Bernoulli **21**(1), 2015, 420–436
DOI: 10.3150/13-BEJ573

Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework

CLÉMENT DOMBRY

Laboratoire de Mathématiques de Besancon, CNRS UMR 6623, Université de Franche-Comté, 16 route de Gray, 25030 Besancon Cedex, France. E-mail: clement.dombry@univ-fcomte.fr

The maximum likelihood method offers a standard way to estimate the three parameters of a generalized extreme value (GEV) distribution. Combined with the block maxima method, it is often used in practice to assess the extreme value index and normalization constants of a distribution satisfying a first order extreme value condition, assuming implicitly that the block maxima are exactly GEV distributed. This is unsatisfactory since the GEV distribution is a good approximation of the block maxima distribution only for blocks of large size. The purpose of this paper is to provide a theoretical basis for this methodology. Under a first order extreme value condition only, we prove the existence and consistency of the maximum likelihood estimators for the extreme value index and normalization constants within the framework of the block maxima method.

Keywords: block maxima method; consistency; extreme value index; maximum likelihood estimator

1. Introduction and results

Estimation of the extreme value index is a central problem in extreme value theory. A variety of estimators are available in the literature, for example among others, the Hill estimator [9], the Pickand's estimator [14], the probability weighted moment estimator introduced by Hosking *et al.* [11] or the moment estimator suggested by Dekkers *et al.* [6]. The monographs by Embrechts *et al.* [8], Beirlant *et al.* [2] or de Haan and Ferreira [5] provide good reviews on this estimation problem. All the above mentioned estimators are explicitly defined and usually, consistency is shown under the domain of attraction condition only while asymptotic normality is more complicated and requires some second order condition. In this paper, we are interested on estimators based on the maximum likelihood method. Let us stress that MLEs are implicitly defined as solutions of the likelihood equations, making existence and consistency a sensible problem, possibly even more difficult than asymptotic normality where the second order condition may ease the proof of existence.

Two different types of maximum likelihood estimators (MLEs) have been introduced, based on the peak over threshold (POT) method and block maxima method respectively. The POT method relies on the fact that, under the extreme value condition, exceedances over high threshold converge to a generalized Pareto distribution (GPD) (see Balkema and de Haan [1]). A MLE within the GPD model has been proposed by Smith [19]. Its theoretical properties under the extreme value condition are quite difficult to analyze due to the absence of an explicit expression of the likelihood equations: existence and consistency have been proven by Zhou [22], asymptotic normality by Drees *et al.* [7]. The block maxima method relies on the approximation of the maxima distribution by a generalized extreme value (GEV) distribution. Computational issues for ML estimation within the GEV model have been considered by Prescott and Walden [15, 16], Hosking [10] and Macleod [12]. Since the support of the GEV distribution depends on the unknown extreme value index γ, the usual regularity conditions ensuring good asymptotic properties are not satisfied. This problem is studied by Smith [18]: asymptotic normality is proven for $\gamma > -1/2$ and consistency for $\gamma > -1$.

It should be stressed that the block maxima method is based on the assumption that the observations come from a distribution satisfying the extreme value condition so that the maximum of a large number of observations follows approximatively a generalized extreme value (GEV) distribution. On the contrary, the properties of the maximum likelihood relies implicitly on the assumption that the block maxima have *exactly* a GEV distribution. In many situations, this strong assumption is unsatisfactory and we shall only suppose that the underlying distribution is in the max-domain of attraction of an extreme value distribution. This is the purpose of the present paper to justify the maximum likelihood method for the block maxima method under an extreme value condition only.

Let us mention that the POT method is often considered as more powerful since it uses all relevant high observations while the block maxima method may retain less high observations but also some lower observations. In the POT method, Zhou [22], Theorem 2.2, has proved that the number n of peaks to be used within a data set of size N must satisfy $n/N \to 0$ in order to ensure consistency. In the block maxima method, for a data set of size $N = nm$ divided into n blocks of size m, only the n block maxima are used. Theorem 2 below states that consistency holds as soon as $m/\log(n) \to +\infty$, or equivalently $n\log(n)/N \to 0$. This extra logarithmic factor explains why the number of observations used for the block maxima is in general of a lower order than for the POT method.

We now recall some basic notions of univariate extreme value theory. The extreme value distribution function with index γ is denoted by G_γ,

$$G_\gamma(x) = \exp\bigl(-(1+\gamma x)^{-1/\gamma}\bigr), \qquad 1+\gamma x > 0.$$

We say that a distribution function F satisfies the extreme value condition with index γ, or equivalently that F belongs to the max-domain of attraction of G_γ if there exist constants $a_m > 0$ and b_m such that

$$\lim_{m\to+\infty} F^m(a_m x + b_m) = G_\gamma(x), \qquad x \in \mathbb{R}. \tag{1}$$

That is commonly denoted by $F \in D(G_\gamma)$. The necessary and sufficient conditions for $F \in D(G_\gamma)$ can be presented in different ways, see, for example, de Haan [4] or de Haan and Ferreira [5], Chapter 1. We remind the following simple criterion and choice of normalization constants.

