ASYMPTOTIC BEHAVIOUR OF THE PROBABILITY-WEIGHTED MOMENTS AND PENULTIMATE APPROXIMATION

Jean Diebolt[1], Armelle Guillou[2] and Rym Worms[3]

Abstract. The P.O.T. (Peaks-Over-Threshold) approach consists of using the Generalized Pareto Distribution (GPD) to approximate the distribution of excesses over a threshold. We use the probability-weighted moments to estimate the parameters of the approximating distribution. We study the asymptotic behaviour of these estimators (in particular their asymptotic bias) and also the functional bias of the GPD as an estimate of the distribution function of the excesses. We adapt penultimate approximation results to the case where parameters are estimated.

1. Introduction

Nowadays, it seems common practice to use statistics to describe or even predict the average behaviour of many different phenomena. However, that average behaviour is not the only interesting subject of research in many situations. Indeed, in many problems of considerable interest, it is necessary to estimate the probability of occurence of extreme values. For instance, one is often interested in the probability that the maximum of n random variables exceeds a given threshold or, *vice versa*, one wants to determine a level such that the exceedance probability is below a given small value. For example, a hydraulics engineer has to estimate the necessary height of a dike such that the probability of a flooding in a given year is less than 10^{-4} (*cf.* [4]). More precisely, extreme value theory discusses the conditions under which extremes from a random sample $X_1, ..., X_n$ are attracted to a nondegenerate limit distribution when the sample size n increases to infinity. This condition seems reasonable when a statistical analysis is based solely on the extremes from the given sample. Here we assume that the random variables X_i $(i = 1, ..., n)$ are independent and identically distributed with common distribution function F. We require that the properly centred and normed sample maxima $X_{n,n} = \max\{X_1, ..., X_n\}$ converges in distribution to a non-degenerate limit. Gnedenko (see [11]) showed that this limit distribution is necessarily of extreme value type, *i.e.* for some $\gamma \in \mathbb{R}$ there exists sequences of constants

[1] Université de Marne-la-Vallée, Équipe d'Analyse et de Mathématiques Appliquées, 5 boulevard Descartes, bâtiment Copernic, Champs-sur-Marne, 77454 Marne-la-Vallée Cedex 2, France; e-mail: `diebolt@math.univ-mlv.fr`

[2] Université Paris VI, Laboratoire de Statistique Théorique et Appliquée, Boîte 158, 175 rue du Chevaleret, 75013 Paris, France; e-mail: `guillou@ccr.jussieu.fr`

[3] Université de Rouen, Laboratoire de Mathématiques Raphaël Salem, UMR 6085 du CNRS, Site Colbert, UFR Sciences, 76821 Mont-Saint-Aignan Cedex, France; e-mail: `rym.worms@univ-rouen.fr`

$\sigma_n > 0$ and $\alpha_n \in \mathbb{R}$ such that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{X_{n,n} - \alpha_n}{\sigma_n} \leq x\right) \longrightarrow H_\gamma(x) \qquad (1)$$

for all continuity points of the extreme value distribution function H_γ, defined as

$$H_\gamma(x) = \begin{cases} \exp\left(-(1+\gamma x)^{-\frac{1}{\gamma}}\right), & \text{for } \gamma \neq 0 \text{ and } 1 + \gamma x > 0 \\ \exp(-\exp(-x)), & \text{for } \gamma = 0 \text{ and } x \in \mathbb{R}. \end{cases}$$

The distribution function F is said to belong to the maximum domain of attraction of H_γ. The real-valued parameter γ is referred to as the extreme value index (EVI) of F. Most common continuous distribution functions satisfy this weak condition. Distributions with $\gamma > 0$ are called heavy-tailed, as their tail $\bar{F}$ typically decays slowly as a power function. Examples in this Fréchet class are the Pareto, Burr, Student't, α-stable ($\alpha < 2$) and loggamma distributions. The Gumbel class of distributions with $\gamma = 0$ encompasses the exponential, normal, lognormal, gamma and classical Weibull distributions, the tail of which diminishes exponentially fast. Finally, the Weibull class consists of distributions with $\gamma < 0$, which all have a finite right endpoint $s_+(F) := \sup\{x : F(x) < 1\}$. Examples in this class are the uniform and reverse Burr distributions.

The problem of estimating the so-called extreme value index γ, which determines the behaviour of the underlying d.f. F in its upper tail, has received much attention in the last twenty years. An extensive motivation of this estimation problem can be found in [10]. More recently, de Haan and Rootzén (see [15]) demonstrated how to use an estimate of γ to construct estimators of extreme quantiles. Nevertheless, the generalized extreme value distribution is appropriate when the data consist of a set of maxima. However, there has been some criticism of this approach, because using only maxima leads to the loss of information contained in other large sample values in a given period. This problem is remedied by considering several of the largest order statistics instead of just the largest one, that is considering all values larger than a given threshold. The differences between these values and a given threshold are called excesses over the threshold. Denote by $F_u(x) := \mathbb{P}(X - u \leq x | X > u)$ the distribution of the excesses of X over u, given that u is exceeded, and with $G_{\gamma,\sigma}$ the Generalized Pareto Distribution (GPD) defined, for all $x \geq 0$, by

$$G_{\gamma,\sigma}(x) = \begin{cases} 1 - \left(1 + \dfrac{\gamma x}{\sigma}\right)^{-\frac{1}{\gamma}}, & \text{for } \gamma \neq 0 \text{ and } 1 + \gamma x/\sigma > 0, \\ 1 - \exp\left(-\dfrac{x}{\sigma}\right), & \text{for } \gamma = 0. \end{cases}$$

Pickands' and Balkema and de Haan's result (see [17] and [1]) on the limiting distribution of excesses over a high threshold states that condition (1) holds if and only if

$$\lim_{u\to s_+(F)} \sup_{0 < x < s_+(F) - u} \left| F_u(x) - G_{\gamma,\sigma(u)}(x) \right| = 0$$

for some positive scaling function $\sigma(u)$ depending on u.

Thus, if, for some n, one fixes a high threshold u_n and selects from a sample $X_1, ..., X_n$ only those observations $X_{i_1}, ..., X_{i_{N_n}}$ that exceed u_n, a GPD with parameters γ and $\sigma_n = \sigma(u_n)$ is likely to be a good approximation for the distribution F_{u_n} of the N_n excesses $Y_j := X_{i_j} - u_n$, $j = 1, ..., N_n$. This is called the Peaks-Over-Threshold (P.O.T.) method.

As for the generalized extreme value (GEV) distribution, quite a number of techniques have been proposed to fit a GPD $G_{\gamma,\sigma}$ to the excesses $Y_1, ..., Y_{N_n}$. Maximum Likelihood Estimation (MLE, see [19]) of the Generalized Pareto Distribution has previously been considered in the literature, but Hosking and Wallis (see [16]) showed, using computer simulation, that, Maximum Likelihood Estimation, although asymptotically the most efficient method, does not clearly display its efficiency even in samples as large as 500, and that, in such cases, estimators derived by the method of probability-weighted moments are more reliable. This motivates in particular the fact

that, in this paper, we use this last mentioned method (PWM) in order to estimate the parameters γ and σ_n. Nevertheless, note that, in a forthcoming paper, we will derive the more general conditions on the estimators under which our results hold. These PWM estimators, denoted by $\hat{\gamma}_n$ and $\hat{\sigma}_n$ (see Sect. 2.1 for definitions), are based on the empirical distribution function of the excesses

$$\mathbb{A}_{n,u_n}(x) = \frac{1}{N_n} \sum_{j=1}^{N_n} \mathbb{I}_{\{Y_j \leq x\}}.$$

It is of course very important to measure the error between $\bar{F}_{u_n} := 1 - F_{u_n}$ (unknown) and its estimator $\bar{G}_{\hat{\gamma}_n,\hat{\sigma}_n} := 1 - G_{\hat{\gamma}_n,\hat{\sigma}_n}$. This error can be splitted into two parts: an approximation error and an estimation error. The first one, also called *bias of approximation*, is justified by the fact that the distribution of the excesses over u_n is only approximated by a GPD, which implies a systematic error studied in [20]. Since the distribution of the excesses over u_n is F_{u_n} and not G_{γ,σ_n}, the error due to the estimation of (γ, σ_n) is also divided into an approximation error due to the bias of approximation and a random term due to fluctuations.

Note that N_n follows a binomial distribution $\mathcal{B}(n, 1 - F(u_n))$. We suppose, in all the sequel that, $n(1 - F(u_n)) \to \infty$ as $n \to \infty$, that is $N_n \to \infty$ in probability. In such a case, $\frac{N_n}{n(1-F(u_n))} \to 1$ in probability as $n \to \infty$.

Let

$$F_{u_n}^*(y) = F_{u_n}(\sigma_n y) \quad \text{and} \quad \mathbb{A}_{n,u_n}^*(y) = \mathbb{A}_{n,u_n}(\sigma_n y)$$

for all $y \in \mathbb{R}_+$. We will study the asymptotic behaviour of $\hat{\gamma}_n$ and $\hat{\sigma}_n$ and more specifically the difference

$$\bar{F}_{u_n}(x) - \bar{G}_{\hat{\gamma}_n,\hat{\sigma}_n}(x)$$

where $\hat{\gamma}_n$ and $\hat{\sigma}_n$ are the PWM estimators introduced by Hosking and Wallis (in [16]). Note that the main difference with Drees (see [7]) is that we consider the estimation of the two parameters γ and σ_n simultaneously.

