \right) \xrightarrow{d} \mathcal{L}.$$

Conditions (6) and (7) impose that the bias induced by the slowly varying function vanishes at infinity. On the contrary, the bias term due to $\hat{\theta}_n$ cannot always be cancelled, depending on the asymptotic order of (β_n), see Section 4 for examples. Finally, condition $\tau_n \to 1$ is a stronger version of (4).

3 Examples of Weibull tail-distribution

In this section, we give some examples of distributions satisfying assumptions **(A.1)** and **(A.2)**.

Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, $\sigma > 0$. From [13], Table 3.4.4, we have $V(x) = x^{1/2}\ell(x)$ and an asymptotic expansion of the slowly varying function is given by:

$$\ell(x) = 2^{1/2}\sigma - \frac{\sigma}{2^{3/2}} \frac{\log x}{x} + O(1/x).$$

Therefore $\theta = 1/2$, $\rho = -1$ and $b(x) = \log(x)/(4x)$.

Gamma distribution $\Gamma(\beta,\alpha)$, $\alpha,\beta>0$. We use the following parameterization of the density

$$f(x)=\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\exp\left(-\beta x\right).$$

From [13], Table 3.4.4, we obtain $V(x)=x\ell(x)$ with

$$\ell(x)=\frac{1}{\beta}+\frac{\alpha-1}{\beta}\frac{\log x}{x}+O(1/x).$$

We thus have $\theta=1$, $\rho=-1$ and $b(x)=(1-\alpha)\log(x)/x$.

Weibull distribution $\mathcal{W}(\alpha,\lambda)$, $\alpha,\lambda>0$. The inverse cumulative hazard function is $V(x)=\lambda x^{1/\alpha}$, and then $\theta=1/\alpha$, $\ell(x)=\lambda$ for all $x>0$. Therefore $b(x)=0$ and we use the usual convention $\rho=-\infty$.

Extended Weibull distribution $\mathcal{E}W(\alpha,\beta)$, $\alpha\in(0,1)$, $\beta\in\mathbb{R}$. This distribution is introduced in [22]. Its distribution tail is given by:

$$1-F(x)=r(x)\exp(-x^\alpha), \tag{8}$$

where $r\in\mathcal{R}_\beta$. It follows that $V(x)=x^{1/\alpha}\ell(x)$ and the following asymptotic expansion holds:

$$\ell(x)=1+\frac{\beta}{\alpha^2}\frac{\log x}{x}+O(1/x).$$

It is easily seen that $\theta=1/\alpha$, $\rho=-1$ and $b(x)=-\beta\log(x)/(\alpha^2x)$. In [22], it is remarked that this model encompasses the normal and gamma distributions.
The parameters θ and ρ as well as the auxiliary function $b(x)$ associated to these distributions are summarized in Table 1.

4 Some estimates of the Weibull tail-coefficient

Asymptotically unbiased estimates. Beirlant *et al* [2] propose the following estimate

$$\hat\theta_n^{\mathrm{BBTV}}=\frac{\log(n/k_n)}{X_{n-k_n+1,n}}\frac{1}{k_n-1}\sum_{i=1}^{k_n-1}(X_{n-i+1,n}-X_{n-k_n+1,n}). \tag{9}$$

It is proved that (see [2], Theorem 3.2(i)),

$$k_n^{1/2}(\hat\theta_n^{\mathrm{BBTV}}-\theta)\xrightarrow{d}\mathcal{N}(0,\theta^2), \tag{10}$$

for any sequence (k_n) satisfying

$$k_n\to\infty \text{ and } k_n^{1/2}\max\{b(\log(1/c_n)),1/\log(1/c_n)\}\to 0. \tag{11}$$

The same asymptotic result (10) is also obtained for the estimate proposed in [16]:

$$\hat\theta_n^{\mathrm{G}}=\sum_{i=1}^{k_n-1}(\log(X_{n-i+1,n})-\log(X_{n-k_n+1,n}))\Bigg/\sum_{i=1}^{k_n-1}\left(\log_2\left(n/i\right)-\log_2\left(n/k_n\right)\right),$$

where $\log_2(x)=\log(\log(x))$, $x>1$. Moreover, this convergence holds under the same conditions (11), see [16], Theorem 2. Then, in both cases we are in situation **(S.2)**, with $\alpha_n=k_n^{1/2}$, $\beta_n=0$ and $\mathcal{L}=\mathcal{N}(0,\theta^2)$. Denote by $\hat x_{p_n}^{\mathrm{BBTV}}$ and $\hat x_{p_n}^{\mathrm{G}}$ the estimates obtained by plugging respectively $\hat\theta_n=\hat\theta_n^{\mathrm{BBTV}}$ and $\hat\theta_n=\hat\theta_n^{\mathrm{G}}$ in (3). The following corollary can be given:

Corollary 1 *Assume* **(A.1)** *and* **(A.2)** *hold and (11) is verified. If* $\tau_n \to 1$, *then, for* $\hat{x}_{p_n} = \hat{x}^{BBTV}_{p_n}$ *or* $\hat{x}_{p_n} = \hat{x}^{G}_{p_n}$,

$$\frac{\log(1/c_n) k_n^{1/2}}{\log(c_n/p_n)} \left(\frac{\hat{x}_{p_n}}{x_{p_n}} - 1 \right) \xrightarrow{d} \mathcal{N}(0, \theta^2).$$

In the Weibull distribution case, conditions reduce to

$$k_n \to \infty, \; k_n^{1/2} / \log(1/c_n) \to 0, \; \log(1/p_n) \sim \log(1/c_n).$$

Possible choices are $k_n = \varepsilon_n (\log n)^2$ and $\log(1/p_n) \sim \log(n)$, where $\varepsilon_n \to 0$ and $\varepsilon_n (\log n)^2 \to \infty$. For Extended-Weibull distributions (including gamma and Gaussian distributions, see Section 3), conditions are

$$k_n \to \infty, \; k_n^{1/2} \log_2(1/c_n) / \log(1/c_n) \to 0, \; \log(1/p_n) \sim \log(1/c_n).$$

Possible choices are $k_n = \varepsilon_n (\log n)^2$, $\log(1/p_n) \sim \log(n)$, where $\varepsilon_n (\log_2 n)^2 \to 0$, $\varepsilon_n (\log n)^2 \to \infty$. Estimates $\hat{x}^{\mathrm{BBTV}}_{p_n}$ and $\hat{x}^{\mathrm{G}}_{p_n}$ are compared on simulated data in Section 6.

