



**HAL**  
open science

# Optimal rates of convergence in the Weibull model based on kernel-type estimators

Cécile Mercadier, Philippe Soulier

► **To cite this version:**

Cécile Mercadier, Philippe Soulier. Optimal rates of convergence in the Weibull model based on kernel-type estimators. 2011. hal-00565628v1

**HAL Id: hal-00565628**

**<https://hal.science/hal-00565628v1>**

Submitted on 14 Feb 2011 (v1), last revised 20 Sep 2011 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Optimal rates of convergence in the Weibull model based on kernel-type estimators

Cécile Mercadier\*

Philippe Soulier†

## Abstract

Let  $F$  be a distribution function in the maximal domain of attraction of the Gumbel distribution and such that  $-\log(1 - F(x)) = x^{1/\theta}L(x)$  for a positive real number  $\theta$ , called the Weibull tail index, and a slowly varying function  $L$ . It is well known that the estimators of  $\theta$  have a very slow rate of convergence. We establish here a sharp optimality result in the minimax sense, that is when  $L$  is treated as an infinite dimensional nuisance parameter belonging to some functional class. We also establish the rate optimal asymptotic property of a data-driven choice of the sample fraction that is used for estimation.

**Keywords:** Weibull tail index; Rates of convergence; Kernel-type estimators; Optimal sample fraction; Sequential procedure.

**MSC2010 Subject Classification:** primary 62G32; 62G20; secondary 62G05; 62F35.

## 1 Introduction

Let  $F$  be a probability distribution function on  $[0, \infty)$  such that  $1 - F(x) = \exp(-H(x))$  for  $x \geq 0$  and  $H^\leftarrow$  is regularly varying at infinity with positive index  $\theta$ , i.e.

$$H^\leftarrow(t) = t^\theta \ell(t)$$

where  $f^\leftarrow$  denotes the left-continuous inverse of a nondecreasing function  $f$ , and the function  $\ell$  is slowly varying at infinity, i.e. for all  $x > 0$ ,

$$\lim_{t \rightarrow \infty} \ell(xt)/\ell(t) = 1.$$

The quantile function  $Q = F^\leftarrow$  can then be expressed as

$$Q(t) = H^\leftarrow(-\log(1 - t)) = (-\log(1 - t))^\theta \ell(-\log(1 - t)).$$

Such probability distributions  $F$  are usually called Weibull type distributions and the parameter  $\theta$  is called the Weibull tail index or coefficient. These distributions belong to the maximum domain of attraction of the Gumbel distribution. Note that  $\theta = 1/2$  for the positive part of the Gaussian distribution and  $\theta = 1$  for the  $\Gamma(a, \lambda)$  distribution. The estimation

---

\*Université de Lyon; CNRS; Université Lyon 1, Institut Camille Jordan, 43 Boulevard du 11 novembre 1918, F-69622, Villeurbanne Cedex, France

†Laboratoire MODAL'X, Département de Mathématiques, Université Paris Ouest-Nanterre, 200 avenue de la République, 92001 Nanterre Cedex, France

of the parameter  $\theta$  has been considered by [Beirlant et al. \[1996\]](#), [Embrechts et al. \[1997\]](#) and [Girard \[2004\]](#) among other references. These estimators are based on linear combinations of log-spacings between upper order statistics. They have been proved to be consistent with a logarithmic rate of convergence, under semiparametric assumptions, i.e. when the slowly varying function  $\ell$  is considered as an infinite dimensional nuisance parameter. Nevertheless, this logarithmic rate of convergence has never been proved to be optimal in the minimax sense.

To obtain the minimax rate of convergence of an estimator of  $\theta$  in a semiparametric setting is the first goal of this note. A lower bound is obtained in [Section 2](#) and an upper bound for a kernel estimator is derived in [Section 3](#). As often when the rate of convergence is logarithmic, the two bounds are matching. This was observed by [Drees \[1998\]](#) in the context of estimation of the tail index of a regularly varying distribution. In [Section 4](#), we discuss the choice of the number of upper order statistics used to compute the estimator. Several approaches have been considered, based on the asymptotic mean squared error. Pitfalls of these methods have recently been highlighted by [Asimit et al. \[2010\]](#). We suggest a different approach, elaborating on an idea of [Drees and Kaufmann \[1998\]](#), which was itself inspired by the so-called Lepsky method which is familiar in nonparametric estimation. This method is shown to have optimal asymptotic properties. Unfortunately but unsurprisingly, this data-driven estimator has a very poor finite sample behaviour. Improving on this method is the subject of future research.

## 2 Lower bound

Let  $b$  be a function defined on  $[1, \infty]$  such that  $\lim_{x \rightarrow \infty} b(x) = 0$ . Let  $\mathcal{L}(b)$  be the class of locally bounded slowly varying functions  $\ell$  on  $[0, \infty)$  that can be expressed as

$$\ell(x) = \exp \int_1^x \frac{\eta(s)}{s} ds, \quad (1)$$

with  $|\eta(s)| \leq b(s)$  for  $s \geq 1$ , and such that  $x \mapsto x^\theta \ell(x)$  is nondecreasing on  $[0, \infty)$  for any  $\theta > 0$ .

**Theorem 2.1.** *Let  $\theta_2 > \theta_1 > 0$  be positive real numbers and let  $b$  be a decreasing function defined on  $[1, \infty]$  such that  $\lim_{x \rightarrow \infty} b(x) = 0$ . Then*

$$\liminf_{n \rightarrow \infty} \inf_{\hat{\theta}_n} \sup_{\theta \in (\theta_1, \theta_2)} \sup_{\ell \in \mathcal{L}(b)} b^{-1}(\log(n)) \mathbb{E}_{\theta, \ell}^{(n)}[|\hat{\theta}_n - \theta|] \geq 1, \quad (2)$$

where  $\mathbb{E}_{\theta, \ell}^{(n)}$  denotes expectation with respect to the law of an  $n$ -sample from a distribution with quantile function  $Q(t) = (-\log(1-t))^\theta \ell(-\log(1-t))$ , and the infimum is taken over all sequences of estimators of  $\theta$ .

*Example 1.* If  $b(x) = x^\rho$  with  $\rho < 0$ , then the lower bound for the rate of convergence of estimators of  $\theta$  is  $\log^\rho(n)$ .

*Example 2.* If  $b(x) = \log^{-1}(x)$ , then the lower bound is  $\log \log(n)$ .

*Example 3.* Consider a distribution function with tail  $\bar{F}(x) = c_1 x^\delta e^{-c_2 x^\beta}$  with  $c_i > 0$ ,  $\beta > 0$  and  $\delta \neq 0$ . Then the quantile function can be expressed as  $Q(t) = z^{1/\beta} \ell(z)$  with  $z = -c_2^{-1} \log\{(1-t)/c_1\}$  and  $\ell(z) \sim 1 - \beta^{-2} c_2^{-1} \delta z^{-1} \log(z)$  as  $z \rightarrow \infty$ . Thus  $\ell$  is in the class  $\mathcal{L}(b)$  with  $b(z) = Cz^{-1} \log(z)$  for large enough  $C$  and  $z \geq e$ . The uniform rate of convergence over this class is  $C^{-1} \log(n) / \log \log(n)$ . See [Kluppelberg and Villaseñor \[1993\]](#) and [Asimit et al. \[2010\]](#).