In what follows, we suppose that F is twice differentiable and that its inverse F^{-1} exists. Let V and A be two functions defined as

$$V(t) = \bar{F}^{-1}(\mathrm{e}^{-t}) \quad \text{and} \quad A(t) = \frac{V''(\ln t)}{V'(\ln t)} - \gamma.$$

We suppose the following first and second order conditions:

$$\lim_{t \to +\infty} A(t) = 0, \tag{2}$$

and

$$A \text{ is of constant sign at } \infty \text{ and there exists } \rho \leq 0 \text{ such that } |A| \in RV_\rho. \tag{3}$$

Under these assumptions, it is proved in [20] (Th. 1.4, p. 43) that as $u_n \to s_+(F)$

$$\bar{F}_{u_n}(\sigma_n y) - \bar{G}_\gamma(y) = a_n D_{\gamma,\rho}(y) + o(a_n), \text{ as } n \to +\infty, \tag{4}$$

for all y, when

$$\sigma_n := \sigma(u_n) = V'\left(V^{-1}(u_n)\right), \quad a_n := A\left(\mathrm{e}^{V^{-1}(u_n)}\right)$$

$$\bar{G}_\gamma(y) := 1 - G_{\gamma,1}(y),$$

and

$$D_{\gamma,\rho}(y) := \begin{cases} C_{0,\rho}(y), & \text{if } \gamma = 0, \\ C_{\gamma,\rho}\left(\dfrac{1}{\gamma}\ln(1 + \gamma y)\right), & \text{if } \gamma \neq 0, \end{cases}$$

where

$$C_{\gamma,\rho}(y) := \mathrm{e}^{-(1+\gamma)y} I_{\gamma,\rho}(y) \quad \text{and} \quad I_{\gamma,\rho}(y) := \int_0^y \mathrm{e}^{\gamma u} \int_0^u \mathrm{e}^{\rho s} \mathrm{d}s \mathrm{d}u.$$

We assume also that

$$\lim_{n\to\infty} \sqrt{n\left(1 - F(u_n)\right)}\, a_n = \lambda \in \mathbb{R}. \tag{5}$$

This is equivalent to suppose that $\sqrt{N_n}\, a_n$ tends to λ in probability, as $n \to \infty$.

The main result of this paper is the following. When ρ is equal to 0, the error due to the fact that $\bar{F}_{u_n}$ is replaced by $\bar{G}_{\hat{\gamma}_n, \hat{\sigma}_n}$ is of smaller order than the same error in the case $\rho \neq 0$. This result is closely linked (see Sect. 2.5 for further comments) to the penultimate approximation for the distribution of the excesses established in [21] (Gomes and de Haan in [13], generalizing [3, 12] and [14], studied penultimate approximation for the distribution of the maximum). At first sight, it can appear a bit strange since it is well known that, if we consider only the problem of the estimation of the index γ, the smaller $|\rho|$, the more difficult it is to estimate the index. This problem of bias in the estimation of the index has been widely studied recently in the literature and justified in particular the work on regression model by Feuerverger and Hall (see [9]) and Beirlant *et al.* (see [2]). This paper proves that, on the contrary, if we consider the problem of the estimation of the tail distribution, we do not need to construct asymptotically unbiased estimators, which is essential in the other estimation problem.

The remainder of our paper is organized as follows. In Section 2, we introduce the PWM estimators of Hosking and Wallis and study their asymptotic properties, from which we derive the asymptotic normality of $(\hat{\gamma}_n, \hat{\sigma}_n)$. Then, we establish our functional result and adapt penultimate approximation results to the case where the parameters are estimated by the PWM estimators. In Section 3, we provide some simulated examples in order to illustrate our main result. Finally, in Section 4, we describe the formalism used. The details of the proofs are postponed to the Appendix.

2. PROBABILITY-WEIGHTED MOMENTS ESTIMATORS

2.1. **Definitions**

Define, as in [16], the two first probability weighted moments of the GPD by,

$$\mu_0(\gamma, \sigma) = \int_0^\infty x\, dG_{\gamma, \sigma}(x) \ \text{ and } \ \mu_1(\gamma, \sigma) = \int_0^\infty x\left(1 - G_{\gamma, \sigma}(x)\right) dG_{\gamma, \sigma}(x). \tag{6}$$

Note that

$$\mu_0(\gamma, \sigma) = \int_0^\infty \left(1 - G_{\gamma, \sigma}(x)\right) dx \ \text{ and } \ \mu_1(\gamma, \sigma) = \int_0^\infty \frac{1}{2}\left(1 - G_{\gamma, \sigma}(x)\right)^2 dx. \tag{7}$$

When $\gamma < 1$,

$$\mu_0(\gamma, \sigma) = \frac{\sigma}{1 - \gamma} \ \text{ and } \ \mu_1(\gamma, \sigma) = \frac{\sigma}{2(2 - \gamma)}, \tag{8}$$

so that

$$\gamma = 2 - \frac{\mu_0(\gamma, \sigma)}{\mu_0(\gamma, \sigma) - 2\mu_1(\gamma, \sigma)} \ \text{ and } \ \sigma = \frac{2\,\mu_0(\gamma, \sigma)\,\mu_1(\gamma, \sigma)}{\mu_0(\gamma, \sigma) - 2\mu_1(\gamma, \sigma)}. \tag{9}$$

If we now replace $\mu_0(\gamma, \sigma)$ and $\mu_1(\gamma, \sigma)$ by their empirical moment estimators, one obtains the probability-weighted moments estimators $\hat{\gamma}_n$ and $\hat{\sigma}_n$ defined as:

$$\widehat{\gamma}_n = 2 - \frac{\widehat{\mu}_{0,n}}{\widehat{\mu}_{0,n} - 2\widehat{\mu}_{1,n}} \ \text{ and } \ \widehat{\sigma}_n = \frac{2\,\widehat{\mu}_{0,n}\widehat{\mu}_{1,n}}{\widehat{\mu}_{0,n} - 2\widehat{\mu}_{1,n}}, \tag{10}$$

where

$$\widehat{\mu}_{0,n} = \int_0^\infty x \, d\mathbb{A}_{n,\,u_n}(x) = \sigma_n \int_0^\infty y \, d\mathbb{A}_{n,\,u_n}^*(y) = \sigma_n \int_0^\infty \left[1 - \mathbb{A}_{n,\,u_n}^*(y) \right] dy \qquad (11)$$

and

$$\widehat{\mu}_{1,n} = \int_0^\infty \frac{1}{2} \left[1 - \mathbb{A}_{n,\,u_n}(x) \right]^2 dx = \sigma_n \int_0^\infty \frac{1}{2} \left[1 - \mathbb{A}_{n,\,u_n}^*(y) \right]^2 dy. \qquad (12)$$

As mentioned in the Introduction, Hosking and Wallis give formulae for the approximate standard errors of these estimators. They compare their approach to the MLE approach and come to the conclusion that in case $\gamma \geq 0$ the method of probability-weighted moments offers a viable alternative. In the sequel, we note

$$\mu_0 = \mu_0(\gamma, 1) \text{ and } \mu_1 = \mu_1(\gamma, 1).$$

We are going to study the asymptotic behaviour of $(\widehat{\gamma}_n, \widehat{\sigma}_n/\sigma_n)$ as $n \to +\infty$. Note that Drees (in [7]) has already studied the asymptotic behaviour of $\widehat{\gamma}_n$ using a different approach.

2.2. Asymptotic behaviour of $\widehat{\mu}_{0,n}$ and $\widehat{\mu}_{1,n}$

Theorem 1. *Under assumptions (2, 3) with $-1 < \gamma < 1/2$, and if*

$$\lim_{n \to \infty} \sqrt{n \left(1 - F(u_n) \right)} \, a_n = \lambda, \ \ \lambda \in \mathbb{R}, \quad as \ n \to +\infty, \qquad (13)$$

we have, for almost all sequences $k_n \to +\infty$, conditionally on $N_n = k_n$,

$$\sqrt{k_n} \begin{pmatrix} \dfrac{\widehat{\mu}_{0,n}}{\sigma_n} - \mu_0 \\[2mm] \dfrac{\widehat{\mu}_{1,n}}{\sigma_n} - \mu_1 \end{pmatrix} \xrightarrow{d} \mathcal{N} \left(\lambda C, \, \Gamma \right),$$

where

$$\Gamma = \begin{pmatrix} \dfrac{1}{(1-\gamma)^2(1-2\gamma)} & \dfrac{1}{2(1-\gamma)^2(2-\gamma)} \\[3mm] \dfrac{1}{2(1-\gamma)^2(2-\gamma)} & \dfrac{1}{(2-\gamma)^2(3-2\gamma)} \end{pmatrix} \ \ and \ \ C = \begin{pmatrix} \dfrac{1}{(1-\gamma)(1-\gamma+|\rho|)} \\[3mm] \dfrac{1}{2(2-\gamma)(2-\gamma+|\rho|)} \end{pmatrix}.$$

Moreover, if $\lambda = 0$, then

$$\begin{pmatrix} \dfrac{\widehat{\mu}_{0,n}}{\sigma_n} - \mu_0 \\[2mm] \dfrac{\widehat{\mu}_{1,n}}{\sigma_n} - \mu_1 \end{pmatrix} = a_n C + o(a_n) + \frac{1}{\sqrt{k_n}} \begin{pmatrix} \xi_n^0 \\[1mm] \xi_n^1 \end{pmatrix} + \mathcal{O}_P \left(\frac{1}{k_n} \right),$$

where $\begin{pmatrix} \xi_n^0 \\ \xi_n^1 \end{pmatrix}$ converges in distribution to a $\mathcal{N}(0, \Gamma)$ distribution.

Proof.

• We need the following lemma which will be proved in Appendix 5.1.

Lemma 1.