Asymptotically biased estimates. In [6], another estimate of the Weibull tail coefficient is proposed. It is defined by

$$\hat{\theta}^{\mathrm{B}}_n = \frac{1}{k_n} \sum_{i=1}^{k_n - 1} \frac{\log(X_{n-i+1,n})}{\log_2(n/i)},$$

and the author proved that

$$k_n^{1/2} \log(1/c_n) \log_2(1/c_n) \left(\hat{\theta}^{\mathrm{B}}_n - \beta_n - \theta \right) \xrightarrow{d} \mathcal{N}(0, 2\theta^2), \tag{12}$$

for any sequence (k_n) satisfying

$$k_n \to \infty, \; k_n/n \to 0, \text{ and } \frac{\log(1/c_n) \log_2(1/c_n) \log(n)}{k_n^{1/2} \log_2(n)} \to 0, \tag{13}$$

where β_n is a bias coefficient that we do not reproduce here. It is shown in [16], Paragraph 4.1, that this bias term is not asymptotically negligible in the case of a Weibull distribution (*i.e.* when ℓ is a constant function) since

$$\beta_n k_n^{1/2} \log(1/c_n) \log_2(1/c_n) \to \infty.$$

It can be shown that this corresponds to situation **(S.1)**. Thus, denoting by $\hat{x}^{\mathrm{B}}_{p_n}$ the estimate obtained by plugging $\hat{\theta}^{\mathrm{B}}_n$ in (3) yields the following corollary:

Corollary 2 *Assume* **(A.1)** *and* **(A.2)** *hold and (6), (13) are verified. If* $\tau_n \to 1$, *then*

$$\log(1/c_n) k_n^{1/2} \tau_n^{-\beta_n} \left(\frac{\hat{x}^{B}_{p_n}}{x_{p_n}} - \tau_n^{\beta_n} \right) \xrightarrow{d} \mathcal{N}(0, \theta^2).$$

5 Other estimates of extreme quantiles

Most of extreme quantile estimates are dedicated to a specific class of distributions. To describe these classes, it is convenient to introduce the notion of domain of attraction. First, define the excesses above u_n (see equation (2)) on the basis of the $X_i > u_n$ by $Y_i = X_i - u_n$. The tail distribution of the excesses is thus defined by

$$1 - F_{u_n}(x) = \frac{1 - F(x + u_n)}{1 - F(u_n)}, \; x \geq 0.$$

Pickand's theorem [23, 14] states that under some regularity conditions, this tail distribution can be approximated by a generalized Pareto distribution (GPD), which tail distribution is defined by

$$1 - G(x;\sigma,\xi) = \begin{cases} \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ \exp\left(-\frac{x}{\sigma}\right) & \text{if } \xi = 0 \end{cases} \tag{14}$$

with $x > 0$ if $\xi \geq 0$ and $x \in [0, -\sigma/\xi[$ otherwise. The parameter ξ is called the extreme value index, it describes the asymptotic behaviour of the tail of the distribution: When $\xi > 0$, then the distribution F is heavy tailed, it belongs to the domain of attraction of Fréchet (noted DA(Fréchet)). If $\xi = 0$, then the distribution F has a tail decreasing exponentially fast, it belongs to the domain of attraction of Gumbel (noted DA(Gumbel)). Let us highlight that the Weibull-tail distributions described by **(A.1)** belong to this domain of attraction. Finally, when $\xi < 0$, then the distribution F has a finite end point, it belongs to the domain of attraction of Weibull (noted DA(Weibull)).

Hill estimate. The Hill estimate [20] of the extreme value index ξ is defined by:

$$H_n^{(1)} = \frac{1}{k_n - 1} \sum_{i=1}^{k_n - 1} (\log(X_{n-i+1,n}) - \log(X_{n-k_n+1,n})).$$

The associated Hill estimate for extreme quantiles is [25]:

$$\hat{x}_{p_n}^{\mathrm{H}} = X_{n-k_n+1,n}(c_n/p_n)^{H_n^{(1)}},$$

which is dedicated to the DA(Fréchet). We can note that $\hat{x}_{p_n}^{\mathrm{H}}$ can be obtained by choosing in (3):

$$\hat{\theta}_n^{\mathrm{H}} = \frac{\log(c_n/p_n)}{\log(\tau_n)} H_n^{(1)}.$$

GPD estimates. They are defined by

$$\hat{x}_{p_n}^{\mathrm{GPD}} = X_{n-k_n+1,n} - \frac{\hat{\sigma}_n}{\hat{\xi}_n}\left(1 - (c_n/p_n)^{\hat{\xi}_n}\right), \tag{15}$$

where $\hat{\sigma}_n$ and $\hat{\xi}_n$ are estimators of the corresponding GPD parameters, see (14). Let us focus on two approaches. $\hat{x}_{p_n}^{\mathrm{DEH}}$ is obtained by plugging into (15):

$$\hat{\sigma}_n = X_{n-k_n+1,n} H_n^{(1)} \phi(H_n^{(3)}), \quad \text{and} \quad \hat{\xi}_n = H_n^{(3)}$$

where $\phi(t) = 1$ if $t \geq 0$ and $\phi(t) = 1 - t$ otherwise. $H_n^{(3)}$ is the Generalized Hill estimate [7], defined as

$$H_n^{(3)} = H_n^{(1)} + 1 - \frac{1}{2}\left[1 - \frac{(H_n^{(1)})^2}{H_n^{(2)}}\right]^{-1},$$

where

$$H_n^{(2)} = \frac{1}{k_n - 1} \sum_{i=1}^{k_n - 1} \left[\log(X_{n-i+1,n}) - \log(X_{n-k_n+1,n})\right]^2.$$

$\hat{x}_{p_n}^{\mathrm{HW}}$ is obtained by estimating the GPD parameters by the probability weighted moments [21]:

$$\hat{\sigma}_n = \frac{2\hat{\nu}_0\hat{\nu}_1}{\hat{\nu}_0 - 2\hat{\nu}_1} \quad \text{and} \quad \hat{\xi}_n = 2 - \frac{\hat{\nu}_0}{\hat{\nu}_0 - 2\hat{\nu}_1},$$

where

$$\hat{\nu}_r = \frac{1}{k_n - 1} \sum_{i=1}^{k_n - 1} (1 - p_i)^r \left(X_{n-k_n+1+i,n} - X_{n-k_n+1,n}\right), \quad r \in \{0, 1\}, \quad \text{with } p_i = \frac{i - 0.35}{k_n - 1}.$$

Both estimates $\hat{x}_{p_n}^{\mathrm{DEH}}$ and $\hat{x}_{p_n}^{\mathrm{HW}}$ cover all the domains of attraction.