### 3 Kernel estimates of the Weibull tail index

Let  $X_1, X_2, \dots, X_n$  be a sequence of independent and identically distributed random variables with cumulative distribution function  $F$ . We denote by  $X_{1,n}, \dots, X_{n,n}$  their associated order statistics. Following Csörgő et al. [1985], the empirical quantile process,  $Q_n$  is then defined by

$$Q_n(t) = X_{k,n} \Leftrightarrow \frac{k-1}{n} < t \leq \frac{k}{n}, \quad 1 \leq k \leq n.$$

Let  $K$  denote a bounded nonnegative nonincreasing kernel with compact support in  $[0, \infty)$  and such that  $vK'(v)$  is bounded. Without loss of generality, we can assume that  $K$  has compact support in  $[0, 1]$ . Let also  $\lambda = \lambda_n > 0$  be a tuning parameter such that  $\lim_{n \rightarrow \infty} \lambda_n = 0$ . The kernel estimate of the Weibull tail index  $\theta$  is defined by

$$\hat{\theta}_n = \frac{\int_0^1 \log Q_n(1 - v\lambda) d\{vK(v)\}}{\int_0^1 \frac{K(v)}{(-\log(v\lambda))} dv}. \quad (3)$$

If  $K$  is moreover right continuous, we can write

$$\hat{\theta}_n = \frac{\sum_{j=1}^{\lfloor n\lambda \rfloor} \frac{j}{n\lambda} K\left(\frac{j}{n\lambda}\right) \{\log X_{n-j+1,n} - \log X_{n-j,n}\}}{\int_0^1 \frac{K(v)}{(-\log(v\lambda))} dv}$$

where we set  $X_{0,n} = 1$ .

*Example 4.* For  $\lambda = k/n$  and  $K(v) = \mathbf{1}_{0 \leq v < 1}$  we find that  $\hat{\theta}_n = \hat{\theta}_n^{(1)}$  where the latter has been introduced by Gardes and Girard [2006].

**Theorem 3.1.** *If  $\lambda$  is chosen such that*

$$\lim_{n \rightarrow \infty} \frac{b(-\log \lambda)}{b(\log n)} = 1, \quad \lim_{n \rightarrow \infty} b(\log n) \sqrt{n\lambda} = \infty, \quad (4)$$

then

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\theta, \ell}(|\hat{\theta}_n - \theta| \leq b(\log n)) = 1, \quad (5)$$

uniformly with respect to  $\ell \in \mathcal{L}(b)$ .

*Remark 1.* It is possible to choose  $\lambda$  as in (4). See [Drees, 1998, Proposition 2.1]. Note that Condition (4) implies that  $\lim_{n \rightarrow \infty} n\lambda = \infty$ .

Gathering Theorems 2.1 and 3.1 yields the following exact rate of convergence.

$$\liminf_{n \rightarrow \infty} \inf_{\hat{\theta}_n} \sup_{\theta \in (\theta_1, \theta_2)} \sup_{\ell \in \mathcal{L}(b)} \mathbb{P}_{\theta, \ell}^{(n)}(|\hat{\theta}_n - \theta| \leq b(\log n)) = 1. \quad (6)$$

*Remark 2.* The bound (6) is a minimax bound. This means that it is a uniform bound over a certain class of distributions. The function  $b$  describes the a priori information on the class of distribution from which the observed data is assumed to be sampled.

*Example 5.* If  $b(x) = x^\rho$  with  $\rho < 0$ , then the assumptions are satisfied for  $\lambda(n) = k(n)/n$  where  $k(n) = \log(n)^q$  with  $q + 2\rho > 0$ .

*Example 6.* If  $b(x) = \log^{-1}(x)$ , then the result holds for  $\lambda(n) = k(n)/n$  where  $k(n) = \log(n)^\alpha n^\beta$  with  $\alpha \in \mathbb{R}$  if  $\beta \in (0, 1)$  and  $\alpha > 0$  if  $\beta = 0$ .

## 4 Data driven choice of the threshold

Recently, [Asimit et al. \[2010\]](#) discussed methods to select the sample fraction  $k_n = \lambda n$  that may be employed in estimating the Weibull tail index. In particular, they noticed that there is no value  $k_n$  minimizing the asymptotic mean squared error. This is confirmed by our results. Indeed, [Theorem 3.1](#) shows that if  $k_n$  is suitably chosen, the variance of the kernel estimator, denoted by  $\hat{\theta}_{n,k_n}$ , turns out to be asymptotically negligible with respect to its squared bias, and [Theorem 2.1](#) shows that this choice is optimal in the sharp asymptotic minimax sense over certain functional classes. This kind of situation is well known in minimax nonparametric or semiparametric estimation. See e.g. [Butucea and Tsybakov \[2007\]](#). Adapting the idea of [Drees and Kaufmann \[1998\]](#), which was itself inspired by the so-called Lepsky method, we propose here a different choice of the sample fraction.

In this section, for simplicity of exposition, we consider the rectangular kernel  $K = \mathbf{1}_{[0,1]}$  and restrict the tuning parameter to the discrete grid  $\lambda = k/n$ . The resulting estimator of  $\theta$  is denoted by  $\hat{\theta}_{n,k}$ :

$$\hat{\theta}_{n,k} = \frac{\frac{1}{k} \sum_{j=1}^k j \{ \log X_{n-j+1,n} - \log X_{n-j,n} \}}{\frac{n}{k} \int_0^{k/n} \frac{dv}{-\log v}}$$

Let  $r_n$  be an increasing sequence. Define  $\hat{k}_n(r_n)$  by

$$\hat{k}_n(r_n) = \min\{k = 1, \dots, n-1 \mid \max_{1 \leq i \leq k} \sqrt{i} |\hat{\theta}_{n,k} - \hat{\theta}_{n,i}| > r_n\}. \quad (7)$$

**Theorem 4.1.** *Let  $\eta \in \mathcal{L}(b)$  be regularly varying at infinity with index  $\nu < 0$ . If  $\sqrt{\log \log(n)} = o(r_n)$  and  $\log(r_n) = o(\log n)$ , then  $\lim_{n \rightarrow \infty} \mathbb{P}(|\hat{\theta}_{n,\hat{k}_n(r_n)} - \theta| \leq b(\log(n))) = 1$ .*

This result means that  $\hat{\theta}_{n,\hat{k}_n(r_n)}$  is asymptotically optimal over the class  $\mathcal{L}(b)$  restricted to regularly varying functions. If we further restrict the class  $\mathcal{L}(b)$  to nonincreasing functions, we obtain an asymptotic confidence interval.

**Corollary 4.2.** *Under the assumptions of [Theorem 4.1](#), if moreover  $\eta$  is nonincreasing, then  $(er_n \log(n)/\{2\eta(\log n)\})^{-2} \hat{k}_n(r_n) \rightarrow_P 1$ , hence*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( |\hat{\theta}_{n,\hat{k}_n(r_n)} - \theta| \leq \frac{er_n \log n}{2\sqrt{\hat{k}_n(r_n)}} \right) = 1.$$

### Numerical illustration

The theoretical result state that the data-driven estimator  $\hat{\theta}_{n,\hat{k}_n}$  has nice asymptotic properties for any choice of the sequence  $r_n$  which satisfies very weak constraints. Unfortunately, this is not the case in practice. We have run simulations for large to very large sample sizes, which reveal an extremely poor performance of this data-driven choice.