(1) *We have*

$$\int_0^\infty \left(\bar{F}_{u_n}^*(y) - \bar{G}_\gamma(y) \right) \, \mathrm{d}y \; = \; \frac{a_n}{(1-\gamma)(1-\gamma+|\rho|)} \; + \; o(a_n) \,. \tag{14}$$

(2) *Moreover, we have*

$$\frac{1}{2} \int_0^{+\infty} \left[(\bar{F}_{u_n}^*(y))^2 - (\bar{G}_\gamma(y))^2 \right] \, \mathrm{d}y \; = \; \frac{a_n}{2(2-\gamma)(2-\gamma+|\rho|)} \; + \; o(a_n) \,. \tag{15}$$

• Write

$$\frac{\widehat{\mu}_{0,n}}{\sigma_n} - \mu_0 \; = \; \int_0^\infty \left(1 - \mathbb{A}_{n,\,u_n}^*(y) \right) \, \mathrm{d}y - \int_0^\infty \left(1 - G_\gamma(y) \right) \, \mathrm{d}y. \tag{16}$$

Let α_{k_n} be the uniform empirical process based on k_n i.i.d. random variables uniformly distributed on $[0,\,1]$. Then, conditionally on $N_n = k_n$, we have

$$\mathbb{A}_{n,\,u_n}^* \; - \; F_{u_n}^* \; \stackrel{d}{=} \; \frac{1}{\sqrt{k_n}} \, \alpha_{k_n} \circ F_{u_n}^*. \tag{17}$$

Therefore, it follows that

$$\frac{\widehat{\mu}_{0,n}}{\sigma_n} - \mu_0 \stackrel{d}{=} \int_0^\infty \left(\bar{F}_{u_n}^*(y) - \bar{G}_\gamma(y) \right) \, \mathrm{d}y \; - \; \frac{1}{\sqrt{k_n}} \int_0^\infty \alpha_{k_n} \left(F_{u_n}^*(y) \right) \, \mathrm{d}y. \tag{18}$$

Lemma 1 yields

$$\int_0^\infty \left(\bar{F}_{u_n}^*(y) - \bar{G}_\gamma(y) \right) \, \mathrm{d}y = \frac{a_n}{(1-\gamma)(1-\gamma+|\rho|)} \; + \; o(a_n). \tag{19}$$

In order to study the second term in the right-hand side of (18), we need the following lemma established in Einmahl and Mason (see [8]).

Lemma 2. *On a rich enough probability space, there exist a sequence of probabilistically equivalent versions $\widetilde{\alpha}_n(.)$ of the uniform empirical process $\alpha_n(.)$ and a fixed Brownian bridge $\mathbb{B}$ such that, for all ν satisfying $0 \le \nu < 1/4$, we have*

$$\sup_{1/n \le t \le 1-1/n} \frac{n^\nu \, |\widetilde{\alpha}_n(t) - \mathbb{B}(t)|}{(t(1-t))^{1/2-\nu}} \; = \; \mathcal{O}_P(1). \tag{20}$$

Let $\tilde{\alpha}_n$ be such a processus for which we can apply (20). We prove in Appendix 5.2 that for $\gamma < 1/2$, $\int_0^\infty \tilde{\alpha}_{k_n} \left(F_{u_n}^*(y) \right) \mathrm{d}y$ converges in probability to $\int_0^1 h_\gamma(t) \, \mathrm{d}\mathbb{B}(t)$, where

$$h_\gamma(t) = \begin{cases} \dfrac{t^{-\gamma} - 1}{\gamma} & \text{if} \quad \gamma \neq 0, \\[2mm] \ln\left(1/t\right) & \text{if} \quad \gamma = 0. \end{cases} \tag{21}$$

It follows that

$$\frac{\widehat{\mu}_{0,n}}{\sigma_n} - \mu_0 \stackrel{d}{=} \frac{a_n}{(1-\gamma)(1-\gamma+|\rho|)} \; + \; o(a_n) \; + \; \frac{1}{\sqrt{k_n}} \xi_n^0, \tag{22}$$

where

$$\xi_n^0 \xrightarrow{P} -\int_0^1 h_\gamma(t) \, \mathrm{d}\mathbb{B}(t).$$

• Now, similarly we have

$$\frac{\widehat{\mu}_{1,n}}{\sigma_n} - \mu_1 = \int_0^{+\infty} \frac{1}{2} \left(1 - \mathbb{A}_{n,u_n}^*(y)\right)^2 \, \mathrm{d}y - \int_0^{+\infty} \frac{1}{2} \left(1 - G_\gamma(y)\right)^2 \, \mathrm{d}y$$

$$\stackrel{d}{=} \int_0^{+\infty} \frac{1}{2} \left(\bar{F}_{u_n}^*(y) - \frac{1}{\sqrt{k_n}} \alpha_{k_n}\left(F_{u_n}^*(y)\right)\right)^2 \, \mathrm{d}y - \frac{1}{2} \int_0^{+\infty} \left(1 - G_\gamma(y)\right)^2 \, \mathrm{d}y$$

$$\stackrel{d}{=} \frac{1}{2} \int_0^{+\infty} \left(\left(\bar{F}_{u_n}^*(y)\right)^2 - \left(\bar{G}_\gamma(y)\right)^2\right) \, \mathrm{d}y - \frac{1}{\sqrt{k_n}} \int_0^{+\infty} \bar{F}_{u_n}^*(y) \, \alpha_{k_n}\left(F_{u_n}^*(y)\right) \, \mathrm{d}y$$

$$+ \frac{1}{2k_n} \int_0^{+\infty} \left(\alpha_{k_n}\left(F_{u_n}^*(y)\right)\right)^2 \, \mathrm{d}y.$$

We are going to study separately the three terms of the last equality.

Lemma 1 yields

$$\frac{1}{2} \int_0^{+\infty} \left(\left(\bar{F}_{u_n}^*(y)\right)^2 - \left(\bar{G}_\gamma(y)\right)^2\right) \, \mathrm{d}y = \frac{a_n}{2(2-\gamma)(2-\gamma+|\rho|)} + o(a_n).$$

The same proof as in Appendix 5.2, shows that for $\gamma < 1/2$, $\int_0^{+\infty} \bar{F}_{u_n}^*(y) \, \tilde{\alpha}_{k_n}\left(F_{u_n}^*(y)\right) \, \mathrm{d}y$ converges in probability to $\int_0^1 k_\gamma(t) \mathrm{d}\mathbb{B}(t)$, where

$$k_\gamma(t) = -\frac{t^{1-\gamma}}{1-\gamma}. \tag{23}$$

We prove in Appendix 5.3 that

$$\frac{1}{k_n} \int_0^{+\infty} \left(\tilde{\alpha}_{k_n}\left(F_{u_n}^*(y)\right)\right)^2 \, \mathrm{d}y = \mathcal{O}_P\left(\frac{1}{k_n}\right).$$

Therefore, it follows that

$$\frac{\widehat{\mu}_{1,n}}{\sigma_n} - \mu_1 \stackrel{d}{=} \frac{a_n}{2(2-\gamma)(2-\gamma+|\rho|)} + o(a_n) + \frac{1}{\sqrt{k_n}} \xi_n^1 + \mathcal{O}_P\left(\frac{1}{k_n}\right), \tag{24}$$

where

$$\xi_n^1 \xrightarrow{P} -\int_0^1 k_\gamma(t) \, \mathrm{d}\mathbb{B}(t). \tag{25}$$

Finally, as $\lim_{n\to\infty} \sqrt{n(1-F(u_n))} \, a_n = \lambda$ and $N_n/(n(1-F(u_n))) \xrightarrow{P} 1$, it follows that, conditionally on $N_n = k_n$,

$$\sqrt{k_n} \left(\begin{array}{c} \dfrac{\widehat{\mu}_{0,n}}{\sigma_n} - \mu_0 \\[2mm] \dfrac{\widehat{\mu}_{1,n}}{\sigma_n} - \mu_1 \end{array}\right) = \lambda \left(\begin{array}{c} c_0 \\ c_1 \end{array}\right) + \left(\begin{array}{c} \xi_n^0 \\ \xi_n^1 \end{array}\right) + o_P(1), \tag{26}$$

where

$$\begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = \begin{pmatrix} \dfrac{1}{(1-\gamma)(1-\gamma+|\rho|)} \\ \dfrac{1}{2(2-\gamma)(2-\gamma+|\rho|)} \end{pmatrix} \tag{27}$$

and

$$\begin{pmatrix} \xi_n^0 \\ \xi_n^1 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Y_0 \\ Y_1 \end{pmatrix} = \begin{pmatrix} -\displaystyle\int_0^1 h_\gamma(t)\,\mathrm{d}\mathbb{B}(t) \\ -\displaystyle\int_0^1 k_\gamma(t)\,\mathrm{d}\mathbb{B}(t) \end{pmatrix} \quad \text{which is a } \ \mathcal{N}(0,\,\Gamma)\text{ distribution,} \tag{28}$$

with

$$\begin{aligned}
\Gamma &= \begin{pmatrix} \mathrm{Var}(Y_0) & \mathrm{Cov}(Y_0,\,Y_1) \\ \mathrm{Cov}(Y_0,\,Y_1) & \mathrm{Var}(Y_1) \end{pmatrix} \\
&= \begin{pmatrix} \dfrac{1}{(1-\gamma)^2(1-2\gamma)} & \dfrac{1}{2(1-\gamma)^2(2-\gamma)} \\ \dfrac{1}{2(1-\gamma)^2(2-\gamma)} & \dfrac{1}{(2-\gamma)^2(3-2\gamma)} \end{pmatrix}.
\end{aligned} \tag{29}$$

$\square$

2.3. Asymptotic behaviour of $\widehat{\gamma}_n$ and $\widehat{\sigma}_n$

Corollary 1. *Under the same assumptions as Theorem 1, if we suppose moreover that*[4]

$$n\,(1 - F(u_n))\,a_n \to \infty \quad as \ \ n \to +\infty, \tag{30}$$

then, for almost all sequences $k_n \to +\infty$, we have, conditionally on $N_n = k_n$, that

$$\sqrt{k_n} \begin{pmatrix} \widehat{\gamma}_n - \gamma \\ \dfrac{\widehat{\sigma}_n}{\sigma_n} - 1 \end{pmatrix} \xrightarrow{d} \mathcal{N}(\lambda B,\, \Sigma),$$

where

$$\Sigma = \frac{1}{(1-2\gamma)(3-2\gamma)} \begin{pmatrix} (1-\gamma)(2-\gamma)^2(1-\gamma+2\gamma^2) & (2-\gamma)(2-6\gamma+7\gamma^2-2\gamma^3) \\ (2-\gamma)(2-6\gamma+7\gamma^2-2\gamma^3) & 7-18\gamma+11\gamma^2-2\gamma^3 \end{pmatrix}$$

and

$$B = \begin{pmatrix} \dfrac{(2-\gamma)(1-\gamma)}{(1-\gamma+|\rho|)(2-\gamma+|\rho|)} \\ \dfrac{|\rho|}{(1-\gamma+|\rho|)(2-\gamma+|\rho|)} \end{pmatrix}. \tag{31}$$

---

[4]This condition is trivial in the case $\lambda \neq 0$.

