Bias-reduced extreme quantiles estimators of Weibull distributions

Jean Diebolt, Laurent Gardes, Stéphane Girard, Armelle Guillou

BIAS-REDUCED EXTREME QUANTILE ESTIMATORS OF WEIBULL TAIL-DISTRIBUTIONS

Jean Diebolt[(1)], Laurent Gardes[(2)], Stéphane Girard[(2)] and Armelle Guillou[(3)]

[(1)] *CNRS, Université de Marne-la-Vallée*
Équipe d'Analyse et de Mathématiques Appliquées
5, boulevard Descartes, Batiment Copernic
Champs-sur-Marne
77454 Marne-la-Vallée Cedex 2, France

[(2)] *INRIA Rhône-Alpes, team Mistis,*
Inovallée, 655, av. de l'Europe, Montbonnot,
38334 Saint-Ismier cedex, France

[(3)] *Université Paris VI*
Laboratoire de Statistique Théorique et Appliquée
Boîte 158
175 rue du Chevaleret
75013 Paris, France

Abstract. *In this paper, we consider the problem of estimating an extreme quantile of a Weibull tail-distribution. The new extreme quantile estimator has a reduced bias compared to the more classical ones proposed in the literature. It is based on an exponential regression model that was introduced in Diebolt et al. (2008). The asymptotic normality of the extreme quantile estimator is established. We also introduce an adaptive selection procedure to determine the number of upper order statistics to be used. A simulation study as well as an application to a real data set are provided in order to prove the efficiency of the above mentioned methods.*

Key words and phrases. Weibull tail-distribution, extreme quantile, bias-reduction, least-squares approach, asymptotic normality.

AMS Subject classifications. 62G05, 62G20, 62G30.

1 Introduction

Let $X_1, ..., X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables with distribution function F. In the present paper, we assume that F is a Weibull tail-distribution, which means that

$$1 - F(x) = \exp(-H(x)) \qquad \text{with} \qquad H^{-1}(x) := \inf\{t : H(t) \geq x\} = x^{\theta}\ell(x), \tag{1}$$

where $\theta > 0$ denotes the Weibull tail-coefficient and ℓ is a slowly varying function at infinity satisfying

$$\frac{\ell(\lambda x)}{\ell(x)} \longrightarrow 1, \text{ as } x \to \infty, \text{ for all } \lambda > 0. \tag{2}$$

Based on the limited sample $X_1, ..., X_n$, the question is how to obtain a good estimate for a quantile of order $1 - p_n$, $p_n \to 0$ defined by

$$x_{p_n} = \inf\{y : F(y) \geq 1 - p_n\},$$

such that the quantile to be estimated is situated on the border of or beyond the range of the data. Extrapolation outside the sample occurs for instance in reliability (Ditlevsen, 1994), hydrology (Smith, 1991), and finance (Embrechts et al., 1997). Beirlant et al. (1996) investigated this estimation problem and proposed the following estimator of x_{p_n}:

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

where $X_{1,n} \leq ... \leq X_{n,n}$ denote the order statistics associated to the original sample and $\widetilde{\theta}_n$ is an estimator of θ. One can use for instance the estimator introduced in Diebolt et al. (2008):

$$\widetilde{\theta}_n = \frac{1}{k_n}\sum_{i=1}^{k_n} i\log(n/i)\left(\log(X_{n-i+1,n}) - \log(X_{n-i,n})\right). \tag{4}$$

We refer to Gardes and Girard (2005) for a study of the properties of (3). In the preceding equations, k_n denotes an intermediate sequence, i.e. a sequence such that $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. See Broniatowski (1993), Beirlant et al. (1995, 1996), Girard (2004), and Gardes and Girard (2006,2008) for other contributions to the estimation of θ and Beirlant et al. (2006) for Local Asymptotic Normality (LAN) results. Denoting $\tau_n = \log(1/p_n)/\log(n/k_n)$, the estimator (3) can be rewritten as

$$\widetilde{x}_{p_n} = X_{n-k_n+1,n}\,\tau_n^{\widetilde{\theta}_n}.$$

It appears that the extreme quantile of order $1 - p_n$ is estimated through an ordinary quantile of order $1 - k_n/n$ with a multiplicative correction $\tau_n^{\widetilde{\theta}_n}$.
It will appear in the next section that $\widetilde{x}_{p_n}$ exhibits a bias depending on the rate of convergence to 1 of the ratio of the slowly varying function ℓ in (2). In order to quantify this bias, a second-order condition is required. This assumption can be expressed as follows:
Assumption $(R_\ell(b, \rho))$**.** *There exists a constant $\rho < 0$ and a rate function b satisfying $b(x) \to 0$ as $x \to \infty$, such that for all $\varepsilon > 0$ and $1 < A < \infty$, we have*

$$\sup_{\lambda\in[1,A]}\left|\frac{\log(\ell(\lambda x)/\ell(x))}{b(x)K_\rho(\lambda)} - 1\right| \leq \varepsilon, \quad \textit{for } x \textit{ sufficiently large,}$$

with $K_\rho(\lambda) = \int_1^\lambda t^{\rho-1}dt$.
It can be shown that necessarily $|b|$ is regularly varying with index ρ (see e.g. Geluk and de Haan, 1987). In this paper, we focus on the case where the convergence (2) is slow, and thus when the bias term in $\widetilde{\theta}_n$ and therefore in $\widetilde{x}_{p_n}$ is large. This situation is described by the following assumption:

$$x|b(x)| \to \infty \text{ as } x \to \infty, \tag{5}$$

which is fulfilled by Gamma, Gaussian and $\mathcal{D}$ distributions, see Table 1. The $\mathcal{D}$ distribution is an adaptation of Hall's class (Hall and Welsh, 1985) to the framework of Weibull tail-distributions, see the appendix for its definition. The methodology that we propose in order to reduce the bias of $\widetilde{x}_{p_n}$ is to use the following regression model proposed by Diebolt et al. (2008) for the log-spacings of upper order statistics:

$$\begin{aligned} Z_j &:= j \log(n/j)\left(\log(X_{n-j+1,n}) - \log(X_{n-j,n})\right) \\ &= \left(\theta + b(\log(n/k_n))\left(\frac{\log(n/k_n)}{\log(n/j)}\right)\right) f_j + o_{\mathbb{P}}\left(b(\log(n/k_n))\right), \end{aligned} \tag{6}$$

for $1 \le j \le k_n$, where $(f_1, ..., f_{k_n})$ is a vector of independent and standard exponentially distributed random variables and the $o_{\mathbb{P}}-$term is uniform in j. This exponential regression model is similar to the ones proposed by Beirlant et al. (1999, 2002) and Feuerverger and Hall (1999) in the case of Pareto-type distributions. The model (6) allows us to generate bias-corrected estimates $\widehat{\theta}_n$ for θ through a Least-Square (LS) estimation of θ and $b(\log(n/k_n))$. The resulting LS estimates are then the following:

$$\begin{cases} \widehat{\theta}_n = \overline{Z}_{k_n} - \widehat{b}(\log(n/k_n))\,\overline{x}_{k_n} \\ \widehat{b}(\log(n/k_n)) = \displaystyle\sum_{j=1}^{k_n}(x_j - \overline{x}_{k_n})Z_j \Big/ \sum_{j=1}^{k_n}(x_j - \overline{x}_{k_n})^2 \end{cases}$$

where $x_j = \log(n/k_n)/\log(n/j)$, $\overline{x}_{k_n} = \frac{1}{k_n}\sum_{j=1}^{k_n} x_j$ and $\overline{Z}_{k_n} = \frac{1}{k_n}\sum_{j=1}^{k_n} Z_j$. The asymptotic normality of the LS-estimator $\widehat{\theta}_n$ is established in Diebolt et al. (2008). Now, in order to refine $\widetilde{x}_{p_n}$, we can use the additional information about the slowly varying function ℓ that is provided by the LS-estimates for θ and b. To this aim, condition $(R_\ell(b,\rho))$ is used to approximate the ratio $F^{-1}(1-p_n)/X_{n-k_n+1,n}$, noting that

