REVSTAT – Statistical Journal
Volume 8, Number 1, June 2010, 21–35

A FOLDING METHOD FOR EXTREME QUANTILES ESTIMATION

Authors: Armelle Guillou
– Institut de Recherche Mathématique Avancée,
UMR 7501, Université de Strasbourg et CNRS,
7 rue René Descartes, 67084 Strasbourg Cedex, France
guillou@math.u-strasbg.fr

Philippe Naveau
– Laboratoire des Sciences du Climat et de l'Environnement,
Orme des Merisiers / Bat.701 C.E. Saclay, 91191 Gif-sur-Yvette, France
philippe.naveau@lsce.ipsl.fr

Alexandre You
– Université Paris VI, LSTA,
Boîte 158, 175 rue du Chevaleret, 75013 Paris, France
alexandreyou@yahoo.fr

Received: June 2009 Revised: February 2010 Accepted: March 2010

Abstract:

• In order to estimate extreme quantiles from independent and identically distributed random variables, we propose and study a novel folding procedure that improves quantile estimates obtained from the classical Peaks-Over-Threshold method (POT) used in Extreme Value Theory. The idea behind the folding approach is to connect the part of a distribution above a given threshold with the one below it. A simplified version of this approach was studied by You *et al.* (2010). In this paper, an extension based on two thresholds is proposed to better combine the folding scheme with the POT approach. Simulations indicate that this new strategy leads to improved extreme quantiles estimates for finite samples. Asymptotic normality of the folded POT estimators is also derived.

Key-Words:

• *extreme quantile estimation; peaks-over-thresholds; generalized Pareto distribution; folding; generalized probability-weighted moments estimators.*

AMS Subject Classification:

• 62G32, 62G20.

1. MOTIVATION

The study of extremes has grown steadily since the pioneering work of Fisher & Tippett (1928). One of the most famous approaches in Extreme Value Theory is the Peaks-Over-Threshold (POT) method which can be described as follows. Let $\aleph := \{X_1, ..., X_n\}$ be a sample of independent random variables from an unknown distribution function F and consider the N_{u_n} exceedances above a fixed threshold u_n, that is $Y_1, ..., Y_{N_{u_n}}$ where $Y_j := X_{i_j} - u_n$, when $X_{i_j} > u_n$. According to Pickands (1975), for a large class of underlying distributions F, as the threshold u_n increases, the distribution of the exceedances F_{u_n} given by $F_{u_n}(t) := \mathbb{P}\big(X - u_n \le t \,|\, X > u_n\big)$ asymptotically converges to a Generalized Pareto Distribution (GPD) defined as

$$G_{\gamma,\sigma_n}(x) \;=\; \begin{cases} 1 - \big(1 + \gamma\, x/\sigma_n\big)^{-1/\gamma} & \text{if } \gamma \neq 0\,, \\ 1 - \exp\big(-x/\sigma_n\big) & \text{if } \gamma = 0\,, \end{cases}$$

where $\sigma_n = \sigma(u_n) > 0$ and $x \ge 0$ if $\gamma \ge 0$ and $0 \le x < -\sigma_n/\gamma$ if $\gamma < 0$. This result leads to the so-called POT estimator of a high quantile $x_p := F^{\leftarrow}(1-p)$ with $F^{\leftarrow}$ the inverse function of F

$$\widehat{x}_p(u_n) \;=\; u_n + \frac{\widehat{\sigma}}{\widehat{\gamma}} \left\{ \left(\frac{p}{1 - F_n(u_n)} \right)^{-\widehat{\gamma}} - 1 \right\}, \tag{1.1}$$

where $(\widehat{\gamma}, \widehat{\sigma})$ are some estimators of the parameters (γ, σ_n) and $F_n(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \le x\}$ denotes the empirical distribution function.

Recently, You *et al.* (2010) proposed a so-called folding procedure to improve the estimation of x_p. This approach is inspired by perfect sampling techniques used in simulation studies (Corcoran & Schneider, 2003) and the idea is to connect the lower and upper parts of a distribution. More precisely the explicit formulation of this folding transformation and its fundamental property are encapsulated in their Proposition 1 which is recalled below.

Proposition 1.1. *Let X be a random variable with an absolutely continuous distribution function F, u a real number such that $u < \tau_F$ where $\tau_F = \sup\big\{x \in \mathbb{R} : F(x) < 1\big\}$ is the right endpoint of F and H another absolutely continuous distribution function with the same support as F and such that $H(u) \ge F(u)$. Define the following random variable*

$$X^{(H,F)}(u) := \begin{cases} H^{\leftarrow}\left(\dfrac{\overline{F}(u)}{F(u)}\, F(X) + F(u) \right) & \text{if } X < u\,, \\ X & \text{if } X \ge u\,, \end{cases} \tag{1.2}$$

where $\overline{F} := 1 - F$ and $H^{\leftarrow}$ is the inverse function of H. Then

$$\mathbb{P}\big(X^{(H,F)}(u) > x\big) \;=\; \mathbb{P}\big(X > x | X > u\big) + \frac{F(u)}{\overline{F}(u)} \big(\overline{H}(x) - \overline{F}(x)\big)\,, \quad x > u\,. \tag{1.3}$$