Moreover, if $\lambda = 0$, then

$$\sqrt{k_n} \left(\begin{array}{c} \widehat{\gamma}_n - \gamma \\ \dfrac{\widehat{\sigma}_n}{\sigma_n} - 1 \end{array} \right) - a_n B \overset{d}{\longrightarrow} \mathcal{N}(0, \Sigma).$$

Proof. Define the mapping Ψ, from $\mathbb{R}_+^2 \cap \{(a, b) \in \mathbb{R}^2 : a > 2b\}$ into $]-\infty, 1[\times\mathbb{R}_+$, by

$$\Psi(a, b) = \left(2 - \frac{a}{a - 2b}, \frac{2ab}{a - 2b} \right).$$

Then,

$$(\gamma, 1) = \Psi(\mu_0, \mu_1)$$

and

$$\left(\widehat{\gamma}_n, \frac{\widehat{\sigma}_n}{\sigma_n} \right) = \Psi \left(\frac{\widehat{\mu}_{0,n}}{\sigma_n}, \frac{\widehat{\mu}_{1,n}}{\sigma_n} \right).$$

The proof of Theorem 1 implies (see (26, 27) and (28)) that, conditionally on $N_n = k_n$,

$$\begin{aligned} \left(\begin{array}{c} \dfrac{\widehat{\mu}_{0,n}}{\sigma_n} \\ \dfrac{\widehat{\mu}_{1,n}}{\sigma_n} \end{array} \right) &\overset{d}{=} \left(\begin{array}{c} \mu_0 \\ \mu_1 \end{array} \right) + a_n \left(\begin{array}{c} c_0 \\ c_1 \end{array} \right) + \frac{1}{\sqrt{k_n}} \left(\begin{array}{c} \xi_n^0 \\ \xi_n^1 \end{array} \right) + o(a_n) + \mathcal{O}_P \left(\frac{1}{k_n} \right) \\ &\overset{d}{=} \left(\begin{array}{c} \mu_0 \\ \mu_1 \end{array} \right) + \left(\begin{array}{c} \alpha_n \\ \beta_n \end{array} \right). \end{aligned} \tag{32}$$

A Taylor expansion yields

$$\begin{aligned} \left(\begin{array}{c} \widehat{\gamma}_n - \gamma \\ \dfrac{\widehat{\sigma}_n}{\sigma_n} - 1 \end{array} \right) = D\Psi(\mu_0, \mu_1) \left(\begin{array}{c} \alpha_n \\ \beta_n \end{array} \right) &+ \frac{1}{2} \left(\alpha_n \quad \beta_n \right) D^2 \Psi(\mu_0, \mu_1) \left(\begin{array}{c} \alpha_n \\ \beta_n \end{array} \right) \\ &+ o(a_n^2) + \mathcal{O}_P \left(\frac{a_n}{k_n} \right) + \mathcal{O}_P \left(\frac{1}{k_n^2} \right), \end{aligned}$$

where

$$D\Psi(\mu_0, \mu_1) = \frac{2}{(2\mu_1 - \mu_0)^2} \left(\begin{array}{cc} \mu_1 & -\mu_0 \\ -2\mu_1^2 & \mu_0^2 \end{array} \right). \tag{33}$$

Since $\lim_{n\to\infty} \sqrt{n(1 - F(u_n))}\, a_n = \lambda$ and $n(1 - F(u_n))a_n \to \infty$, we have

$$\frac{1}{2} \left(\alpha_n \quad \beta_n \right) D^2 \Psi(\mu_0, \mu_1) \left(\begin{array}{c} \alpha_n \\ \beta_n \end{array} \right) = \mathcal{O}_P \left(\frac{1}{k_n} \right).$$

Moreover,

$$D\Psi(\mu_0, \mu_1) \left(\begin{array}{c} \alpha_n \\ \beta_n \end{array} \right) = a_n \left(\begin{array}{c} b_0 \\ b_1 \end{array} \right) + \frac{1}{\sqrt{k_n}} D\Psi(\mu_0, \mu_1) \left(\begin{array}{c} \xi_n^0 \\ \xi_n^1 \end{array} \right) + o_P(a_n),$$

where

$$\begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \frac{2}{(2\mu_1 - \mu_0)^2} \begin{pmatrix} \mu_1 c_0 - \mu_0 c_1 \\ -2\mu_1^2 c_0 + \mu_0^2 c_1 \end{pmatrix} \tag{34}$$

and

$$D\Psi(\mu_0, \mu_1) \begin{pmatrix} Y_0 \\ Y_1 \end{pmatrix} \quad \text{is a } \mathcal{N}(0, \Sigma) \text{ distribution,}$$

with

$$\Sigma = D\Psi(\mu_0, \mu_1) \, \Gamma \, (D\Psi(\mu_0, \mu_1))^T \,. \tag{35}$$

Thus,

$$\sqrt{k_n} \begin{pmatrix} \widehat{\gamma}_n - \gamma \\ \dfrac{\widehat{\sigma}_n}{\sigma_n} - 1 \end{pmatrix} \xrightarrow{d} \mathcal{N}(\lambda B, \Sigma), \tag{36}$$

with $B = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}$. $\qquad\square$

2.4. Functional result

Theorem 2. *Under the same assumptions as Corollary 1, for almost all sequences $k_n \to +\infty$, we have, conditionally on $N_n = k_n$, that*

(i) *the process*

$$\sqrt{k_n} \left(\bar{F}^*_{u_n}(x) \ - \ \bar{G}_{\widehat{\gamma}_n, \widehat{\sigma}_n/\sigma_n}(x) \right)$$

converges in distribution to

$$\lambda B(x) + \mathbb{Z}(x) \,,$$

where

$$B(x) = D_{\gamma, \rho}(x) + \mathcal{L}_1(\gamma, \rho) \, \frac{\partial G_\gamma}{\partial \gamma}(x) + \mathcal{L}_2(\gamma, \rho) \, \frac{\partial G_\gamma}{\partial x}(x) \tag{37}$$

with

$$\mathcal{L}_1(\gamma, \rho) = \frac{(2 - \gamma)(1 - \gamma)}{(1 - \gamma + |\rho|)(2 - \gamma + |\rho|)} \ \text{and} \ \mathcal{L}_2(\gamma, \rho) = -\frac{x|\rho|}{(1 - \gamma + |\rho|)(2 - \gamma + |\rho|)},$$

and $\mathbb{Z}(x)$ is a centered Gaussian process equal to

$$\frac{\partial G_\gamma}{\partial \gamma}(x) \int_0^1 \ell_\gamma^{(1)}(t) \, d\mathbb{B}(t) \ - \ x \frac{\partial G_\gamma}{\partial x}(x) \int_0^1 \ell_\gamma^{(2)}(t) \, d\mathbb{B}(t),$$

with

$$\ell_\gamma^{(1)}(t) = \begin{cases} (2 - \gamma) \left(2\gamma(2 - \gamma)t + (1 - \gamma)^2 \right) \dfrac{t^{-\gamma}}{\gamma} & \text{if} \quad \gamma \neq 0, \\ 8t - 2\ln t & \text{if} \quad \gamma = 0, \end{cases}$$

and

$$\ell_\gamma^{(2)}(t) = \begin{cases} -\left(2\gamma(2-\gamma)^2 t + (1-\gamma)^3\right) \dfrac{t^{-\gamma}}{\gamma(1-\gamma)} & \text{if} \quad \gamma \neq 0, \\ \ln t - 8t & \text{if} \quad \gamma = 0. \end{cases}$$

Moreover, if $\lambda = 0$, then

$$\sqrt{k_n}\left(\bar{F}_{u_n}^*(x) - \bar{G}_{\widehat{\gamma}_n, \widehat{\sigma}_n/\sigma_n}(x)\right) - a_n B(x)$$

converges in distribution to $\mathbb{Z}(x)$;

(ii) *when $\rho = 0$, the bias $B(x)$ (in (37)) is equal to 0.*

Remark 1. $B(x)$ is called the *functional approximation error*, it contains the approximation error between $\bar{F}_{u_n}^*(x)$ and $\bar{G}_\gamma(x)$ (see (4)) and the approximation error due to the estimation of the unknown parameters γ and σ_n.

