Estimation of the Weibull tail-coefficient with linear combination of upper order statistics

Laurent Gardes, Stéphane Girard

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Estimation of the Weibull tail-coefficient with linear combination of upper order statistics

Laurent Gardes — Stéphane Girard

N° 5571

Mai 2005

Thème COG

ISSN 0249-6399 ISRN INRIA/RR--5571--FR+ENG

Estimation of the Weibull tail-coefficient with linear combination of upper order statistics

Laurent Gardes* , Stéphane Girard †

Thème COG — Systèmes cognitifs
Projet Mistis

Rapport de recherche n° 5571 — Mai 2005 — 27 pages

Abstract: We present a new family of estimators of the Weibull tail-coefficient. The Weibull tail-coefficient is defined as the regular variation coefficient of the inverse failure rate function. Our estimators are based on a linear combination of log-spacings of the upper order statistics. Their asymptotic normality is established and illustrated for two particular cases of estimators in this family. Their finite sample performances are presented on a simulation study.

Key-words: Weibull tail coefficient, extreme values, order statistics, regular variations.

* Laurent Gardes, Université Grenoble 2, LabSAD, BP 47, 38040 Grenoble cedex 9, France, Laurent.Gardes@upmf-grenoble.fr

† Stéphane Girard, Université Grenoble 1, SMS/LMC-IMAG, BP 53, 38041 Grenoble cedex 9, France, Stephane.Girard@imag.fr

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38334 Montbonnot Saint Ismier (France)
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Estimation du coefficient de queue de type Weibull basée sur une combinaison linéaire des plus grandes statistiques d'ordre

Résumé : Nous présentons une nouvelle famille d'estimateurs du coefficient de queue de type Weibull. Ce paramètre est l'indice de variation régulière de l'inverse du taux de panne. Nos estimateurs sont construits à partir d'une combinaison linéaire du logarithme des espacements entre les plus grandes statistiques d'ordre. La normalité asymptotique de ce type d'estimateur est établie et illustrée sur deux cas particuliers d'estimateurs choisis dans cette famille. Les performances des estimateurs sont présentées sur quelques simulations.

Mots-clés : Coefficient de queue de type Weibull, valeurs extrêmes, statistiques d'ordre, variations régulières.

1 Introduction

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables with cumulative distribution function F. We address the problem of estimating the Weibull tail-coefficient $\theta > 0$ defined when the distribution tail satisfies

(A.1) $1 - F(x) = \exp(-H(x))$, $x \geq x_0 \geq 0$, $H^{\leftarrow}(t) = \inf\{x,\ H(x) \geq t\} = t^{\theta}\ell(t)$,

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 failure rate function $H^{\leftarrow}$ is said to be regularly varying at infinity with index θ and this property is denoted by $H^{\leftarrow} \in \mathcal{R}_{\theta}$. As a comparison, Pareto type distributions satisfy $(1/(1-F))^{\leftarrow} \in \mathcal{R}_{\gamma}$, and $\gamma > 0$ is the so-called extreme value index. We refer to [7] for more information on regular variation theory. We also assume a second order condition on ℓ:

(A.2) There exist $\rho \leq 0$ and $b(x) \to 0$ such that uniformly locally on $\lambda \geq 1$

$$\log\left(\frac{\ell(\lambda x)}{\ell(x)}\right) \sim b(x)K_{\rho}(\lambda), \text{ when } x \to \infty,$$

with $K_{\rho}(\lambda) = \int_1^{\lambda} u^{\rho-1}du$.

It can be shown [11] that necessarily $|b| \in \mathcal{R}_{\rho}$. The second order parameter $\rho \leq 0$ 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 estimators. It is used in [14, 13, 3] to prove the asymptotic normality of estimators of the extreme value index γ.

When ℓ is a constant function, this problems reduce to estimating the shape parameter of a Weibull distribution. In this context, simple and efficient methods exist, see for instance [2], Chapter 4 for a review on this topic. Distributions with non-constant slowly varying functions include for instance normal, gamma and extended Weibull distributions (see Section 3 for their definitions). Such distributions are of great use to model large claims in non-life insurance [5]. Dedicated estimation methods have been proposed since the relevant information on the Weibull tail-coefficient is only contained in the extreme upper part of the sample. A first direction was investigated in [6] where an estimator based on the record values is proposed. Another family of approaches [8, 4, 12] consists of using the k_n upper order statistics $X_{n-k_n+1,n} \leq \ldots \leq X_{n,n}$ where (k_n) is an intermediate sequence of

integers such that $1 \leq k_n < n$. Our family of estimators is

$$\hat{\theta}_n(\alpha) = \frac{\sum_{i=1}^{k_n-1} \alpha_{i,n}(\log(X_{n-i+1,n}) - \log(X_{n-k_n+1,n}))}{\sum_{i=1}^{k_n-1} \alpha_{i,n}(\log_2(n/i) - \log_2(n/k_n))}, \ \alpha_{i,n} = W(i/k_n) + \varepsilon_{i,n}, \qquad (1)$$

where we have defined $\log_2(t) = \log(\log(t))$, $t > 1$, and $\varepsilon_{i,n}$, $i = 1, \ldots, k_n - 1$ is a non-random sequence. $W : [0,1] \to \mathbb{R}$ is a smooth score function, verifying

(A.3) W has a continuous derivative W' on $(0,1)$,

(A.4) There exist $M > 0$, $0 \leq q < 1/2$ and $p < 1$ such that, for all $x \in (0,1)$: $|W(x)| \leq Mx^{-q}$ and $W'(x) \leq Mx^{-p-q}$.

Let us highlight that **(A.3)** and **(A.4)** are classical assumptions when studying linear combinations of order statistics (see for instance [17]). We refer to [9, 19] for similar works in the context of the estimation of the extreme-value index.

In Section 2 we state the asymptotic normality of these estimators. Some examples of distributions satisfying **(A.1)** and **(A.2)** are given in Section 3. In Section 4, we provide two examples of weights α verifying **(A.3)** and **(A.4)**. The first one leads to the estimator of θ proposed by Girard [12]. The second one gives rise to a new estimator for Weibull tail-distributions. The behavior of these two estimators is investigated on finite sample situations in Section 5. Finally, proofs are given in Section 6.