Theorem 1. *Let* $U = (\frac{1}{1-F})^{\leftarrow}$ *be the left continuous inverse function of* $1/(1-F)$. *Then* $F \in D(G_\gamma)$ *if and only if there exists a function* $a(t) > 0$ *such that*

$$\lim_{t\to+\infty} \frac{U(tx)-U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \qquad \textit{for all } x > 0.$$

Then, a possible choice for the function $a(t)$ is given by

$$a(t) = \begin{cases} \gamma U(t), & \gamma > 0, \\ -\gamma\big(U(\infty) - U(t)\big), & \gamma < 0, \\ U(t) - t^{-1}\int_0^t U(s)\,\mathrm{d}s, & \gamma = 0, \end{cases}$$

and a possible choice for the normalization constants in (1) *is*

$$a_m = a(m) \quad \textit{and} \quad b_m = U(m).$$

In the sequel, we will always use the normalization constants (a_m) and (b_m) given in Theorem 1. According to the convergence of types theorem (see, e.g., Theorem 14.2 in Billingsley [3]), the normalization constants are unique up to asymptotic equivalence in the following sense: if (a'_m) and (b'_m) are such that $F^m(a'_m x + b'_m) \to G_\gamma(x)$ for all $x \in \mathbb{R}$, then

$$\lim_{m\to+\infty} \frac{a'_m}{a_m} = 1 \quad \text{and} \quad \lim_{m\to+\infty} \frac{b'_m - b_m}{a_m} = 0. \tag{2}$$

The log-likelihood of the extreme value distribution G_γ is given by

$$g_\gamma(x) = -(1 + 1/\gamma)\log(1 + \gamma x) - (1 + \gamma x)^{-1/\gamma},$$

if $1 + \gamma x > 0$ and $-\infty$ otherwise. For $\gamma = 0$, the formula is interpreted as $g_0(x) = -x - \exp(-x)$. The three parameter extreme value distribution with shape γ, location μ and scale $\sigma > 0$ has distribution function $x \mapsto G_\gamma(\frac{x-\mu}{\sigma})$. The corresponding log-likelihood is

$$\ell_{(\gamma,\mu,\sigma)}(x) = g_\gamma\left(\frac{x-\mu}{\sigma}\right) - \log\sigma.$$

The set-up of the block maxima method is the following. We consider independent and identically distributed (i.i.d.) random variables $(X_i)_{i\ge1}$ with common distribution function $F \in D(G_{\gamma_0})$ and corresponding normalization sequences (a_m) and (b_m) as in Theorem 1. We divide the sequence $(X_i)_{i\ge1}$ into blocks of length $m \ge 1$ and define the kth block maximum by

$$M_{k,m} = \max(X_{(k-1)m+1}, \dots, X_{km}).$$

Clearly, the normalized block maximum $(M_{k,m} - b_m)/a_m$ has distribution function $F^m(a_m x + b_m)$ and equation (1) suggests that the distribution of $M_{k,m}$ is approximately a GEV distribution with parameters (γ_0, b_m, a_m) and this is standard to estimate these parameters by the maximum likelihood method. The log-likelihood of the n-sample $(M_{1,m}, \dots, M_{n,m})$ is

$$L_n(\gamma,\mu,\sigma) = \frac{1}{n}\sum_{k=1}^n \ell_{(\gamma,\mu,\sigma)}(M_{k,m}).$$

In general, L_n has no global maximum, leading us to the following weak notion: we say that $(\widehat{\gamma}_n, \widehat{\mu}_n, \widehat{\sigma}_n)$ is a MLE if L_n has a *local* maximum at $(\widehat{\gamma}_n, \widehat{\mu}_n, \widehat{\sigma}_n)$. One sometimes uses the more precise term *pseudo*-MLE to emphasize the fact that the maximum is local and not global. Clearly, a MLE solves the likelihood equations

$$\nabla L_n = 0 \qquad \text{with } \nabla L_n = \left(\frac{\partial L_n}{\partial \gamma}, \frac{\partial L_n}{\partial \mu}, \frac{\partial L_n}{\partial \sigma}\right). \tag{3}$$

Conversely, any solution of the likelihood equations with a negative definite Hessian matrix is a MLE.

For the purpose of asymptotics, we let the length of the blocks $m = m(n)$ depend on the sample size n. Our main result is the following theorem, stating the existence of consistent MLEs.

Theorem 2. *Suppose $F \in D(G_{\gamma_0})$ with $\gamma_0 > -1$ and assume that*

$$\lim_{n\to+\infty} \frac{m(n)}{\log n} = +\infty. \tag{4}$$

Then there exists a sequence of estimators $(\widehat{\gamma}_n, \widehat{\mu}_n, \widehat{\sigma}_n)$ and a random integer $N \geq 1$ such that

$$\mathbb{P}\big[(\widehat{\gamma}_n, \widehat{\mu}_n, \widehat{\sigma}_n) \text{ is a MLE for all } n \geq N\big] = 1 \tag{5}$$

and

$$\widehat{\gamma}_n \xrightarrow{a.s.} \gamma_0, \qquad \frac{\widehat{\mu}_n - b_m}{a_m} \xrightarrow{a.s.} 0 \quad \text{and} \quad \frac{\widehat{\sigma}_n}{a_m} \xrightarrow{a.s.} 1 \qquad \text{as } n \to +\infty. \tag{6}$$

The condition $\gamma_0 > -1$ is natural and agrees with Smith [18]: it is easy to see that the likelihood equation (3) has no solution with $\gamma \leq -1$ so that no consistent MLE exists when $\gamma_0 < -1$ (see Remark 4 below). Condition (4) states that the block length $m(n)$ grows faster than logarithmically in the sample size n, which is not very restrictive. Let us mention a few further remarks on this condition.