$$X_{n-k_n+1,n} \stackrel{d}{=} F^{-1}(U_{n-k_n+1,n}),$$

with $U_{1,n} \le ... \le U_{n,n}$ the order statistics of a uniform $(0,1)$ sample of size n,

$$\begin{aligned} \frac{x_{p_n}}{X_{n-k_n+1,n}} &\stackrel{d}{=} \frac{F^{-1}(1-p_n)}{F^{-1}(U_{n-k_n+1,n})} \\ &= \frac{(-\log(p_n))^\theta}{(-\log(1-U_{n-k_n+1,n}))^\theta}\,\frac{\ell(-\log(p_n))}{\ell(-\log(1-U_{n-k_n+1,n}))} \\ &\stackrel{d}{=} \frac{(-\log(p_n))^\theta}{(-\log(U_{k_n,n}))^\theta}\,\frac{\ell(-\log(p_n))}{\ell(-\log(U_{k_n,n}))} \end{aligned}$$

$$\simeq \left(\frac{\log(1/p_n)}{\log(n/k_n)}\right)^{\theta} \exp\left\{b\,(\log(n/k_n))\,\frac{\left(\frac{\log(1/p_n)}{\log(n/k_n)}\right)^{\rho}-1}{\rho}\right\}.$$

The last step follows by replacing $U_{k_n,n}$ with k_n/n. Hence, we arrive at the following estimator for extreme quantiles

$$\widehat{x}_{p_n} = X_{n-k_n+1,n}\left(\frac{\log(1/p_n)}{\log(n/k_n)}\right)^{\widehat{\theta}_n} \exp\left\{\widehat{b}\,(\log(n/k_n))\,\frac{\left(\frac{\log(1/p_n)}{\log(n/k_n)}\right)^{\widehat{\rho}_n}-1}{\widehat{\rho}_n}\right\},$$

or equivalently,

$$\widehat{x}_{p_n} = X_{n-k_n+1,n}\,\tau_n^{\widehat{\theta}_n} \exp\left(\widehat{b}(\log(n/k_n))K_{\widehat{\rho}_n}(\tau_n)\right).$$

Here, $\widehat{\rho}_n$ is an arbitrary estimator of ρ. It will appear in the next section (see Theorem 1(ii)) that, if τ_n converges to a constant value $\tau > 1$, one can even choose $\widehat{\rho}_n = \rho^{\#}$ a constant value, for instance the canonical value $\rho^{\#} = -1$, as suggested by Feuerverger and Hall (1999). Note that the estimator (3) can be seen as a particular case of $\widehat{x}_{p_n}$ obtained by neglecting the bias-term. In the following, we use the LS-estimators of θ and b defined previously. The study of the asymptotic properties of the extreme quantile estimators $\widehat{x}_{p_n}$ and $\widetilde{x}_{p_n}$ is the aim of Section 2. An adaptive selection procedure for k_n is also proposed. A simulation study as well as a real data set are provided in Sections 3 and 4. Proofs are postponed to Section 5.

2 Bias-reduced extreme quantile estimator

The asymptotic normality of our bias-reduced extreme quantile estimator $\widehat{x}_{p_n}$ is established in the following theorem.

Theorem 1 *Suppose (1) holds together with ($R_\ell(b,\rho)$) and (5). We assume that*

$$k_n \to \infty, \frac{\sqrt{k_n}}{\log(n/k_n)} b\,(\log(n/k_n)) \to \lambda \in \mathbb{R}, \tag{7}$$

and if $\lambda = 0$,

$$\frac{\sqrt{k_n}}{\log(n/k_n)} \to \infty \text{ and } \frac{\log^2(k_n)}{\log(n/k_n)} \to 0. \tag{8}$$

Under the additional condition that

$$|\widehat{\rho}_n - \rho| \log(\tau_n) = O_{\mathbb{P}}(1), \tag{9}$$

we have

(i) *if $\tau_n \to \infty$*

$$\frac{\sqrt{k_n}}{\log(n/k_n)\log(\tau_n)}\left(\log(\widehat{x}_{p_n}) - \log(x_{p_n})\right) \xrightarrow{d} \mathcal{N}(0,\theta^2),$$

(ii) *if $\tau_n \to \tau, \tau > 1$, and if we replace $\widehat{\rho}_n$ by a canonical choice $\rho^{\#} < 0$, then*

$$\frac{\sqrt{k_n}}{\log(n/k_n)}\left(\log(\widehat{x}_{p_n}) - \log(x_{p_n})\right) \xrightarrow{d} \mathcal{N}\left(\lambda\mu(\tau), \theta^2\sigma^2(\tau)\right),$$

with

$$\sigma^2(\tau) = \left(K_{\rho^\#}(\tau) - \log(\tau)\right)^2,$$

and

$$\mu(\tau) = \left(K_{\rho^\#}(\tau) - K_\rho(\tau)\right).$$

In the following remark we provide some possible choices for the sequences (k_n) and (p_n).

Remark 1 *Suppose (1) holds together with $(R_\ell(b,\rho))$ and (5). Then, choosing*

$$k_n = \left(\lambda \frac{\log(n)}{b(\log(n))}\right)^2,\ \lambda \neq 0,\ p_n = n^{-\tau},\ \tau > 1,\ \textit{and}\ \widehat{\rho}_n = \rho^\# < 0,$$

Theorem 1(ii) applies and thus

$$\frac{1}{b(\log(n))}\left(\log(\widehat{x}_{p_n}) - \log(x_{p_n})\right) \xrightarrow{d} \mathcal{N}\left(\mu(\tau), \left(\frac{\theta}{\lambda}\right)^2 \sigma^2(\tau)\right).$$

Clearly, the faster b converges to 0, the faster $\widehat{x}_{p_n}$ converges to x_{p_n}.

As a comparison, one can establish similar results for $\widetilde{x}_{p_n}$.

Theorem 2 *Suppose (1) holds together with $(R_\ell(b,\rho))$ and (5). We assume that*

$$k_n \to \infty,\ \sqrt{k_n} b(\log(n/k_n)) \to \lambda \in \mathbb{R}, \liminf \tau_n > 1,$$

and if $\lambda = 0$, $\log(k_n)/\log(n) \to 0$, we have:

$$\frac{\sqrt{k_n}}{\log(\tau_n)}\left\{\log(\widetilde{x}_{p_n}) - \log(x_{p_n}) - b(\log(n/k_n))\left(\frac{\log(\tau_n)}{k_n}\sum_{j=1}^{k_n}\left(\frac{\log(n/j)}{\log(n/k_n)}\right)^\rho - K_\rho(\tau_n)\right)\right\} \xrightarrow{d} \mathcal{N}(0,\theta^2).$$

The next corollary allows an asymptotic comparison of $\widehat{x}_{p_n}$ and $\widetilde{x}_{p_n}$.

Corollary 1 *Under the assumptions of Theorem 2, we have*

(i) *if $\tau_n \to \infty$*

$$\frac{\sqrt{k_n}}{\log(\tau_n)}\left(\log(\widetilde{x}_{p_n}) - \log(x_{p_n})\right) \xrightarrow{d} \mathcal{N}(\lambda,\theta^2),$$

(ii) *if $\tau_n \to \tau, \tau > 1$, then*

$$\sqrt{k_n}\left(\log(\widetilde{x}_{p_n}) - \log(x_{p_n})\right) \xrightarrow{d} \mathcal{N}\left(\lambda\tilde{\mu}(\tau), \theta^2\tilde{\sigma}^2(\tau)\right),$$

with

$$\tilde{\sigma}^2(\tau) = \log^2(\tau),$$

and

$$\tilde{\mu}(\tau) = \left(\log(\tau) - K_\rho(\tau)\right).$$

In the situation where $\tau_n \to \infty$, clearly $\log(\widehat{x}_{p_n})$ is asymptotically unbiased whereas $\log(\widetilde{x}_{p_n})$ is biased. When $\tau_n \to \tau$, $\tau > 1$ both estimates are asymptotically biased but the bias of $\log(\widehat{x}_{p_n})$ can be smaller than the one of $\log(\widetilde{x}_{p_n})$ if $\rho^\#$ is close to ρ.