2 Asymptotic normality

In this section, we state our main result on the limiting behavior of $\hat{\theta}_n(\alpha)$. Its proof is postponed to Section 6. In the sequel, we note $\|\varepsilon\|_{n,\infty} = \max_{i=1,\ldots,k_n-1} |\varepsilon_{i,n}|$. We also define

$$\begin{aligned}
\mu(W) &= \int_0^1 W(x)\log(1/x)dx, \\
\sigma^2(W) &= \int_0^1 \int_0^1 W(x)W(y)\frac{\min(x,y)(1-\max(x,y))}{xy}dxdy, \\
\sigma(\theta, W) &= \theta\frac{\sigma(W)}{\mu(W)}.
\end{aligned}$$

Theorem 1 *Suppose* **(A.1)**–**(A.4)** *hold. Then*

$$k_n^{1/2}(\hat{\theta}_n(\alpha) - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\theta, W)),$$

for any sequence (k_n) *satisfying* $k_n \to \infty$, $k_n/n \to 0$ *and*

$$k_n^{1/2} \max\{b(\log(n/k_n)), 1/\log(n/k_n), \|\varepsilon\|_{n,\infty}\} \to 0. \tag{2}$$

Some examples of application of this result are given in Section 4, Corollary 1 and Corollary 2.

3 Some examples of Weibull tail-distributions

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 [10], Table 3.4.4, we have $H^{\leftarrow}(x) = x^{1/2}\ell(x)$ 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 $\Gamma(\alpha, \beta)$, $\alpha, \beta > 0$. We use the following parameterization of the density

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

From [10], Table 3.4.4, we obtain $H^{\leftarrow}(x) = x\ell(x)$ 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 failure rate function is $H^{\leftarrow}(x) = \lambda x^{1/\alpha}$, and then $\ell(x) = \lambda$ 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 [15]. Its distribution tail is given by:

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

where $r \in \mathcal{R}_{\beta}$. It follows that $H^{\leftarrow}(x) = x^{1/\alpha}\ell(x)$ 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 $b(x) = -\beta\log(x)/(\alpha^2 x)$. In [15], 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 examples of estimators

First, we show in Paragraph 4.1, that our family of estimators (1) encompasses the Hill type estimator $\hat{\theta}_n^G$ proposed in [12]. Moreover, it will appear in Corollary 1 that the asymptotic normality of $\hat{\theta}_n^G$ stated in [12], Theorem 2 is just a consequence of our main result Theorem 1. Second, in Paragraph 4.2, we use the framework of Section 1 and Section 2 to exhibit a new estimator of the Weibull tail-coefficient and to establish its asymptotic normality in Corollary 2.

4.1 Girard estimator

Girard [12] proposes the following estimator of the Weibull tail-coefficient:

$$\hat{\theta}_n^G = \frac{\displaystyle\sum_{i=1}^{k_n-1}\left(\log(X_{n-i+1,n}) - \log(X_{n-k_n+1,n})\right)}{\displaystyle\sum_{i=1}^{k_n-1}\left(\log_2(n/i) - \log_2(n/k_n)\right)}. \tag{3}$$

Clearly, $\hat{\theta}_n^G$ is a particular case of $\hat{\theta}_n(\alpha)$ with $\alpha_{i,n} = 1$ for all $i = 1,\ldots,k_n-1$. Thus, we have $W(x) = 1$ for all $x \in [0,1]$ and $\varepsilon_{i,n} = 0$ for all $i = 1,\ldots,k_n$. The asymptotic normality of $\hat{\theta}_n^G$ is then a direct consequence of Theorem 1. Remarking that in this case $\sigma^2(W) = 1$ and $\mu(W) = 1$, yields the following result,

Corollary 1 *Suppose* **(A.1)** *and* **(A.2)** *hold. Then,*

$$k_n^{1/2}(\hat{\theta}_n^G - \theta) \xrightarrow{d} \mathcal{N}(0,\theta^2),$$

for any sequence (k_n) *satisfying* $k_n \to \infty,\ k_n/n \to 0$ *and*

$$k_n^{1/2} \max\{b(\log(n/k_n)), 1/\log(n/k_n)\} \to 0.$$

which is exactly Theorem 2 in [12].

4.2 Zipf estimator for Weibull tail-distribution

We propose a new estimator of the Weibull tail-coefficient based on a quantile plot (QQ-plot) adapted to our situation. It consists of drawing the pairs $(\log_2(n/i), \log(X_{n-i+1,n}))$ for $i = 1, \ldots, n-1$. The resulting graph should be approximatively linear (with slope θ), at least for the large values of i. Thus, we introduce $\hat{\theta}_n^Z$ the least square estimator of θ based on the k_n largest observations:

$$\hat{\theta}_n^Z = \frac{\sum_{i=1}^{k_n-1} (\log_2(n/i) - \tau_n) \log(X_{n-i+1,n})}{\sum_{i=1}^{k_n-1} (\log_2(n/i) - \tau_n) \log_2(n/i)}, \tag{4}$$

where

$$\tau_n = \frac{1}{k_n - 1} \sum_{i=1}^{k_n-1} \log_2(n/i).$$

This estimator is similar to the Zipf estimator for the extreme value index proposed by Kratz and Resnick [16] and Schultze and Steinebach [18]. One can prove that $\hat{\theta}_n^Z$ belongs to family (1) with a score function $W(x) = -(\log(x) + 1)$ and thus apply Theorem 1 to obtain the asymptotic normality of $\hat{\theta}_n^Z$:

Corollary 2 *Suppose* **(A.1)** *and* **(A.2)** *hold. Then,*

$$k_n^{1/2}(\hat{\theta}_n^Z - \theta) \xrightarrow{d} \mathcal{N}(0, 2\theta^2),$$

for any sequence (k_n) *satisfying* $k_n \to \infty,\ k_n/n \to 0$ *and*

$$k_n^{1/2} \max\{b(\log(n/k_n)), \log^2(k_n)/\log(n/k_n)\} \to 0. \tag{5}$$

Its proof is given in Section 6.