Remark 1. A control of the block size is needed, as the following observations show. The log-likelihood $L_n(\gamma, \mu, \sigma)$ is finite if and only if

$$\min_{1\leq k\leq n} \left(1 + \gamma \frac{M_{k,m} - \mu}{\sigma}\right) > 0,$$

so that any MLE $(\widehat{\gamma}_n, \widehat{\mu}_n, \widehat{\sigma}_n)$ must satisfy

$$\min_{1\leq k\leq n} \left(1 + \widehat{\gamma}_n \frac{M_{k,m} - \widehat{\mu}_n}{\widehat{\sigma}_n}\right) > 0 \qquad \text{a.s.}$$

Hence, for a consistent MLE satisfying equation (6), one gets

$$\liminf_{n\to+\infty} \min_{1\leq k\leq n} \left(1 + \gamma_0 \frac{M_{k,m} - b_m}{a_m}\right) \geq 0 \qquad \text{a.s.,}$$

so that

$$\lim_{n\to+\infty} \mathbb{P}\left[\min_{1\le k\le n}\left(1+\gamma_0\frac{M_{k,m}-b_m}{a_m}\right) > -\varepsilon\right] = 1 \qquad \text{for all } \varepsilon > 0. \tag{7}$$

Equation (7) is thus a necessary condition for consistency. It is trivial in the case $\gamma_0 \le 0$. If $\gamma_0 > 0$, it holds

$$\mathbb{P}\left[\min_{1\le k\le n}\left(1+\gamma_0\frac{M_{k,m}-b_m}{a_m}\right) > -\varepsilon\right] = \left(1 - F^m(-\varepsilon b_m)\right)^n$$

so that equation (7) is equivalent to

$$\lim_{n\to+\infty} nF^m(-\varepsilon b_m) = 0 \qquad \text{for all } \varepsilon > 0. \tag{8}$$

This necessary condition compares the growth rates of n and $m(n)$, taking the left tail of F into account. Using this, one can easily construct an example violating (8) so that no consistent MLE exists: take for instance $\gamma_0 > 0$, $F \in D(G_{\gamma_0})$ such that $\lim_{x\to-\infty} F(x)\log|x| = 1$ and $m(n) = (\log n)^p$ with $p \in (0,1)$.

Remark 2. The preceding remark raises the question whether condition (4) is sharp or not. To this aim, we compare (in the case $\gamma_0 > 0$) the sufficient condition (4) and the necessary condition (8). One can show that for any $\gamma_0 > 0$ and any sequence $m(n)$ satisfying

$$\lim_{n\to+\infty} m(n) = +\infty \quad \text{and} \quad \lim_{n\to+\infty} \frac{m(n)}{\log n} = 0,$$

there exists $F \in D(G_{\gamma_0})$ (with $F(x) \to 0$ slowly enough as $x \to -\infty$) such that condition (8) fails and hence no consistent MLE exists. The intermediate case

$$\lim_{n\to+\infty} \frac{m(n)}{\log n} = c \in (0,+\infty)$$

remains open, as well as the case $\gamma_0 \le 0$ where no simple necessary condition is available.

Remark 3. It shall be stressed that condition (4) appears only in the proof of Lemma 4 below. Under stronger assumptions on the distribution $F \in D(G_{\gamma_0})$, Lemma 4 can be proved without condition (4): this is for instance the case if F is a Pareto distribution function, then the proof of Lemma 4 and Theorem 2 goes through under the minimal condition $\lim_{n\to+\infty} m(n) = +\infty$. See Remark 5 in the Appendix for more details.

The structure of the paper is as follows. We gather in Section 2 some preliminaries on properties of the GEV log-likelihood and of the empirical distribution associated to normalized block maxima. Section 3 is devoted to the proof of Theorem 2, which relies on an adaptation of Wald's method for proving the consistency of M-estimators. Some technical computations (proof of Lemma 4) involving regular variation theory are postponed to the Appendix.

2. Preliminaries

2.1. Properties of the GEV log-likelihood

We gather in the following proposition some basic properties of the GEV log-likelihood. We note x_γ^- and x_γ^+ the left and right end point of the domain g_γ, that is,

$$\left(x_\gamma^-, x_\gamma^+\right) = \{x \in \mathbb{R};\ 1 + \gamma x > 0\}.$$

Clearly, it is equal to $(-\infty, -1/\gamma)$, $\mathbb{R}$ and $(-1/\gamma, +\infty)$ when $\gamma < 0$, $\gamma = 0$ and $\gamma > 0$, respectively.

Proposition 1. *The function g_γ is infinitely differentiable on its domain.*

1. *If $\gamma \le -1$, g_γ is strictly increasing on its domain and*

$$\lim_{x \to x_\gamma^-} g_\gamma(x) = -\infty, \qquad \lim_{x \to x_\gamma^+} g_\gamma(x) = \begin{cases} +\infty & \text{if } \gamma < -1, \\ 0 & \text{if } \gamma = -1. \end{cases}$$

2. *If $\gamma > -1$, g_γ is increasing on $(x_\gamma^-, x_\gamma^*]$ and decreasing on $[x_\gamma^*, x_\gamma^+)$, where*

$$x_\gamma^* = \frac{(1+\gamma)^{-\gamma} - 1}{\gamma}.$$

Furthermore,

$$\lim_{x \to x_\gamma^-} g_\gamma(x) = \lim_{x \to x_\gamma^+} g_\gamma(x) = -\infty$$

and g_γ reaches its maximum $g_\gamma(x_\gamma^*) = (1+\gamma)(\log(1+\gamma) - 1)$ uniquely.