We have generated  $N$  replications of a  $n$ -sample from the distribution function  $F$  corresponding to the function  $\eta(x) = x^{-1/2-3/2u}$  through [\(1\)](#), for  $u \in [0, 1]$ , so that  $\eta$  is in the class  $\mathcal{L}(b)$  for  $b(x) = x^{-1/2}$ . The value of the parameter  $u$  is uniformly distributed in  $[0, 1]$  and changes for each replication. We have computed the  $N$  sequences  $\{\hat{\theta}_{n,k} \mid k = 1, \dots, n-1\}$ , which are

plotted in black in Figures 1 and 2. The sample fraction  $\hat{k}_n(r_n)$  is computed for each of these  $N$  sequences, according to the algorithm (7). The  $N$  values are represented by vertical green lines. The true value of the parameter  $\theta$  is given by the horizontal blue line.

The estimator is severely biased and the confidence interval suggested in Corollary 4.1 is very large and nearly useless. Its half-length  $l_n = (e r_n \log n)/(2\sqrt{\hat{k}_n(r_n)})$  is given by Corollary 4.2. An improvement of this method is needed, but we think that it is better to look for genuinely semiparametric methods rather than to artificially restrict the model. For instance Asimit et al. [2010] suggest to use the model of Example 3, but this is a parametric model, which incurs a risk of misspecification.

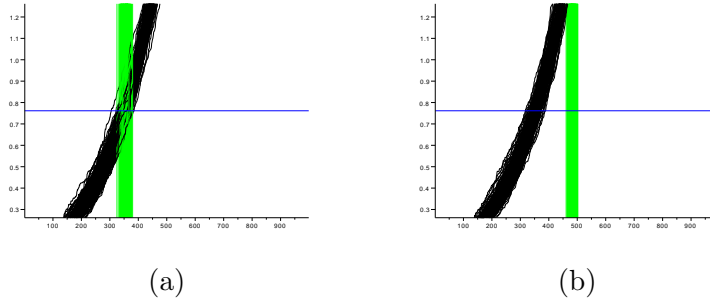


Figure 1: Sample properties of  $\hat{k}_n(r_n)$  for  $\theta = 0.761461$ ,  $n = 1000$ ,  $b(\log n) \sim 0.38$ . (a)  $r_n = \log n$  and half-length  $l_n \sim 3.5$ . (b)  $r_n = (\log n)^{3/2}$  and half-length  $l_n \sim 8.75$ .

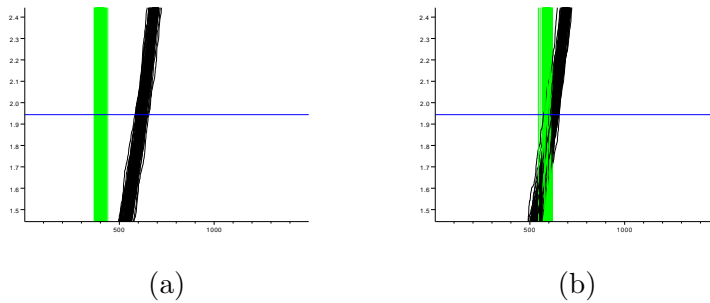


Figure 2: Sample properties of  $\hat{k}_n(r_n)$  for  $\theta = 1.9439654$ ,  $n = 1500$ , and  $b(\log n) \sim 0.37$ . (a)  $r_n = \log n$  and half-length  $l_n \sim 3.63$ . (b)  $r_n = (\log n)^{3/2}$  and half-length  $l_n \sim 8.03$ .

## 5 Proofs

The proof of Theorem 2.1 follows a very standard path to obtain minimax risk bounds. It also makes use of some computations in Beirlant et al. [2006]. The proof of Theorem 3.1 is very much inspired from similar results for the Hill estimator in Csörgő et al. [1985]. The proof of Theorem 4.1 is an adaptation of the proof of Drees and Kaufmann [1998, Corollary 1]. We include all details for the sake of self-containedness.

We preface the proofs by recalling some results on the uniform empirical process. We deduce from these results two lemmas that will be useful. Recall that the quantile function  $Q$  and

the empirical quantile process  $Q_n$  are defined by

$$\begin{aligned} Q(t) &= (-\log(1-t))^\theta \ell(-\log(1-t)), \\ Q_n(t) &= X_{n,k} \Leftrightarrow \frac{k-1}{n} < t \leq \frac{k}{n}, \quad 1 \leq k \leq n. \end{aligned}$$

Let  $U_n$  denote the uniform quantile process, i.e.  $Q_n(t) = Q(U_n(t))$  for any  $t \in [0, 1]$ .

**Lemma 5.1** (Csörgő et al. [1986]). *There exists a sequence of Brownian bridges  $B_n$  such that for any  $\nu \in [0, 1/2)$ ,*

$$\sup_{1/n \leq u \leq 1} u^{\nu-1/2} |\sqrt{n}\{1-u-U_n(1-u)\} - B_n(u)| = O_P(n^{-\nu}). \quad (8)$$

**Lemma 5.2** (Csörgő et al. [1985], Lemma 13). *For  $D > 1$ , define  $A_n(D) = \{u \in [1/n, 1] : u/D \leq 1 - U_n(1-u) \leq Du\}$ . Then*

$$\lim_{D \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(A_n(D)) = 1.$$

**Lemma 5.3** (Mason [1983], Lemma 14). *For any  $\nu \in [0, 1/2)$ ,*

$$\sup_{1/n \leq u \leq 1} n^\nu u^{\nu-1} |1-u-U_n(1-u)| = O_P(1).$$

**Lemma 5.4** (Csörgő and Horváth [1993], Theorem 4.2.1).

$$\sup_{1/n \leq u \leq 1} n^{1/2} u^{-1/2} |1-u-U_n(1-u)| = O_P\left(\sqrt{\log \log(n)}\right).$$

Denote

$$\Delta_n(u) = \log(-\log\{1-U_n(1-u)\}) - \log(-\log u).$$

Combining Lemmas 5.2 and 5.4 we obtain

**Lemma 5.5.**

$$\sup_{1/n \leq u \leq 1} \sqrt{u} \log u \Delta_n(u) = O_P\left(\sqrt{\frac{\log \log n}{n}}\right).$$

*Proof.* By a Taylor expansion, we have

$$\Delta_n(u) = \frac{1-u-U_n(1-u)}{\alpha_n(u) \log(\alpha_n(u))}$$

with  $\alpha_n(u)$  such that

$$u \wedge \{1-U_n(1-u)\} \leq \alpha_n(u) \leq u \vee \{1-U_n(1-u)\}.$$

Applying Lemmas 5.2 and 5.4, we get for any  $\epsilon > 0$ , that there exists a constant  $D > 1$  large enough such that

$$\begin{aligned} D^{-1}u &\leq \alpha_n(u) \leq Du, \\ |1-u-U_n(1-u)| &\leq D\sqrt{(\log \log n)/n}\sqrt{u}, \\ |u \log(u) / (\alpha_n(u) \log \alpha_n(u))| &\leq D, \end{aligned}$$

for all  $u \in (1/n, 1)$  with probability exceeding  $1 - \epsilon$ . Thus, we obtain for large enough  $D$  that  $\max_{1/n \leq u \leq 1} |\sqrt{u} \log(u) \Delta_n(u)| \leq D^2 \sqrt{(\log \log n)/n}$  with probability exceeding  $1 - \epsilon$ .  $\square$

**Lemma 5.6.**

$$\log\{-\log(1 - U_{n,n})\} - \log \log n = O_P(1/\log n) , \quad (9)$$

$$\int_0^{1/n} \Delta_n(u) du = O_P(1/\{n \log n\}) . \quad (10)$$

*Proof.* To prove (9), we apply the convergence in distribution of  $T_n := -\log(1 - U_{n,n}) - \log(n)$  to the Gumbel law combined with a Taylor expansion:

$$\log\{-\log(1 - U_{n,n})\} - \log \log n = T_n/\log n + o(T_n/\log n) .$$

To prove (10), we write

$$\int_0^{1/n} \Delta_n(u) du = \frac{1}{n} (\log\{-\log(1 - U_{n,n})\} - \log \log n) + \int_0^{1/n} \log \log n - \log(-\log u) du .$$