5 Numerical experiments

The finite sample performance of the estimators $\hat{\theta}_n^Z$ and $\hat{\theta}_n^G$ are investigated on 5 different distributions: $\Gamma(0.5, 1)$, $\Gamma(1.5, 1)$, $\mathcal{N}(1.2, 1)$, $\mathcal{W}(2.5, 2.5)$ and $\mathcal{W}(0.4, 0.4)$. We limit ourselves

to these two estimators, since it is shown in [12] that $\hat{\theta}_n^G$ gives better results than the other approaches [8, 4]. In each case, $N = 200$ samples $(\mathcal{X}_{n,i})_{i=1,\ldots,N}$ of size $n = 500$ were simulated. On each sample $(\mathcal{X}_{n,i})$, the estimates $\hat{\theta}_{n,i}^Z(k_n)$ and $\hat{\theta}_{n,i}^G(k_n)$ are computed for $k_n = 2, \ldots, 250$. Finally, the Hill-type plots are built by drawing the points

$$\left(k_n, \frac{1}{N}\sum_{i=1}^{N} \hat{\theta}_{n,i}^Z(k_n)\right) \text{ and } \left(k_n, \frac{1}{N}\sum_{i=1}^{N} \hat{\theta}_{n,i}^G(k_n)\right).$$

We also present the associated MSE (mean square error) plots obtained by plotting the points

$$\left(k_n, \frac{1}{N}\sum_{i=1}^{N} \left(\hat{\theta}_{n,i}^Z(k_n) - \theta\right)^2\right) \text{ and } \left(k_n, \frac{1}{N}\sum_{i=1}^{N} \left(\hat{\theta}_{n,i}^G(k_n) - \theta\right)^2\right).$$

The results are presented on figures 1–5. It appears that for both estimates the sign of the bias is driven by function b in **(A.2)**. It is appealing that, in all plots, the graphs obtained with $\hat{\theta}_n^Z$ are smoother than these associated with $\hat{\theta}_n^G$, making the choice of k_n less difficult in practice. The results obtained with the two estimators are very similar on Weibull distributions (figure 4 and figure 5), especially in terms of mean square error. For other Gamma and Gaussian distributions (figures 1–3), $\hat{\theta}_n^Z$ gives better results in terms of bias and mean square error.

6 Proofs

For the sake of simplicity, in the following, we note k for k_n. We first quote a lemma providing classical results on the asymptotic behavior of exponential order statistics (see [12], Lemma 1 for a detailed proof).

Lemma 1 *Let $\{E_{1,n}, \ldots, E_{n,n}\}$ be the order statistics generated by n independent standard exponential random variables. Suppose $k \to \infty$ and $k/n \to 0$. Then,*

(i) $E_{n-i+1,n}/\log(n/i) \xrightarrow{P} 1$, *uniformly on $i = 1, \ldots, k$, and*

(ii) $k^{1/2}\left(E_{n-k+1,n} - \log(n/k)\right) \xrightarrow{d} \mathcal{N}(0,1)$.

In order to be self-contained, we quote a lemma which is quite useful when dealing with linear combinations of order statistics. It summarized some results on L-statistics established in [17].

Lemma 2 *Let $Y_{1,n} \leq \ldots \leq Y_{n,n}$ be order statistics associated to n independent random variables with common distribution function F. Let J be a continuous function defined on $[0,1]$. If for some $r \geq 2$, $s \geq 2$, $\delta > 0$ and $M > 0$,*

$$|J(x)| < Mx^{1/r-1/2}(1-x)^{1/s-1/2}, \text{ for all } x \in (0,1),$$

and,

$$|J'(x)| < Mx^{-3/2+1/r+\delta}(1-x)^{-3/2+1/s+\delta},$$

and if $\int_0^1 x^{1/r}(1-x)^{1/s}dF^{-1}(x) < \infty$, *then,*

$$n^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right)Y_{i,n} - \mu\right) \xrightarrow{d} \mathcal{N}(0,\sigma^2),$$

where $\mu = \int_0^1 J(x)F^{-1}(x)dx$ *and* $\sigma^2 = \int_0^1\int_0^1 J(x)J(y)(\min(x,y)-xy)dF^{-1}(x)dF^{-1}(y)$.

The proofs of the following lemmas are postponed to Appendix. The next lemma presents an expansion of $\hat{\theta}_n(\alpha)$.

Lemma 3 *Suppose* $k \to \infty$ *and* $k/n \to 0$. *Under* **(A.1)** *and* **(A.2)**, *the following expansions hold:*

$$\hat{\theta}_n(\alpha) \stackrel{d}{=} \frac{T_n^{(2)}}{T_n^{(1)}} \stackrel{d}{=} \frac{\theta T_n^{(3,0)} + (1+o_P(1))b(E_{n-k+1,n})T_n^{(3,\rho)}}{T_n^{(1)}},$$

where we have defined

$$\begin{aligned}
T_n^{(1)} &= \frac{1}{k-1}\sum_{i=1}^{k-1}\alpha_{i,n}(\log_2(n/i) - \log_2(n/k)),\\
T_n^{(2)} &= \frac{1}{k-1}\sum_{i=1}^{k-1}\alpha_{i,n}(\log(X_{n-i+1,n}) - \log(X_{n-k+1,n})),\\
T_n^{(3,\rho)} &= \frac{1}{k-1}\sum_{i=1}^{k-1}\alpha_{k-i,n}K_\rho\left(1+\frac{F_{i,k-1}}{E_{n-k+1,n}}\right),\ \rho \le 0,
\end{aligned}$$

and where

- $E_{n-k+1,n}$ *is the* $(n-k+1)$*th order statistics associated to* n *independent standard exponential variables.*
- $\{F_{1,k-1},\ldots,F_{k-1,k-1}\}$ *are ordered statistics independent from* $E_{n-k+1,n}$ *and generated by* $k-1$ *independent standard exponential variables.*

The following lemma provides an expansion of

$$\tau_n = \frac{1}{k-1}\sum_{i=1}^{k-1}(\log_2(n/i) - \log_2(n/k)),$$

which frequently appears in the proofs.

Lemma 4 *If $k \to \infty$ and $k/n \to 0$ as $n \to \infty$ then,*

$$\tau_n = \frac{1}{\log(n/k)} \left(1 + O\left(\frac{\log(k)}{k}\right) + O\left(\frac{1}{\log(n/k)}\right)\right).$$

The next lemmas are dedicated to the study of the different terms appearing in Lemma 3. First, we focus on the non-random term $T_n^{(1)}$.

Lemma 5 *Suppose $k \to \infty$ and $k/n \to 0$. Under* **(A.1)**–**(A.4)**, *the following expansion hold:*

$$T_n^{(1)} = \frac{\mu(W)}{\log(n/k)} \left\{1 + O\left(\log(k) k^{q-1}\right) + O\left(\frac{1}{\log(n/k)}\right) + O\left(\|\varepsilon\|_{n,\infty}\right)\right\}.$$

Second, we focus on the random term $T_n^{(3,\rho)}$.

Lemma 6 *Suppose $k \to \infty$ and $k/n \to 0$. Under* **(A.1)**–**(A.4)**, *the following expansion hold for all $\rho \leq 0$:*

$$T_n^{(3,\rho)} \stackrel{d}{=} \frac{\mu(W)}{E_{n-k+1,n}} \left\{1 + \frac{\sigma(W)}{\mu(W)} k^{-1/2} \xi_n + O_P\left(\frac{1}{\log(n/k)}\right) + O_P\left(\|\varepsilon\|_{n,\infty}\right)\right\},$$

where $\xi_n \stackrel{d}{\to} N(0,1)$.