Let us now introduce the empirical adapted Asymptotic Mean Squared Error (AMSE*) of $\widetilde{x}_{p_n}$ defined as

$$AMSE^*(\widetilde{x}_{p_n}) = \text{Asymptotic}\,\mathbb{E}\left(\log(\widetilde{x}_{p_n}) - \log(x_{p_n})\right)^2.$$

Note that we use an adapted version of the AMSE since it takes into account the fact that the distribution of the quantile estimators is found to be closer to a lognormal distribution than to the asymptotic normal distribution. This is classical in the literature, see for instance Matthys *et al.* (2004). As a consequence of Theorem 2, we have

$$AMSE^*(\widetilde{x}_{p_n}) = \theta^2\frac{\log^2(\tau_n)}{k_n} + b^2(\log(n/k_n))\left\{\frac{\log(\tau_n)}{k_n}\sum_{j=1}^{k_n}\left(\frac{\log(n/j)}{\log(n/k_n)}\right)^{\rho} - K_\rho(\tau_n)\right\}^2. \quad (10)$$

We can now take benefit of the estimation of $b(\log(n/k_n))$ by estimating the AMSE* given in (10) by:

$$\widehat{AMSE}^*(\widetilde{x}_{p_n}) = (\widehat{\theta}_n)^2\frac{\log^2(\tau_n)}{k_n} + (\widehat{b}(\log(n/k_n)))^2\left\{\frac{\log(\tau_n)}{k_n}\sum_{j=1}^{k_n}\left(\frac{\log(n/j)}{\log(n/k_n)}\right)^{-1} + \tau_n^{-1} - 1\right\}^2.$$

Note that, in the latter formula, we replaced ρ by a canonical choice ($\rho = -1$) instead of estimating this parameter. In fact, this second-order parameter is difficult to estimate in practice, and we can easily check by simulations that fixing its value does not influence much the result (see e.g. Beirlant et al., 1999, 2002 or Feuerverger and Hall, 1999). Then, the intermediate sequence k_n can be selected by minimizing the previous quantity:

$$\hat{k}_n = \arg\min_{k_n} \widehat{AMSE}^*(\widetilde{x}_{p_n}).$$

This adaptive procedure for selecting the number of upper order statistics is in the same spirit as the one proposed by Matthys *et al.* (2004) in the case of the Weissman estimator (Weissman, 1978). In order to illustrate the usefulness of the bias reduction and of the selection procedure, we provide a simulation study in the next section.

3 A small simulation study

First, the finite sample performance of the estimators $\widetilde{x}_{p_n}$ and $\widehat{x}_{p_n}$ are investigated on 4 different distributions: $|\mathcal{N}(0,1)|$, $\Gamma(0.25, 0.25)$, $\mathcal{D}(1, 0.5)$ and $\mathcal{W}(0.25, 0.25)$. It is shown in Gardes and Girard (2005) that $\widetilde{x}_{p_n}$ gives better results than the other approaches (Hosking and Wallis, 1987; Breiman et al., 1990; Beirlant et al., 1995). This explains why $\widehat{x}_{p_n}$ is only compared to the estimator $\widetilde{x}_{p_n}$. In the following, we take $p_n := p_n(\tau) = n^{-\tau}$ with $\tau = 2$ and 4 and we choose $\widehat{\rho}_n = -1$. We simulate $N = 500$ samples $(\mathcal{X}_{n,i})_{i=1,\ldots,N}$ of size $n = 500$. On each sample $(\mathcal{X}_{n,i})$, the estimates $\widehat{x}_{p_n(\tau),i}$, are computed for $\tau = 2, 4$ and for $k_n = 2, \ldots, 360$. We present the plots obtained by drawing the points

$$(k_n, \text{med}_i(\log(\widehat{x}_{p_n(\tau),i}))) \text{ for } \tau = 2 \text{ and } 4,$$

where $\text{med}_i(\log(\widehat{x}_{p_n(\tau),i}))$ is the median value of $\log(\widehat{x}_{p_n(\tau),i})$, $i = 1, \ldots, N$. We also present the associated MSE plots

$$\left(k_n, \frac{1}{N}\sum_{i=1}^{N}\left(\log(\widehat{x}_{p_n(\tau),i}) - \log(x_{p_n(\tau)})\right)^2\right) \text{ for } \tau = 2 \text{ and } 4.$$

The same procedure is achieved for the estimator $\widetilde{x}_{p_n}$. Results are presented on figures 1–4. For the $|\mathcal{N}(0,1)|$, $\Gamma(0.25, 0.25)$ and $\mathcal{D}(1, 0.5)$ distributions, the bias of $\log(\widehat{x}_{p_n})$ is smaller than the one of $\log(\widetilde{x}_{p_n})$. Let us highlight that, for the latter distribution, a significant bias reduction is obtained although $\widehat{\rho}_n \neq \rho$. Moreover, bias reduction is usually associated with an increase in variance. However, as illustrated in panel (b) of figures 1–4, our estimator $\widehat{x}_{p_n}$ is still competitive in an adapted MSE sense. Note that for the $\mathcal{W}(0.25, 0.25)$ distribution, Theorem 1 does not apply since $xb(x) = 0$. In this case, the behavior of $\log(\widetilde{x}_{p_n})$ is slightly better.
Second, we investigate the behavior of the adaptive procedure for selecting the number of upper order statistics in $\widetilde{x}_{p_n}$. For $i = 1, \ldots, N$ and $\tau = 2, 4$, we denote by

$$\widehat{k}_{n,i}^{esti} = \arg\min_{k_n \in [1,n]} \widehat{AMSE}^*(\widetilde{x}_{p_n(\tau),i}),$$

the value selected on the sample $(\mathcal{X}_{n,i})$. As a comparison, we introduce the value that would be obtained by minimizing the true AMSE*:

$$k_n^{opt} = \arg\min_{k_n \in [1,n]} AMSE^*(\widetilde{x}_{p_n(\tau)}).$$

Figures 5–8 contain the paired boxplots of the log-quantile estimators $\log(\widetilde{x}_{p_n(\tau),i})$ at the adaptively selected values $\widehat{k}_{n,i}^{esti}$ on the right and at the sample fraction k_n^{opt} with smallest AMSE* on the left. The horizontal line indicates the true value of $\log(x_{p_n})$. The method proposed for the adaptive choice of k_n, while not achieving the goal of minimizing the MSE, does lead to quantile estimators that exhibit good bias properties while the variance is inflated in comparison to the asymptotically optimal values.

4 Real data

Here, the good performance of the adaptive selection procedure is illustrated through the analysis of extreme events on the Nidd river data set. This data set is standard in extreme value studies (see e.g. Hosking and Wallis, 1987, or Davison and Smith, 1990.) It consists in 154 exceedances of the level 65 m^3s^{-1} by the river Nidd (Yorkshire, England) during the period 1934-1969 (35 years). In environmental studies, the most common quantity of interest is the N-year return level, defined as the level which is exceeded on average once in N years. Here, we focus on the estimation of the 50- and 100- year return levels. According to Hosking and Wallis (1987), the Nidd data may reasonably be assumed to come from a distribution in the Gumbel maximum domain of attraction. This result was confirmed in Diebolt et al. (2008) who have shown that one could consider Weibull tail-distributions as a possible model for such data. The Weibull tail-coefficient is estimated at 0.91. We obtained $321.5m^3s^{-1}$ as an estimation of the 50-year return level and $359m^3s^{-1}$ as an estimation of the 100-year return level. Note that these results are in accordance with the results obtained by profile-likelihood or Bayesian methods, see Diebolt et al. (2005) or Davison and Smith (1990).