Remark 4. According to Proposition 1, for $\gamma \le -1$, the function g_γ is strictly increasing with a positive derivative g'_γ on its domain. On the other hand, the partial derivative of the log-likelihood L_n with respect to μ is given by

$$\frac{\partial L_n}{\partial \mu}(\gamma, \mu, \sigma) = -\frac{1}{\sigma} \sum_{k=1}^{n} g'_\gamma\left(\frac{M_{k,m} - \mu}{\sigma}\right).$$

This entails that $\frac{\partial L_n}{\partial \mu} < 0$ when $\gamma \le -1$ and that equation (3) has no solution in $(-\infty, -1] \times \mathbb{R} \times (0, +\infty)$. Hence any MLE $(\widehat{\gamma}_n, \widehat{\mu}_n, \widehat{\sigma}_n)$ satisfies $\widehat{\gamma}_n > -1$ and no consistent MLE does exist if $\gamma_0 < -1$. The limit case $\gamma_0 = -1$ is more difficult to analyze and is disregarded in this paper.

2.2. Normalized block maxima

In view of equation (1), we define the normalized block maxima

$$\widetilde{M}_{k,m} = \frac{M_{k,m} - b_m}{a_m}, \qquad k \ge 1,$$

and the corresponding likelihood

$$\widetilde{L}_n(\gamma, \mu, \sigma) = \frac{1}{n} \sum_{k=1}^{n} \ell_{(\gamma,\mu,\sigma)}(\widetilde{M}_{k,m}).$$

It should be stressed that the normalization sequences (a_m) and (b_m) are unknown so that the normalized block maxima $\widetilde{M}_{k,m}$ and the likelihood $\widetilde{L}_n$ cannot be computed from the data only. However, they will be useful in our theoretical analysis since they have good asymptotic properties. The following simple lemma will be useful.

Lemma 1. *$(\widehat{\gamma}_n, \widehat{\mu}_n, \widehat{\sigma}_n)$ is a MLE if and only if $\widetilde{L}_n$ has a local maximum at $(\widehat{\gamma}_n, (\widehat{\mu}_n - b_m)/a_m, \widehat{\sigma}_n/a_m)$.*

Proof. The GEV likelihood satisfies the scaling property

$$\ell_{(\gamma,\mu,\sigma)}\big((x-b)/a\big) = \ell_{(\gamma,a\mu+b,a\sigma)}(x) + \log a$$

so that

$$L_n(\gamma, \mu, \sigma) = \widetilde{L}_n\left(\gamma, \frac{\mu - b_m}{a_m}, \frac{\sigma}{a_m}\right) - \log a_m.$$

Hence the local maximizers of L_n and $\widetilde{L}_n$ are in direct correspondence and the lemma follows. □

2.3. Empirical distributions

The likelihood function $\tilde{L}_n$ can be seen as a functional of the empirical distribution defined by

$$\mathbb{P}_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{\widetilde{M}_{k,m}},$$

where δ_x denotes the Dirac mass at point $x \in \mathbb{R}$. For any measurable $f : \mathbb{R} \to [-\infty, +\infty)$, we note $\mathbb{P}_n[f]$ the integral with respect to $\mathbb{P}_n$, that is,

$$\mathbb{P}_n[f] = \frac{1}{n} \sum_{k=1}^{n} f(\widetilde{M}_{k,m}).$$

With these notations, it holds

$$\widetilde{L}_n(\gamma, \mu, \sigma) = \mathbb{P}_n[\ell_{(\gamma,\mu,\sigma)}].$$

The empirical process is defined by

$$\mathbb{F}_n(t) = \mathbb{P}_n\big((-\infty, t]\big) = \frac{1}{n} \sum_{k=1}^{n} 1_{\{\widetilde{M}_{k,m} \le t\}}, \qquad t \in \mathbb{R}.$$

In the case of an i.i.d. sequence, the Glivenko–Cantelli theorem states that the empirical process converges almost surely uniformly to the sample distribution function. According to the general theory of empirical processes (see, e.g., Shorack and Wellner [17], Theorem 1, page 106), this result can be extended to triangular arrays of i.i.d. random variables. Equation (1) entails the following result.

Lemma 2. *Suppose* $F \in D(G_{\gamma_0})$ *and* $\lim_{n\to+\infty} m(n) = +\infty$. *Then*,

$$\sup_{t\in\mathbb{R}} \left|\mathbb{F}_n(t) - G_{\gamma_0}(t)\right| \xrightarrow{a.s.} 0 \qquad \text{as } n \to +\infty.$$

This entails that $\mathbb{P}_n \Rightarrow G_{\gamma_0}$ almost surely, where $\Rightarrow$ denotes weak convergence, that is,

$$\mathbb{P}_n[f] \xrightarrow{\text{a.s.}} G_{\gamma_0}[f] := \int_{\mathbb{R}} f(t)\,\mathrm{d}G_{\gamma_0}(t) \qquad \text{as } n \to +\infty$$

for all bounded and continuous functions $f : \mathbb{R} \to \mathbb{R}$. The following lemma dealing with more general functions will be useful.