A very important special case occurs when H is chosen to be equal to F in (1.2). In this context, the random variable $X^{(F,F)}(u)$ has the same probability distribution as the conditional variable $\big[X \,|\, X > u\big]$, the latter being the variable of interest for the aforementioned POT method. We call $X^{(F,F)}(u)$ the folded transformation of X and we denote it as $X^{(F)}(u) := X^{(F,F)}(u)$. In practice, F is unknown and the folding transformation cannot be applied directly. One must substitute the unknown F by suitable proxies. The choice of a proxy is especially sensitive for the inverse function $F^{\leftarrow}$. This explains the introduction of H in the definition of $X^{(H,F)}(u)$. To study the effect of choosing the proxy H instead of F in the folding procedure, we introduce the difference $\Delta_u^{(H,F)}(x) := \big|\mathbb{P}\big(X^{(H,F)}(u) \le x\big) - \mathbb{P}\big(X^{(F,F)}(u) \le x\big)\big|$ for $x > u$. According to Proposition 1.1, we can write

$$\Delta_u^{(H,F)}(x) \;=\; \frac{F(u)}{\overline{F}(u)}\,\big|\overline{H}(x) - \overline{F}(x)\big|\,.$$

Second-order extreme value theory (e.g. de Haan & Ferreira, 2006) provides the necessary tools to characterize the behavior of $\Delta_u^{(H,F)}(x)$ for a specific H.

Proposition 1.2. *Assume that F satisfies the following second-order condition. There exists some positive function $a(\cdot)$ and some positive or negative function $A(\cdot)$ with $\lim_{t\to\infty} A(t) = 0$ such that*

$$\lim_{t\to\infty} \frac{1}{A(t)} \left(\frac{U(tx) - U(t)}{a(t)} - D_\gamma(x)\right) = B(x)\ , \qquad x > 0\ , \tag{1.4}$$

where $U := (1/\overline{F})^{\leftarrow}$, $D_\gamma(x) := \frac{x^\gamma - 1}{\gamma}$ if $\gamma \neq 0$ and $\log x$ if $\gamma = 0$, and B is some function that is not a multiple of D_γ. If the tail distribution $\overline{H}(x) := 1 - H(x)$ is chosen to behave as a GPD tail such that

$$\overline{H}(x) \;=\; \overline{F}(u)\,\overline{G}_{\gamma,\sigma(u)}(x - u)\ , \tag{1.5}$$

with $x > u$ and $\sigma(u) = a\big(1/\overline{F}(u)\big)$, then for all y satisfying $1 + \gamma y > 0$, we have

$$\begin{aligned} \lim_{u\to\tau_F} \frac{1}{|\alpha(u)|}\,\Delta_u^{(H,F)}\big(u + \sigma(u)\,y\big) \;&=\; \lim_{u\to\tau_F} \left|\frac{1}{\alpha(u)}\left(\frac{\overline{F}\big(u + \sigma(u)\,y\big)}{\overline{F}(u)} - \overline{G}_\gamma(y)\right)\right| \\ &=\; \big(\overline{G}_\gamma(y)\big)^{1+\gamma}\,\big|B_{\gamma,\rho}\big(1/\overline{G}_\gamma(y)\big)\big| \end{aligned} \tag{1.6}$$

where $\overline{G}_\gamma := \overline{G}_{\gamma,1}$, $\alpha(u) := A\big(1/\overline{F}(u)\big)$ and

$$B_{\gamma,\rho}(x) := \begin{cases} \dfrac{1}{\rho}\left(\dfrac{x^{\gamma+\rho} - 1}{\gamma+\rho} - \dfrac{x^\gamma - 1}{\gamma}\right) & \text{if } \gamma \neq 0,\ \rho \neq 0\,, \\[2ex] \dfrac{1}{\gamma}\left(x^\gamma \log x - \dfrac{x^\gamma - 1}{\gamma}\right) & \text{if } \rho = 0 \neq \gamma\,, \\[2ex] \dfrac{1}{\rho}\left(\dfrac{x^\rho - 1}{\rho} - \log x\right) & \text{if } \rho \neq 0 = \gamma\,, \\[2ex] \dfrac{1}{2}\,(\ln x)^2 & \text{if } \rho = \gamma = 0\,. \end{cases}$$

The first equality in (1.6) tells us that choosing H as a GPD approximation, see (1.5), implies that the rate of convergence of $\Delta_u^{(H,F)}$ towards zero is identical to the one obtained by working with exceedances. The second equality in (1.6) simply restates the result derived by de Haan & Ferreira (2006, p. 48) about the relationship between the rate of convergence and the second-order auxiliary function $A(\cdot)$. The main consequence of Proposition 1.2 is that a GPD can be viewed as the appropriate choice for the distribution function H. In real applications, we do not know the parameters of such a GPD and a first estimation has to be given before implementing our folding procedure. This also means that any reasonable GPD estimation procedure can be used to initialize our algorithm, the better the estimation of $\sigma(u)$ and γ, the better the efficiency of the folding procedure. Still, our main goal in this paper is not to compare all existing GPD estimation methods (e.g., Smith, 1987; Greenwood *et al.*, 1979) and to find the best one (if one could do that). Instead, our aim is to study our folding approach with a specific estimation method for which we have experience with (Diebolt *et al.*, 2004, 2007).

At this stage, our approach can be viewed as the mixing of two elements, the folding procedure described by Proposition 1.1 and the POT method. Each element is associated with a particular threshold choice. For the sake of simplicity, You *et al.* (2010) considered that both thresholds were equal. This is not necessary. One threshold could be chosen for computing the preliminary GPD parameters estimates and another one for the folding transformation itself. In this paper, we follow this path. We propose and study a novel folding approach based on two thresholds u_n and u'_n such that $u_n = o(u'_n)$. Compared to a conventional approach and to our past folding procedure, simulations clearly indicate that this new double-threshold folding approach significantly reduces the mean squared error of extreme quantile estimates, particularly for small and moderate sample sizes (see Section 4). Asymptotic properties of our GPD parameters estimators are derived (see Section 3). The proof of our results are postponed to the appendix. Results presented in Sections 3 and 4 solely focus on heavy tailed distributions because our previous study (You *et al.*, 2010) indicated that the folding gain is the strongest for this type of tails.