The result follows from (9) and  $\int_0^{1/n} \log \log n - \log(-\log u) du = O(1/(n \log n))$ .  $\square$

## 5.1 Proof of Theorem 2.1

Denote  $s_n = \log(n)$  and  $\delta_n = b(\log(n))$ . For  $\epsilon \in \{-1, 1\}$ , define the function  $\ell_n^\epsilon \in \mathcal{L}(b)$  by

$$\ell_n^\epsilon(x) = \exp \left\{ \epsilon \delta_n \int_1^{s_n \wedge x} \frac{ds}{s} \right\} = (s_n \wedge x)^{\epsilon \delta_n} .$$

These functions belong to  $\mathcal{L}(b)$  since  $b$  is nonincreasing, thus, defining  $\eta_n^\epsilon = \epsilon \delta_n \mathbf{1}_{[1, s_n]}$  yields (1) with  $|\eta_n| \leq b$ . Fix  $\theta_0 > 0$  and define  $\theta_n^\epsilon = \theta_0 + \epsilon \delta_n$ ,  $h_n^\epsilon(z) = z^{1/\theta_n^\epsilon} \ell_n^\epsilon(z)$ . Let  $F_n^\epsilon$  be the probability distribution function such that  $(F_n^\epsilon)^\leftarrow(t) = h_n^\epsilon(-\log(1-t))$ . Then  $-\log[1 - F_n^\epsilon(x)] = \{x \ell_n^\epsilon(x)\}^{1/\theta_n^\epsilon}$  with

$$\ell_n^\epsilon(x) = (x \wedge t_n)^{\epsilon \delta_n / \theta_0}$$

and  $t_n = s_n^{\theta_0}$ . Define finally  $g_n^\epsilon(x) = \{x \ell_n^\epsilon(x)\}^{1/\theta_n^\epsilon}$ ,  $g_0(x) = x^{1/\theta_0}$  and  $F_0(x) = 1 - e^{-g_0(x)}$ . Then,  $s_n = g_0(t_n)$  and  $\bar{F}^\epsilon(t_n) = e^{-s_n}$  and

$$g_n^\epsilon(x) = \begin{cases} g_0(x) & \text{if } x \leq t_n , \\ x^{1/\theta_n^\epsilon} s_n^{\epsilon \delta_n / \theta_n^\epsilon} & \text{if } x > t_n . \end{cases}$$

Let  $\mathbb{P}_+^n$  and  $\mathbb{P}_-^n$  be the law of a  $n$  sample of i.i.d. random variables with distribution  $F_n^+$  and  $F_n^-$ , respectively, and denote  $\mathbb{E}_+^n$  and  $\mathbb{E}_-^n$  the corresponding expectations. For  $\zeta \in (0, 1)$ , denote  $A = \{d\mathbb{P}_+^n/d\mathbb{P}_-^n \geq \zeta\}$ . Then, for any  $\theta$  and  $\ell \in \mathcal{L}(b)$ , for any estimator  $\theta_n$ ,

$$\begin{aligned} \mathbb{E}_{\theta, \ell}^n[|\hat{\theta}_n - \theta|] &\geq \frac{1}{2} \mathbb{E}_+^n[|\hat{\theta}_n - \theta_n^+|] + \frac{1}{2} \mathbb{E}_-^n[|\hat{\theta}_n - \theta_n^-|] = \frac{1}{2} \mathbb{E}_-^n \left[ |\hat{\theta}_n - \theta_n^-| + \frac{d\mathbb{P}_+^n}{d\mathbb{P}_-^n} |\hat{\theta}_n - \theta_n^+| \right] \\ &\geq \frac{1}{2} \mathbb{E}_-^n \left[ \left\{ |\hat{\theta}_n - \theta_n^-| + \frac{d\mathbb{P}_+^n}{d\mathbb{P}_-^n} |\hat{\theta}_n - \theta_n^-| \right\} \mathbf{1}_A \right] \geq \frac{1}{2} \mathbb{E}_-^n \left[ \left\{ |\hat{\theta}_n - \theta_n^-| + \zeta |\hat{\theta}_n - \theta_n^+| \right\} \mathbf{1}_A \right] \\ &\geq \frac{\zeta}{2} \mathbb{E}_0^n \left[ \left\{ |\hat{\theta}_n - \theta_0| + |\hat{\theta}_n - \theta_n| \right\} \mathbf{1}_A \right] \geq \frac{\zeta |\theta_n^+ - \theta_n^-|}{2} \mathbb{P}_-^n(A) = \zeta \delta_n \mathbb{P}_-^n(A) . \end{aligned} \quad (11)$$



In order to bound the last term, let

$$K(\mathbb{P}_-^n, \mathbb{P}_+^n) = \mathbb{E}_-^n \left[ \log \left( \frac{d\mathbb{P}_-^n}{d\mathbb{P}_+^n} \right) \right]$$

denote the Kullback-Leiber distance. By Pinsker's inequality, [Tsybakov, 2004, Lemme 2.5], we get

$$\begin{aligned} \mathbb{P}_-^n(A) &= 1 - \mathbb{P}_-^n \left( \frac{d\mathbb{P}_+^n}{d\mathbb{P}_-^n} \leq \zeta \right) = 1 - \mathbb{P}_-^n \left( \log \left( \frac{d\mathbb{P}_+^n}{d\mathbb{P}_-^n} \right) \geq \log(1/\zeta) \right) \\ &\geq 1 - \frac{1}{\log(1/\zeta)} \mathbb{E}_-^n \left[ \log \left( \frac{d\mathbb{P}_+^n}{d\mathbb{P}_-^n} \right)_+ \right] \geq 1 - \frac{1}{\sqrt{2} \log(1/\zeta)} \sqrt{K(\mathbb{P}_-^n, \mathbb{P}_+^n)}. \end{aligned}$$

Since we consider an i.i.d. sample, we have

$$K(\mathbb{P}_-^n, \mathbb{P}_+^n) = nK(F_n^-, F_n^+)$$

and there only remains to prove that  $\liminf_{n \rightarrow \infty} nK(F_n^-, F_n^+) = 0$ .