Proof. (i) Write

$$\bar{F}_{u_n}^*(x) - \bar{G}_{\widehat{\gamma}_n}(\sigma_n x/\widehat{\sigma}_n) = \bar{F}_{u_n}^*(x) - \bar{G}_\gamma(\sigma_n x/\widehat{\sigma}_n) + \bar{G}_\gamma(\sigma_n x/\widehat{\sigma}_n) - \bar{G}_{\widehat{\gamma}_n}(\sigma_n x/\widehat{\sigma}_n), \tag{38}$$

with

$$\begin{aligned} \bar{F}_{u_n}^*(x) - \bar{G}_\gamma(\sigma_n x/\widehat{\sigma}_n) &= \bar{F}_{u_n}^*(x) - \bar{G}_\gamma(x) + x\left(\frac{\sigma_n}{\widehat{\sigma}_n} - 1\right)\frac{\partial G_\gamma}{\partial x}(x) + \mathcal{O}_P\left(\frac{1}{k_n}\right) \\ &= a_n D_{\gamma,\rho}(x) - x\left(\frac{\widehat{\sigma}_n}{\sigma_n} - 1\right)\frac{\partial G_\gamma}{\partial x}(x) + o_P(a_n) \end{aligned}$$

and

$$\bar{G}_\gamma(\sigma_n x/\widehat{\sigma}_n) - \bar{G}_{\widehat{\gamma}_n}(\sigma_n x/\widehat{\sigma}_n) = (\widehat{\gamma}_n - \gamma)\frac{\partial G_\gamma}{\partial \gamma}(\sigma_n x/\widehat{\sigma}_n) + \mathcal{O}_P\left(\frac{1}{k_n}\right).$$

The result follows from the facts that, on the first hand (see Cor. 1),

$$\widehat{\gamma}_n - \gamma = a_n b_0 + \frac{\eta_n^0}{\sqrt{k_n}} + o_P(a_n) \quad \text{and} \quad \frac{\widehat{\sigma}_n}{\sigma_n} - 1 = a_n b_1 + \frac{\eta_n^1}{\sqrt{k_n}} + o_P(a_n),$$

with

$$\begin{pmatrix} b_0 \\ b_1 \end{pmatrix} \text{ given by (34)} \quad \text{and} \quad \begin{pmatrix} \eta_n^0 \\ \eta_n^1 \end{pmatrix} \xrightarrow{d} D\Psi(\mu_0, \mu_1) \begin{pmatrix} Y_0 \\ Y_1 \end{pmatrix},$$

and, on the other hand (see (28) and (33)),

$$\begin{aligned} D\Phi(\mu_0, \mu_1)\begin{pmatrix} Y_0 \\ Y_1 \end{pmatrix} &= \frac{2}{(2\mu_1 - \mu_0)^2}\begin{pmatrix} \displaystyle\int_0^1 (\mu_1 h_\gamma(t) - \mu_0 k_\gamma(t))\, d\mathbb{B}(t) \\ \displaystyle\int_0^1 (-2\mu_1^2 h_\gamma(t) + \mu_0^2 k_\gamma(t))\, d\mathbb{B}(t) \end{pmatrix} \\ &= \begin{pmatrix} \displaystyle\int_0^1 \ell_\gamma^{(1)}(t)\, d\mathbb{B}(t) \\ \displaystyle\int_0^1 \ell_\gamma^{(2)}(t)\, d\mathbb{B}(t) \end{pmatrix}. \end{aligned}$$

(ii) When $\gamma = 0$, it is easy to check that $D_{\gamma,\rho}(x) = \dfrac{x^2}{2}e^{-x} = -\dfrac{\partial G_\gamma}{\partial \gamma}(x)\mid_{\gamma=0}$, whereas if $\gamma \neq 0$, $I_{\gamma,0}(x) =$
$-\dfrac{e^{\gamma x} - 1 - \gamma x e^{\gamma x}}{\gamma^2}$ and hence $C_{\gamma,0}(x) = e^{-(1+\gamma)x}\dfrac{1 + \gamma x e^{\gamma x} - e^{\gamma x}}{\gamma^2}$. Therefore, as for $\gamma = 0$,

$$D_{\gamma,0}(x) = C_{\gamma,0}\left(\frac{1}{\gamma}\ln\left(1 + \gamma x\right)\right) = -\frac{\partial G_\gamma}{\partial \gamma}(x). \tag{39}$$

Taking $\rho = 0$ in (37) and using the fact that $D_{\gamma,0}(x) = -\dfrac{\partial G_\gamma}{\partial \gamma}(x)$, we derive that $B(x) = 0$, which is not the case if $\rho < 0$. $\qquad\square$

Let us give now some comments concerning this theorem:

- this result means that, in case $\rho = 0$, the error due to the fact that $\bar{F}^*_{u_n}(x)$ is replaced by $\bar{G}_{\widehat{\gamma}_n,\widehat{\sigma}_n/\sigma_n}(x)$ is of smaller order than the error made when one takes $\widehat{\gamma}_n$ in order to estimate γ. As noticed in the Introduction, this result is surprising taking into account the fact that in the case of the estimation of the index γ, the bias goes up when ρ tends to 0;
- the second order parameter ρ is zero for many usual distributions in the Gumbel domain of attraction ($\gamma = 0$): *e.g.* the normal, lognormal, gamma and classical Weibull distributions. Hence, our result applies to all these distributions;
- this result is closely linked to penultimate approximation established in [21], which is developed in the next section.

2.5. Penultimate approximation

It is established in [21] (Th. 1) that penultimate approximation is only possible in the case $\rho = 0$. That is, $F^*_{u_n}(x)$ is approximated by $G_{\gamma+a_n}(x)$ rather than by $G_\gamma(x)$. When ρ is zero, it is proved in [21] (Th. 3) that under some additional conditions (satisfied by many usual distributions), when $\gamma > -1$,

$$\bar{F}_{u_n}(\sigma_n y) - \bar{G}_{\gamma+a_n}(y) = b_n D_\gamma(y) + o(b_n) \ \text{ as } \ n \to +\infty, \tag{40}$$

where $b_n = o(a_n)$ is defined by

$$b_n = e^{V^{-1}(u_n)}A'\left(e^{V^{-1}(u_n)}\right) \tag{41}$$

and

$$D_\gamma(y) = \begin{cases} C_0(y) & \text{if} \quad \gamma = 0, \\ C_\gamma\left(\dfrac{1}{\gamma}\ln\left(1 + \gamma y\right)\right) & \text{if} \quad \gamma \neq 0, \end{cases}$$

where

$$C_\gamma(y) = e^{-(1+\gamma)y}M_\gamma(y) \ \text{ and } \ M_\gamma(y) = \frac{1}{2}\int_0^y u^2 e^{\gamma u}\,du. \tag{42}$$

The same proof as for Corollary 1 shows that the error made by replacing $\gamma + a_n$ by $\widehat{\gamma}_n$ is equal to

$$\frac{\gamma^2 - 3\gamma + 3}{2 - \gamma}\,b_n + o(b_n),$$

where b_n is given by (41).

Now a similar demonstration as the one for Theorem 2 implies that, in the case $\rho = 0$, the functional approximation error (analogous of B), which is of smaller order that a_n, becomes

$$b_n \left(D_\gamma(x) + \frac{\gamma^2 - 3\gamma + 3}{2 - \gamma} \frac{\partial G_\gamma}{\partial \gamma}(x) - \frac{\gamma - 1}{2 - \gamma} x \frac{\partial G_\gamma}{\partial x}(x) \right) + o(b_n),$$

which is equal to

$$b_n \frac{xe^{-x}}{2} \left(1 - \frac{3x}{2} + \frac{x^2}{2} \right) + o(b_n),$$

when $\gamma = 0$.

Example. Let us consider the classical Weibull distribution function defined by

$$\bar{F}(x) = \exp\left(-x^\beta\right), \quad x \geq 0,$$

with $\beta > 0$. For this distribution, $\gamma = 0$ and, when $\beta \neq 1$, we have $V(t) = t^{1/\beta}$. Hence $A(t) = \left(\frac{1}{\beta} - 1\right) \frac{1}{\ln t}$ and the second order parameter ρ is equal to 0. This yields

$$\sigma_n = \frac{1}{\beta} u_n^{1-\beta}$$

$$a_n = \left(\frac{1}{\beta} - 1\right) u_n^{-\beta}$$

$$b_n = \left(1 - \frac{1}{\beta}\right) u_n^{-2\beta}.$$

Let $k_n = n(1 - F(u_n))$ be the everedged number of excesses over u_n. It follows from the assumption $\sqrt{n(1 - F(u_n))} a_n \to \lambda$ that k_n is of order $(\ln n)^2$.

3. SIMULATIONS

Next we report on a small sample study performed in order to illustrate our theoretical results. With this aim, we consider 100 samples of size $n = 500$ drawn from a $\text{Burr}(\beta, \tau, \lambda)$ distribution, defined as

$$\bar{F}(x) := 1 - F(x) = \left(\frac{\beta}{\beta + x^\tau}\right)^\lambda, \quad x > 0,$$

where $\beta, \lambda, \tau > 0$.

Note that for such a distribution $\gamma = \frac{1}{\lambda \tau}$ and, if $\tau \neq 1$, $\rho = -1/\lambda$. We choose $k_n = n/10$ as suggested in [5] and [6] and $(\beta, \tau, \lambda) = (1, \frac{1}{2}, 8)$; $(1, \frac{2}{5}, 10)$ and $(1, \frac{1}{3}, 12)$. For these models $\gamma = \frac{1}{4}$ and $\rho = -\frac{1}{8}$, $-\frac{1}{10}$ and $-\frac{1}{12}$ respectively.