2. A NEW FOLDING PROCEDURE WITH TWO THRESHOLDS

Suppose that the variable $\big[X - u_n | \, X > u_n\big]$ approximatively follows a $\mathrm{GPD}(\gamma, \sigma_n)$ for some large threshold u_n. The thresholding stability property of the GPD basically means that $\big[X - u'_n | \, X > u'_n\big]$ can also be approximated by a $\mathrm{GPD}\big(\gamma,\, \sigma_n + \gamma(u'_n - u_n)\big)$ for any $u'_n > u_n$. In other words, the tail $\overline{F}(t) = \overline{F}_{u'_n}(t - u'_n)\, \overline{F}(u'_n)$ can be approximated by $\overline{G}_{\gamma, \sigma_n + \gamma(u'_n - u_n)}(t - u'_n)\, \overline{F}(u'_n)$. In terms of inverse distributions, this approximation can be expressed as $\overline{F}^{\leftarrow}(t) \simeq$

$\overline{G}^{\leftarrow}_{\gamma,\sigma_n+\gamma(u'_n-u_n)}\Big(\frac{t}{\overline{F}(u'_n)}\Big)+u'_n$. According to Proposition 1.1, the folded variable $X^{(F)}(u'_n)$ can be rewritten as

$$X^{(F)}(u'_n) = \overline{F}^{\leftarrow}\left(\overline{F}(u'_n)\left[1-\frac{F(X)}{F(u'_n)}\right]\right), \qquad \text{if } X < u'_n .$$

By plugging the approximation for $\overline{F}^{\leftarrow}$ in the expression of $X^{(F)}(u'_n)$, it is natural to define the following folded variables

$$\widehat{X}^{(F)}_i(u'_n) = \begin{cases} \overline{G}^{\leftarrow}_{\widehat{\gamma},\widehat{\sigma}'}\left(1-\dfrac{F_n(X_i)}{F_n(u'_n)}\right)+u'_n, & \text{if } F_n(X_i) < F_n(u'_n), \\ X_i, & \text{if } F_n(X_i) \geq F_n(u'_n), \end{cases} \tag{2.1}$$

where $\widehat{\widehat{\sigma}}' := \widehat{\sigma}+\widehat{\gamma}(u'_n-u_n)$ and $(\widehat{\gamma},\widehat{\sigma})$ are estimated from the exceedances $(Y_1,...,Y_{N_{u_n}})$. Note that the folding transformation given by (2.1) does not depend on the numerical values of the observations X_i when $X_i < u'_n$ but only on their ranks because $nF_n(X_i)$ is equal to the rank of the observation X_i.

Equation (2.1) allows us to describe our new folding procedure as follows:

Step 1. Select one threshold u_n and estimate $(\widehat{\gamma},\widehat{\sigma})$ of the GPD parameters (γ,σ_n) from the exceedances above u_n. Select a second threshold $u'_n > u_n$ and calculate $\widehat{\widehat{\sigma}}' := \widehat{\sigma}+\widehat{\gamma}(u'_n-u_n)$.

Step 2. Build the folded version $\aleph^{(F)} := \left\{\widehat{X}^{(F)}_1(u'_n),...,\widehat{X}^{(F)}_n(u'_n)\right\}$ using transformation (2.1).

Step 3. Estimate the GPD parameters $\big(\widehat{\gamma}^{(F)},\widehat{\sigma}^{(F)}\big)$ from the folded sample $\aleph^{(F)}$.

Step 4. Compute the POT extreme quantile estimator $\widehat{x}^{(F)}_p(u'_n)$ (according to (1.1)) with the estimates $\big(\widehat{\gamma}^{(F)},\widehat{\sigma}^{(F)}\big)$.

In steps 1 and 3, any reasonable GPD estimator of (γ,σ_n) could be used. Here we implement the generalized probability-weighted moments (GPWM) (Diebolt *et al.*, 2004, 2007). This is an extension of the classical probability-weighted moments method (Greenwood *et al.*, 1979) and it can be described as follows. Let ω be a continuous function, null at zero, and which admits a right derivative at zero. The GPWM is defined as $\nu_\omega = \mathbb{E}\big[Z\,\omega\big(\overline{G}_{\gamma,\sigma}(Z)\big)\big]$ where Z follows a GPD(γ,σ) with $\gamma < 2$. If we denote by W the primitive of ω null at zero, then an integration by parts allows us to write ν_ω as $\nu_\omega = \int_0^\infty W\big(\overline{G}_{\gamma,\sigma}(x)\big)\,dx$. Diebolt *et al.* (2004, 2007) proposed and studied $\widehat{\nu}_{\omega,n} = \int_0^\infty W\big(\overline{F}_{n,u_n}(x)\big)\,dx$ as an estimator of ν_ω, where $\overline{F}_{n,u_n}$ corresponds to the exceedances empirical survival function defined by $\overline{F}_{n,u_n}(x) = \frac{1}{N_{u_n}}\sum_{i=1}^{N_{u_n}} \mathbb{1}\big\{X_i-u_n > x\big\}$. To estimate (γ,σ_n) by implementing a method-of-moments approach, two GPWMs are needed. In this work, the two functions ω_1 and ω_2 are equal to $\omega_1(x) = x$ and $\omega_2(x) = x^{3/2}$ (see Diebolt *et al.*, 2004, for a justification of these choices).