For  $x > t_n$ , the log-likelihood between  $F_n^+$  and  $F_n^-$  (see Beirlant et al. [2006, Equation (6.4)]) is given by

$$\log \frac{dF_n^+}{dF_n^-} = \log(\theta_n^-/\theta_n^+) + (\theta_n^-/\theta_n^+ - 1) \log(g_n^-/s_n) + g_n^- \times \left\{ 1 - (g_n^-/s_n)^{\theta_n^-/\theta_n^+ - 1} \right\}.$$

Since  $\log(dF_n^+/dF_n^-) = 0$  for  $x \leq t_n$ , we obtain

$$\begin{aligned} K(F_n^-, F_n^+) &= \int_{t_n}^{\infty} \log(dF_n^-/dF_n^+) dF_n^- \\ &= -\log(\theta_n^-/\theta_n^+) \bar{F}_n^-(t_n) - (\theta_n^-/\theta_n^+ - 1) \int_{t_n}^{\infty} \log(g_n^-/s_n) dF_n^- \\ &\quad - \int_{t_n}^{\infty} g_n^- \times \left\{ 1 - (g_n^-/s_n)^{\theta_n^-/\theta_n^+ - 1} \right\} dF_n^- \\ &= -I - (\theta_n^-/\theta_n^+ - 1) \times II - III. \end{aligned}$$

We make the change of variable  $u = g_n^-(x)$  in the above integrals separately.

$$\begin{aligned} I &= -e^{-s_n} (2\delta_n/\theta_0 + O(\delta_n^2)), \\ II &= \int_{s_n}^{\infty} \log(u/s_n) e^{-u} du = e^{-s_n} \int_0^{\infty} \log(1+v/s_n) e^{-v} dv \sim s_n^{-1} e^{-s_n}, \\ III &= \int_{s_n}^{\infty} u \left\{ 1 - (u/s_n)^{\theta_n^-/\theta_n^+ - 1} \right\} e^{-u} du \\ &= e^{-s_n} \int_0^{\infty} (v+s_n) \left\{ 1 - (1+v/s_n)^{\theta_n^-/\theta_n^+ - 1} \right\} e^{-v} dv \sim \{2\delta_n/\theta_0 + O(\delta_n^2)\} e^{-s_n}. \end{aligned}$$

Summing the three terms yields

$$K(F_n^-, F_n^+) = O\left(\frac{\delta_n}{s_n} \vee \delta_n^2\right) e^{-s_n}.$$

Since  $s_n = \log(n)$ , it holds that  $\lim_{n \rightarrow \infty} nK(F_n^-, F_n^+) = 0$ . Thus  $\liminf_n \mathbb{P}_-^n(A) = 1$ . Plugging this limit into (11) yields

$$\liminf_{n \rightarrow \infty} \sup_{\theta \in (\theta_1, \theta_2)} \sup_{\ell \in \mathcal{L}(b)} b^{-1}(\log(n)) \mathbb{E}_{\theta, \ell} [|\hat{\theta}_n - \theta|] \geq \zeta \liminf_{n \rightarrow \infty} \mathbb{P}_-^n(A) = \zeta.$$

Since this is true for all  $\zeta \in (0, 1)$ , (2) is proved.

## 5.2 Proof of Theorem 3.1

Let us define  $L(t) = \log(-\log(t))$ . Then  $\log Q(t) = \theta L(1-t) + \log \ell(-\log(1-t))$  and

$$\hat{\theta}_{n,\lambda} = \theta + \theta \xi_{n,\lambda} + \beta_{n,\lambda} + \tilde{\beta}_{n,\lambda}, \quad (12)$$

with

$$\begin{aligned} g(\lambda) &= \int_0^1 L(v\lambda) d\{vK(v)\} = \int_0^1 \frac{K(v)}{-\log(v\lambda)} dv \sim \frac{\int_0^1 K(t) dt}{\log(1/\lambda)}, \\ \xi_{n,\lambda} &= \frac{1}{g(\lambda)} \int_0^1 \{L(1 - U_n(1 - v\lambda)) - L(v\lambda)\} d\{vK(v)\}, \\ \beta_{n,\lambda} &= \frac{1}{g(\lambda)} \int_0^1 \int_1^{-\log(\lambda v)} \frac{\eta(s)}{s} ds d\{vK(v)\} = \frac{1}{g(\lambda)} \int_0^1 \frac{\eta(-\log(v\lambda))}{-\log(v\lambda)} K(v) dv, \\ \tilde{\beta}_{n,\lambda} &= \frac{1}{g(\lambda)} \int_0^1 \int_{-\log(\lambda v)}^{-\log\{1 - U_n(1 - v\lambda)\}} \frac{\eta(s)}{s} ds d\{vK(v)\}. \end{aligned}$$

The term  $\beta_{n,\lambda}$  is a deterministic bias. Since  $b$  is decreasing, a bound for  $\beta_{n,\lambda}$  is given by

$$|\beta_{n,\lambda}| \leq b(-\log(\lambda)),$$

uniformly with respect to  $\ell \in \mathcal{L}(b)$ . We will prove in Subsection 5.2.1 that  $\sqrt{n\lambda} \xi_{n,\lambda}$  converges weakly to a centered Gaussian distribution and in Subsection 5.2.2 that  $\tilde{\beta}_{n,\lambda} = o_P(b(\log n))$ . These results, together with the decomposition (12) and Assumption (4), yield (5).

### 5.2.1 The stochastic term $\xi_{n,\lambda}$

Let  $\{B_n\}$  be a sequence of Brownian bridges such that Eq. (8) in Lemma 5.1 holds. Informally, we can write

$$L(1 - U_n(1 - \lambda v)) - L(\lambda v) \approx L'(\lambda v) \{1 - \lambda v - U_n(1 - \lambda v)\} \approx \frac{B_n(\lambda v)}{\sqrt{n\lambda v \log(\lambda v)}}.$$

Thus, we hope that we can prove that

$$\int_0^1 \{L(1 - U_n(1 - \lambda v)) - L(\lambda v)\} d\{vK(v)\},$$

has the the same asymptotic distribution as

$$\frac{1}{\sqrt{n\lambda}} \int_0^1 \frac{B_n(\lambda v)}{v \log(\lambda v)} d\{vK(v)\} \sim_d \frac{1}{\sqrt{n\lambda} \log(1/\lambda)} \mathbf{N} \left( 0, 2 \int_0^1 K(t) d\{tK(t)\} \right).$$

We will prove that this approximation is valid by splitting the integral into two parts. Recall that we have defined  $\Delta_n(u) = L(1 - U_n(1 - u)) - L(u)$ .

**Lemma 5.7.** *Under the assumptions of Theorem 3.1,*

$$\sqrt{n\lambda} \log(1/\lambda) \int_{1/(\lambda n)}^1 \Delta_n(\lambda v) d\{vK(v)\} \rightarrow_d \mathbf{N} \left( 0, 2 \int_0^1 K(t) d\{tK(t)\} \right), \quad (13)$$

$$\sqrt{n\lambda} \log(1/\lambda) \int_{1/(\lambda n)}^1 |\Delta_n(\lambda v)| |d\{vK(v)\}| = O_P(1). \quad (14)$$

*Proof of Lemma 5.7.* Define

$$\begin{aligned}\delta_{n,\lambda} &= \sqrt{n\lambda} \log(1/\lambda) \int_{1/(\lambda n)}^1 \left\{ \Delta_n(\lambda v) - \frac{B_n(\lambda v)}{\sqrt{n\lambda v \log(\lambda v)}} \right\} d\{vK(v)\}, \\ \zeta_{n,\lambda} &= \frac{\log(1/\lambda)}{\sqrt{\lambda}} \int_{1/(\lambda n)}^1 \frac{B_n(\lambda v)}{v \log(\lambda v)} d\{vK(v)\}.\end{aligned}$$

Then  $\sqrt{n\lambda} \log(1/\lambda) \int_{1/(\lambda n)}^1 \Delta_n(\lambda v) d\{vK(v)\} = \delta_{n,\lambda} + \zeta_{n,\lambda}$ . We will prove that  $\delta_{n,\lambda} = o_P(1)$  and  $\zeta_{n,\lambda}$  converges weakly to the desired limit. Denote  $\tilde{B}_n(u) = \sqrt{n}\{1 - u - U_n(1 - u)\}$ . The function  $L$  is differentiable with  $L'(u) = 1/(u \log u)$ , hence

$$\Delta_n(\lambda v) = L(1 - U_n(1 - \lambda v)) - L(\lambda v) = L'(\alpha_n(\lambda v)) \frac{\tilde{B}_n(\lambda v)}{\sqrt{n}} = \frac{\tilde{B}_n(\lambda v)}{\sqrt{n\alpha_n(\lambda v) \log(\alpha_n(\lambda v))}}$$

with  $\alpha_n(u) = u + \vartheta n^{-1/2} \tilde{B}_n(u)$ , for some  $\vartheta \in (0, 1)$  and

$$(\lambda v) \wedge \{1 - U_n(1 - \lambda v)\} \leq \alpha_n(\lambda v) \leq (\lambda v) \vee \{1 - U_n(1 - \lambda v)\}. \quad (15)$$