5 Proofs

Proof of Theorem 1. We decompose our quantile estimator as follows:

$$\log(\widehat{x}_{p_n}) - \log(x_{p_n}) \quad = \quad \log(X_{n-k_n+1,n}) + \widehat{\theta}_n \log(\tau_n)$$

$$
\begin{aligned}
&+ \quad \widehat{b}\left(\log(n/k_n)\right) K_{\widehat{\rho}_n}(\tau_n) - \log\left((-\log(p_n))^{\theta}\ell(-\log(p_n))\right) \\
&\stackrel{d}{=} \quad \theta\left\{\log\left(-\log\left(U_{k_n,n}\right)\right) - \log\log(n/k_n)\right\} \\
&+ \quad \left(\widehat{\theta}_n - \theta\right)\log(\tau_n) \\
&+ \quad \left\{\log\ell\left(-\log\left(U_{k_n,n}\right)\right) - \log\ell\left(\log(n/k_n)\right)\right\} \\
&+ \quad \left\{\log\ell\left(\log(n/k_n)\right) - \log\ell\left(-\log(p_n)\right) + b\left(\log(n/k_n)\right) K_{\rho}(\tau_n)\right\} \\
&+ \quad \left(\widehat{b}\left(\log(n/k_n)\right) - b\left(\log(n/k_n)\right)\right) K_{\widehat{\rho}_n}(\tau_n) \\
&+ \quad b\left(\log(n/k_n)\right)\left\{K_{\widehat{\rho}_n}(\tau_n) - K_{\rho}(\tau_n)\right\} \\
&:= \quad \sum_{j=1}^{6} B_{j,k_n}.
\end{aligned}
$$

We successively discuss each of the terms B_{j,k_n}, $j = 1, ..., 6$. First concerning B_{1,k_n}, remark that

$$
\log\left(-\log(U_{k_n,n})\right) - \log\log(n/k_n) \stackrel{d}{=} \log\left(\frac{T_{n-k_n+1,n}}{\log(n/k_n)}\right),
$$

where $T_{j,n}$ denotes the order statistics from an i.i.d. standard exponential sample of size n. Since it is well known that

$$
\sqrt{k_n}\left(T_{n-k_n+1,n} - \log(n/k_n)\right) \stackrel{d}{\longrightarrow} \mathcal{N}(0,1), \tag{11}
$$

we clearly have

$$
\frac{\sqrt{k_n}}{\log(n/k_n)\log(\tau_n)} B_{1,k_n} = O_{\mathbb{P}}\left(\frac{1}{\log^2(n/k_n)\log\tau_n}\right) = o_{\mathbb{P}}(1). \tag{12}
$$

Remark now that

$$
\frac{\sqrt{k_n}}{\log(n/k_n)\log(\tau_n)} B_{2,k_n} = \frac{\sqrt{k_n}}{\log(n/k_n)}\left(\widehat{\theta}_n - \theta\right) \stackrel{d}{\longrightarrow} \mathcal{N}(0,\theta^2), \tag{13}
$$

by Theorem 3.1 in Diebolt et al. (2008). Next, using $(R_\ell(b,\rho))$ and (11), we get

$$
\begin{aligned}
B_{3,k_n} &\stackrel{d}{=} \log\left(\frac{\ell(T_{n-k_n+1,n})}{\ell(\log(n/k_n))}\right) \\
&= K_{\rho}\left(\frac{T_{n-k_n+1,n}}{\log(n/k_n)}\right) b(\log(n/k_n))(1 + o_{\mathbb{P}}(1)) \\
&= \left(\frac{T_{n-k_n+1,n}}{\log(n/k_n)} - 1\right) b(\log(n/k_n))(1 + o_{\mathbb{P}}(1)) \\
&= O_{\mathbb{P}}\left(\frac{b(\log(n/k_n))}{\sqrt{k_n}\log(n/k_n)}\right),
\end{aligned} \tag{14}
$$

so that under (7),

$$
\frac{\sqrt{k_n}}{\log(n/k_n)\log(\tau_n)} B_{3,k_n} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}\log(n/k_n)\log(\tau_n)}\right) = o_{\mathbb{P}}(1). \tag{15}
$$

Next, under $(R_\ell(b,\rho))$ one has (for a suitably chosen b) that for all $\epsilon > 0$ (see Drees, 1998, Lemma 2.1)

$$\sup_{t>1} t^{-(\epsilon+\rho)} \left| \frac{\log \ell(tx) - \log \ell(x)}{b(x)} - K_\rho(t) \right| \longrightarrow 0.$$

Hence, we conclude, choosing $\epsilon < -\rho$, that

$$\left| \frac{B_{4,k_n}}{b(\log(n/k_n))} \right| = \left| \frac{\log \ell\left(\log(n/k_n)\right) - \log \ell(\tau_n \log(n/k_n))}{b(\log(n/k_n))} + K_\rho(\tau_n) \right| \longrightarrow 0, \tag{16}$$

which implies that, under (7), we have

$$\frac{\sqrt{k_n}}{\log(n/k_n)\log(\tau_n)} B_{4,k_n} = o\left(\frac{1}{\log(\tau_n)}\right) = o(1). \tag{17}$$

Next, we can check that, according to (6),

$$B_{5,k_n} = K_{\widehat{\rho}_n}(\tau_n) \frac{1}{k_n} \sum_{j=1}^{k_n} \beta_{j,n}(f_j - 1),$$

where

$$\beta_{j,n} := \frac{(x_j - \overline{x}_{k_n})(\theta + b(\log(n/k_n))x_j)}{\frac{1}{k_n}\sum_{i=1}^{k_n}(x_i - \overline{x}_{k_n})^2}.$$

A direct application of Lyapounov's theorem, combined with Lemma 7.3 in Diebolt et al. (2008) yields

$$\frac{\sqrt{k_n}}{\log(n/k_n)} \frac{1}{k_n} \sum_{j=1}^{k_n} \beta_{j,n}(f_j - 1) \xrightarrow{d} \mathcal{N}(0, \theta^2).$$

Therefore

$$\frac{\sqrt{k_n}}{\log(n/k_n)\log(\tau_n)} B_{5,k_n} = \frac{K_{\widehat{\rho}_n}(\tau_n)}{\log(\tau_n)} \xi_{1,n} \quad \text{with} \quad \xi_{1,n} \xrightarrow{d} \mathcal{N}(0, \theta^2). \tag{18}$$

Finally, following the method of proof of Lemma 1 in de Haan and Rootzén (1993), we find that

$$\begin{aligned}
&\left| \int_1^{\tau_n} s^{\rho-1} (s^{\widehat{\rho}_n - \rho} - 1)\, ds - (\widehat{\rho}_n - \rho) \int_1^{\tau_n} s^{\rho-1} \log(s)\, ds \right| \\
&\quad \leq |\widehat{\rho}_n - \rho| \int_1^{\tau_n} s^{\rho-1} \log(s)\, (s^{|\widehat{\rho}_n - \rho|} - 1)\, ds \\
&\quad \leq |\widehat{\rho}_n - \rho| \left(\int_1^{\tau_n} s^{\rho-1} \log(s)\, ds \right) \left(\tau_n^{|\widehat{\rho}_n - \rho|} - 1 \right),
\end{aligned}$$

where the first inequality comes from the fact that $\left|\frac{e^x - 1}{x} - 1\right| \leq e^{|x|} - 1$. Hence

$$\begin{aligned}
\frac{\sqrt{k_n}}{\log(n/k_n)\log(\tau_n)} B_{6,k_n} &= \frac{\sqrt{k_n}}{\log(n/k_n)} b\left(\log(n/k_n)\right) \\
&\times (\widehat{\rho}_n - \rho) \frac{\int_1^{\tau_n} x^{\rho-1} \log(x)\, dx}{\log(\tau_n)} \left\{ 1 + O\left(\tau_n^{|\widehat{\rho}_n - \rho|} - 1 \right) \right\}.
\end{aligned}$$