3. ASYMPTOTIC NORMALITY

Before stating our main result, we need to prove the asymptotic normality of the pair $(\widehat{\gamma}, \widehat{\widehat{\sigma}}'/\sigma'_n)$.

Lemma 3.1. *Let F be three times differentiable such that its inverse $F^{\leftarrow}$ exists. Let V and M be two functions defined as $V(t) = \overline{F}^{\leftarrow}(e^{-t})$ and $M(t) = \frac{V''(\ln t)}{V'(\ln t)} - \gamma$. Suppose the following conditions hold*

$$M \text{ is of constant sign at } \infty \tag{3.1}$$

and

$$\exists\, \rho < 0:\ |M| \in RV_\rho \text{ with a normalized slowly varying function (see Bingham et al., 1987)}. \tag{3.2}$$

Then, for $\gamma \in (0, 3/2)$ and for all C^1-functions ω_1 and ω_2, null at 0, conditionally on $\{N_{u_n} = k_n\}$ and $\{N_{u'_n} = k'_n\}$ with $u_n = o(u'_n)$, and for all intermediate sequences $k_n > k'_n \to \infty$ such that $\sqrt{k_n}\, a_n \to \lambda \in \mathbb{R}$, we have that

$$\sqrt{k_n} \begin{pmatrix} \widehat{\gamma} - \gamma \\ \frac{\widehat{\widehat{\sigma}}'}{\sigma'_n} - 1 \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\lambda \begin{pmatrix} 1 \\ \frac{1}{\gamma} \end{pmatrix} B_1,\ \begin{pmatrix} 1 \\ \frac{1}{\gamma} \end{pmatrix} \Sigma_{1,1} \begin{pmatrix} 1 \\ \frac{1}{\gamma} \end{pmatrix}'\right),$$

where

$$\begin{cases} a_n := M\left(e^{V^{\leftarrow}(u_n)}\right), \qquad \sigma'_n := V'\left(V^{\leftarrow}(u'_n)\right), \\ \mathbf{B} := \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} = \mathbf{A}\,\mathbf{C} \qquad \text{where } \mathbf{A} := DT_{(\omega_1,\omega_2)}(\nu^1_{\omega_1}, \nu^1_{\omega_2}), \\ \mathbf{C} := \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} \frac{\phi_1(\gamma+\rho)-\phi_1(\gamma)}{\rho} \\ \frac{\phi_2(\gamma+\rho)-\phi_2(\gamma)}{\rho} \end{pmatrix} \quad \text{where } \phi_j(\gamma) := \int_0^1 W_j(u)\, u^{-\gamma-1}\, du,\ \ j \in \{1,2\}, \\ \mathbf{\Sigma} := \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{1,2} & \Sigma_{2,2} \end{pmatrix} = \mathbf{A}\,\mathbf{\Gamma}\,\mathbf{A}^{\mathrm{T}}, \end{cases}$$

and $\mathbf{\Gamma}$ is the variance-covariance matrix of the pair $(Y_{\omega_1}, Y_{\omega_2})$ defined as

$$Y_{\omega_1} = \int_0^1 t^{-\gamma-1}\, \omega_1(t)\, \mathbb{B}(t)\, dt \qquad \text{and} \qquad Y_{\omega_2} = \int_0^1 t^{-\gamma-1}\, \omega_2(t)\, \mathbb{B}(t)\, dt$$

with $\mathbb{B}$ a Brownian bridge on $[0,1]$.

The case $\rho = 0$ is excluded from Lemma 3.1. This corresponds to $M(t) = \ell(t)$, a slowly varying function, since in that case the limiting distribution depends explicitely on $\ell(\cdot)$, and the two sequences u_n and u'_n. Also this restriction is not really a problem since most of the classical distributions in the Fréchet domain of attraction have a second order parameter $\rho < 0$ (except the loggamma).

Now, we can establish our main asymptotic result which shows that, in the case where $\gamma \in (0, 3/2)$, the estimators based on the double-threshold folding approach have a similar asymptotic normality as the one derived in You *et al.* (2010) in case of one threshold.

Theorem 3.1. *Under the same assumptions stated in Lemma 3.1, we have*

$$\sqrt{k_n}\begin{pmatrix} \dfrac{\widehat{\nu}^{(F)}_{\omega_1,n}}{\sigma'_n} - \nu^1_{\omega_1} \\ \dfrac{\widehat{\nu}^{(F)}_{\omega_2,n}}{\sigma'_n} - \nu^1_{\omega_2} \end{pmatrix} \xrightarrow{d} \mathcal{N}\big(\lambda\, B_1 \mathbb{F},\, \mathbb{F}\,\Sigma_{1,1}\mathbb{F}'\big)$$

where

$$\mathbb{F} := \begin{pmatrix} -\dfrac{1}{\lambda}\displaystyle\int_0^1 u^{-\gamma}\ln u\,\omega_1(u)\,du \\ -\dfrac{1}{\lambda}\displaystyle\int_0^1 u^{-\gamma}\ln u\,\omega_2(u)\,du \end{pmatrix}$$

and

$$\nu^1_{\omega_j} = \int_0^\infty W_j\big(\overline{G}_{\gamma,1}(x)\big)\,dx \qquad \textit{for} \;\; j = 1, 2\,.$$

Note that this convergence in distribution does not hold in case $\gamma \le 0$.