Define  $R_{n,\lambda}$  by

$$R_{n,\lambda} = \sqrt{n\lambda} \log(1/\lambda) \int_{1/(\lambda n)}^1 \left\{ \Delta_n(\lambda v) - \frac{\tilde{B}_n(\lambda v)}{\sqrt{n\lambda v \log(\lambda v)}} \right\} d\{vK(v)\}.$$

Applying Lemmas 5.2 and 5.3, for any  $\epsilon > 0$  and  $\nu \in (1/4, 1/2)$ , we can find  $D > 1$  and  $A > 0$  such that, with probability exceeding  $1 - \epsilon$ ,

$$D^{-1}u \leq \alpha_n(u) \leq Du, \quad |\tilde{B}_n(u)| \leq An^{-\nu}u^{1-\nu}. \quad (16)$$

Since the function  $u \rightarrow -u \log u$  is increasing near zero, we obtain, with probability greater than  $1 - \epsilon$ ,

$$\begin{aligned}|R_{n,\lambda}| &\leq AD\sqrt{\lambda} \log(1/\lambda) n^{1/2-\nu} \\ &\quad \times \int_{1/(\lambda n)}^1 \frac{(\lambda v)^{1-\nu} |\lambda v \log(\lambda v) - \alpha_n(\lambda v) \log(\alpha_n(\lambda v))|}{\lambda^2 v^2 |\log(\lambda v) \log(\lambda v/D)|} |d\{vK(v)\}|. \quad (17)\end{aligned}$$

Note now that the function  $u \rightarrow -u \log u$  is concave and increasing near zero with derivative  $\log(1/u) - 1$ , so (15) and (16) imply that, with probability greater than  $1 - \epsilon$ ,

$$\begin{aligned}|\lambda v \log(\lambda v) - \alpha_n(v) \log(\alpha_n(v))| &\leq |\lambda v \log(\lambda v) - (1 - U_n(1 - \lambda v)) \log(1 - U_n(1 - \lambda v))| \\ &\leq |\log(\lambda v \wedge \{1 - U_n(1 - \lambda v)\})| |1 - \lambda v - U_n(1 - \lambda v)| \\ &\leq An^{-\nu} (\lambda v)^{1-\nu} |\log(\lambda v \wedge \{1 - U_n(1 - \lambda v)\})| \\ &\leq An^{-\nu} (\lambda v)^{1-\nu} |\log(\lambda v/D)|.\end{aligned}$$

Plugging this bound into (17) yields, for  $\nu \in (0, 1/2)$ ,

$$|R_{n,\lambda}| \leq A^2 D (n\lambda)^{1/2-2\nu} \int_{1/(\lambda n)}^1 v^{-2\nu} |d\{vK(v)\}| \leq C (n\lambda)^{1/2-2\nu},$$

for some constant  $C$ , with probability greater than  $1 - \epsilon$ . Thus  $R_{n,\lambda} = o_P(1)$ .

Denote now

$$S_{n,\lambda} = \sqrt{n\lambda} \log(1/\lambda) \int_{1/(\lambda n)}^1 \frac{\tilde{B}_n(\lambda v) - B_n(\lambda v)}{\sqrt{n\lambda v \log(\lambda v)}} d\{vK(v)\}.$$

Applying Lemma 5.1, for any  $\epsilon \in (0, 1)$ , there exists  $A$  such that, with probability exceeding  $1 - \epsilon$ ,

$$|S_{n,\lambda}| \leq A(n\lambda)^{-\nu} \int_{1/(\lambda n)}^1 v^{-1/2-\nu} d\{vK(v)\}.$$

Since the integral is convergent, we get that  $S_{n,\lambda} = O_P((n\lambda)^{-\nu}) = o_P(1)$ . Altogether, we obtain that  $\delta_{n,\lambda} = o_P(1)$ .

We must now prove that  $\zeta_{n,\lambda}$  converges weakly to the desired limit. Since  $\{\zeta_{n,\lambda}\}$  is a sequence of centered Gaussian random variables, we only need to prove that its variance converges. We compute

$$\begin{aligned} \text{var}(\zeta_{n,\lambda}) &= 2 \frac{\log^2(1/\lambda)}{\lambda} \int_{1/\lambda n}^1 \frac{d\{tK(t)\}}{t \log(\lambda t)} \int_{1/\lambda n}^t \frac{\lambda s(1-\lambda t)}{s \log(\lambda s)} d\{sK(s)\} \\ &\sim 2 \int_{1/\lambda n}^1 \frac{d\{tK(t)\}}{t} \int_{1/\lambda n}^t d\{sK(s)\} \sim 2 \int_0^1 K(t) d\{tK(t)\}. \end{aligned}$$

This proves (13). To prove (14), it only remains to prove that

$$\frac{\log(1/\lambda)}{\sqrt{\lambda}} \int_{1/(\lambda n)}^1 \frac{|B_n(\lambda v)|}{v \log(\lambda v)} d\{vK(v)\} = O_P(1). \quad (18)$$

Since  $\mathbb{E}[|B_n(\lambda v)|] = \sqrt{2/\pi} \sqrt{\lambda v}$ , we have

$$\begin{aligned} \frac{\log(1/\lambda)}{\sqrt{\lambda}} \mathbb{E} \left[ \int_{1/(\lambda n)}^1 \frac{|B_n(\lambda v)|}{v |\log(\lambda v)|} d\{vK(v)\} \right] &= \frac{\sqrt{2} \log(1/\lambda)}{\sqrt{\pi}} \int_{1/(\lambda n)}^1 \frac{d\{vK(v)\}}{v^{1/2} |\log(\lambda v)|} \\ &\leq \sqrt{\frac{2}{\pi}} \int_0^1 v^{-1/2} d\{vK(v)\}. \end{aligned}$$

This proves (18), which concludes the proof of Lemma 5.7.  $\square$

**Lemma 5.8.** *Under the assumptions of Theorem 3.1,  $\sqrt{n\lambda} \log(1/\lambda) \int_0^{1/(\lambda n)} \Delta_n(\lambda v) d\{vK(v)\} = o_P(1)$ .*

*Proof of Lemma 5.8.* Note that if  $v\lambda < 1/n$ , then  $U_n(1 - v\lambda) = U_{n,n}$ . Thus,

$$\begin{aligned} &\sqrt{n\lambda} \log(1/\lambda) \int_0^{1/(\lambda n)} \{L(1 - U_{n,n}) - L(v\lambda)\} d\{vK(v)\} \\ &= \frac{\sqrt{n\lambda} \log(1/\lambda) \{L(1 - U_{n,n}) - L(1/n)\}}{\lambda n} K(1/\{\lambda n\}) \end{aligned} \quad (19)$$

$$+ \sqrt{n\lambda} \log(1/\lambda) \int_0^{1/\{\lambda n\}} \{L(1/n) - L(v\lambda)\} d\{vK(v)\}. \quad (20)$$