If $\tau_n \to \infty$, this implies that $\widehat{\rho}_n \xrightarrow{\mathbb{P}} \rho$ by the assumption (9) and therefore

$$\frac{\sqrt{k_n}}{\log(n/k_n)\log(\tau_n)} B_{6,k_n} = o_{\mathbb{P}}(1). \tag{19}$$

Combining (12), (13) and (15) with (17)-(19), Theorem 1(i) follows. If $\tau_n \to \tau, \tau > 1$, then the normalization factor $\log(\tau_n) \to \log(\tau) \neq 0$ can be omitted in (12), (15) and (17) while preserving the negligeability of these terms. Besides, we can replace $\widehat{\rho}_n$ with any canonical choice, for instance $\rho^{\#} < 0$, and therefore

$$\begin{aligned} \frac{\sqrt{k_n}}{\log(n/k_n)} B_{6,k_n} &= \frac{\sqrt{k_n}}{\log(n/k_n)} b\left(\log(n/k_n)\right)\left(K_{\rho^{\#}}(\tau_n) - K_{\rho}(\tau_n)\right) \\ &\longrightarrow \lambda\mu(\tau) := \lambda\left(K_{\rho^{\#}}(\tau) - K_{\rho}(\tau)\right). \end{aligned} \tag{20}$$

The limiting distribution is then given by (13) and (18) with a bias term due to (20). To conclude with the second part of our Theorem 1, we have to establish the limiting distribution of

$$U_n := \frac{\sqrt{k_n} K_{\rho^{\#}}(\tau_n)}{\log(n/k_n)}\left(\widehat{b}\left(\log(n/k_n)\right) - b\left(\log(n/k_n)\right)\right) + \frac{\sqrt{k_n}\log(\tau_n)}{\log(n/k_n)}\left(\widehat{\theta}_n - \theta\right).$$

To this aim, remark that

$$U_n = \frac{k_n^{-\frac{1}{2}}}{\log(n/k_n)} \sum_{j=1}^{k_n} \omega_{j,n}(f_j - 1) + o_{\mathbb{P}}(1),$$

where

$$\omega_{j,n} = \beta_{j,n} K_{\rho^{\#}}(\tau_n) + \alpha_{j,n}\log(\tau_n)$$

and

$$\alpha_{j,n} = \left(\theta + b\left(\log(n/k_n)\right) x_j\right)\left(1 - \frac{x_j - \overline{x}_{k_n}}{\frac{1}{k_n}\sum_{i=1}^{k_n}(x_i - \overline{x}_{k_n})^2}\overline{x}_{k_n}\right).$$

Using Lemma 7.3 in Diebolt et al. (2008), direct computations lead to

$$\sum_{j=1}^{k_n} \mathrm{Var}\left(\omega_{j,n}(f_j - 1)\right) = \sum_{j=1}^{k_n} \omega_{j,n}^2 \sim \theta^2 \left(\log(n/k_n)\right)^2 k_n \sigma^2(\tau)$$

and

$$\sum_{j=1}^{k_n} \mathbb{E}\left(\omega_{j,n}(f_j - 1)\right)^4 = 9\sum_{j=1}^{k_n} \omega_{j,n}^4 \sim C k_n \left(\log(n/k_n)\right)^4,$$

where C is a suitable constant. Therefore a direct application of Lyapounov's theorem yields

$$U_n \xrightarrow{d} \mathcal{N}\left(0, \theta^2\sigma^2(\tau)\right),$$

which achieves the proof of the second part of Theorem 1. □

Proof of Theorem 2. We have:

$$\begin{aligned}
&\log(\widetilde{x}_{p_n}) - \log(x_{p_n}) - b(\log(n/k_n))\log(\tau_n)\frac{1}{k_n}\sum_{j=1}^{k_n}\left(\frac{\log(n/j)}{\log(n/k_n)}\right)^{\rho} + b(\log(n/k_n))\frac{\tau_n^{\rho}-1}{\rho} \\
= \;& \theta\{\log(-\log(U_{k_n,n})) - \log\log(n/k_n)\} \\
+ \;& \left(\widetilde{\theta}_n - \theta - b(\log n/k_n)\frac{1}{k_n}\sum_{j=1}^{k_n}\left(\frac{\log(n/j)}{\log(n/k_n)}\right)^{\rho}\right)\log(\tau_n) \\
+ \;& \{\log\ell(-\log(U_{k_n,n})) - \log\ell(\log(n/k_n))\} \\
+ \;& \left\{\log\ell(\log(n/k_n)) - \log\ell(-\log(p_n)) + b(\log(n/k_n))\frac{\tau_n^{\rho}-1}{\rho}\right\} \\
:= \;& B_{1,k_n} + B_{7,k_n} + B_{3,k_n} + B_{4,k_n}.
\end{aligned}$$

From (12), we have

$$\frac{\sqrt{k_n}}{\log(\tau_n)}B_{1,k_n} = O_{\mathrm{P}}\left(\frac{1}{\log(n/k_n)\log(\tau_n)}\right) = o_{\mathrm{P}}(1),$$

and Theorem 2.2 in Diebolt et al. (2008) states that

$$\frac{\sqrt{k_n}}{\log(\tau_n)}B_{7,k_n} \xrightarrow{d} \mathcal{N}(0,\theta^2).$$

From (14), we have

$$B_{3,k_n} = O_{\mathrm{P}}\left(\frac{b(\log(n/k_n))}{\sqrt{k_n}\log(n/k_n)}\right),$$

and thus $\sqrt{k_n}b(\log(n/k_n)) \to \lambda \in \mathbb{R}$ entails

$$\frac{\sqrt{k_n}}{\log(\tau_n)}B_{3,k_n} = O_{\mathrm{P}}\left(\frac{1}{\sqrt{k_n}\log(n/k_n)\log(\tau_n)}\right) = o_{\mathrm{P}}(1).$$

Finally, (16) implies that

$$B_{4,k_n} = o_{\mathrm{P}}(b(\log(n/k_n))),$$

which, combined with $\sqrt{k_n}b(\log(n/k_n)) \to \lambda \in \mathbb{R}$ yields

$$\frac{\sqrt{k_n}}{\log(\tau_n)}B_{4,k_n} = o_{\mathrm{P}}\left(\frac{1}{\log(\tau_n)}\right) = o_{\mathrm{P}}(1),$$

which achieves the proof of Theorem 2. □

Appendix

Diebolt et al. (2008) introduced the class of distributions $\mathcal{D}(\alpha,\beta)$ with distribution function given by

$$1 - F(x) = \exp(-H(x)) \text{ where } H^{-1}(x) := x^{1/\alpha}(1 + x^{-\beta}),$$

α and β being two parameters such that $0 < \alpha$, $0 < \beta < 1$ and $\alpha\beta \leq 1$. Under these conditions, the above class of distributions fulfill assumptions (1) with $(R_\ell(b,\rho))$ and

(5) where $\theta = 1/\alpha$, $\rho = -\beta$, $\ell(x) = 1 + x^{-\beta}$ and $b(x) = -\beta x^{-\beta}$. It is thus possible to obtain distributions with arbitrary $\theta > 0$ and $-1 < \rho < 0$. These results are summarized in Table 1.

Acknowledgement

The authors are grateful to the referees for a careful reading of the paper that led to significant improvements of the earlier draft.