Lemma 3. *Suppose* $F \in D(G_{\gamma_0})$ *and* $\lim_{n\to+\infty} m(n) = +\infty$. *Then*, *for all upper semi-continuous functions* $f : \mathbb{R} \to [-\infty, +\infty)$ *bounded from above*,

$$\limsup_{n\to+\infty} \mathbb{P}_n[f] \le G_{\gamma_0}[f] \qquad a.s.$$

Proof. Let M be an upper bound for f. The function $\tilde{f} = M - f$ is non-negative and lower semicontinuous. Clearly,

$$\mathbb{P}_n[f] = M - \mathbb{P}_n[\tilde{f}] \quad \text{and} \quad G_{\gamma_0}[f] = M - G_{\gamma_0}[\tilde{f}],$$

whence it is enough to prove that

$$\liminf_{n\to+\infty} \mathbb{P}_n[\tilde{f}] \ge G_{\gamma_0}[\tilde{f}] \qquad \text{a.s.}$$

To see this, we use the relation $\mathbb{P}_n[\tilde{f}] = \int_0^1 \tilde{f}(\mathbb{F}_n^{\leftarrow}(u))\,\mathrm{d}u$ where $\mathbb{F}_n^{\leftarrow}$ is the left-continuous inverse function

$$\mathbb{F}_n^{\leftarrow} = \inf\left\{t \in \mathbb{R}; \mathbb{F}_n(t) \ge u\right\}, \qquad u \in (0, 1).$$

Lemma 2 together with the continuity of the distribution function G_{γ_0} entail that almost surely, $\mathbb{F}_n^{\leftarrow}(u) \to G_{\gamma_0}^{\leftarrow}(u)$ for all $u \in (0, 1)$ as $n \to +\infty$. Using the fact that $\tilde{f}$ is lower semi-continuous, we obtain

$$\liminf_{n\to+\infty} \tilde{f}\left(\mathbb{F}_n^{\leftarrow}(u)\right) \ge \tilde{f}\left(G_{\gamma_0}^{\leftarrow}(u)\right), \qquad u \in (0, 1).$$

On the other hand, according to Fatou's lemma,

$$\liminf_{n\to+\infty} \int_0^1 \tilde{f}\left(\mathbb{F}_n^{\leftarrow}(u)\right)\mathrm{d}u \ge \int_0^1 \liminf_{n\to+\infty} \tilde{f}\left(\mathbb{F}_n^{\leftarrow}(u)\right)\mathrm{d}u.$$

Combining the two inequalities,

$$\liminf_{n\to+\infty} \mathbb{P}_n[\tilde{f}] \geq G_{\gamma_0}[\tilde{f}] \qquad \text{a.s.} \qquad \square$$

The next lemma plays a crucial role in our proof of Theorem 2. Its proof is quite technical and is postponed to the Appendix.

Lemma 4. *Suppose $F \in D(G_{\gamma_0})$ with $\gamma_0 > -1$ and assume condition (4) is satisfied. Then,*

$$\lim_{n\to+\infty} \mathbb{P}_n[g_{\gamma_0}] = G_{\gamma_0}[g_{\gamma_0}] \qquad a.s. \tag{9}$$

It shall be stressed that Lemma 4 is the only part in the proof of Theorem 2 where condition (4) is needed (see Remarks 3 and 5).

3. Proof of Theorem 2

We introduce the short notation $\Theta = (-1, +\infty) \times \mathbb{R} \times (0, +\infty)$. A generic point of Θ is denoted by $\theta = (\gamma, \mu, \sigma)$.

The restriction $\widetilde{L}_n : \Theta \to [-\infty, +\infty)$ is continuous, so that for any compact $K \subset \Theta$, $\widetilde{L}_n$ is bounded and reaches its maximum on K. We can thus define $\widetilde{\theta}_n^K = (\widetilde{\gamma}_n^K, \widetilde{\mu}_n^K, \widetilde{\sigma}_n^K)$ such that

$$\widetilde{\theta}_n^K = \arg\max_{\theta\in K} \widetilde{L}_n(\theta). \tag{10}$$

By convention, if the argmax is not unique, the smallest one with respect to the lexicographic order is chosen. The following proposition is the key in the proof of Theorem 2.

Proposition 2. *Let $\theta_0 = (\gamma_0, 0, 1)$ and $K \subset \Theta$ be a compact neighborhood of θ_0. Under the assumptions of Theorem 2,*

$$\lim_{n\to+\infty} \widetilde{\theta}_n^K = \theta_0 \qquad a.s.$$

The proof of Proposition 2 relies on an adaptation of Wald's method for proving the consistency of M-estimators (see Wald [21] or van der Vaart [20], Theorem 5.14). The standard theory of M-estimation is designed for i.i.d. samples, while we have to deal with the triangular array $\{(\widetilde{M}_{k,m})_{1\leq k\leq n}, n \geq 1\}$. We first state two lemmas.