ET estimate. This estimate, introduced in [5], is dedicated to the DA(Gumbel). It is obtained by setting $\hat{\xi}_n = 0$ and

$$\hat{\sigma}_n = \frac{1}{k_n - 1} \sum_{i=1}^{k_n - 1} (X_{n-i+1,n} - X_{n-k_n+1,n}), \tag{16}$$

in (15) leading to

$$\hat{x}^{\text{ET}}_{p_n} = X_{n-k_n+1,n} + \hat{\sigma}_n \log(c_n/p_n). \tag{17}$$

Weak consistency and asymptotic normality of this estimate are established in [8, 9].

Beirlant *et al* estimate. In [2], the following estimate of extreme quantiles for Weibull tail-distribution is proposed:

$$\hat{x}^{\text{BBTV},2}_{p_n} = X_{n-k_n+1,n} \left(1 + \frac{\hat{\sigma}_n \log(c_n/p_n)}{\hat{\theta}^{\text{BBTV}}_n X_{n-k_n+1,n}} \right)^{\hat{\theta}^{\text{BBTV}}_n}, \tag{18}$$

where $\hat{\sigma}_n$ is defined by (16). Let us point out that the expression of this estimate can be simplified by remarking that

$$1 + \frac{\hat{\sigma}_n \log(c_n/p_n)}{\hat{\theta}^{\text{BBTV}}_n X_{n-k_n+1,n}} = \tau_n,$$

and thus $\hat{x}^{\text{BBTV},2}_{p_n} = \hat{x}^{\text{BBTV}}_{p_n}$, the estimate studied in Corollary 1. The authors show that $\hat{x}^{\text{ET}}_{p_n}$ can be seen as the first order Taylor expansion of $\hat{x}^{\text{BBTV}}_{p_n}$.

6 Numerical experiments

First, the definition of the extrapolation parameter introduced in [12] is recalled. This parameter is quite useful to compare the performances of extreme quantiles estimates. Next, our estimator $\hat{x}^{\text{G}}_{p_n}$ is compared to other estimates of extreme quantiles through a simulation study.

Extrapolation parameter. The extrapolation parameter w_n is an empirical measure of the maximal distance between the largest observation $X_{n,n}$ and the largest extreme quantile x_{p_n} for which the relative error

$$D_{p_n} = \left| \frac{\hat{x}_{p_n} - x_{p_n}}{x_{p_n} - \bar{x}_n} \right|$$

of the estimate $\hat{x}_{p_n}$ remains smaller than 0.3. The introduction of the empirical mean $\bar{x}_n$ in the relative error yields a measure which is invariant with respect to both scale and position parameters. More precisely, to estimate this extrapolation parameter, the authors compute D_{p_n} for $p_n = 1/\{n[\ln(n)]^\delta\}$ with δ in the finite set $\{0, 0.1, 0.2, 0.35, 0.5, 0.75, 1, 2, 3.5, 5, 7.5, 10, 16\}$. Linear interpolation is then used to estimate the extrapolation parameter by $\hat{w}_n$ such that:

$$D_{\hat{p}_n} \approx 0.3 \text{ with } \hat{p}_n = 1/\{n[\ln(n)]^{\hat{w}_n}\} \text{ and } D_{p_n} \leq 0.3 \text{ for } p_n \geq \hat{p}_n.$$

When $D_{p_n} \leq 0.3$ for all the values of δ, then the extrapolation parameter $\hat{w}_n$ is set to 15.

Simulations. The finite sample behavior of our estimator $\hat{x}^{\text{G}}_{p_n}$ is investigated on 12 different distributions classified as follows:

- DA(Fréchet): Pareto distribution with index $\alpha = 2$,
- DA(Weibull): $\mathcal{U}[0, 1]$,
- DA(Gumbel) involves two subclasses of distributions:
  - non Weibull tail-distribution: standard log-normal distribution,

– Weibull tail-distributions:
$\mathcal{W}(0.25,1)$, $\mathcal{W}(0.5,1)$, $\mathcal{W}(0.7,1)$ (Super-exponential distributions),
$\mathcal{W}(1.5,1)$, $\mathcal{W}(2,1)$ (Sub-exponential distributions),
$\mathcal{W}(1,1)$, $\Gamma(2,1)$ (Exponential-like distributions).

For the above mentioned distributions, $N = 1000$ samples $(\mathcal{X}_{n,i})_{i=1,\dots,N}$ of size n were generated. On each sample $\mathcal{X}_{n,i}$, the estimation $\hat{w}_{n,i}$ of the extrapolation parameter is computed for $n = 100, 200, \dots, 1000$. The parameter k_n is chosen as a realization of an uniform distribution on the interval $[n/10, n/5]$. This permits to take into account of the fact that, for most methods, the user does not know how to choose the optimal value of k_n. A part of our future work will consist in proposing a solution to select the best parameter k_n for $\hat{x}^{\mathrm{G}}_{p_n}$. Finally, we present the plot obtained by drawing the points $(n, \tilde{w}_n)$, where $\tilde{w}_n$ is the median value of the sample $(\hat{w}_{n,i})_{i=1,\dots,N}$. The same work is achieved with the estimators $\hat{x}^{\mathrm{HW}}_{p_n}$, $\hat{x}^{\mathrm{ET}}_{p_n}$ and $\hat{x}^{\mathrm{BBTV}}_{p_n}$. The results obtained with estimator $\hat{x}^{\mathrm{H}}_{p_n}$ are not presented here since it appeared that its extrapolation power is always close to zero. Furthermore, the behavior of the estimates $\hat{x}^{\mathrm{HW}}_{p_n}$ and $\hat{x}^{\mathrm{DEH}}_{p_n}$ is quite similar (or if not, the extrapolation parameter of $\hat{x}^{\mathrm{HW}}_{p_n}$ is larger than the one of $\hat{x}^{\mathrm{DEH}}_{p_n}$) and thus, we choose not to present the results associated to the estimate $\hat{x}^{\mathrm{DEH}}_{p_n}$.