References

[1] Beirlant, J., Bouquiaux, C., Werker, B., (2006), Semiparametric lower bounds for tail index estimation, *Journal of Statistical Planning and Inference*, **136**, 705–729.

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

[4] Beirlant, J., Dierckx, G., Guillou, A., Starica, C., (2002), On exponential representations of log-spacings of extreme order statistics, *Extremes*, **5**, 157–180.

[5] Beirlant, J., Teugels, J., Vynckier, P., (1996), *Practical analysis of extreme values*, Leuven university press, Leuven.

[6] Breiman, L., Stone, C. J., Kooperberg, C. (1990), Robust confidence bounds for extreme upper quantiles, *Journal of Computational Statistics and Simulation*, **37**, 127–149.

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

[8] Davison, A., Smith, R., (1990), Models for exceedances over high thresholds, *Journal of the Royal Statistical Society Ser. B*, **52**, 393–442.

[9] Diebolt, J., El-Aroui, M., Garrido, M., Girard, S., (2005), Quasi-conjugate Bayes estimates for GPD parameters and application to heavy tails modelling, *Extremes*, **8**, 57–78.

[10] Diebolt, J., Gardes, L., Girard, S., Guillou, A., (2008), Bias-reduced estimators of the Weibull-tail coefficient, *Test*, **17**, 311–331.

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

[12] Drees, H., (1998), On smooth statistical tail functionals, *Scandinavian Journal of Statistics*, **25**, 187–210.

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

[14] Feuerverger, A., Hall, P., (1999), Estimating a Tail Exponent by Modelling Departure from a Pareto Distribution, *Annals of Statistics*, **27**, 760–781.

[15] Gardes, L., Girard, S., (2005), Estimating extreme quantiles of Weibull tail-distributions, *Communication in Statistics - Theory and Methods*, **34**, 1065–1080.

[16] Gardes, L., Girard, S., (2006), Comparison of Weibull tail-coefficient estimators, *REVSTAT - Statistical Journal*, **4**, 163–188.

[17] Gardes, L., Girard, S., (2008), Estimation of the Weibull tail-coefficient with linear combination of upper order statistics, *Journal of Statistical Planning and Inference*, **138**, 1416–1427.

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

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

[20] de Haan, L., Rootzén, H., (1993), On the estimation of high quantiles, *Journal of Statistical Planning and Inference*, **35**, 1–13.

[21] Hall, P., Welsh, A.H., (1985), Adaptive estimates of parameters of regular variation, *Annals of Statistics*, **13**, 331–341.

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

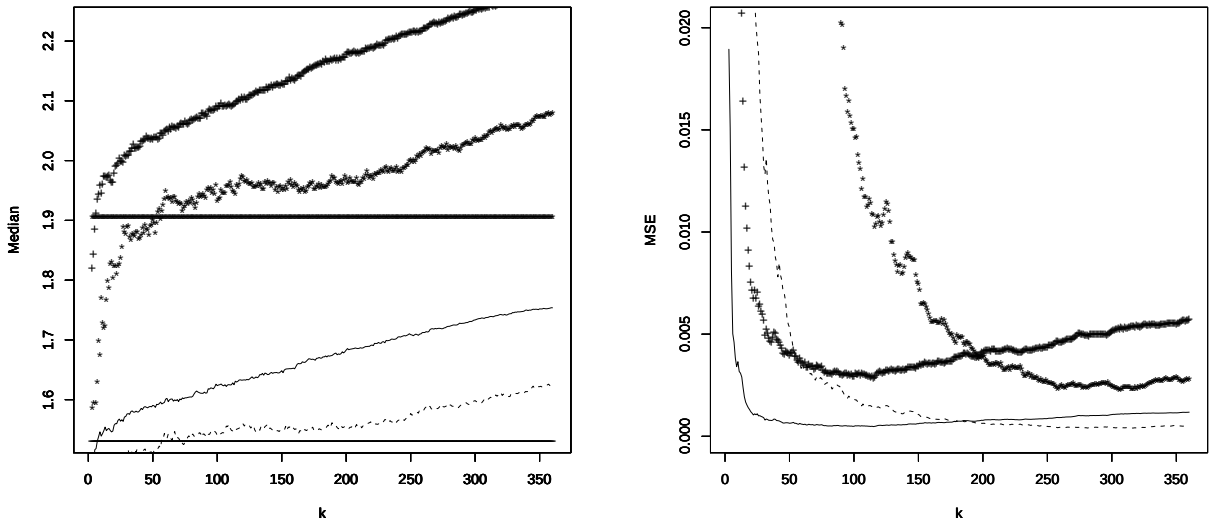
[23] Matthys, G., Delafosse, E., Guillou, A., Beirlant, J., (2004), Estimating catastrophic quantile levels for heavy-tailed distributions, *Insurance Mathematics & Economics*, **34**, 517–537.

[24] Smith, J., (1991), Estimating the upper tail of flood frequency distributions, *Water Resources Research*, **23**, 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.

Distribution	θ	$b(x)$	ρ
Absolute Gaussian $\lvert\mathcal{N}\rvert(\mu, \sigma^2)$	$1/2$	$\frac{1}{4}\frac{\log x}{x}$	-1
Gamma $\Gamma(\alpha \neq 1, \beta)$	1	$(1-\alpha)\frac{\log x}{x}$	-1
Weibull $\mathcal{W}(\alpha, \lambda)$	$1/\alpha$	0	$-\infty$
$\mathcal{D}(\alpha, \beta)$	$1/\alpha$	$-\beta x^{-\beta}$	$-\beta$

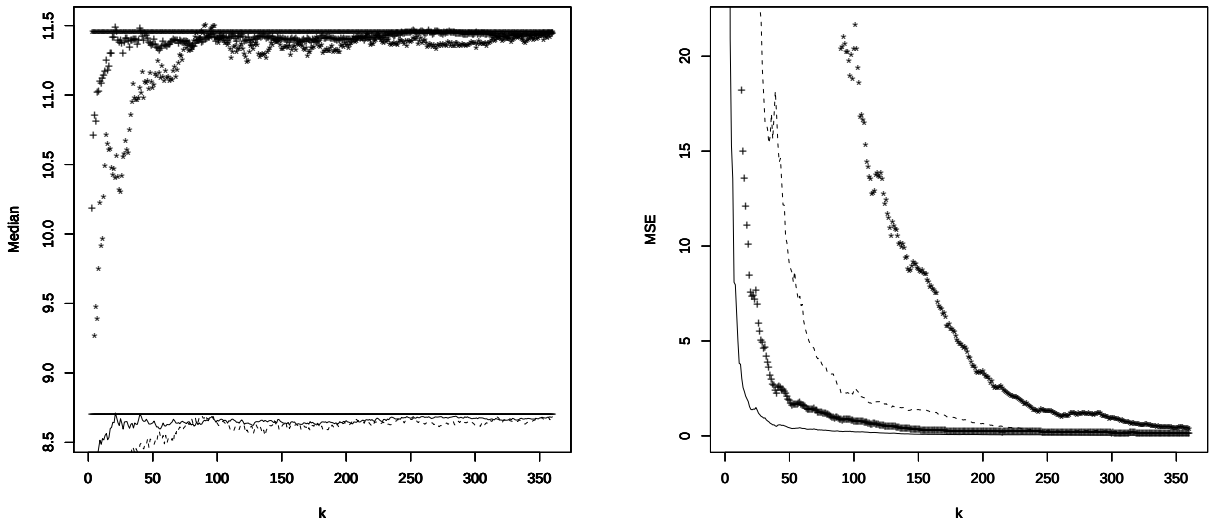
Table 1: Parameters θ, ρ and the function $b(x)$ associated to some distributions