Lemma 5. *For all $\theta \in \Theta$, $G_{\gamma_0}[\ell_\theta] \leq G_{\gamma_0}[\ell_{\theta_0}]$ and the equality holds if and only if $\theta = \theta_0$.*

Proof. The quantity $G_{\gamma_0}[\ell_{\theta_0} - \ell_\theta]$ is the Kullback–Leibler divergence of the GEV distributions with parameters θ_0 and θ and is known to be non-negative (see van der Vaart [20], Section 5.5). It vanishes if and only if the two distributions agree. This occurs if and only if $\theta = \theta_0$ because the GEV model is identifiable. $\square$

Lemma 6. *For $B \subset \Theta$, define*

$$\ell_B(x) = \sup_{\theta \in B} \ell_\theta(x), \qquad x \in \mathbb{R}.$$

Let $\theta \in \Theta$ and $B(\theta, \varepsilon)$ be the open ball in Θ with center θ and radius $\varepsilon > 0$. Then,

$$\lim_{\varepsilon \to 0} G_{\gamma_0}[\ell_{B(\theta,\varepsilon)}] = G_{\gamma_0}[\ell_\theta].$$

Proof. Proposition 1 implies

$$\ell_\theta(x) = g_\gamma\big((x-\mu)/\sigma\big) - \log\sigma \le g_\gamma\big(x_\gamma^*\big) - \log\sigma.$$

One can deduce that if B is contained in $(-1, \bar{\gamma}] \times [\bar{\sigma}, +\infty) \times \mathbb{R}$ for some $\bar{\gamma} > -1$ and $\bar{\sigma} > 0$, then there exists $M(\bar{\gamma}, \bar{\sigma})$ such that

$$\ell_\theta(x) \le M(\bar{\gamma}, \bar{\sigma}) \qquad \text{for all } \theta \in B, x \in \mathbb{R}.$$

Hence, there exists $M > 0$ such that the function $M - \ell_{B(\theta,\varepsilon)}$ is non-negative for ε small enough. The continuity of $\theta \mapsto \ell_\theta(x)$ on Θ implies

$$\lim_{\varepsilon \to 0} \ell_{B(\theta,\varepsilon)}(x) = \ell_\theta(x) \qquad \text{for all } x \in \mathbb{R}.$$

Then, Fatou's lemma entails

$$G_{\gamma_0}\Big[\liminf_{\varepsilon \to 0}(M - \ell_{B(\theta,\varepsilon)})\Big] \le \liminf_{\varepsilon \to 0} G_{\gamma_0}[M - \ell_{B(\theta,\varepsilon)}],$$

whence we obtain

$$\limsup_{\varepsilon \to 0} G_{\gamma_0}[\ell_{B(\theta,\varepsilon)}] \le G_{\gamma_0}[\ell_\theta].$$

On the other hand, $\theta \in B(\theta, \varepsilon)$ implies $G_{\gamma_0}[\ell_{B(\theta,\varepsilon)}] \ge G_{\gamma_0}[\ell_\theta]$. We deduce

$$\lim_{\varepsilon \to 0} G_{\gamma_0}[\ell_{B(\theta,\varepsilon)}] = G_{\gamma_0}[\ell_\theta].$$

□

Proof of Proposition 2. In view of Lemmas 5 and 6, for each $\theta \in K$ such that $\theta \neq \theta_0$, there exists $\varepsilon_\theta > 0$ such that

$$G_{\gamma_0}[\ell_{B(\theta,\varepsilon_\theta)}] < G_{\gamma_0}[\ell_{\theta_0}].$$

Fix $\delta > 0$. The set $\Delta = \{\theta \in K; \|\theta - \theta_0\| \ge \delta\}$ is compact and is covered by the open balls $\{B(\theta, \varepsilon_\theta), \theta \in \Delta\}$. Let $B_i = B(\theta_i, \varepsilon_{\theta_i})$, $1 \le i \le p$, be a finite subcover. Using the relation $\widetilde{L}_n(\theta) = \mathbb{P}_n[\ell_\theta]$, we see that

$$\sup_{\theta \in \Delta} \widetilde{L}_n(\theta) \le \max_{1 \le i \le p} \mathbb{P}_n[\ell_{B_i}].$$

The function ℓ_{B_i} is upper semi-continuous and bounded from above, so that Lemma 3 entails

$$\limsup_{n\to+\infty} \mathbb{P}_n[\ell_{B_i}] \le G_{\gamma_0}[\ell_{B_i}] \qquad \text{a.s.},$$

whence

$$\limsup_{n\to+\infty} \sup_{\theta\in\Delta} \widetilde{L}_n(\theta) \le \max_{1\le i\le p} G_{\gamma_0}[\ell_{B_i}] < G_{\gamma_0}[\ell_{\theta_0}] \qquad \text{a.s.} \tag{11}$$

According to Lemma 4, $\mathbb{P}_n[\ell_{\theta_0}] \xrightarrow{\text{a.s.}} G_{\gamma_0}[g_{\gamma_0}]$, so that

$$\liminf_{n\to+\infty} \sup_{\theta\in K} \widetilde{L}_n(\theta) \ge G_{\gamma_0}[\ell_{\theta_0}] \qquad \text{a.s.} \tag{12}$$

Since $\widetilde{\theta}_n^K$ realizes the maximum of $\widetilde{L}_n$ over K, equations (11) and (12) together entail that $\widetilde{\theta}_n^K \in K \setminus \Delta$ for large n. Equivalently, $\|\widetilde{\theta}_n^K - \theta_0\| < \delta$ for large n. Since δ is arbitrary, this proves the convergence $\widetilde{\theta}_n^K \xrightarrow{\text{a.s.}} \theta_0$ as $n \to +\infty$. □

Proof of Theorem 2. Let $K \subset \Theta$ be a compact neighborhood of θ_0 as in Proposition 2 and define $\widetilde{\theta}_n^K$ by equation (10). We prove that Theorem 2 holds true with the sequence of estimators

$$(\widehat{\gamma}_n, \widehat{\mu}_n, \widehat{\sigma}_n) = \big(\widetilde{\gamma}_n^K, a_m\widetilde{\mu}_n^K + b_m, a_m\widetilde{\sigma}_n^K\big), \qquad n \ge 1.$$