We take as explicit versions of $\widehat{\mu}_{0,n}$ and $\widehat{\mu}_{1,n}$ the following expressions (see [16] for the choice of $\widehat{\mu}_{1,n}$)

$$\widehat{\mu}_{0,n} = \frac{1}{N_n} \sum_{i=1}^{N_n} Y_i \text{ and } \widehat{\mu}_{1,n} = \frac{1}{N_n} \sum_{i=1}^{N_n} p_i Y_i,$$

where $p_i = 1 - \frac{i - 0.35}{N_n}$. For each distribution and each sample we compute the PWM estimators

$$\widehat{\gamma}_n^j = 2 - \frac{\widehat{\mu}_{0,n}^j}{\widehat{\mu}_{0,n}^j - 2\widehat{\mu}_{1,n}^j} \quad \text{and} \quad \widehat{\sigma}_n^j = \frac{2\widehat{\mu}_{0,n}^j \widehat{\mu}_{1,n}^j}{\widehat{\mu}_{0,n}^j - 2\widehat{\mu}_{1,n}^j},$$

for $j = 1, \ldots, 100$, and we plot in Figure 1

$$\left| \bar{F}_{u_n}(x) - \frac{1}{100} \sum_{j=1}^{100} \bar{G}_{\widehat{\gamma}_n^j, \widehat{\sigma}_n^j}(x) \right|.$$

From this figure it is clear that as ρ tends to 0, the quantity $|\bar{F}_{u_n}(x) - \frac{1}{100} \sum_{j=1}^{100} \bar{G}_{\widehat{\gamma}_n^j, \widehat{\sigma}_n^j}(x)|$ becomes small. This
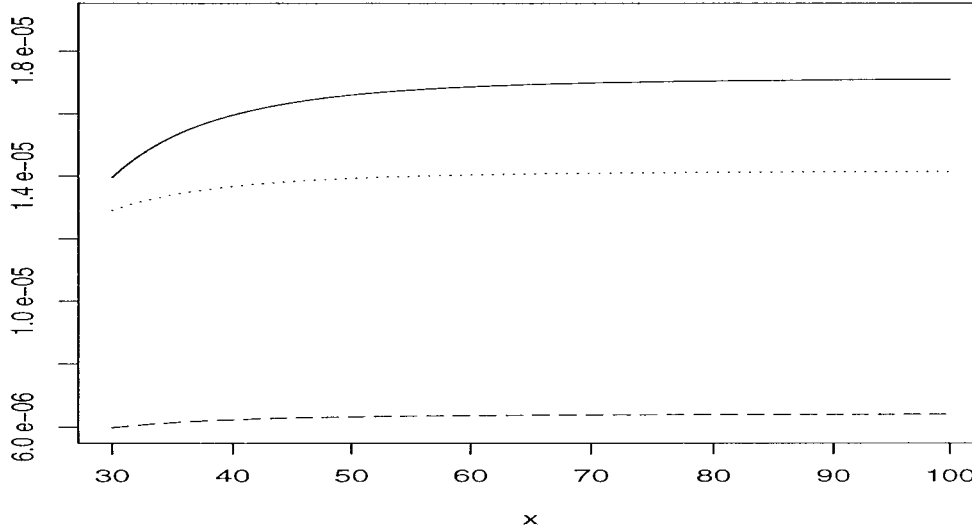

FIGURE 1. Bias in the case of a Burr$(1, \frac{1}{2}, 8)$ distribution (full line), of a Burr$(1, \frac{2}{5}, 10)$ distribution (dotted line) and of a Burr$(1, \frac{1}{3}, 12)$ distribution (dashed line).

corroborate the theory obtained in this paper, and proves that if we are interested in the tail distribution, we do not have to use asymptotically unbiased estimators, which is on the other hand essential in the case of the estimation of an extreme value index.

4. Sketch of the formalism

We denote by $T(\bar{F})$ a class of estimators of (γ, σ) such that for all γ in some interval and $\sigma > 0$,

$$T\left(\bar{G}_{\gamma, \sigma}\right) = \left(T_1\left(\bar{G}_{\gamma, \sigma}\right), T_2\left(\bar{G}_{\gamma, \sigma}\right)\right) = (\gamma, \sigma). \tag{43}$$

We restrict our study to estimators $T(\bar{F})$ satisfying (43) which are scale invariant for all $\bar{F}$ in some large class of distributions and all $\sigma > 0$,

$$T_1\left(\bar{F}\left(\frac{\bullet}{\sigma}\right)\right) = T_1\left(\bar{F}\right) \quad \text{and} \quad T_2\left(\bar{F}\left(\frac{\bullet}{\sigma}\right)\right) = \sigma\, T_2\left(\bar{F}\right). \tag{44}$$

If condition (44) is satisfied, then in particular for all γ and $\sigma > 0$,

$$T_1\left(\bar{G}_{\gamma, \sigma}\right) = T_1\left(\bar{G}_{\gamma}\right) = \gamma \quad \text{and} \quad T_2\left(\bar{G}_{\gamma, \sigma}\right) = \sigma\, T_2\left(\bar{G}_{\gamma}\right) = \sigma, \quad \text{thus} \quad T_2\left(\bar{G}_{\gamma}\right) = 1. \tag{45}$$

Consequently

$$\partial\, T_1\left(\bar{G}_{\gamma}\right)/\partial \gamma = 1 \quad \text{and} \quad \partial\, T_2\left(\bar{G}_{\gamma}\right)/\partial \gamma = 0. \tag{46}$$

We suppose moreover a form of Hadamard differentiability: there exist linear forms $DT(\bar{G}_\gamma)$, defined on a suitable space, such that under assumption (5), we have

$$\lim_{n \to +\infty} \frac{T\left(\bar{F}^*_{u_n}\right) - T\left(\bar{G}_\gamma\right)}{a_n} = DT\left(\bar{G}_\gamma\right)[D_{\gamma, \rho}], \tag{47}$$

$$\lim_{n \to +\infty} \sqrt{N_n}\left(T\left(\bar{F}^*_{u_n} + \frac{1}{\sqrt{N_n}}\alpha_{N_n}\left(F^*_{u_n}\right)\right) - T\left(\bar{F}^*_{u_n}\right)\right) = DT\left(\bar{G}_\gamma\right)\left[\mathbb{B} \circ \bar{G}_\gamma\right], \tag{48}$$

in probability, and

$$\lim_{n \to +\infty} \frac{T\left(\bar{G}_{\gamma + a_n}\right) - T\left(\bar{G}_\gamma\right)}{a_n} = DT\left(\bar{G}_\gamma\right)\left[\frac{\partial \bar{G}_\gamma}{\partial \gamma}\right]. \tag{49}$$

We defined σ_n as in the Introduction. Thus, denoting by $(\gamma^b_n, \sigma^b_n) = T(\bar{F}_{u_n})$ the values of γ and σ which could be obtained by replacing $\bar{G}_{\gamma, \sigma_n}$ by the excesses distribution, we have

$$\left(\gamma^b_n - \gamma, \frac{\sigma^b_n}{\sigma_n} - 1\right) = T\left(\bar{F}^*_{u_n}\right) - T\left(\bar{G}_\gamma\right) = a_n\,DT\left(\bar{G}_\gamma\right)[D_{\gamma, \rho}] + o(a_n)$$

$$= a_n\,(\mathcal{L}_1(\gamma, \rho), \mathcal{L}_2(\gamma, \rho)) + o(a_n). \tag{50}$$

This heuristic reasoning shows that the approximation error due to the estimation of the parameters γ and σ_n is of order a_n. We can now infer, always heuristically, the principal term in the asymptotic expansion (in terms of a_n) of $\bar{F}^*_{u_n} - \bar{G}_{\gamma^b_n, \sigma^b_n/\sigma_n}$, by calculating the first order approximation between $\bar{G}_{\gamma^b_n, \sigma^b_n/\sigma_n}$ and $\bar{G}_{\gamma, 1}$. This principal term is the function denoted by B in Section 2.4 and called the functional approximation error. It contains not only the approximation error between $\bar{F}^*_{u_n}(x)$ and $\bar{G}_\gamma(x)$ (see (47)) but also the approximation error due to the estimation of the unknown parameters γ and σ_n (see (50)). It is given by

$$B(x) = D_{\gamma, \rho}(x) + \mathcal{L}_1(\gamma, \rho)\frac{\partial G_\gamma}{\partial \gamma}(x) - \mathcal{L}_2(\gamma, \rho)\,x\frac{\partial G_\gamma}{\partial x}(x). \tag{51}$$

In the special case $\rho = 0$, we find that $\mathcal{L}_1(\gamma, 0) = 1$ and $\mathcal{L}_2(\gamma, 0) = 0$ which imply that the function B is equal to zero.

5. APPENDIX

5.1. **Proof of Lemma 1**

5.1.1.

Our aim is to prove that

$$\int_0^\infty \frac{\bar{F}^*_{u_n}(y) - \bar{G}_\gamma(y)}{a_n}\,\mathrm{d}y \text{ converges to } \frac{1}{(1 - \gamma)(1 - \gamma + |\rho|)}, \text{ as } n \to \infty. \tag{52}$$

We only treat the case[5] $\gamma > 0$.

Since

$$\bar{G}_\gamma(y) = \bar{G}_0\left(\frac{1}{\gamma}\ln\left(1 + \gamma y\right)\right)$$

and

$$\bar{F}^*_{u_n}(y) = \bar{G}_0\left(V^{-1}(u_n + \sigma_n y) - V^{-1}(u_n)\right),$$

---

[5]The cases $\gamma = 0$ and $\gamma < 0$ are similar.

the left-hand side of (52) becomes

$$\int_0^\infty \frac{\bar{F}_{u_n}^*(y) - \bar{G}_\gamma(y)}{a_n} \, \mathrm{d}y = \frac{1}{a_n} \int_0^\infty \exp\left(V^{-1}(u_n) - V^{-1}(u_n + \sigma_n y)\right) \mathrm{d}y$$
$$- \frac{1}{a_n} \int_0^\infty \exp\left(-\frac{1}{\gamma} \ln(1 + \gamma y)\right) \mathrm{d}y.$$

Taking s equal to $V^{-1}(u_n + \sigma_n y) - V^{-1}(u_n)$ in the first (resp. $\ln(1 + \gamma y)/\gamma$ in the second) term of the right-hand side of the preceding formula, we obtain

$$\int_0^\infty \frac{\bar{F}_{u_n}^*(y) - \bar{G}_\gamma(y)}{a_n} \, \mathrm{d}y = \frac{1}{a_n} \int_0^\infty \mathrm{e}^{-s} \left(\frac{V'\left(s + V^{-1}(u_n)\right)}{V'\left(V^{-1}(u_n)\right)} - \mathrm{e}^{\gamma s}\right) \mathrm{d}s.$$

It is well-known (see (1.22) in [20]) that under (2) and (3),

$$\lim_{n \to +\infty} \frac{1}{a_n} \left(\frac{V'\left(s + V^{-1}(u_n)\right)}{V'\left(V^{-1}(u_n)\right)} - \mathrm{e}^{\gamma s}\right) = \mathrm{e}^{\gamma s} \int_0^s \mathrm{e}^{\rho t} \, \mathrm{d}t,$$

and for all $\epsilon > 0$ and n sufficiently large,

$$\frac{\mathrm{e}^{-s}}{a_n} \left|\frac{V'\left(s + V^{-1}(u_n)\right)}{V'\left(V^{-1}(u_n)\right)} - \mathrm{e}^{\gamma s}\right| \le \mathrm{C} \; \mathrm{e}^{(\gamma + \epsilon - 1)s} \int_0^s \mathrm{e}^{\rho t} \, \mathrm{d}t, \text{ for some constant } C > 0. \qquad (53)$$

Since $\rho \le 0$, the right-hand side of (53) is integrable for ϵ sufficiently small and $\gamma < 1/2$. Then the result follows by applying Lebesgue dominated convergence.