Results are collected in Figures 1-3. Figure 1 is dedicated to the results obtained on super-exponential distributions in DA(Gumbel) and Figure 2 is dedicated to the results obtained on sub-exponential distributions in DA(Gumbel). On the upper part of Figure 3 are presented the results obtained on exponential-like distributions, and on its lower part, the results obtained on non DA(Gumbel) distributions are collected. It appears that for the Weibull tail-distributions $\hat{x}^{\mathrm{G}}_{p_n}$ always gives better results than the other estimates. The best extrapolation power is obtained for sub-exponential distributions. Let us note that, in case of exponential-like distributions, the ET estimator $\hat{x}^{ET}_{p_n}$ is competitive with $\hat{x}^{\mathrm{G}}_{p_n}$ and $\hat{x}^{\mathrm{BBTV}}_{p_n}$. In case of heavy tail distributions (DA(Fréchet)) all the considered methods share a very poor extrapolation power. In case of distributions with finite endpoint (DA(Weibull)), unsurprisingly, only $\hat{x}^{\mathrm{HW}}_{p_n}$ is satisfying.

Conclusion and further work. As a conclusion, our opinion is that estimator $\hat{x}^{\mathrm{G}}_{p_n}$ yields very good results for estimating quantiles of Weibull tail-distributions. It would be interesting to extend its range of application to a larger class of distributions in DA(Gumbel). A first step could be to adapt this method to the case $H \in \mathcal{R}_0$ in order to include for instance the log-normal distribution. From a practical point of view, an automatic selection method for k_n should be developed. Various approaches are proposed in the extreme index estimation context, see [18, 11, 17] among others, and could be adapted to our problem.

Appendix: Proofs

Lemma 1 *Suppose $k_n \to \infty$ and $k_n/n \to 0$. Then, under* **(A.1)**,

$$\log(1/c_n) k_n^{1/2} \left(\frac{X_{n-k_n+1,n}}{u_n} - 1 \right) \xrightarrow{d} \mathcal{N}(0, \theta^2),$$

where u_n is defined by (2): $u_n = V(\log(1/c_n))$.

Proof : The first step of the proof consists of showing that

$$\Lambda_n := k_n^{1/2} \frac{X_{n-k_n+1,n} - u_n}{V'(\log(1/c_n))}$$

converges in distribution to a standard Gaussian variable. In this aim, introduce $E_{n-k_n+1,n}$ the $(n-k_n+1)$th order statistic generated by n independent standard exponential random variables. We thus have

$$\Lambda_n \stackrel{d}{=} k_n^{1/2} \frac{V(E_{n-k_n+1,n}) - V(\log(1/c_n))}{V'(\log(1/c_n))}.$$

Defining $\xi_n = k_n^{1/2}(E_{n-k_n+1,n} - \log(1/c_n))$ yields

$$\Lambda_n \stackrel{d}{=} \xi_n + k_n^{1/2} \int_0^{k_n^{-1/2}\xi_n} \left(\frac{V'(s+\log(1/c_n))}{V'(\log(1/c_n))} - 1 \right) ds,$$

and ξ_n converges in distribution to a standard Gaussian distribution, see for instance Lemma 1(ii) in [16]. Now, **(A.1)** implies that $V' \in \mathcal{R}_{\theta-1}$ and thus

$$\frac{V'(s+\log(1/c_n))}{V'(\log(1/c_n))} \to 1 \text{ as } n \to \infty,$$

uniformly locally on $s \in [0, k_n^{-1/2}\xi_n]$. It follows that $\Lambda_n \stackrel{d}{=} \xi_n(1 + o_P(1))$. The second step of the proof is based on the property of the regularly varying functions

$$\frac{\log(1/c_n)V'(\log(1/c_n))}{V(\log(1/c_n))} \to \theta \text{ as } n \to \infty,$$

which gives the result. ■

Lemma 2 *Suppose $k_n \to \infty$ and $k_n/n \to 0$. Then, under* **(A.1)***, the two following expansions hold:*

(i) $\hat{x}_{p_n}/\tilde{x}_{p_n} = (1+\tau_n^{(1)})(1+\tau_n^{(2)})$,

(ii) $\hat{x}_{p_n}/x_{p_n} = (1+\tau_n^{(1)})(1+\tau_n^{(2)})(1+\tau_n^{(3)})$,

where we have defined

- $\tau_n^{(1)} = \dfrac{\xi_n}{k_n^{1/2}\log(1/c_n)}$ *where* $\xi_n \xrightarrow{d} \mathcal{N}(0,\theta^2)$,
- $\tau_n^{(2)} = \exp\left((\hat{\theta}_n - \theta)\log(\tau_n)\right) - 1$,
- $\tau_n^{(3)} = \dfrac{\ell(\log(1/c_n))}{\ell(\log(1/p_n))} - 1$.

Proof : The proof is straightforward since (5) leads to

$$\frac{\hat{x}_{p_n}}{\tilde{x}_{p_n}} = \frac{X_{n-k_n+1,n}}{u_n} (\tau_n)^{\hat{\theta}_n - \theta}.$$

Defining $1+\tau_n^{(1)} = X_{n-k_n+1,n}/u_n$ gives (i) with Lemma 1. Besides,

$$\frac{\tilde{x}_{p_n}}{x_{p_n}} = \frac{V(\log(1/c_n))}{V(\log(1/p_n))} \left(\frac{\log(1/p_n)}{\log(1/c_n)}\right)^\theta = \frac{\ell(\log(1/c_n))}{\ell(\log(1/p_n))},$$

which yields (ii). ■

Proof of Theorem 1. The proof is based on the expansion given in Lemma 2(ii): First, $\tau_n^{(1)} \xrightarrow{P} 0$. Then, since $\hat{\theta}_n \xrightarrow{P} \theta$ and, τ_n remains in a compact subset of $(0, +\infty)$, we have

$$\tau_n^{(2)} = (\hat{\theta}_n - \theta)\log(\tau_n)(1 + o_P(1)) \xrightarrow{P} 0. \tag{19}$$