According to Lemma 1, $(\widehat{\gamma}_n, \widehat{\mu}_n, \widehat{\sigma}_n)$ is a MLE if and only if $\widetilde{L}_n$ has a local maximum at $\widetilde{\theta}_n^K = (\widetilde{\gamma}_n^K, \widetilde{\mu}_n^K, \widetilde{\sigma}_n^K)$. Since $\widetilde{\theta}_n^K = \arg\max_{\theta\in K} \widetilde{L}_n(\theta)$, this is the case as soon as $\widetilde{\theta}_n^K$ lies in the interior set $\mathrm{int}(K)$ of K. Proposition 2 implies the almost surely convergence $\widetilde{\theta}_n^K \xrightarrow{\text{a.s.}} \theta_0$ which is equivalent to equation (6). Furthermore, since $\theta_0 \in \mathrm{int}(K)$, this implies $\widetilde{\theta}_n^K \in \mathrm{int}(K)$ for large n so that $(\widehat{\gamma}_n, \widehat{\mu}_n, \widehat{\sigma}_n)$ is a MLE for large n. This proves equation (5). □

Appendix: Proof of Lemma 4

We will use the following criterion to prove equation (9).

Lemma 7. *Suppose $F \in D(G_{\gamma_0})$ and $\lim_{n\to+\infty} m(n) = +\infty$. We note $Y_m = g_{\gamma_0}(a_m^{-1}(M_{1,m} - b_m))$. If there exists a sequence $(\alpha_n)_{n\ge1}$ and $p > 2$ such that*

$$\sum_{n\ge1} n\mathbb{P}\big(|Y_m| > \alpha_n\big) < +\infty \quad \textit{and} \quad \sup_{n\ge1} \mathbb{E}\big[|Y_m|^p 1_{\{|Y_m|\le\alpha_n\}}\big] < +\infty,$$

then equation (9) holds true.

Proof. We note $\mu = G_{\gamma_0}[g_{\gamma_0}]$ and we define

$$Y_{k,m} = g_{\gamma_0}\big(a_m^{-1}(M_{k,m} - b_m)\big) \quad \text{and} \quad S_n = \sum_{k=1}^n Y_{k,m}.$$

With these notations, (9) is equivalent to $n^{-1}S_n \overset{\text{a.s.}}{\to} \mu$. We introduce the truncated variables

$$\widetilde{Y}_{k,m} = Y_{k,m}1_{\{|Y_{k,m}|\le\alpha_n\}} \quad \text{and} \quad \widetilde{S}_n = \sum_{k=1}^n \widetilde{Y}_{k,m}.$$

Clearly,

$$\begin{aligned}\mathbb{P}[\widetilde{S}_n \neq S_n] &\le \mathbb{P}\big[\widetilde{Y}_{k,m} \neq Y_{k,m} \text{ for some } k \in \{1,\dots,n\}\big] \\ &\le n\mathbb{P}\big[|Y_m| > \alpha_n\big],\end{aligned}$$

so that $\sum_{n\ge1} n\mathbb{P}[|Y_m| > \alpha_n] < +\infty$ entails $\sum_{n\ge1} \mathbb{P}[\widetilde{S}_n \neq S_n] < +\infty$. By the Borel–Cantelli lemma, this implies that the sequences $(\widetilde{S}_n)_{n\ge1}$ and $(S_n)_{n\ge1}$ coincide eventually, whence $n^{-1}S_n \overset{\text{a.s.}}{\to} \mu$ if and only if $n^{-1}\widetilde{S}_n \overset{\text{a.s.}}{\to} \mu$. We now prove this last convergence.

We first prove that $\mathbb{E}[\widetilde{Y}_{1,m}] \to \mu$. Indeed, by the continuous mapping theorem, the convergence (1) implies $Y_{1,m} \Rightarrow g_{\gamma_0}(Z)$ with $Z \sim G_{\gamma_0}$, where $\Rightarrow$ stands for weak convergence. Since $\mathbb{P}[\widetilde{Y}_{1,m} \neq Y_{1,m}]$ converges to 0 as $n \to +\infty$, it also holds $\widetilde{Y}_{1,m} \Rightarrow g_{\gamma_0}(Z)$. Together with the condition $\sup_{n\ge1} \mathbb{E}[|\widetilde{Y}_{1,m}|^p] < \infty$, this entails $\mathbb{E}[\widetilde{Y}_{1,m}] \to \mathbb{E}[g_{\gamma_0}(Z)] = \mu$.

Next, Theorem 2.10 in Petrov [13] provides the upper bound

$$\mathbb{E}\big[\big|\widetilde{S}_n - \mathbb{E}[\widetilde{S}_n]\big|^p\big] \le C(p)n^{p/2}\mathbb{E}\big[\big|\widetilde{Y}_{1,m} - \mathbb{E}[\widetilde{Y}_{1,m}]\big|^p\big]$$

for some constant $C(p) > 0$ depending only on p. Equivalently,

$$\mathbb{E}\big[\big|n^{-1}\widetilde{S}_n - \mu_n\big|^p\big] \le C(p)n^{-p/2}\mathbb{E}\big[|\widetilde{Y}_{1,m} - \mu_n|^p\big]$$

with $\mu_n = \mathbb{E}[\widetilde{Y}_{1,m}]$. Furthermore,

$$\mathbb{E}\big[|\widetilde{Y}_{1,m} - \mu_n|^p\big] \le 2^{p-1}\big(\mathbb{E}\big[|\widetilde{Y}_{1,m}|^p\big] + |\mu_n|^p\big)$$

is uniformly bounded by some constant $C > 0$. By the Markov inequality, for all $\varepsilon > 0$,

$$\mathbb{P}\big[\big|n^{-1}\widetilde{S}_n - \mu_n\big| \ge \varepsilon\big] \le \varepsilon^{-p}\mathbb{E}\big[\big|n^{-1}\widetilde{S}_n - \mu_n\big|^p\big] \le \varepsilon^{-p}C(p)Cn^{-p/2}.$$