4. A SIMULATION STUDY

The aim of this section is to illustrate the superiority of the double-threshold folding over the conventional (Diebolt *et al.*, 2007) and the simple folding approaches (You *et al.*, 2010), in particular in terms of the mean squared error for small and moderate sample sizes. Simulations were performed for four sample sizes $n = 100$, 500, 1 000 and 5 000 from a Burr $(1, 2, 0.5)$ distribution defined by $\overline{F}(x) = (1+x^2)^{-1/2}$ and from a standard Fréchet distribution defined by $\overline{F}(x) = 1 - e^{-1/x}$, respectively. For these two distributions, $\gamma = 1$ and $\rho < 0$. For each value of n, 5 000 samples were generated and k_n was chosen such that the condition $\sqrt{k_n}\,a_n \to \lambda$ was satisfied, which corresponds to $k_n \simeq c_1 n^{4/5}$ for the Burr distribution and to $k_n \simeq c_2 n^{2/3}$ for the Fréchet distribution. The threshold u_n was chosen from $u_n = F^{\leftarrow}\big(1 - \frac{k_n}{n}\big)$. Three return levels for three return periods, $t = 100, 200$ and 1 000, were computed. Concerning the choice of the second threshold for our double-threshold folding method, we selected $k'_n = c_3\, n^{3/5}$ for the Burr distribution and $k'_n = c_4\, n^{1/2}$ for the Fréchet distribution such that $u_n = o(u'_n)$. Tables 1–3 and 4–6 display the bias and the root mean squared error (RMSE) of the quantile x_p for the Burr and Fréchet distributions, respectively.

Table 1: Burr(1, 2, 0.5) distribution — Bias (RMSE) of the return level estimates corresponding to a return period $t = 100$.

n	k_n	k'_n	Conventional	Folding with one threshold	Folding with two thresholds
100	80	16	35.7 (240.4)	10.7 (183.0)	1.6 (103.8)
500	288	42	16.4 (195.4)	4.0 (106.2)	1.5 (46.7)
1 000	502	63	8.5 (78.8)	2.3 (54.1)	1.3 (35.3)
5 000	1 820	166	2.6 (21.1)	1.1 (16.8)	0.3 (16.9)

Table 2: Same as Table 1 but for the return period $t = 200$.

n	k_n	k'_n	Conventional	Folding with one threshold	Folding with two thresholds
100	80	16	117.0 (840.1)	45.6 (626.3)	11.2 (299.2)
500	288	42	54.2 (752.4)	15.9 (395.7)	6.9 (140.9)
1 000	502	63	27.1 (297.9)	8.1 (192.9)	5.1 (112.7)
5 000	1 820	166	6.7 (67.1)	2.1 (49.4)	0.8 (40.3)

Table 3: Same as Table 1 but for the return period $t = 1\,000$.

n	k_n	k'_n	Conventional	Folding with one threshold	Folding with two thresholds
100	80	16	1 877.5 (16 252.5)	998.0 (11 748.4)	346.5 (3 785.8)
500	288	42	955.4 (17 449.2)	385.8 (8 783.8)	167.4 (2 029.8)
1 000	502	63	468.8 (6 702.8)	199.6 (3 890.0)	119.1 (1 820.5)
5 000	1 820	166	73.9 (990.5)	22.2 (589.4)	19.1 (504.8)

Table 4: Standard Fréchet distribution — Bias (RMSE) of the return level estimates corresponding to the return period $t = 100$.

n	k_n	k'_n	Conventional	Folding with one threshold	Folding with two thresholds
100	64	15	48.8 (173.2)	15.4 (112.1)	15.0 (93.9)
500	188	33	17.2 (70.9)	7.5 (50.6)	8.5 (47.1)
1 000	300	47	12.2 (48.5)	6.8 (38.0)	6.8 (36.2)
5 000	877	106	3.5 (17.3)	2.5 (15.7)	1.6 (13.9)

Table 5: Same as Table 4 but for the return period $t = 200$.

n	k_n	k'_n	Conventional	Folding with one threshold	Folding with two thresholds
100	64	15	158.0 (581.7)	56.4 (353.3)	46.3 (286.3)
500	188	33	55.9 (242.7)	25.5 (156.0)	27.7 (140.8)
1 000	300	47	39.5 (163.3)	22.3 (117.2)	23.0 (111.0)
5 000	877	106	11.8 (57.6)	8.6 (49.6)	6.8 (42.0)

Table 6: Same as Table 4 but for the return period $t = 1\,000$.

n	k_n	k'_n	Conventional	Folding with one threshold	Folding with two thresholds
100	64	15	2 304.7 (10 052.4)	935.6 (5 396.5)	638.6 (3 214.3)
500	188	33	785.8 (4 459.5)	367.6 (2 206.0)	370.6 (1 845.0)
1 000	300	47	536.3 (2 789.3)	298.5 (1 580.8)	309.6 (1 480.8)
5 000	877	106	151.0 (899.2)	110.1 (664.6)	99.4 (549.3)

These tables clearly show that the double-threshold folding improves considerably the RMSE, compared to the single-threshold folding and the conventional approach. This gain is emphasized for small and moderate sample sizes and for large return periods.