5.1.2.

Since $\lim_{y \to +\infty} y \left(\bar{F}_{u_n}^*(y)\right)^2 = 0$, for $\gamma < \frac{1}{2}$, it follows by integration by parts that the quantity of interest can be written as

$$\frac{1}{2} \int_0^{+\infty} \left(\left(\bar{F}_{u_n}^*(y)\right)^2 - \left(\bar{G}_\gamma(y)\right)^2\right) \, \mathrm{d}y = \int_0^{+\infty} y \, \bar{F}_{u_n}^*(y) \, \mathrm{d}F_{u_n}^*(y) - \int_0^{+\infty} y \, \bar{G}_\gamma(y) \, \mathrm{d}G_\gamma(y).$$

We are going now to prove that

$$\frac{1}{a_n} \left(\int_0^\infty y \, \bar{F}_{u_n}^*(y) \, \mathrm{d}F_{u_n}^*(y) - \int_0^\infty y \, \bar{G}_\gamma(y) \, \mathrm{d}G_\gamma(y)\right) \text{ converges to } \frac{1}{2(2 - \gamma)(2 - \gamma + |\rho|)}, \text{ as } n \to \infty.$$

Again we only treat the case $\gamma > 0$. Write

$$\int_0^\infty y \, \bar{F}_{u_n}^*(y) \, \mathrm{d}F_{u_n}^*(y) - \int_0^\infty y \, \bar{G}_\gamma(y) \, \mathrm{d}G_\gamma(y) = \int_0^\infty y \, \left(\bar{G}_\gamma(y) - \bar{F}_{u_n}^*(y)\right) \, \mathrm{d}\bar{G}_\gamma(y)$$
$$+ \int_0^\infty y \, \bar{F}_{u_n}^*(y) \, \mathrm{d}\left(\bar{G}_\gamma(y) - \bar{F}_{u_n}^*(y)\right).$$

We study separately the two terms of the right-hand side.

Similar computations as in Appendix 5.1.1 prove that

$$\int_0^\infty y \left(\bar{G}_\gamma(y) - \bar{F}_{u_n}^*(y) \right) \mathrm{d}\bar{G}_\gamma(y) = \int_0^\infty \mathrm{e}^{-s} \frac{V \left(s + V^{-1}(u_n) \right) - u_n}{V' \left(V^{-1}(u_n) \right)} \, \mathrm{d}B_{u_n}(s)$$
$$+ \int_0^\infty \mathrm{e}^{-2s} \left(\frac{V \left(s + V^{-1}(u_n) \right) - u_n}{V' \left(V^{-1}(u_n) \right)} - \frac{\mathrm{e}^{\gamma s} - 1}{\gamma} \right) \mathrm{d}s,$$

where

$$B_{u_n}(s) = \bar{G}_0(s) - \bar{G}_0 \left(\frac{1}{\gamma} \ln \left(1 + \gamma \frac{V \left(s + V^{-1}(u_n) \right) - u_n}{V' \left(V^{-1}(u_n) \right)} \right) \right),$$

and similarly

$$\int_0^\infty y \, \bar{F}_{u_n}^*(y) \, \mathrm{d} \left(\bar{G}_\gamma(y) - \bar{F}_{u_n}^*(y) \right) = - \int_0^\infty \mathrm{e}^{-s} \frac{V \left(s + V^{-1}(u_n) \right) - u_n}{V' \left(V^{-1}(u_n) \right)} \, \mathrm{d}B_{u_n}(s).$$

Consequently it follows that

$$\frac{1}{a_n} \left(\int_0^\infty y \bar{F}_{u_n}^*(y) \, \mathrm{d}F_{u_n}^*(y) - \int_0^\infty y \bar{G}_\gamma(y) \, \mathrm{d}G_\gamma(y) \right) = \int_0^\infty \frac{\mathrm{e}^{-2s}}{a_n} \left(\frac{V \left(s + V^{-1}(u_n) \right) - u_n}{V' \left(V^{-1}(u_n) \right)} - \frac{\mathrm{e}^{\gamma s} - 1}{\gamma} \right) \mathrm{d}s.$$

It is well-known (see (1.23) in [20]) that under (2) and (3),

$$\lim_{n \to +\infty} \frac{1}{a_n} \left(\frac{V \left(s + V^{-1}(u_n) \right) - u_n}{V' \left(V^{-1}(u_n) \right)} - \frac{\mathrm{e}^{\gamma s} - 1}{\gamma} \right) = \int_0^s \mathrm{e}^{\gamma z} \int_0^z \mathrm{e}^{\rho t} \, \mathrm{d}t \, \mathrm{d}z,$$

and, for all $\epsilon > 0$ and n sufficiently large,

$$\frac{\mathrm{e}^{-2s}}{a_n} \left[\frac{V \left(s + V^{-1}(u_n) \right) - u_n}{V' \left(V^{-1}(u_n) \right)} - \frac{\mathrm{e}^{\gamma s} - 1}{\gamma} \right] \leq \mathrm{C} \, \mathrm{e}^{(\epsilon-2)s} \int_0^s \mathrm{e}^{\gamma z} \int_0^z \mathrm{e}^{\rho t} \, \mathrm{d}t \, \mathrm{d}z.$$

The result now follows, once again, by applying Lebesgue dominated convergence.

5.2. Proof that $\int_0^\infty \tilde{\alpha}_{k_n} \left(F_{u_n}^*(y) \right) \mathrm{d}y$ converges to $\int_0^1 h_\gamma(t) \, \mathrm{d}\mathbb{B}(t)$

Write $\bar{F}_{u_n}^*(y) = \bar{G}_0 \left(V^{-1}(u_n + \sigma_n y) - V^{-1}(u_n) \right)$ and take s equal to $V^{-1}(u_n + \sigma_n y) - V^{-1}(u_n)$ in order to obtain that

$$\int_0^\infty \tilde{\alpha}_{k_n} \left(F_{u_n}^*(y) \right) \mathrm{d}y = \int_0^\infty \tilde{\alpha}_{k_n} \left(1 - \mathrm{e}^{-s} \right) \frac{V' \left(s + V^{-1}(u_n) \right)}{V' \left(V^{-1}(u_n) \right)} \, \mathrm{d}s$$
$$= \int_0^\infty \tilde{\alpha}_{k_n} (1 - \mathrm{e}^{-s}) \left(\frac{V' \left(s + V^{-1}(u_n) \right)}{V' \left(V^{-1}(u_n) \right)} - \mathrm{e}^{\gamma s} \right) \mathrm{d}s + \int_0^\infty \tilde{\alpha}_{k_n} (1 - \mathrm{e}^{-s}) \mathrm{e}^{\gamma s} \, \mathrm{d}s. \quad (54)$$

We study separately the two terms of the right-hand side of (54).

• Note that

$$\int_0^\infty \tilde{\alpha}_{k_n} \left(1 - \mathrm{e}^{-s} \right) \mathrm{e}^{\gamma s} \, \mathrm{d}s = \int_0^1 \tilde{\alpha}_{k_n} (1 - t) \frac{\mathrm{d}t}{t^{1+\gamma}} \quad (55)$$

is well-defined. The right-hand side of (55) can be splitted into three terms as follows:

$$
\int_0^1 \tilde{\alpha}_{k_n}(1-t)\frac{\mathrm{d}t}{t^{1+\gamma}} = \int_0^{1/k_n} \tilde{\alpha}_{k_n}(1-t)\frac{\mathrm{d}t}{t^{1+\gamma}} + \int_{1/k_n}^{1-1/k_n} \tilde{\alpha}_{k_n}(1-t)\frac{\mathrm{d}t}{t^{1+\gamma}}
$$
$$
+ \int_{1-1/k_n}^1 \tilde{\alpha}_{k_n}(1-t)\frac{\mathrm{d}t}{t^{1+\gamma}}. \tag{56}
$$

- Let $M > \sqrt{2}$ and $\epsilon > 0$. It follows from Jaeschke theorem (see Th. 1, p. 600, Chap. 16 in [18]) that there exists n_0 such that for all $n \geq n_0$ and t in $]0, 1[$,

$$
\tilde{\alpha}_{k_n}(1-t) \leq M \sqrt{\ln\ln k_n\ t(1-t)}, \text{ with probability } > 1 - \epsilon.
$$

Then we derive that

$$
\int_0^{1/k_n} \tilde{\alpha}_{k_n}(1-t)\frac{\mathrm{d}t}{t^{1+\gamma}} \leq M \sqrt{\ln\ln k_n} \int_0^{1/k_n} \sqrt{t(1-t)}\frac{\mathrm{d}t}{t^{1+\gamma}}, \text{ with probability } > 1 - \epsilon.
$$