We are now in position to prove Theorem 1 and Corollary 2.

Proof of Theorem 1. From Lemma 3, we have

$$\begin{aligned} k^{1/2}(\hat{\theta}_n(\alpha) - \theta) &\stackrel{d}{=} \theta k^{1/2} \left(\frac{T_n^{(3,0)}}{T_n^{(1)}} - 1\right) + k^{1/2} b(E_{n-k+1,n}) \frac{T_n^{(3,\rho)}}{T_n^{(1)}} (1 + o_P(1)) \\ &=: T_n^{(4,1)} + T_n^{(4,2)}. \end{aligned}$$

Lemma 5 and Lemma 6 yield for all $\rho \leq 0$:

$$\begin{aligned} \frac{T_n^{(3,\rho)}}{T_n^{(1)}} \stackrel{d}{=} \frac{\log(n/k)}{E_{n-k+1,n}} \Bigg\{ 1 &+ \frac{\sigma(W)}{\mu(W)} k^{-1/2} \xi_n \\ &+ O_P\left(\frac{1}{\log(n/k)}\right) + O_P\left(\|\varepsilon\|_{n,\infty}\right) + O\left(k^{q-1} \log(k)\right) \Bigg\}. \end{aligned}$$

Lemma 1(ii) entails that

$$\frac{\log(n/k)}{E_{n-k+1,n}} \stackrel{d}{=} 1 + O_P\left(\frac{k^{-1/2}}{\log(n/k)}\right).$$

Consequently, we have

$$\frac{T_n^{(3,\rho)}}{T_n^{(1)}} \stackrel{d}{=} 1 + \frac{\sigma(W)}{\mu(W)} k^{-1/2} \xi_n + O_P\left(\frac{1}{\log(n/k)}\right) + O_P\left(\|\varepsilon\|_{n,\infty}\right) + O\left(k^{q-1}\log(k)\right), \qquad (6)$$

and thus

$$\begin{aligned} T_n^{(4,1)} &\stackrel{d}{=} \theta \frac{\sigma(W)}{\mu(W)} \xi_n + O_P\left(\frac{k^{1/2}}{\log(n/k)}\right) + O_P\left(k^{1/2}\|\varepsilon\|_{n,\infty}\right) + O\left(k^{q-1/2}\log(k)\right) \\ &\stackrel{d}{\to} N(0, \sigma^2(\theta, W)), \end{aligned}$$

with (2) and since $q < 1/2$. Equation (6) also implies that

$$\begin{aligned} T_n^{(4,2)} &\stackrel{d}{=} k^{1/2} b(E_{n-k+1,n})(1 + o_P(1)) \\ &\stackrel{d}{=} k^{1/2} b(\log(n/k))(1 + o_P(1)) \\ &\stackrel{P}{\to} 0. \end{aligned}$$

with (2) and after remarking that $b(E_{n-k+1,n})/b(\log(n/k))$ converges to 1 in probability, since $E_{n-k+1,n}/\log(n/k)$ converges to 1 in probability (see Lemma 1(i)) and $|b| \in \mathcal{R}_\rho$. The result is proved. ■

Proof of Corollary 2. First remark that (4) can be rewritten as

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

where

$$\begin{aligned} \alpha_{i,n}^Z &= \log(n/k_n)\left(\log_2(n/i) - \tau_n\right) \\ &= \log(n/k)\left(\log\left(1 + \frac{\log(k/i)}{\log(n/k)}\right) - \tau_n\right) \\ &= \log(k/i) + O\left(\frac{\log^2(k)}{\log(n/k)}\right) - \log(n/k)\tau_n, \\ &= \log(k/i) - 1 + O\left(\frac{\log^2(k)}{\log(n/k)}\right) + O\left(\frac{\log(k)}{k}\right), \end{aligned}$$

uniformly on $i = 1, \ldots, k$ with Lemma 4. Therefore, we have $\alpha_{i,n}^Z = W(i/k) + \varepsilon_{i,n}$ with $W(x) = -(\log(x) + 1)$ and $\varepsilon_{i,n} = O(\log^2(k)/\log(n/k)) + O(\log(k)/k)$, uniformly on $i =$

$1, \ldots, k$. Then, it is easy to check that W satisfies conditions **(A.3)** and **(A.4)** and that condition (2) reduces to condition (5). We conclude the proof by remarking that

$$\mu(W) = \int_0^1 \log(x)(\log(x)+1)dx = 1$$

and

$$\sigma^2(W) = \int_0^1 \int_0^1 (\log(x)+1)(\log(y)+1)\frac{\min(x,y)(1-\max(x,y))}{xy}dxdy = 2.$$

■

References

[1] Arnold, B.C, Balakrishnan, N., Nagaraja H.N., (1992), *A First course in order statistics*, Wiley and sons.

[2] Bain, L., Engelhardt, M., (1991), *Statistical analysis of reliability and life-testing models. Theory and methods. 2nd ed.*, Statistics: Textbooks and Monographs Series, **115**, Marcel Dekker, Inc.

[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., 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.

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

[6] Berred, M., (1991), Record values and the estimation of the Weibull tail-coefficient. *Comptes-Rendus de l'Académie des Sciences* **t. 312**, Série I, 943–946.

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

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

[9] Csörgö, S., Deheuvels, P., Mason, D., (1985), Kernel estimates of the tail index of a distribution, *The Annals of Statistics*, **13**, 1050–1077.

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

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

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

[13] 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.

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

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

[16] Kratz, M., Resnick, S., (1996), The QQ-estimator and heavy tails, *Stochastic Models*, **12**, 699–724.

[17] Mason, D.M., (1981), Asymptotic normality of linear combinations of order statistics with a smooth score function, *Annals of Statistics*, **9**(4), 899–908.

[18] Schultze, J., Steinebach, J., (1996), On least squares estimates of an exponential tail coefficient, *Statistics and Decisions*, **14**, 353–372.

[19] Viharos, L. (1999), Weighted least-squares estimators of tail indices, *Probability and Mathematical Statistics*, **19**, 249–265.

Appendix: proof of lemmas

In the sequel, we note $J(x) = W(1-x)$ for $x \in (0,1)$.