APPENDIX: DETAILED PROOFS

Proof of Proposition 1.2: The proof is mainly a consequence of Theorem 2.3.8 in de Haan & Ferreira (2006), which states that, if (1.4) holds, then

$$\lim_{u\to\tau_F} \frac{1}{\alpha(u)} \left(\frac{\overline{F}\big(u+\sigma(u)\,y\big)}{\overline{F}(u)} - \overline{G}_\gamma(y) \right) = \big(\overline{G}_\gamma(y)\big)^{1+\gamma} B_{\gamma,\rho}\left(\frac{1}{\overline{G}_\gamma(y)}\right)$$

for all y such that $1+\gamma\, y > 0$.

From (1.5), it follows that

$$\begin{aligned}
\lim_{u\to\tau_F} \frac{\Delta_u^{(H,F)}\big(u+\sigma(u)\,y\big)}{|\alpha(u)|} &= \lim_{u\to\tau_F} F(u) \times \lim_{u\to\tau_F} \left| \frac{1}{\alpha(u)} \left(\overline{G}_\gamma(y) - \frac{\overline{F}\big(u+\sigma(u)\,y\big)}{\overline{F}(u)} \right) \right| \\
&= \big(\overline{G}_\gamma(y)\big)^{1+\gamma} \left| B_{\gamma,\rho}\left(\frac{1}{\overline{G}_\gamma(y)}\right) \right|. \qquad \square
\end{aligned}$$

Proof of Lemma 3.1: First, note that the assumption $\sqrt{k_n}\, a_n \to \lambda \in \mathbb{R}$ can be rewritten as

$$\sqrt{k_n}\, M\left(\frac{1}{\overline{F}(u_n)}\right) \longrightarrow \lambda \in \mathbb{R}\,. \tag{A.1}$$

Now, let $\sigma_n = V'\big(V^{\leftarrow}(u_n)\big)$. We have

$$\begin{aligned}
\sqrt{k_n}\left(\frac{\widehat{\sigma}'}{\sigma'_n} - 1\right) &= \sqrt{k_n}\left(\frac{\widehat{\sigma} + \widehat{\gamma}(u'_n - u_n)}{\sigma'_n} - 1\right) \\
&= \sqrt{k_n}\left(\frac{\widehat{\sigma}}{\sigma_n} - 1\right)\frac{\sigma_n}{\sigma'_n} + \left[\frac{\gamma}{\widehat{\gamma}}\,\sqrt{k_n}\left(\frac{\sigma_n}{\gamma\, u_n} - 1\right) - \frac{\sqrt{k_n}}{\widehat{\gamma}}\,(\widehat{\gamma} - \gamma)\right]\frac{\widehat{\gamma}\, u_n}{\sigma_n}\,\frac{\sigma_n}{\sigma'_n} \\
&\quad + \sqrt{k_n}\left(\frac{\gamma\, u'_n}{\sigma'_n} - 1\right)\frac{\widehat{\gamma}}{\gamma} + \frac{1}{\gamma}\,\sqrt{k_n}\,(\widehat{\gamma} - \gamma) \\
&=: Q_{1,n} + Q_{2,n} + Q_{3,n} + \frac{1}{\gamma}\,\sqrt{k_n}\,(\widehat{\gamma} - \gamma)\,.
\end{aligned}$$

We know that $\sqrt{k_n}\left(\frac{\widehat{\sigma}}{\sigma_n} - 1\right)$ is asymptotically normal (Diebolt *et al.*, 2007) and

$$\frac{\sigma_n}{\sigma'_n} \sim \frac{\gamma\, u_n}{\gamma\, u'_n} \longrightarrow 0\,. \tag{A.2}$$

Therefore, it is clear that

$$Q_{1,n} \xrightarrow{\mathbb{P}} 0\,. \tag{A.3}$$

Now, remark that

$$\begin{aligned}
\sqrt{k_n}\left(\frac{\sigma_n}{\gamma\, u_n} - 1\right) &= \sqrt{k_n}\left(\frac{V'\big(-\ln \overline{F}(u_n)\big)}{\gamma\, V\big(-\ln \overline{F}(u_n)\big)} - 1\right) \\
&= \frac{1}{\gamma}\,\sqrt{k_n}\, M\left(\frac{1}{\overline{F}(u_n)}\right)\left[\frac{M\left(\frac{1}{\overline{F}(u_n)}\right)}{\frac{V'\left(-\ln \overline{F}(u_n)\right)}{V\left(-\ln \overline{F}(u_n)\right)} - \gamma}\right]^{-1}.
\end{aligned}$$

To conclude with this term, we have to use the following lemma.

Lemma A.1 (Worms, 2000, p. 19)**.** *Suppose that $M(t) \to 0$ and $\frac{tM'(t)}{M(t)} \to \rho$ as $t \to \infty$. Then*

(i) *if $\gamma > 0$, we have*

$$\lim_{t\to\infty} M(e^t)\Big/\left[\frac{V'(t)}{V(t)} - \gamma\right] = \frac{\gamma + \rho}{\gamma}$$

and

$$\lim_{t\to\infty} M(e^t)\Big/\left[\frac{V(t)}{V'(t)} - \frac{1}{\gamma}\right] = -\gamma(\gamma + \rho)\,;$$

(ii) *if* $\gamma < 0$, *we have*

$$\lim_{t\to\infty}\left[\frac{V(\infty)-V(t)}{V'(t)}+\frac{1}{\gamma}\right]\Big/M(e^t) = \frac{1}{\gamma(\gamma+\rho)}$$

and

$$\lim_{t\to\infty}\left[\frac{V'(t)}{V(\infty)-V(t)}+\gamma\right]\Big/M(e^t) = -\frac{\gamma}{\gamma+\rho}\,.$$