(a) Median as a function of k_n

(b) MSE as a function of k_n

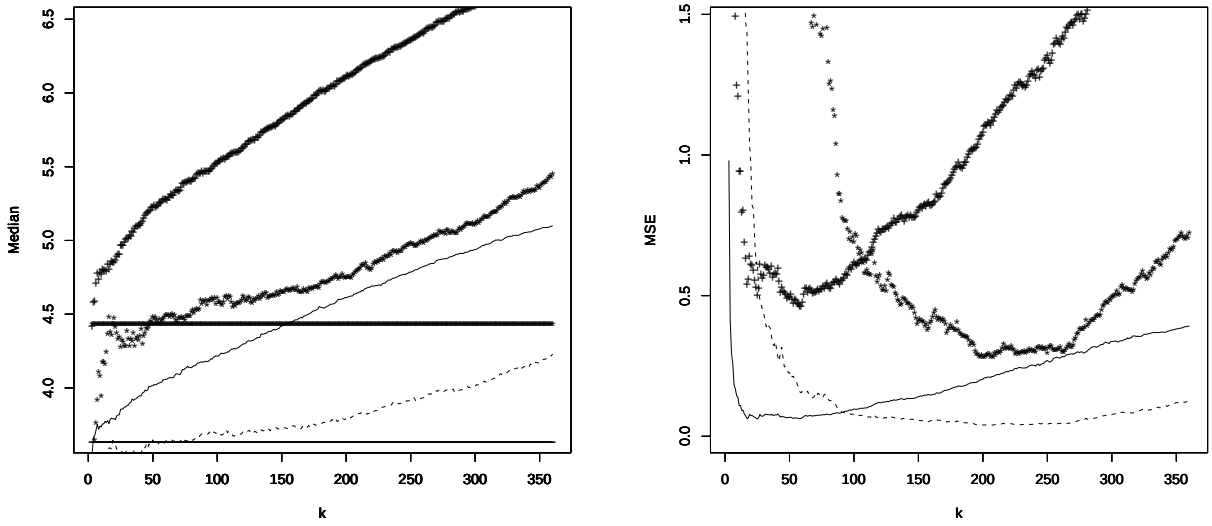
Figure 1: Comparison of $\log(\widetilde{x}_{p_n})$ ($\tau = 2$: continuous line, $\tau = 4$: $+ + +$) and $\log(\widehat{x}_{p_n})$ ($\tau = 2$: dashed line, $\tau = 4$: $\star \star \star$) for the $|\mathcal{N}(0,1)|$ distribution. The horizontal lines on the left panel represent the true values of $\log(x_{p_n})$ (thin line: $\tau = 2$, thick line: $\tau = 4$).



(a) Median as a function of k_n

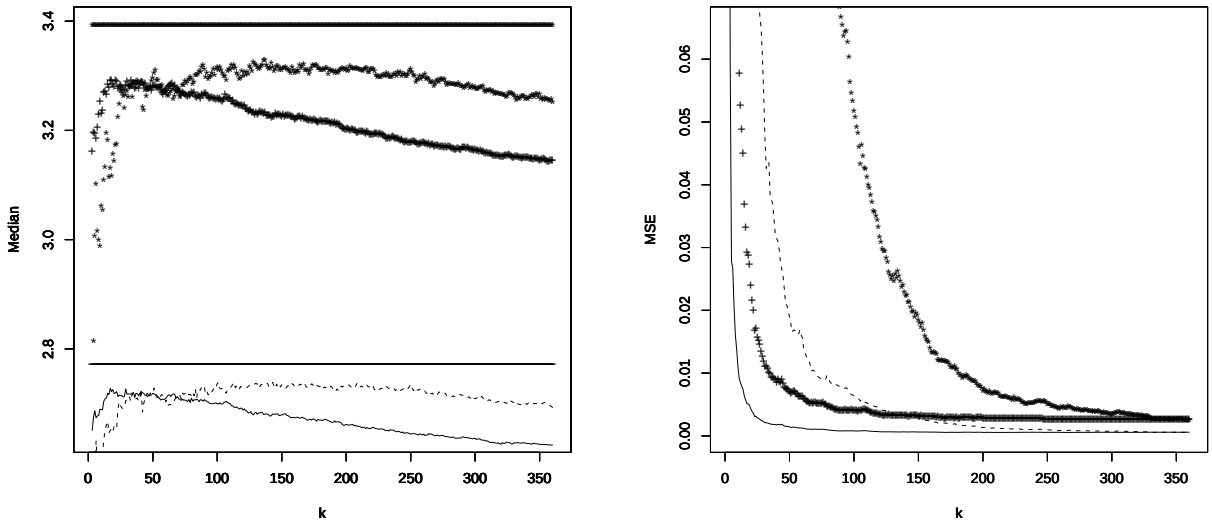
(b) MSE as a function of k_n

Figure 2: Comparison of $\log(\widetilde{x}_{p_n})$ ($\tau = 2$: continuous line, $\tau = 4$: $+ + +$) and $\log(\widehat{x}_{p_n})$ ($\tau = 2$: dashed line, $\tau = 4$: $\star \star \star$) for the $\mathcal{W}(0.25, 0.25)$ distribution. The horizontal lines on the left panel represent the true values of $\log(x_{p_n})$ (thin line: $\tau = 2$, thick line: $\tau = 4$).



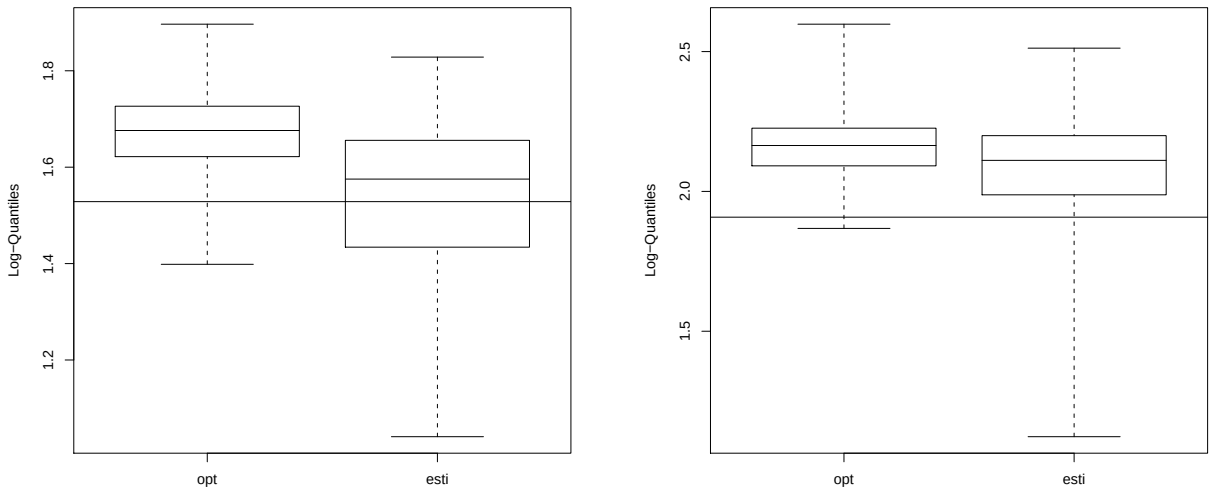
(a) Median as a function of k_n (b) MSE as a function of k_n

Figure 3: Comparison of $\log(\widetilde{x}_{p_n})$ ($\tau = 2$: continuous line, $\tau = 4$: $+ + +$) and $\log(\widehat{x}_{p_n})$ ($\tau = 2$: dashed line, $\tau = 4$: $\star \star \star$) for the $\Gamma(0.25, 0.25)$ distribution. The horizontal lines on the left panel represent the true values of $\log(x_{p_n})$ (thin line: $\tau = 2$, thick line: $\tau = 4$).