Since  $-\log(1 - U_{n,n}) + \log(1/n)$  converges weakly to the standard Gumbel law, we get that  $L(1 - U_{n,n}) - L(1/n) = O_P(1/\log n)$ , thus the term in (19) is  $O_P((n\lambda)^{-1/2} \log(1/\lambda)/\log n)$ . Applying integration by parts, we see that the term in (20) is

$$-\sqrt{n\lambda} \log(1/\lambda) \int_0^{1/(\lambda n)} \frac{K(v)}{(-\log(v\lambda))} dv = O\left((n\lambda)^{-1/2} \log(1/\lambda)/\log n\right) = O\left((n\lambda)^{-1/2}\right).$$

□

In conclusion, we have proved that

$$\sqrt{n\lambda} \xi_{n,\lambda} \rightarrow_d \mathbf{N} \left( 0, \frac{2 \int_0^1 K(t) d\{tK(t)\}}{\left(\int_0^1 K(t) dt\right)^2} \right).$$

### 5.2.2 The term $\tilde{\beta}_{n,\lambda}$

Write  $\tilde{\beta}_{n,\lambda} = \phi_{n,\lambda} + \psi_{n,\lambda}$  with

$$\begin{aligned} \phi_{n,\lambda} &= \frac{1}{g(\lambda)} \int_0^{1/(\lambda n)} \int_{-\log(v\lambda)}^{-\log\{1-U_n(1-v\lambda)\}} \frac{\eta(s)}{s} ds d\{vK(v)\}, \\ \psi_{n,\lambda} &= \frac{1}{g(\lambda)} \int_{1/(\lambda n)}^1 \int_{-\log(v\lambda)}^{-\log\{1-U_n(1-v\lambda)\}} \frac{\eta(s)}{s} ds d\{vK(v)\}. \end{aligned}$$

Note that if  $v\lambda < 1/n$ , then  $U_n(1 - v\lambda) = U_{n,n}$ . Thus, integrating by parts, we obtain

$$\begin{aligned} \phi_{n,\lambda} &= \frac{1}{g(\lambda)} \int_0^{1/(\lambda n)} \int_{-\log v\lambda}^{-\log(1-U_{n,n})} \frac{\eta(s)}{s} ds d\{vK(v)\} \\ &= \frac{K(1/\lambda n)}{\lambda n g(\lambda)} \int_{\log n}^{-\log(1-U_{n,n})} \frac{\eta(s)}{s} ds + \frac{1}{g(\lambda)} \int_0^{1/(\lambda n)} \frac{\eta(-\log(v\lambda))}{-\log(v\lambda)} K(v) dv. \end{aligned}$$

The second term can be bounded by

$$\frac{b(\log n)}{g(\lambda)} \int_0^{1/(\lambda n)} \frac{K(v)}{-\log(v\lambda)} dv = O\left(\frac{b(\log n)}{n\lambda}\right).$$

The first term can be bounded by

$$\frac{b(\log(n) \wedge \{-\log(1 - U_{n,n})\})}{n\lambda g(\lambda)} \left| \log \frac{-\log(1 - U_{n,n})}{\log n} \right|$$

The weak convergence of  $-\log(1 - U_{n,n}) - \log n$  to the standard Gumbel law implies that  $-\log(1 - U_{n,n}) = O_P(\log n)$  and  $\log\{-\log(1 - U_{n,n})/\log n\} = O_P(1/\log n)$ . Altogether, we obtain that  $\phi_{n,\lambda} = O_P(b(\log n)/\{n\lambda\})$ .

Consider now  $\psi_{n,\lambda}$ . For  $1/(\lambda n) \leq v \leq 1$ , it holds that

$$-\log(1 - U_n(1 - v\lambda)) \geq -\log(1 - U_n(1 - \lambda)) \rightarrow_P +\infty,$$

since  $\lambda \rightarrow 0$ . It also holds that  $-\log(v\lambda) \geq -\log \lambda \rightarrow +\infty$ . Thus, defining

$$b_n = b(\{-\log \lambda\} \wedge \{-\log(1 - U_n(1 - \lambda))\})$$

we get  $b_n \rightarrow_P 0$ . This and the bound (14) in Lemma 5.7 yield

$$|\psi_{n,\lambda}| \leq \frac{b_n}{g(\lambda)} \int_{1/(\lambda n)}^1 \left| \log \frac{\log(1 - U_n(1 - \lambda v))}{\log(\lambda v)} \right| d\{vK(v)\} = o_P(1/\sqrt{n\lambda}).$$

Thus, under Assumption (4), we conclude that

$$\tilde{\beta}_{n,\lambda} = O_P\left(\frac{b(\log n)}{\sqrt{n\lambda}}\right) + o_P\left(\frac{1}{\sqrt{n\lambda}}\right) = o_P(b(\log n)).$$

### 5.3 Proof of Theorem 4.1

Define  $\beta_{n,k}$  and  $k_n^*$  by

$$\beta_{n,k} = \frac{n}{kg(k/n)} \int_0^{k/n} \frac{\eta(-\log v)}{-\log(v)} dv$$

$$k_n^*(r_n) = \min\{k = 1, \dots, n-1 \mid \max_{1 \leq i \leq k} \sqrt{i} |\beta_{n,k} - \beta_{n,i}| > r_n\}.$$

If the function  $\eta$  is regularly varying at infinity with index  $\nu < 0$ , it can be proved as in [Drees and Kaufmann, 1998, p. 156] that

$$k_n^*(r_n) \sim \{(e/2) r_n \log(n)/\eta(\log n)\}^2. \quad (21)$$

This implies that  $k_n^*(r_n)$  is an intermediate sequence and satisfies condition (4), thus, by Theorem 3.1,  $\lim_{n \rightarrow \infty} \mathbb{P}(|\hat{\theta}_{n,k_n^*(r_n)} - \theta| \leq b(\log n)) = 1$ . Thus, the proof will be concluded by showing that  $\lim_{n \rightarrow \infty} \mathbb{P}(\hat{k}_n(r_n) = k_n^*(r_n)) = 1$ . Corollary 4.2 will follow by noting that if  $\eta$  is nondecreasing, we can replace  $b$  by  $\eta$ .

Let  $r_n^-$  and  $r_n^+$  be two sequences such that  $r_n^- < r_n < r_n^+$ ,  $\sqrt{\log \log n} = o(r_n^-)$ ,  $\log(r_n^+) = o(\log n)$  and  $\sqrt{\log \log n} = o(|r_n - r_n^\pm|)$ . Then  $k_n^*(r_n^-)$  and  $k_n^*(r_n^+)$  are intermediate sequences by (21). Applying the definition of  $\hat{k}_n(r_n)$  and  $k_n^*(r_n^\pm)$  and Lemma 5.9, we obtain straightforwardly

$$\begin{aligned} \mathbb{P}(k_n^*(r_n^-) \leq \hat{k}_n(r_n)) &= \mathbb{P}\left(\max_{1 \leq k \leq k_n^*(r_n^-)} \max_{1 \leq i \leq k} \sqrt{i} |\hat{\theta}_{n,i} - \hat{\theta}_{n,k}| \leq r_n\right) \\ &\geq \mathbb{P}\left(r_n^- + 2 \max_{1 \leq i \leq k_n^*(r_n^-)} \sqrt{i} |\hat{\theta}_{n,i} - \theta - \beta_{n,i}| \leq r_n\right) \rightarrow 1. \end{aligned}$$

Conversely,

$$\begin{aligned} \mathbb{P}(k_n^*(r_n^+) \geq \hat{k}_n(r_n)) &\geq \mathbb{P}\left(\max_{1 \leq i \leq k_n^*(r_n^+)} \sqrt{i} |\hat{\theta}_{n,i} - \hat{\theta}_{n,k_n^*(r_n^+)}| > r_n\right) \\ &\geq \mathbb{P}\left(r_n^+ - 2 \max_{1 \leq i \leq k_n^*(r_n^+)} \sqrt{i} |\hat{\theta}_{n,i} - \theta - \beta_{n,i}| > r_n\right) \rightarrow 1. \end{aligned}$$

To conclude, it suffices to note that (21) implies that  $k_n^*(r_n^+)/k_n^*(r_n^-) \rightarrow 1$ .