Proof of Lemma 3. Let us consider

$$T_n^{(2)} =: \frac{1}{k-1}\sum_{i=1}^{k-1} \alpha_{i,n}(\log(X_{n-i+1,n}) - \log(X_{n-k+1,n})),$$

and let $E_{1,n}, \ldots, E_{n,n}$ be ordered statistics generated by n independent standard exponential random variables. Under **(A.1)**, we have

$$\begin{aligned} T_n^{(2)} &\stackrel{d}{=} \frac{1}{k-1}\sum_{i=1}^{k-1} \alpha_{i,n}(\log H^{\leftarrow}(E_{n-i+1,n}) - \log H^{\leftarrow}(E_{n-k+1,n})) \\ &\stackrel{d}{=} \theta\frac{1}{k-1}\sum_{i=1}^{k-1} \alpha_{i,n}\log\left(\frac{E_{n-i+1,n}}{E_{n-k+1,n}}\right) + \frac{1}{k-1}\sum_{i=1}^{k-1} \alpha_{i,n}\log\left(\frac{\ell(E_{n-i+1,n})}{\ell(E_{n-k+1,n})}\right). \end{aligned}$$

Define $x_n = E_{n-k+1,n}$ and $\lambda_{i,n} = E_{n-i+1,n}/E_{n-k+1,n}$. It is clear, in view of Lemma 1(i) that $x_n \stackrel{P}{\to} \infty$ and $\lambda_{i,n} \stackrel{P}{\to} 1$. Thus, **(A.2)** yields that uniformly in $i = 1, \ldots, k-1$:

$$\begin{aligned} T_n^{(2)} &\stackrel{d}{=} \theta\frac{1}{k-1}\sum_{i=1}^{k-1} \alpha_{i,n}\log\left(\frac{E_{n-i+1,n}}{E_{n-k+1,n}}\right) \\ &+ (1+o_p(1))b(E_{n-k+1,n})\frac{1}{k-1}\sum_{i=1}^{k-1} \alpha_{i,n}K_\rho\left(\frac{E_{n-i+1,n}}{E_{n-k+1,n}}\right). \end{aligned}$$

The Rényi representation of the Exp(1) ordered statistics ([1], p. 72) yields

$$\left\{\frac{E_{n-i+1,n}}{E_{n-k+1,n}}\right\}_{i=1,\ldots,k-1} \stackrel{d}{=} \left\{1 + \frac{F_{k-i,k-1}}{E_{n-k+1,n}}\right\}_{i=1,\ldots,k-1}, \tag{7}$$

where $\{F_{1,k-1}, \ldots, F_{k-1,k-1}\}$ are ordered statistics independent from $E_{n-k+1,n}$ and generated by $k-1$ independent standard exponential variables $\{F_1, \ldots, F_{k-1}\}$. Therefore,

$$\begin{aligned} T_n^{(2)} &\stackrel{d}{=} \theta\frac{1}{k-1}\sum_{i=1}^{k-1} \alpha_{i,n}\log\left(1 + \frac{F_{k-i,k-1}}{E_{n-k+1,n}}\right) \\ &+ (1+o_p(1))b(E_{n-k+1,n})\frac{1}{k-1}\sum_{i=1}^{k-1} \alpha_{i,n}K_\rho\left(1 + \frac{F_{k-i,k-1}}{E_{n-k+1,n}}\right). \end{aligned}$$

Changing i to $k-i$ in the above formula and remarking that $K_0(x) = \log(x)$ conclude the proof. ■

Proof of Lemma 4. We have

$$\begin{aligned}
\tau_n &= \frac{1}{k-1}\sum_{i=1}^{k-1}\log\left(1+\frac{\log(k/i)}{\log(n/k)}\right)\\
&= \frac{1}{\log(n/k)}\frac{1}{k-1}\sum_{i=1}^{k-1}\log(k/i)\\
&+ \frac{1}{k-1}\sum_{i=1}^{k-1}\left(\log\left(1+\frac{\log(k/i)}{\log(n/k)}\right)-\frac{\log(k/i)}{\log(n/k)}\right)\\
&=: -\frac{1}{\log(n/k)}\tau_n^{(1)}+\tau_n^{(2)}.
\end{aligned}$$

The inequality $-x^2/2 \leq \log(1+x)-x \leq 0$, $x \geq 0$ yields:

$$|\tau_n^{(2)}| \leq \frac{1}{2\log^2(n/k)}\frac{1}{k-1}\sum_{i=1}^{k-1}\log^2(i/k) = O\left(\frac{1}{\log^2(n/k)}\right), \tag{8}$$

since

$$\frac{1}{k-1}\sum_{i=1}^{k-1}\log^2(k/i) \to \int_0^1 \log^2(x)dx = 2,$$

as $k \to \infty$. Furthermore, remark that

$$\tau_n^{(1)} = \frac{1}{k-1}\log\left(\prod_{i=1}^{k}\frac{i}{k}\right) = \frac{1}{k-1}\log\left(\frac{k!}{k^k}\right).$$

Using Stirling's formula:

$$k! = \left(\frac{k}{\mathrm{e}}\right)^k\sqrt{2\pi k}(1+o(1)),$$

leads to

$$\tau_n^{(1)} = \frac{1}{k-1}\left(\frac{1}{2}\log(2\pi k)-k+o(1)\right) = -1+O\left(\frac{\log(k)}{2k}\right). \tag{9}$$

Collecting (8) and (9) concludes the proof. ■

Proof of Lemma 5. Since $\alpha_{i,n} = W(i/k) + \varepsilon_{i,n}$, we have

$$\begin{aligned} T_n^{(1)} &= \frac{1}{k-1}\sum_{i=1}^{k-1} W(i/k)\log\left(1+\frac{\log(k/i)}{\log(n/k)}\right) + \frac{1}{k-1}\sum_{i=1}^{k-1}\varepsilon_{i,n}\log\left(1+\frac{\log(k/i)}{\log(n/k)}\right) \\ &=: T_n^{(1,1)} + T_n^{(1,2)}. \end{aligned}$$

The first term can be expanded as

$$\begin{aligned} T_n^{(1,1)} &= \frac{1}{\log(n/k)}\frac{1}{k-1}\sum_{i=1}^{k-1} W(i/k)\log(k/i) \\ &+ \frac{1}{k-1}\sum_{i=1}^{k-1} W(i/k)\left\{\log\left(1+\frac{\log(k/i)}{\log(n/k)}\right) - \frac{\log(k/i)}{\log(n/k)}\right\} \\ &=: \frac{T_n^{(1,1,1)}}{\log(n/k)} + T_n^{(1,1,2)}. \end{aligned}$$