Finally, $\tau_n^{(3)} \to 0$ as $n \to \infty$ since ℓ is slowly varying. ■

Proof of Theorem 2. In both situations **(S.1)** and **(S.2)**, the proof is based on the expansion given in Lemma 2(i), where

$$\tau_n^{(2)} = \exp[(\hat{\theta}_n - \beta_n - \theta)\log(\tau_n)]\tau_n^{\beta_n} - 1.$$

Introducing

$$\varepsilon_n = (\hat{\theta}_n - \beta_n - \theta)\log(\tau_n),$$

we have $\varepsilon_n = o_P(1)$ and therefore

$$\tau_n^{-\beta_n}(1+\tau_n^{(2)}) = 1 + \varepsilon_n + O_P(\varepsilon_n^2).$$

In situation **(S.1)**, this yields

$$\begin{aligned}
\tilde{\Delta}_n &:= \log(1/c_n)k_n^{1/2}\tau_n^{-\beta_n}\left(\frac{\hat{x}_{p_n}}{\tilde{x}_{p_n}} - \tau_n^{\beta_n}\right)\\
&= \log(1/c_n)k_n^{1/2}\tau_n^{(1)}\tau_n^{-\beta_n}(1+\tau_n^{(2)}) + \log(1/c_n)k_n^{1/2}[\tau_n^{-\beta_n}(1+\tau_n^{(2)}) - 1]\\
&= \xi_n(1+O_P(\varepsilon_n)) + \log(1/c_n)k_n^{1/2}\varepsilon_n(1+O_P(\varepsilon_n))\\
&= \xi_n(1+O_P(\varepsilon_n)) + o_P(1),
\end{aligned}$$

and the result follows. The proof in the situation **(S.2)** is similar: Introduce

$$\begin{aligned}
\tilde{\Delta}'_n &:= \frac{\alpha_n}{\log(\tau_n)}\tau_n^{-\beta_n}\left(\frac{\hat{x}_{p_n}}{\tilde{x}_{p_n}} - \tau_n^{\beta_n}\right)\\
&= \frac{\alpha_n}{\log(\tau_n)}\tau_n^{(1)}\tau_n^{-\beta_n}(1+\tau_n^{(2)}) + \frac{\alpha_n}{\log(\tau_n)}[\tau_n^{-\beta_n}(1+\tau_n^{(2)}) - 1]\\
&= \xi_n\frac{\alpha_n}{\log(\tau_n)k_n^{1/2}\log(1/c_n)}(1+O_P(\varepsilon_n)) + \frac{\alpha_n\varepsilon_n}{\log(\tau_n)}(1+O_P(\varepsilon_n))\\
&= \alpha_n(\hat{\theta}_n - \beta_n - \theta)(1+o_P(1)) + o_P(1),
\end{aligned}$$

which gives the result. ∎

Proof of Theorem 3. In the situation **(S.1)**, we define

$$\Delta_n := \log(1/c_n)k_n^{1/2}\tau_n^{-\beta_n}\left(\frac{\hat{x}_{p_n}}{x_{p_n}} - \tau_n^{\beta_n}\right).$$

In view of Lemma 2(ii), it follows that

$$\Delta_n = \tilde{\Delta}_n(1+\tau_n^{(3)}) + \log(1/c_n)k_n^{1/2}\tau_n^{(3)}.$$

Theorem 2 yields $\tilde{\Delta}_n \xrightarrow{d} \mathcal{N}(0,\theta^2)$, and it has already been seen that $\tau_n^{(3)} = o_P(1)$ in the proof of Theorem 1. It only remains to prove that $\log(1/c_n)k_n^{1/2}\tau_n^{(3)} \to 0$ as $n\to\infty$. Taking into account that $\tau_n^{(3)} \sim \log(1+\tau_n^{(3)})$ yields

$$\log(1/c_n)k_n^{1/2}\tau_n^{(3)} = \log(1/c_n)k_n^{1/2}\log\left(\frac{\ell(\log(1/c_n))}{\ell(\log(1/p_n))}\right)(1+o(1)).$$

Now, **(A.2)** implies that

$$\begin{aligned}
\log(1/c_n)k_n^{1/2}\tau_n^{(3)} &= \log(1/c_n)k_n^{1/2}\frac{b(\log(1/c_n))}{\rho}(\tau_n^{\rho} - 1)(1+o(1))\\
&= -k_n^{1/2}\log(\tau_n)\log(1/c_n)b(\log(1/c_n))(1+o(1))
\end{aligned}$$

which converges to 0 with condition (6). In the situation **(S.2)**, the proof is similar. Introduce

$$\begin{aligned}\Delta'_n &:= \frac{\alpha_n}{\log(\tau_n)} \tau_n^{-\beta_n} \left(\frac{\hat{x}_{p_n}}{x_{p_n}} - \tau_n^{\beta_n} \right) \\ &= \tilde{\Delta}'_n (1 + \tau_n^{(3)}) + \frac{\alpha_n}{\log(\tau_n)} \tau_n^{(3)}.\end{aligned}$$

Theorem 2 yields $\tilde{\Delta}'_n \xrightarrow{d} \mathcal{L}$, and thus it remains to prove that $\alpha_n \tau_n^{(3)} / \log(\tau_n) \to 0$.

$$\begin{aligned}\frac{\alpha_n}{\log(\tau_n)} \tau_n^{(3)} &= \frac{\alpha_n}{\log(\tau_n)} \log \left(\frac{\ell(\log(1/c_n))}{\ell(\log(1/p_n))} \right) (1 + o(1)), \\ &= -\frac{\alpha_n}{\log(\tau_n)} \frac{b(\log(1/c_n))}{\rho} (\tau_n^{\rho} - 1)(1 + o(1)) \\ &= -\alpha_n b(\log(1/c_n))(1 + o(1))\end{aligned}$$

which converges to 0 with condition (7). ■

Proof of Corollary 1. Let us prove that condition **(S.2)** holds with $\alpha_n = k_n^{1/2}$, $\beta_n = 0$ and $\mathcal{L} = \mathcal{N}(0, \theta^2)$. In view of (10), it only remains to verify that

$$\log(\tau_n) = o(\alpha_n) \text{ and } \alpha_n = o(\log(\tau_n) \log(1/c_n) k_n^{1/2}).$$