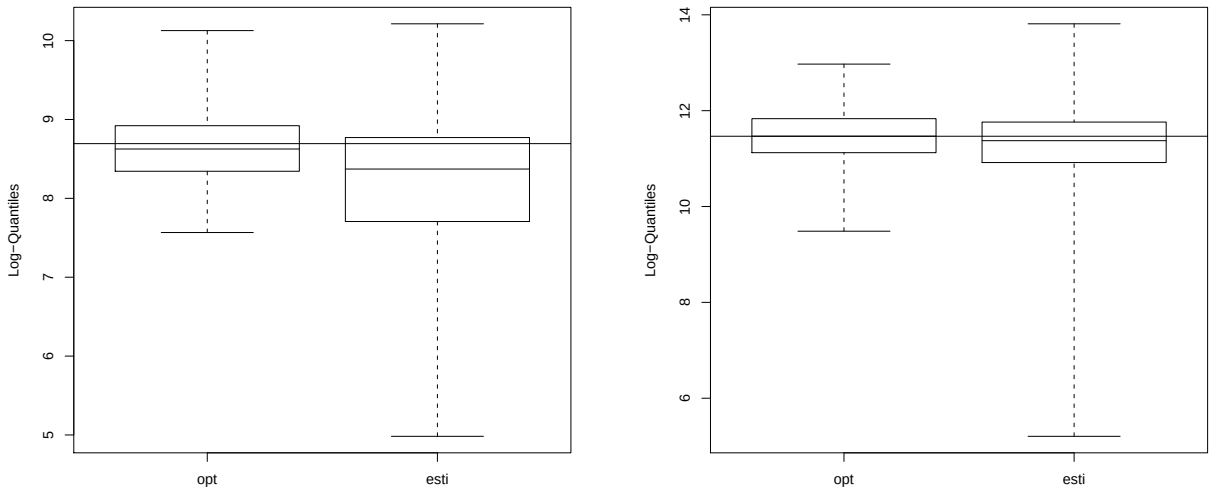
(a) Median as a function of k_n (b) MSE as a function of k_n

Figure 4: Comparison of $\log(\widetilde{x}_{p_n})$ ($\tau = 2$: continuous line, $\tau = 4$: $+ + +$) and $\log(\widehat{x}_{p_n})$ ($\tau = 2$: dashed line, $\tau = 4$: $\star \star \star$) for the $\mathcal{D}(1, 0.5)$ distribution. The horizontal lines on the left panel represent the true values of $\log(x_{p_n})$ (thin line: $\tau = 2$, thick line: $\tau = 4$).



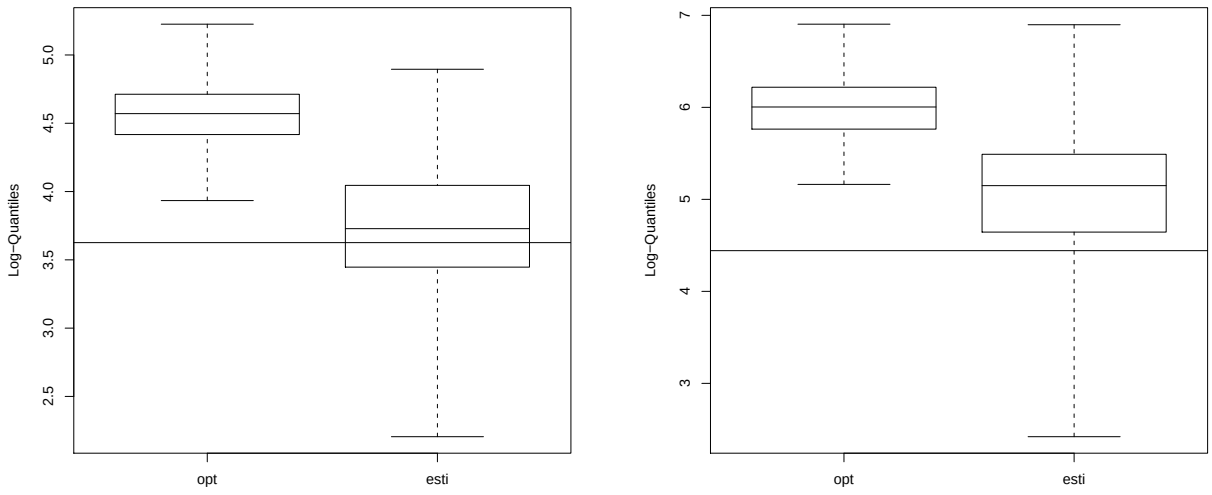
(a) $\tau = 2$ (b) $\tau = 4$

Figure 5: $|\mathcal{N}(0,1)|$ distribution. Boxplots of $\log(\widetilde{x}_{p_n})$ at the optimal value of k_n obtained by minimizing the true AMSE* (left), and at the value of k_n obtained by minimizing the estimated AMSE* (right). The horizontal line indicates the true value of $\log(x_{p_n})$.



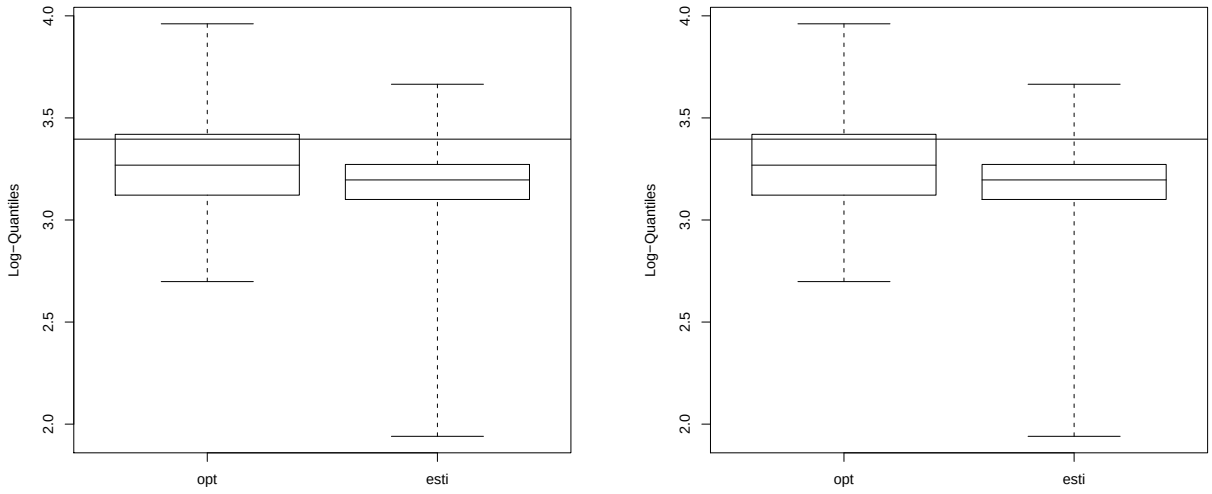
(a) $\tau = 2$ (b) $\tau = 4$

Figure 6: $\mathcal{W}(0.25, 0.25)$ distribution. Boxplots of $\log(\widetilde{x}_{p_n})$ at the optimal value of k_n obtained by minimizing the true AMSE* (left), and at the value of k_n obtained by minimizing the estimated AMSE* (right). The horizontal line indicates the true value of $\log(x_{p_n})$.



(a) $\tau = 2$ (b) $\tau = 4$

Figure 7: $\Gamma(0.25, 0.25)$ distribution. Boxplots of $\log(\widetilde{x}_{p_n})$ at the optimal value of k_n obtained by minimizing the true AMSE* (left), and at the value of k_n obtained by minimizing the estimated AMSE* (right). The horizontal line indicates the true value of $\log(x_{p_n})$.



(a) $\tau = 2$ (b) $\tau = 4$

Figure 8: $\mathcal{D}(1, 0.5)$ distribution. Boxplots of $\log(\widetilde{x}_{p_n})$ at the optimal value of k_n obtained by minimizing the true AMSE* (left), and at the value of k_n obtained by minimizing the estimated AMSE* (right). The horizontal line indicates the true value of $\log(x_{p_n})$.