Let us define $H(x) = W(x)\log(1/x)$, $x \in (0,1)$. Then, the Riemann sum $T_n^{(1,1,1)}$ can be compared to $\mu(W)$ by:

$$\begin{aligned} |T_n^{(1,1,1)} - \mu(W)| &\le \sum_{i=1}^{k-1}\int_{i/k}^{(i+1)/k} |H(i/k) - H(x)|dx + \int_0^{1/k} |H(x)|dx + O(1/k) \\ &= \frac{1}{2k^2}\sum_{i=1}^{k-1}\sup_{i/k\le x\le (i+1)/k} |H'(x)| + \int_0^{1/k} |H(x)|dx + O(1/k) \\ &=: T_n^{(1,1,1,1)} + T_n^{(1,1,1,2)} + O(1/k). \end{aligned}$$

Assumption **(A.4)** implies that there exists $M' > 0$ such that $|H'(x)| \le M' x^{-q-1}$ for all $x \in (0,1]$ and thus,

$$\begin{aligned} T_n^{(1,1,1,1)} &\le \frac{M'}{2k^2}\sum_{i=1}^{k-1}\left(\frac{i}{k}\right)^{-q-1} \\ &\le \frac{M'}{2k}\left(\int_{1/k}^1 t^{-q-1}dt + k^q\right) = \begin{cases} O\left(k^{q-1}\right) & \text{if } q \ne 0, \\ O\left(\log(k)/k\right) & \text{if } q = 0. \end{cases} \end{aligned} \tag{10}$$

Assumption **(A.4)** also yields $|H(x)| \le M x^{-q}\log(1/x)$ for all $x \in (0,1]$ and thus,

$$\begin{aligned} |T_n^{(1,1,1,2)}| &\le M\int_0^{1/k} x^{-q}\log(1/x)dx \\ &= O\left(k^{q-1}\log(k)\right). \end{aligned} \tag{11}$$

Collecting (10) and (11) implies that

$$T_n^{(1,1,1)} = \mu(W) + O(k^{q-1}\log(k)). \tag{12}$$

Besides, the well-known inequality $|\log(1+x) - x| \leq x^2/2$, $x > 0$ and **(A.4)** lead to

$$\begin{aligned} |T_n^{(1,1,2)}| &\leq \frac{1}{k-1}\sum_{i=1}^{k-1} |W(i/k)| \left|\log\left(1 + \frac{\log(k/i)}{\log(n/k)}\right) - \frac{\log(k/i)}{\log(n/k)}\right| \\ &\leq \frac{M}{2\log^2(n/k)} \frac{1}{k-1}\sum_{i=1}^{k-1} (i/k)^{-q}\log^2(k/i) \\ &= O\left(\frac{1}{\log^2(n/k)}\right), \end{aligned} \tag{13}$$

since

$$\frac{1}{k-1}\sum_{i=1}^{k-1} (i/k)^{-q}\log^2(k/i) \to \int_0^1 x^{-q}\log^2(1/x)dx < +\infty,$$

when $k \to \infty$. Finally, $T_n^{(1,2)}$ is bounded by

$$\begin{aligned} |T_n^{(1,2)}| &\leq \|\varepsilon\|_{n,\infty} \frac{1}{k-1}\sum_{i=1}^{k-1} \log\left(1 + \frac{\log(k/i)}{\log(n/k)}\right) = \|\varepsilon\|_{n,\infty}\tau_n \\ &= O\left(\frac{\|\varepsilon\|_{n,\infty}}{\log(n/k)}\right), \end{aligned} \tag{14}$$

by Lemma 4. Collecting (12), (13) and (14) gives the result. ■

Proof of Lemma 6. Since $\alpha_{k-i,n} = J(i/k) + \varepsilon_{k-i,n}$, we have,

$$\begin{aligned} T_n^{(3,\rho)} &= \frac{1}{k-1}\sum_{i=1}^{k-1} J(i/k)K_\rho\left(1 + \frac{F_{i,k-1}}{E_{n-k+1,n}}\right) \\ &+ \frac{1}{k-1}\sum_{i=1}^{k-1} \varepsilon_{k-i,n}K_\rho\left(1 + \frac{F_{i,k-1}}{E_{n-k+1,n}}\right) \\ &=: T_n^{(3,\rho,1)} + T_n^{(3,\rho,2)}. \end{aligned}$$

The first term can be expanded as

$$\begin{aligned} T_n^{(3,\rho,1)} &= \frac{1}{E_{n-k+1,n}} \frac{1}{k-1} \sum_{i=1}^{k-1} J(i/k) F_{i,k-1} \\ &+ \frac{1}{k-1} \sum_{i=1}^{k-1} J(i/k) \left\{ K_\rho \left(1 + \frac{F_{i,k-1}}{E_{n-k+1,n}} \right) - \frac{F_{i,k-1}}{E_{n-k+1,n}} \right\} \\ &=: \frac{T_n^{(3,\rho,1,1)}}{E_{n-k+1,n}} + T_n^{(3,\rho,1,2)}. \end{aligned}$$

Now, **(A.3)** and **(A.4)** imply that the linear combination of exponential order statistics $T_n^{(3,\rho,1,1)}$ satisfies the conditions of Lemma 2 and thus is asymptotically Gaussian. More precisely, we have

$$T_n^{(3,\rho,1,1)} \stackrel{d}{=} \mu(W) + \sigma(W) k^{-1/2} \xi_n, \tag{15}$$

where $\xi_n \stackrel{d}{\to} N(0,1)$,

$$\begin{aligned} \mu(W) &= \int_0^1 W(x) \log(1/x) dx \text{ and} \\ \sigma^2(W) &= \int_0^1 \int_0^1 W(x) W(y) \frac{\min(x,y)(1-\max(x,y))}{xy} dx dy. \end{aligned}$$

The upper bound on $T_n^{(3,\rho,1,2)}$ is obtained by remarking that for all $x \geq 0$,

$$|K_\rho(1+x) - x| \leq \frac{1-\rho}{2} x^2.$$

It follows that

$$\begin{aligned} T_n^{(3,\rho,1,2)} &\leq \frac{1}{k-1} \sum_{i=1}^{k-1} |J(i/k)| \left| K_\rho \left(1 + \frac{F_{i,k-1}}{E_{n-k+1,n}} \right) - \frac{F_{i,k-1}}{E_{n-k+1,n}} \right| \\ &\leq \frac{1-\rho}{2} \frac{1}{E_{n-k+1,n}^2} \frac{1}{k-1} \sum_{i=1}^{k-1} |J(i/k)| F_{i,k-1}^2. \end{aligned}$$