Remarking that $\log(\tau_n) \to 0$ and $\alpha_n \to \infty$ straightforwardly yields $\log(\tau_n) = o(\alpha_n)$. Now,

$$\frac{\alpha_n}{\log(\tau_n) \log(1/c_n) k_n^{1/2}} \sim \frac{1}{\log(c_n/p_n)} \leq \frac{1}{\log(k_n)} = o(1)$$

since $p_n < 1/n$ by definition of an extreme quantile. The conclusion follows. ■

Proof of Corollary 2. It suffices to prove that condition **(S.1)** holds. To this end, consider

$$\begin{aligned}\log(\tau_n) \log(1/c_n) k_n^{1/2} (\hat{\theta}_n^{\mathrm{B}} - \beta_n - \theta) &= k_n^{1/2} \log(1/c_n) \log_2(1/c_n) (\hat{\theta}_n^{\mathrm{B}} - \beta_n - \theta) \frac{\log(\tau_n)}{\log_2(1/c_n)} \\ &= O_P(1) \frac{\log(\tau_n)}{\log_2(1/c_n)},\end{aligned}$$

in view of (12). This converges to 0 in probability since $\tau_n \to 1$ and $c_n \to \infty$. ■

References

[1] Beirlant, J., Dierckx, G., Goegebeur, Y., Matthys, G., (1999), Tail index estimation and an exponential regression model, *Extremes*, **2**, 177–200.

[2] Beirlant J., Broniatowski, M., Teugels, J.L., Vynckier, P., (1995), The mean residual life function at great age: Applications to tail estimation, *Journal of Statistical Planning and Inference*, **45**, 21–48.

[3] Beirlant, J., Teugels, J.L., (1992), Modeling large claims in non-life insurance. *Insurance: Math. Econom.*, **11**, 17–29.

[4] Bingham, N.H., Goldie, C.M., Teugels, J.L., (1987), *Regular variation*, Encyclopedia of Mathematics and its Applications, **27**, Cambridge University Press.

[5] Breiman, L., Stone, C.J., Kooperberg, C., (1990), Robust Confidence Bounds for Extreme Upper Quantiles. *J. Statist. Comput. Simul.*, **37**, 127–149.

[6] Broniatowski, M., (1993), On the estimation of the Weibull tail coefficient, *Journal of Statistical Planning and Inference*, **35**, 349–366.

[7] Dekkers, A.L.M., Einmahl, J.H.J., de Haan, L., (1989), A moment estimator for the index of an extreme-value distribution, *The Annals of Statistics*, **17**, 1833–1855.

[8] Diebolt, J., Girard, S., (1999), Consistency of the ET method and smooth variations, *Comptes-Rendus de l'Académie des Sciences*, t. **329**, Série I, 821–826.

[9] Diebolt, J., Girard, S., (2003), A Note on the asymptotic normality of the ET method for extreme quantile estimation, *Statistics and Probability Letters*, **62**(4), 397–406.

[10] Ditlevsen, O., (1994), Distribution Arbitrariness in Structural Reliability, *Structural Safety and Reliability*, 1241–1247, Balkema, Rotterdam.

[11] Drees,H., Kaufmann, E., (1998), Selecting the optimal sample fraction in univariate extreme value estimation, *Stochastic Processes and their Applications*, **75**(2), 149–172.

[12] El-Aroui, M.A., Diebolt, J., (2002), On the use of the peaks over thresholds method for estimating out-of-sample quantiles, *Computational Statistics & Data Analysis*, **39**, 453–475.

[13] Embrechts, P., Klüppelberg, C., Mikosch, T., (1997), *Modelling extremal events*, Springer.

[14] Galambos, J., (1987), *The asymptotic theory of extreme order statistics*, R.E. Krieger publishing compagny.

[15] Geluk, J.L., de Haan, L., (1987), Regular Variation, Extensions and Tauberian Theorems. *Math Centre Tracts*, **40**, Centre for Mathematics and Computer Science, Amsterdam.

[16] Girard, S., (2004), A Hill type estimate of the Weibull tail-coefficient, *Communication in Statistics - Theory and Methods*, **33**(2), 205–234.

[17] Gomes, I., Oliveira, O., (2001), The bootstrap methodology in statistics of extremes. Choice of the optimal sample fraction. *Extremes*, **4**(4), 331–358.

[18] Guillou, A., Hall, P., (2001), A diagnostic for selecting the threshold in extreme-value analysis, *J. Roy. Statist. Soc. Ser. B*, **63**(2), 293–305.

[19] Häusler, E., Teugels, J.L., (1985), On asymptotic normality of Hill's estimator for the exponent of regular variation, *The Annals of Statistics*, **13**, 743–756.

[20] Hill, B.M., (1975), A simple general approach to inference about the tail of a distribution, *The Annals of Statistics*, **3**, 1163–1174.

[21] Hosking, J., Wallis, J., (1987), Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, **29**, 339–349.

[22] Klüppelberg, C., Villaseñor, J.A., (1993), Estimation of distribution tails - a semiparametric approach, *Deutschen Gesellschaft für Versicherungsmathematik*, **XXI(2)**, 213–235.