Indeed by (A.1), we deduce that

$$\sqrt{k_n}\left(\frac{\sigma_n}{\gamma\,u_n}-1\right) \longrightarrow \frac{\lambda}{\gamma+\rho}\,.$$

Combining this convergence with (A.2) and the fact that $\sqrt{k_n}\,(\widehat{\gamma}-\gamma)$ is asymptotically normal (Diebolt *et al.*, 2007), we deduce that

$$Q_{2,n} \xrightarrow{\mathbb{P}} 0\,. \tag{A.4}$$

Similarly

$$\begin{aligned}\sqrt{k_n}\left(\frac{\gamma\,u'_n}{\sigma'_n}-1\right) &= \gamma\,\sqrt{k_n}\left(\frac{V\big(-\ln\overline{F}(u'_n)\big)}{V'\big(-\ln\overline{F}(u'_n)\big)}-\frac{1}{\gamma}\right)\\ &= \gamma\,\sqrt{k_n}\;M\!\left(\frac{1}{\overline{F}(u_n)}\right)\frac{M\Big(\frac{1}{\overline{F}(u'_n)}\Big)}{M\Big(\frac{1}{\overline{F}(u_n)}\Big)}\left[\frac{M\Big(\frac{1}{\overline{F}(u'_n)}\Big)}{\frac{V(-\ln\overline{F}(u'_n))}{V'(-\ln\overline{F}(u'_n))}-\frac{1}{\gamma}}\right]^{-1}.\end{aligned}$$

Now since $\gamma > 0$ and $|M| \in RV_\rho$ with $\rho < 0$

$$\frac{M\Big(\frac{1}{\overline{F}(u'_n)}\Big)}{M\Big(\frac{1}{\overline{F}(u_n)}\Big)} \longrightarrow 0\,. \tag{A.5}$$

Consequently, using again the abovementioned lemma in Worms (2000), we deduce that

$$Q_{3,n} \xrightarrow{\mathbb{P}} 0\,. \tag{A.6}$$

Finally, going back to (A.3), (A.4) and (A.6), we get

$$\sqrt{k_n}\left(\frac{\widehat{\sigma}'}{\sigma'_n}-1\right) = \frac{1}{\gamma}\,\sqrt{k_n}\,(\widehat{\gamma}-\gamma)+o_{\mathbb{P}}(1)\,.$$

Lemma 3.1 then follows from Diebolt *et al.* (2007). □

Proof of Theorem 3.1: It is a direct application of the proof of Theorem 1 in You *et al.* (2010) combining with our Lemma 3.1 and using the following decomposition: conditionally on $\{N_{u_n} = k_n\}$ and $\{N_{u'_n} = k'_n\}$, we have

$$\begin{aligned}
&\sqrt{k_n}\left(\frac{\widehat{\nu}^{(F)}_{\omega,n}}{\sigma'_n} - \int_0^\infty W\big(\overline{G}_\gamma(x)\big)\,dx\right) = \\
&= \sqrt{k_n}\left(\frac{\widehat{\nu}^{(F)}_{\omega,n}}{\sigma'_n} - \frac{1}{\sigma'_n}\int_0^\infty W\Big(\overline{\widetilde{F}}{}^{(F)}_{n,u'_n}(x)\Big)\,dx\right) \\
&\quad + \left(1-\frac{k'_n}{n}\right)\sqrt{k_n}\int_0^\infty\Big(\overline{G}_{\widehat{\gamma},\frac{\widehat{\widehat{\sigma}}'}{\sigma'_n}}(x) - \overline{G}_\gamma(x)\Big)\,\omega\big(\overline{G}_\gamma(x)\big)\,dx \\
&\quad + \frac{k'_n}{n}\sqrt{k_n}\int_0^\infty\Big(\overline{F}_{n,u'_n}(\sigma'_n x) - \overline{G}_\gamma(x)\Big)\,\omega\big(\overline{G}_\gamma(x)\big)\,dx \\
&\quad + \sqrt{k_n}\int_0^\infty\!\!\int_0^1(1-t)\left(\left(1-\frac{k'_n}{n}\right)\Big(\overline{G}_{\widehat{\gamma},\frac{\widehat{\widehat{\sigma}}'}{\sigma'_n}}(x) - \overline{G}_\gamma(x)\Big) + \frac{k'_n}{n}\Big(\overline{F}_{n,u'_n}(\sigma'_n x) - \overline{G}_\gamma(x)\Big)\right)^2 \\
&\quad\times\omega'\left(\overline{G}_\gamma(x) + \left(\left(1-\frac{k'_n}{n}\right)\Big(\overline{G}_{\widehat{\gamma},\frac{\widehat{\widehat{\sigma}}'}{\sigma'_n}}(x) - \overline{G}_\gamma(x)\Big) + \frac{k'_n}{n}\Big(\overline{F}_{n,u'_n}(\sigma'_n x) - \overline{G}_\gamma(x)\Big)\right)t\right)dt\,dx\,,
\end{aligned}$$

where $\overline{\widetilde{F}}{}^{(F)}_{n,u'_n}(x) = \left(1-\frac{k'_n}{n}\right)\overline{G}_{\widehat{\gamma},\widehat{\widehat{\sigma}}'}(x) + \frac{k'_n}{n}\overline{F}_{n,u'_n}(x)$.