Now, when $k \to \infty$,

$$\frac{1}{k-1} \sum_{i=1}^{k-1} |J(i/k)| F_{i,k-1}^2 = O_P(1) \text{ and } \frac{E_{n-k+1,n}}{\log(n/k)} \stackrel{P}{\to} 1,$$

by Lemma 1(i) and Lemma 2. Thus

$$T_n^{(3,\rho,1,2)} = \frac{1}{E_{n-k+1,n}} O_P \left(\frac{1}{\log(n/k)} \right), \tag{16}$$

and then collecting (15) and (16),

$$T_n^{(3,\rho,1)} = \frac{1}{E_{n-k+1,n}} \left(\mu(W) + \sigma(W) k^{-1/2} \xi_n + O_P\left(\frac{1}{\log(n/k)} \right) \right). \tag{17}$$

Similarly, $T_n^{(3,\rho,2)}$ is bounded by

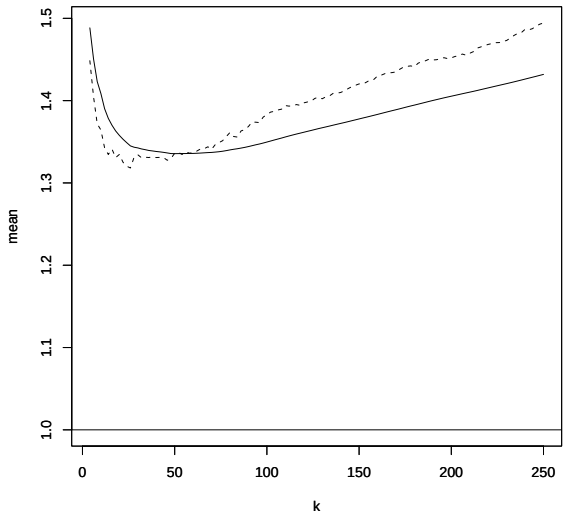
$$\begin{aligned} |T_n^{(3,\rho,2)}| &\leq \|\varepsilon\|_{n,\infty} \frac{1}{k-1} \sum_{i=1}^{k-1} K_\rho \left(1 + \frac{F_{i,k-1}}{E_{n-k+1,n}} \right) \\ &\leq \|\varepsilon\|_{n,\infty} \frac{1}{k-1} \sum_{i=1}^{k-1} \frac{F_{i,k-1}}{E_{n-k+1,n}} \\ &\stackrel{d}{=} \frac{\|\varepsilon\|_{n,\infty}}{E_{n-k+1,n}} \frac{1}{k-1} \sum_{i=1}^{k-1} F_i \\ &= \frac{1}{E_{n-k+1,n}} O_P\left(\|\varepsilon\|_{n,\infty} \right), \end{aligned} \tag{18}$$

by the law of large numbers. Collecting (17) and (18) gives the result. ■

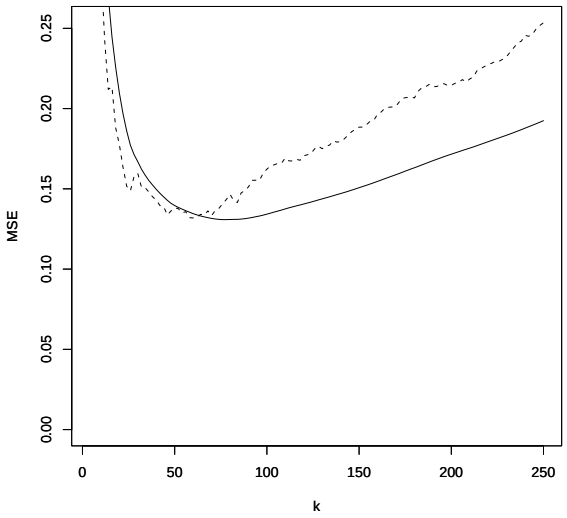
Contents

	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

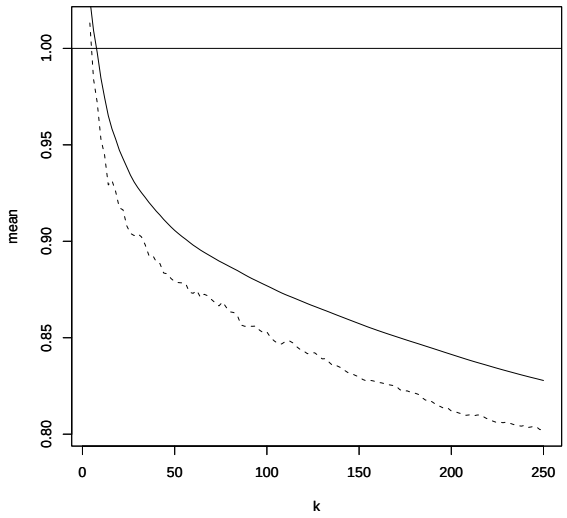


(a) Mean as a function of k_n

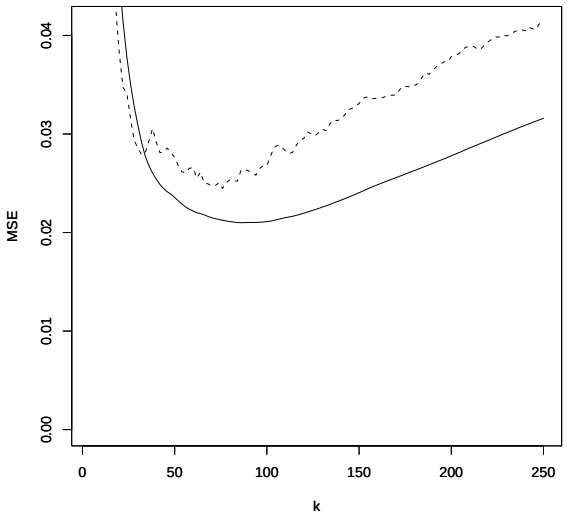


(b) Mean square error as a function of k_n.

Figure 1: Comparison of estimates $\hat{\theta}_n^Z$ (solid line) and $\hat{\theta}_n^G$ (dashed line) for the $\Gamma(0.5, 1)$ distribution. In (a), the straight line is the true value of θ.