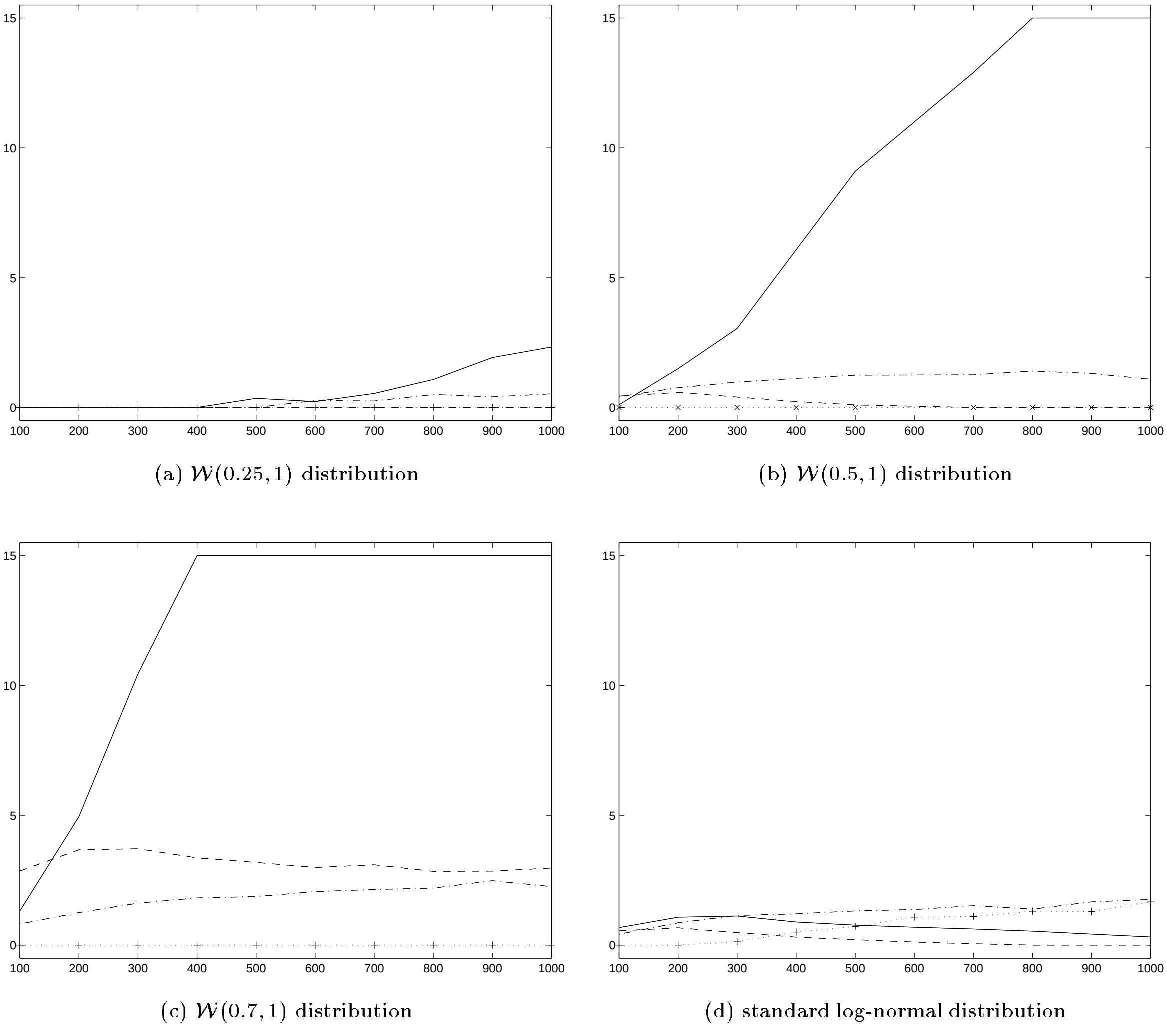
[23] Pickands, J., (1975), Statistical inference using extreme order statistics, *The Annals of Statistics*, **3**, 119–131.

[24] Smith, J., (1991), Estimating the upper tail of flood frequency distributions, *Water Resour. Res.* , **23**(8), 1657–1666.

[25] Weissman, I., (1978), Estimation of parameters and large quantiles based on the k largest observations, *Journal of the American Statistical Association*, **73**, 812–815.

	θ	$b(x)$	ρ
$\mathcal{N}(\mu,\sigma^2)$	$1/2$	$\frac{1}{4}\frac{\log x}{x}$	-1
$\Gamma(\beta,\alpha\neq 1)$	1	$(1-\alpha)\frac{\log x}{x}$	-1
$\mathcal{W}(\alpha,\lambda)$	$1/\alpha$	0	$-\infty$
$\mathcal{E}W(\alpha,\beta\neq 0)$	$1/\alpha$	$-\frac{\beta}{\alpha^2}\frac{\log x}{x}$	-1

Table 1: Parameters θ, ρ and the function $b(x)$ associated to some usual distributions



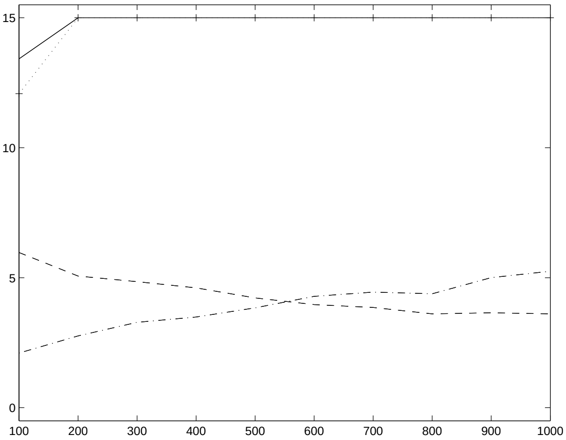
(a) $\mathcal{W}(0.25, 1)$ distribution

(b) $\mathcal{W}(0.5, 1)$ distribution

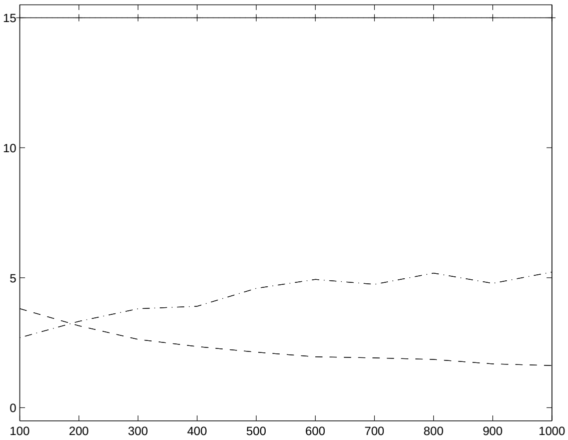
(c) $\mathcal{W}(0.7, 1)$ distribution

(d) standard log-normal distribution

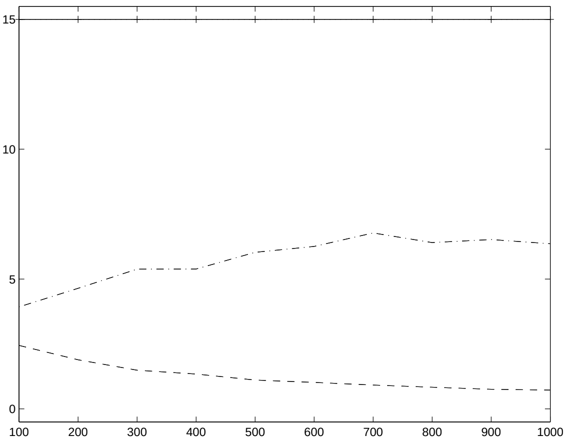
Figure 1: Comparison of estimates $\hat{x}_{p_n}^{\mathrm{G}}$ (solid line), $\hat{x}_{p_n}^{ET}$ (dashed line), $\hat{x}_{p_n}^{BBTV}$ (dotted line) and $\hat{x}_{p_n}^{HW}$ (dash-dot line) on super-exponential distributions.