Since

$$
\int_0^{1/k_n} \sqrt{t(1-t)}\frac{\mathrm{d}t}{t^{1+\gamma}} \leq \int_0^{1/k_n} \sqrt{t}\frac{\mathrm{d}t}{t^{1+\gamma}}
$$

and, for $\gamma < \frac{1}{2}$,

$$
\sqrt{\ln\ln k_n} \int_0^{1/k_n} \sqrt{t}\frac{\mathrm{d}t}{t^{1+\gamma}} = \sqrt{\ln\ln k_n}\ \frac{k_n^{\gamma-\frac{1}{2}}}{\frac{1}{2}-\gamma} \to 0 \text{ as } k_n \to +\infty,
$$

we have

$$
\int_0^{1/k_n} \tilde{\alpha}_{k_n}(1-t)\frac{\mathrm{d}t}{t^{1+\gamma}} \xrightarrow{P} 0 \text{ as } k_n \to +\infty.
$$

- Similarly, we can prove that

$$
\int_{1-1/k_n}^1 \tilde{\alpha}_{k_n}(1-t)\frac{\mathrm{d}t}{t^{1+\gamma}} \xrightarrow{P} 0 \text{ as } k_n \to +\infty.
$$

- Therefore, it remains to study the term

$$
\int_{1/k_n}^{1-1/k_n} \tilde{\alpha}_{k_n}(1-t)\frac{\mathrm{d}t}{t^{1+\gamma}}.
$$

Let $\nu < \frac{1}{4}$, $M > 0$ and $\epsilon > 0$. Then, it follows from (20) that there exists n_1 such that for all $n \geq n_1$ and $t \in [\frac{1}{k_n}, 1 - \frac{1}{k_n}]$,

$$
\tilde{\alpha}_{k_n}(1-t) \leq \mathbb{B}(1-t) + Mk_n^{-\nu}\left(t(1-t)\right)^{1/2-\nu}, \text{ with probability } > 1 - \epsilon.
$$

Hence,

$$\int_{1/k_n}^{1-1/k_n} \tilde{\alpha}_{k_n}(1-t)\frac{\mathrm{d}t}{t^{1+\gamma}} \le \int_{1/k_n}^{1-1/k_n} \mathbb{B}(1-t)\frac{\mathrm{d}t}{t^{1+\gamma}} + Mk_n^{-\nu}\int_{1/k_n}^{1-1/k_n} (t(1-t))^{1/2-\nu}\,\frac{\mathrm{d}t}{t^{1+\gamma}}, \qquad (57)$$

with probability $> 1 - \epsilon$.

Since $\gamma < \frac{1}{2}$, the two integrals in the right-hand side converge according to the law of iterated logarithm for the brownian motion.

Finally

$$\lim_{n\to\infty} \int_{1/k_n}^{1-1/k_n} \tilde{\alpha}_{k_n}(1-t)\frac{\mathrm{d}t}{t^{1+\gamma}} \le \int_0^1 \mathbb{B}(1-t)\frac{\mathrm{d}t}{t^{1+\gamma}} + o_P(1).$$

The same computations can be applied in order to obtain the lower bound. Consequently

$$\lim_{n\to\infty} \int_{1/k_n}^{1-1/k_n} \tilde{\alpha}_{k_n}(1-t)\frac{\mathrm{d}t}{t^{1+\gamma}} = \int_0^1 \mathbb{B}(1-t)\frac{\mathrm{d}t}{t^{1+\gamma}} = \int_0^1 h_\gamma(t)\mathrm{d}\mathbb{B}(t).$$

• It remains to prove that $\int_0^\infty \tilde{\alpha}_{k_n}(1-\mathrm{e}^{-s})(V'(s+V^{-1}(u_n))/V'(V^{-1}(u_n)) - \mathrm{e}^{\gamma s})\,\mathrm{d}s$ tends to 0 in probability, as n tends to $+\infty$. We know that, under assumptions (2) and (3) (see Prop. 1.1 in [20]), for $\epsilon > 0$ and n sufficiently large

$$\left| \frac{V'\left(s+V^{-1}(u_n)\right)}{V'\left(V^{-1}(u_n)\right)} - \mathrm{e}^{\gamma s} \right| \le \mathrm{C}\,|a_n|\mathrm{e}^{(\gamma+\epsilon)s}\int_0^s \mathrm{e}^{\rho t}\,\mathrm{d}t. \qquad (58)$$

Moreover, using Theorems 1 and 3 (p. 604 of [18]), there exist $M > 0$ and $n_0(\omega) \in \mathbb{N}$ almost surely finite such that for $n \ge n_0$, $s > 0$ and $\delta > 0$

$$\left| \tilde{\alpha}_{k_n}\left(1-\mathrm{e}^{-s}\right) \right| \le M\mathrm{e}^{-s/2}(\ln k_n)^{\frac{1}{2}+\delta} \qquad \text{a.s.} \qquad (59)$$

It follows from (58) and (59) that

$$\int_0^\infty \left| \tilde{\alpha}_{k_n}\left(1-\mathrm{e}^{-s}\right) \right| \left| \frac{V'\left(s+V^{-1}(u_n)\right)}{V'\left(V^{-1}(u_n)\right)} - \mathrm{e}^{\gamma s} \right| \mathrm{d}s \le \mathrm{C}\,|a_n|(\ln k_n)^{\frac{1}{2}+\delta}\int_0^\infty \mathrm{e}^{(\gamma+\epsilon-\frac{1}{2})s}\int_0^s \mathrm{e}^{\rho t}\,\mathrm{d}t\,\mathrm{d}s. \qquad (60)$$

Since $\gamma < 1/2$, one can choose ϵ sufficiently small in order to have $\gamma + \epsilon - 1/2 < 0$. Moreover, $a_n(\ln k_n)^{1/2+\delta}$ converges to 0 as n tends to $+\infty$ under the assumptions.

Hence $\int_0^\infty \alpha_{k_n}(1-\mathrm{e}^{-s})(V'(s+V^{-1}(u_n))/V'(V^{-1}(u_n)) - \mathrm{e}^{\gamma s})\,\mathrm{d}s$ converges to 0 in probability.

5.3. Proof that $\frac{1}{k_n}\int_0^{+\infty}\left(\tilde{\alpha}_{k_n}\left(F_{u_n}^*\right)(y)\right)^2\mathrm{d}y = \mathcal{O}_P\left(\frac{1}{k_n}\right)$

Using the same methodology as in 5.2, we have

$$\int_0^{+\infty} \left(\tilde{\alpha}_{k_n}\left(F_{u_n}^*(y)\right)\right)^2\mathrm{d}y = \int_0^{+\infty} \left(\tilde{\alpha}_{k_n}(1-\mathrm{e}^{-s})\right)^2\left(\frac{V'\left(s+V^{-1}(u_n)\right)}{V'\left(V^{-1}(u_n)\right)} - \mathrm{e}^{\gamma s}\right)\mathrm{d}s$$

$$+ \int_0^{+\infty} \left(\tilde{\alpha}_{k_n}(1-\mathrm{e}^{-s})\right)^2\mathrm{e}^{\gamma s}\,\mathrm{d}s, \qquad (61)$$

with the first integral converging to zero in probability. Therefore we have to establish that

$$\int_0^{+\infty} \left(\tilde{\alpha}_{k_n}(1-\mathrm{e}^{-s})\right)^2\mathrm{e}^{\gamma s}\,\mathrm{d}s = \mathcal{O}_P(1).$$

Again this integral can be splitted into three terms. Following exactly the same lines as in 5.2 we can easily prove that these terms tend to zero in probability, which achieves the proof.

References

[1] A. Balkema and L. de Haan, Residual life time at a great age. *Ann. Probab.* **2** (1974) 792-801.

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

[3] J.P. Cohen, Convergence rates for the ultimate and penultimate approximations in extreme-value theory. *Adv. Appl. Prob.* **14** (1982) 833-854.

[4] A.L.M. Dekkers and L. de Haan, On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist.* **17** (1989) 1795-1832.

[5] J. Diebolt, V. Durbec, M.A. El Aroui and B. Villain, Estimation of extreme quantiles: Empirical tools for methods assessment and comparison. *Int. J. Reliability Quality Safety Engrg.* **7** (2000) 75-94.

[6] J. Diebolt and M.A. El Aroui, On the use of Peaks over Threshold methods for estimating out-of-sample quantiles. *Comput. Statist. Data Anal.* (to appear).

[7] H. Drees, A general class of estimators of the extreme value index. *J. Statist. Plann. Inf.* **66** (1998) 95-112.

[8] U. Einmahl and D.M. Mason, Approximation to permutation and exchangeable processes. *J. Theor. Probab.* **5** (1992) 101-126.

[9] A. Feuerverger and P. Hall, Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann. Statist.* **27** (1999) 760-781.

[10] J. Galambos, *Asymptotic theory of extreme order statistics*. Krieger, Malabar, Florida (1978).

[11] B.V. Gnedenko, Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44** (1943) 423-453.

[12] M.I. Gomes, Penultimate limiting forms in extreme value theory. *Ann. Inst. Stat. Math.* **36** (1984) 71-85.

[13] I. Gomes and L. de Haan, Approximation by penultimate extreme value distributions. *Extremes* **2** (2000) 71-85.

[14] M.I. Gomes and D.D. Pestana, *Non standard domains of attraction and rates of convergence*. John Wiley & Sons (1987) 467-477.

[15] L. de Haan and H. Rootzén, On the estimation of high quantiles. *J. Statist. Plann. Infer.* **35** (1993) 1-13.

[16] J. Hosking and J. Wallis, Parameter and quantile estimation for the Generalized Pareto Distribution. *Technometrics* **29** (1987) 339-349.

[17] J. Pickands III, Statistical inference using extreme order statistics. *Ann. Statist.* **3** (1975) 119-131.

[18] G.R. Shorack and J.A. Wellner, *Empirical Processes with Applications to Statistics*. Wiley, New York (1986).

[19] R.L. Smith, Estimating tails of probability distributions. *Ann. Statist.* **15** (1987) 1174-1207.

[20] R. Worms, *Vitesses de convergence pour l'approximation des queues de distributions*. Thèse de doctorat de l'Université de Marne-la-Vallée (2000).

[21] R. Worms, Penultimate approximation for the distribution of the excesses. *ESAIM: P&S* **6** (2002) 21-31.