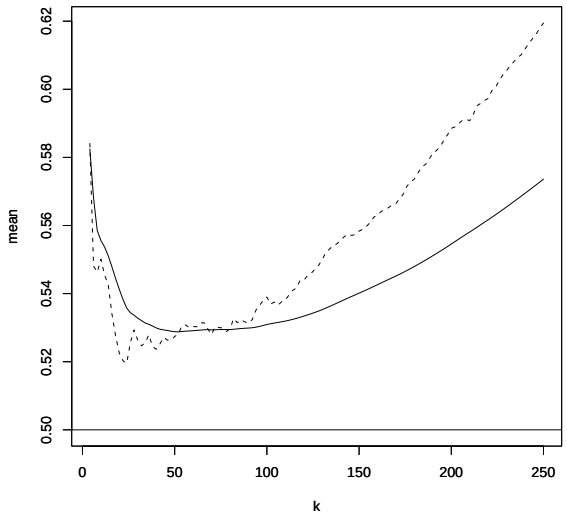


(a) Mean as a function of k_n

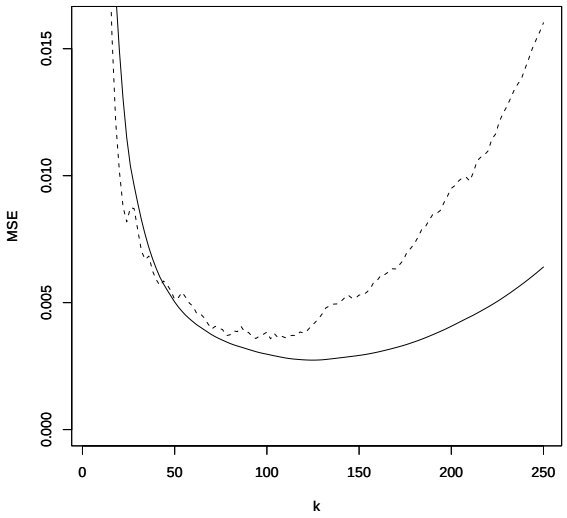


(b) Mean square error as a function of k_n.

Figure 2: Comparison of estimates $\hat{\theta}_n^Z$ (solid line) and $\hat{\theta}_n^G$ (dashed line) for the $\Gamma(1.5, 1)$ distribution. In (a), the straight line is the true value of θ.

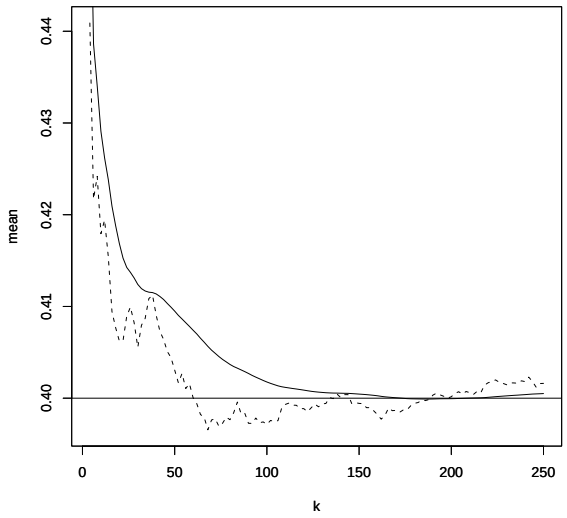


(a) Mean as a function of k_n

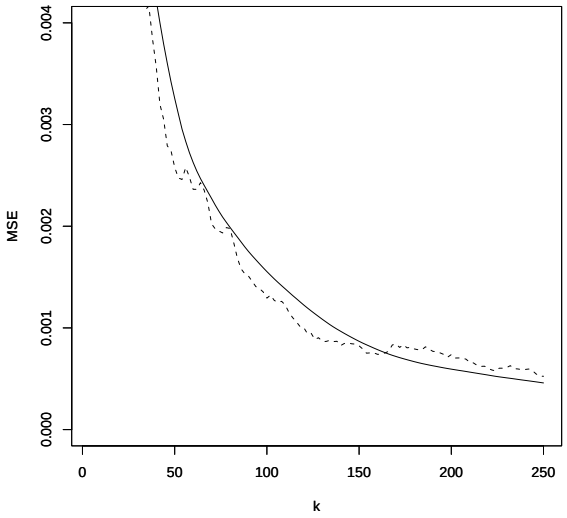


(b) Mean square error as a function of k_n.

Figure 3: Comparison of estimates $\hat{\theta}_n^Z$ (solid line) and $\hat{\theta}_n^G$ (dashed line) for the $\mathcal{N}(1.2, 1)$ distribution. In (a), the straight line is the true value of θ.

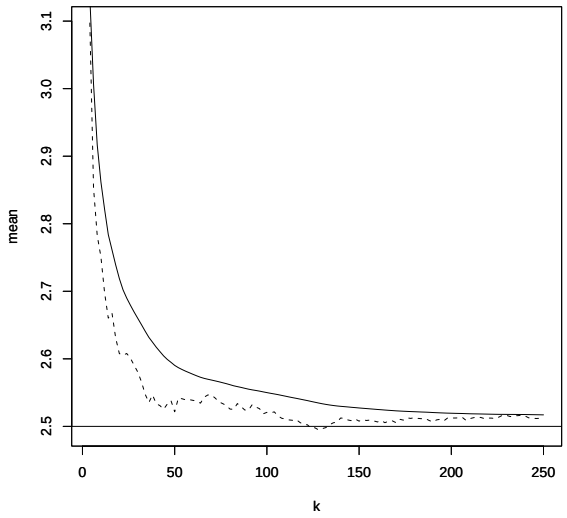


(a) Mean as a function of k_n

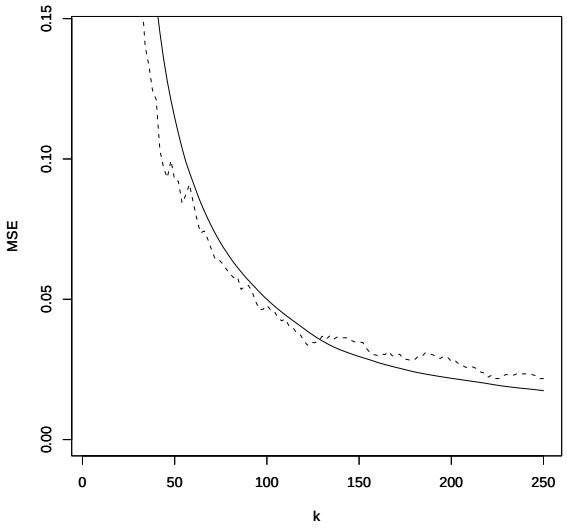


(b) Mean square error as a function of k_n.

Figure 4: Comparison of estimates $\hat{\theta}_n^Z$ (solid line) and $\hat{\theta}_n^G$ (dashed line) for the $\mathcal{W}(2.5, 2.5)$ distribution. In (a), the straight line is the true value of θ.



(a) Mean as a function of k_n



(b) Mean square error as a function of k_n.

Figure 5: Comparison of estimates $\hat{\theta}_n^Z$ (solid line) and $\hat{\theta}_n^G$ (dashed line) for the $\mathcal{W}(0.4, 0.4)$ distribution. In (a), the straight line is the true value of θ.

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399