**Lemma 5.9.** *For any intermediate sequence  $k_n$ ,*

$$\max_{1 \leq i \leq k_n} \sqrt{i} |\hat{\theta}_{n,i} - \theta - \beta_{n,i}| = O_P\left(\sqrt{\log \log n}\right). \quad (22)$$

*Proof.* Applying the decomposition (12), we have  $\hat{\theta}_{n,i} - \theta - \beta_{n,i} = \theta\xi_{n,i} + \tilde{\beta}_{n,i}$ . We will prove (22) by proving separately that

$$\max_{1 \leq i \leq k_n} \sqrt{i} |\xi_{n,i}| = O_P \left( \sqrt{\log \log n} \right), \quad (23)$$

$$\max_{1 \leq i \leq k_n} \sqrt{i} |\tilde{\beta}_{n,i}| = O_P \left( \sqrt{\log \log n} \right). \quad (24)$$

We start by proving (23).

$$\sqrt{i} \xi_{n,i} = \frac{n}{\sqrt{i} g(i/n)} \int_0^{1/n} \Delta_n(u) du + \frac{n}{\sqrt{i} g(i/n)} \int_{1/n}^{i/n} \Delta_n(u) du.$$

Applying Lemma 5.5, applying Lemma 5.6, we obtain

$$\begin{aligned} & \max_{1 \leq i \leq n-1} \frac{n}{\sqrt{i} g(n/i)} \int_{1/n}^{i/n} |\Delta_n(u)| du \\ &= O_P \left( \sqrt{n \log \log n} \right) \times \max_{1 \leq i \leq n-1} \frac{1}{\sqrt{i} g(i/n)} \int_{1/n}^{i/n} \frac{du}{\sqrt{u} \log u} = O_P \left( \sqrt{\log \log n} \right). \end{aligned}$$

Next, since  $U_n(1-u) = U_{n,n}$  for  $u \in [0, 1/n]$ , we have

$$\begin{aligned} \int_0^{1/n} \Delta_n(u) du &= n^{-1} L(1 - U_{n,n}) - \int_0^{1/n} L(u) du \\ &= n^{-1} \{L(1 - U_{n,n}) - L(n)\} + O(1/\{n \log(n)\}) = O_P(1/\{n \log(n)\}). \end{aligned}$$

Thus

$$\max_{1 \leq i \leq n-1} \frac{n}{\sqrt{i} g(i/n)} \int_0^{1/n} \Delta_n(u) du = O_P(1).$$

This proves (23). Consider now (24). As before, we split the integral defining  $\tilde{\beta}_{n,i}$  into two parts. Write  $\tilde{\beta}_{n,i} = \phi_{n,i} + \psi_{n,i}$ , with

$$\begin{aligned} \phi_{n,i} &= \frac{n}{ig(i/n)} \int_0^{1/n} \int_{-\log(u)}^{-\log\{1-U_n(1-u)\}} \frac{\eta(s)}{s} ds du, \\ \psi_{n,i} &= \frac{n}{ig(i/n)} \int_{1/n}^{i/n} \int_{-\log(u)}^{-\log\{1-U_n(1-u)\}} \frac{\eta(s)}{s} ds du. \end{aligned}$$

Since  $U_n(1-u) = U_{n,n}$  for  $u \leq 1/n$ , applying the bound (10) in Lemma 5.6 yields

$$\begin{aligned} \max_{1 \leq i \leq n-1} \sqrt{i} |\phi_{n,i}| &\leq \max_{1 \leq i \leq n-1} \frac{n}{\sqrt{i} g(i/n)} b(\{\log n\} \wedge \{-\log(1 - U_{n,n})\}) O_P(1/(n \log n)) \\ &= O_P(b(\log n)) = o_P(1). \end{aligned}$$

For any intermediate sequence  $k_n$ , we have that  $-\log(1 - k_n/n) \rightarrow_P \infty$ . Thus, noting that  $b$  is nonincreasing and applying Lemma 5.6, we have,

$$\begin{aligned} \max_{1 \leq i \leq k_n} |\psi_{n,i}| &\leq b(\{\log n\} \wedge \{-\log(1 - k_n/n)\}) \max_{1 \leq i \leq k_n} \frac{n}{\sqrt{i} g(i/n)} \int_{1/n}^{i/n} |\Delta_n(u)| du \\ &= o_P \left( \sqrt{\log \log n} \right). \end{aligned}$$

□

## Acknowledgement

This research was partially supported by the ANR grant ANR-08-BLAN-0314-02.

## References

- Alexandru V. Asimit, Deyuan Li, and Liang Peng. Pitfalls in using Weibull tailed distributions. *J. Statist. Plann. Inference*, 140(7):2018–2024, 2010.
- Jan Beirlant, Petra Vynckier, and Jozef Teugels. *Practical analysis of extreme values*. Leuven University Press, 1996.
- Jan Beirlant, Christel Bouquiaux, and Bas Werker. Semiparametric lower bounds for tail index estimation. *Journal of Statistical Planning and Inference*, 136(3):705–729, 2006.
- Cristina Butucea and Alexandre B. Tsybakov. Sharp optimality in density deconvolution with dominating bias. I. *Rossiiskaya Akademiya Nauk. Teoriya Veroyatnostei i ee Primeneniya*, 52(1):111–128, 2007.
- Miklós Csörgő and Lajos Horváth. *Weighted approximations in probability and statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Ltd., Chichester, 1993. With a foreword by David Kendall.
- Miklós Csörgő, Sándor Csörgő, Lajos Horváth, and David M. Mason. Weighted empirical and quantile processes. *Ann. Probab.*, 14(1):31–85, 1986.
- Sándor Csörgő, Paul Deheuvels, and David Mason. Kernel estimates of the tail index of a distribution. *The Annals of Statistics*, 13(3):1050–1077, 1985.
- Holger Drees. Optimal rates of convergence for estimates of the extreme value index. *The Annals of Statistics*, 26(1):434–448, 1998.
- Holger Drees and Edgar Kaufmann. Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Process. Appl.*, 75(2):149–172, 1998.
- Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling extremal events*, volume 33 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997. For insurance and finance.
- Laurent Gardes and Stéphane Girard. Comparison of Weibull tail-coefficient estimators. *REVSTAT*, 4(2):163–188, 2006.
- Stéphane Girard. A Hill type estimator of the Weibull tail-coefficient. *Communications in Statistics. Theory and Methods*, 33(2):205–234, 2004.
- Claudia Klüppelberg and José Villaseñor. Estimation of distribution tails - a semiparametric approach. *Blätter der Deutschen Gesellschaft für Versicherungsmathematik*, 21:213–235, 1993.
- David M. Mason. The asymptotic distribution of weighted empirical distribution functions. *Stochastic Processes and their Applications*, 15(1):99–109, 1983.
- Alexandre Tsybakov. *Introduction à l'estimation non-paramétrique*, volume 41 of *Mathématiques & Applications*. Springer-Verlag, Berlin, 2004.