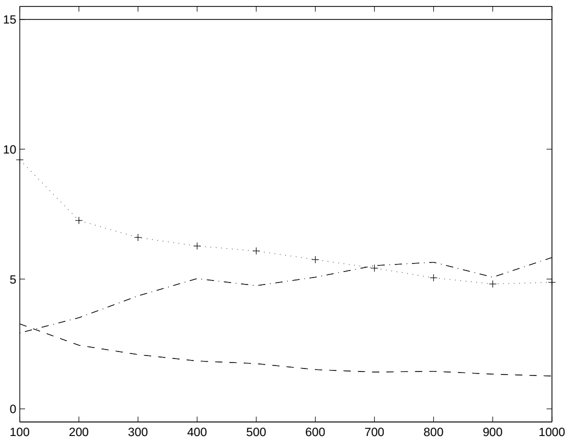
(a) $\mathcal{W}(1.5, 1)$ distribution



(b) $\mathcal{W}(2, 1)$ distribution

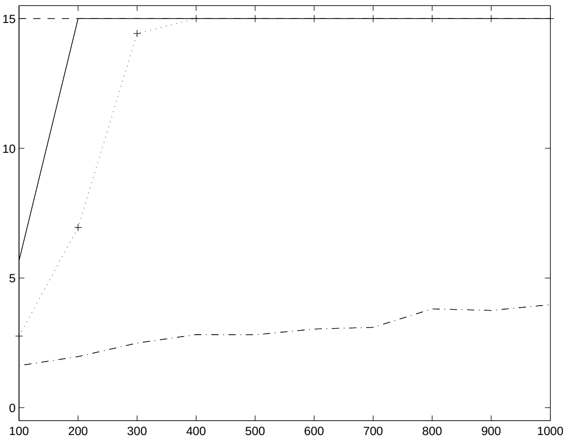


(c) $\mathcal{W}(4, 1)$ distribution

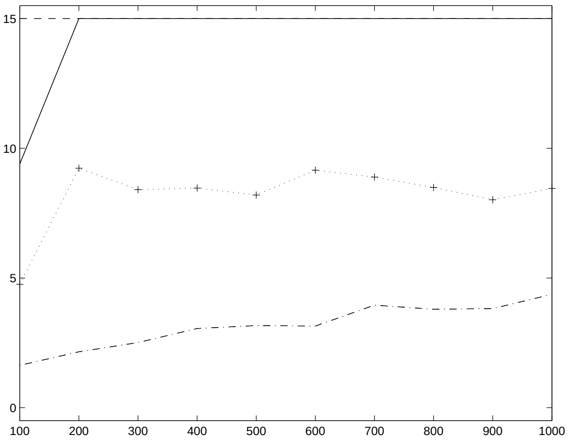


(d) $\mathcal{N}(4, 1)$ distribution

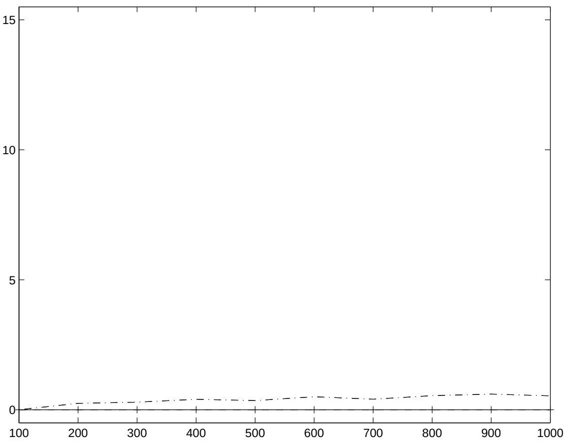
Figure 2: Comparison of estimates $\hat{x}_{p_n}^{\mathrm{G}}$ (solid line), $\hat{x}_{p_n}^{ET}$ (dashed line), $\hat{x}_{p_n}^{BBTV}$ (dotted line) and $\hat{x}_{p_n}^{HW}$ (dash-dot line) on sub-exponential distributions.



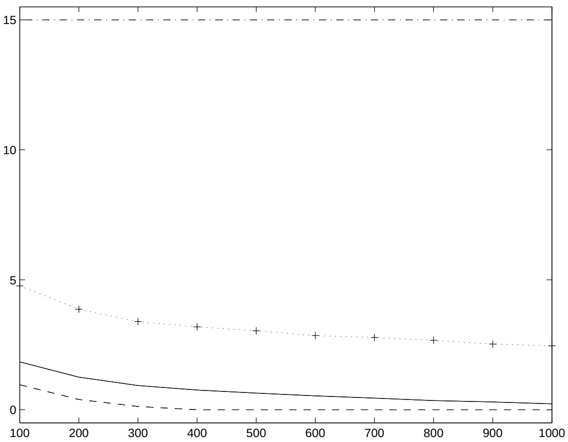
(a) $\mathcal{W}(1,1)$ distribution



(b) $\Gamma(2,1)$ distribution



(c) Pareto distribution ($\alpha = 2$)



(d) $\mathcal{U}(0,1)$ distribution

Figure 3: Comparison of estimates $\hat{x}_{p_n}^{\mathrm{G}}$ (solid line), $\hat{x}_{p_n}^{ET}$ (dashed line), $\hat{x}_{p_n}^{BBTV}$ (dotted line) and $\hat{x}_{p_n}^{HW}$ (dash-dot line) on exponential-like distributions (a, b) and non DA(Gumbel) distributions (c, d).