All the details of the proof are given on the web page `http://www-irma.u-strasbg.fr/~guillou/Proof_folding_thm3-1.pdf`. Now, we will prove that our theorem does not hold in case $\gamma \le 0$. Indeed if $\gamma < 0$, then

$$\begin{aligned}
\sqrt{k_n}\left(\frac{\widehat{\widehat{\sigma}}'}{\sigma'_n} - 1\right) &= \frac{\sigma_n}{\sigma'_n}\left\{\sqrt{k_n}\left(\frac{\widehat{\sigma}}{\sigma_n} - 1\right) - \frac{1}{\gamma}\sqrt{k_n}\,(\widehat{\gamma}-\gamma)\,\frac{-\gamma\big(V(\infty)-u_n\big)}{\sigma_n}\right. \\
&\qquad\left. + \gamma\sqrt{k_n}\left(\frac{V(\infty)-u_n}{\sigma_n} + \frac{1}{\gamma}\right)\right\} \\
&\quad + \sqrt{k_n}\left(\frac{-\widehat{\gamma}\big(V(\infty)-u'_n\big)}{\sigma'_n} - 1\right) \\
&=: \frac{\sigma_n}{\sigma'_n}\,T_{1,n} + T_{2,n}\,.
\end{aligned}$$

Clearly $T_{1,n}$ tends in distribution to a normal distribution, but

$$\frac{\sigma_n}{\sigma'_n} \simeq \frac{V(\infty)-u_n}{V(\infty)-u'_n} = 1 + \frac{u'_n\big(1+o(1)\big)}{V(\infty)-u'_n} \to \infty\,.$$

Therefore $\frac{\sigma_n}{\sigma'_n}T_{1,n}$ tends to infinity, whereas $T_{2,n}$ can be rewritten as

$$T_{2,n} = \sqrt{k_n}(\gamma-\widehat{\gamma})\,\frac{V(\infty)-u'_n}{\sigma'_n} - \frac{\sqrt{k_n}}{\gamma+\rho}\,M\big(e^{V^{\leftarrow}(u_n)}\big)\,\frac{M\big(e^{V^{\leftarrow}(u'_n)}\big)}{M\big(e^{V^{\leftarrow}(u_n)}\big)}\,\big(1+o(1)\big)$$

by Lemma 1.2 (ii) in Worms (2000). This implies that $T_{2,n}$ tends to a normal distribution, since it is the case for the first part of this term, whereas the second one tends to 0.

Now, in case $\gamma = 0$, we can easily find a counter-example of our Lemma 3.1. First, note that

$$\begin{aligned}
\sqrt{k_n}\left(\frac{\widehat{\widehat{\sigma}}'}{\sigma_n'} - 1\right) &= \sqrt{k_n}\left(\frac{\widehat{\sigma} + \widehat{\gamma}(u_n' - u_n)}{\sigma_n'} - 1\right) \\
&= \sqrt{k_n}\left(\frac{\widehat{\sigma}}{\sigma_n} - 1\right)\frac{\sigma_n}{\sigma_n'} + \sqrt{k_n}\,(\widehat{\gamma} - \gamma)\,\frac{u_n' - u_n}{\sigma_n'} + \sqrt{k_n}\,\frac{\sigma_n}{\sigma_n'} - \sqrt{k_n} \\
&=: \widetilde{Q}_{1,n} + \widetilde{Q}_{2,n} + \widetilde{Q}_{3,n} - \sqrt{k_n}\,.
\end{aligned}$$

Now, if we consider an exponential random variable with parameter 1, then $\sigma_n' = \sigma_n = 1$ and $\frac{u_n' - u_n}{\sigma_n'} = u_n'\big(1 + o(1)\big)$ by assumption. This implies that $\widetilde{Q}_{1,n}$ is asymptotically normal, $\widetilde{Q}_{2,n} \xrightarrow{\mathbb{P}} \infty$ and $\widetilde{Q}_{3,n} - \sqrt{k_n} = 0$. Therefore Lemma 3.1 is not valid. □

ACKNOWLEDGMENTS

Part of this work was supported by the "Conseil Scientifique" of the University of Strasbourg, by the ANR AssimilEx project and by the FP7 ACQWA project. The authors are grateful to the referee for the constructive comments.

REFERENCES

[1] Bingham, N.H.; Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*, Cambridge University Press, Cambridge.

[2] Corcoran, J.N. and Schneider, U. (2003). Shift and scale coupling methods for perfect simulation, *Probability in the Engineering and Informational Sciences*, **17**, 277–303.

[3] Diebolt, J.; Guillou, A. and Rached, I. (2004). A new look at probability-weighted moments estimators, *Comptes Rendus de l'Académie des Sciences de Paris*, **338**, Série I, 629–634.

[4] Diebolt, J.; Guillou, A. and Rached, I. (2007). Approximation of the distribution of excesses through a generalized probability-weighted moments method, *Journal of Statistical Planning and Inference*, **137**, 841–857.

[5] Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or the smallest member of a sample, *Proceedings of the Cambridge Philosophical Society*, **24**, 180–190.

[6] GREENWOOD, J.A.; LANDWEHR, J.M.; MATALAS, N.C. and WALLIS, J.R. (1979). Probability-weighted moments: definition and relation to parameters of several distributions expressable in inverse form, *Water Resources Research*, **15**, 1049–1054.

[7] DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory*, New York, Springer-Verlag.

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

[9] SMITH, R.L. (1987). Estimating tails of probability distributions, *Annals of Statistics*, **15**, 1174–1207.

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

[11] YOU, A.; SCHNEIDER, U.; GUILLOU, A. and NAVEAU, P. (2010). Improving extreme quantile estimation via a folding procedure, *Journal of Statistical Planning and Inference* (to appear).