Since $p > 2$, we get that

$$\sum_{n\ge1} \mathbb{P}\big[\big|n^{-1}\widetilde{S}_n - \mu_n\big| \ge \varepsilon\big] < +\infty$$

and the Borel–Cantelli lemma entails $n^{-1}\widetilde{S}_n - \mu_n \overset{\text{a.s.}}{\to} 0$. Since $\mu_n \to \mu$, we deduce $n^{-1}\widetilde{S}_n \overset{\text{a.s.}}{\to} \mu$ which proves the lemma. □

Proof of Lemma 4. We prove that there exists a sequence (α_n) and $p > 2$ satisfying

$$\sum_{n\ge1} n\alpha_n^{-p} < +\infty \tag{13}$$

and

$$\sup_{n\geq 1}\mathbb{E}\big[\big(|Y_m|\wedge\alpha_n\big)^p\big]<+\infty. \tag{14}$$

The Markov inequality yields

$$\mathbb{P}\big[|Y_m|\geq\alpha_n\big]\leq\alpha_n^{-p}\mathbb{E}\big[\big(|Y_m|\wedge\alpha_n\big)^p\big]$$

so that equations (13) and (14) together entail

$$\sum_{n\geq 1}n\mathbb{P}\big(|Y_m|>\alpha_n\big)<+\infty \quad \text{and} \quad \sup_{n\geq 1}\mathbb{E}\big[|Y_m|^p 1_{\{|Y_m|\leq\alpha_n\}}\big]<+\infty.$$

This shows that equations (13) and (14) together imply the assumptions of Lemma 7 and prove Lemma 4.

We first evaluate the quantity $\mathbb{E}[(|Y_m|\wedge\alpha_n)^p]$ from equation (14). Recall that $Y_m = g_{\gamma_0}((M_{1,m}-b_m)/a_m)$. It is well known that the random variable X_i with distribution function F has the same distribution as the random variable $F^{\leftarrow}(V)$, with V a uniform random variable on $(0,1)$. We deduce that the random variable $M_{1,m}=\bigvee_{i=1}^m X_i$ has the same distribution as $F^{\leftarrow}(V_m)$, with V_m a random variable with distribution $mv^{m-1}1_{(0,1)}(v)\,\mathrm{d}v$ (this is the distribution of the maximum of m i.i.d. uniform random variables on $[0,1]$). Hence,

$$\mathbb{E}\big[\big(|Y_m|\wedge\alpha_n\big)^p\big]=\int_0^1\big(\big|g_{\gamma_0}\big(\big(F^{\leftarrow}(v)-b_m\big)/a_m\big)\big|\wedge\alpha_n\big)^p mv^{m-1}\,\mathrm{d}v.$$

The relations $U(x)=F^{\leftarrow}(1-1/x)$ and $b_m=U(m)$ together with the change of variable $v=1-1/(mx)$ yield

$$\mathbb{E}\big[\big(|Y_m|\wedge\alpha_n\big)^p\big]=\int_{1/m}^{\infty}\big(\big|g_{\gamma_0}\big(\widetilde{U}_m(x)\big)\big|\wedge\alpha_n\big)^p\left(1-\frac{1}{mx}\right)^{m-1}x^{-2}\,\mathrm{d}x,$$

where

$$\widetilde{U}_m(x)=\frac{U(mx)-U(m)}{a_m}.$$

We now provide an upper bound for the integral and we use the following estimates. There exists a constant $c>0$ such that for $1+\gamma_0 y>0$

$$\big|g_{\gamma_0}(y)\big|\leq\begin{cases}c(1+\gamma_0 y)^{-1/\gamma_0}, & y<0,\\ (1+1/\gamma_0)\log(1+\gamma_0 y)+1, & y\geq 0.\end{cases}$$

Note that $\widetilde{U}_m(x)$ is positive for $x>1$ and negative for $x<1$. Furthermore, for all $x\geq 1/m$ and $m\geq 2$,

$$\left(1-\frac{1}{mx}\right)^{m-1}\leq\exp\big(-(m-1)/(mx)\big)\leq\exp\big(-1/(2x)\big).$$

Using these estimates, we obtain the following upper bound: for $m \geq m_0$ (m_0 to be precised later),

$$\mathbb{E}\big[\big(|Y_m| \wedge \alpha_n\big)^p\big] \leq I_1 + I_2 + I_3 \tag{15}$$

with

$$\begin{aligned}
I_1 &= \int_{1/m}^{m_0/m} \alpha_n^p \exp\big(-(m-1)/(mx)\big)x^{-2}\,\mathrm{d}x,\\
I_2 &= \int_{m_0/m}^{1} c^p\big(1+\gamma_0\widetilde{U}_m(x)\big)^{-p/\gamma_0}\exp\big(-1/(2x)\big)x^{-2}\,\mathrm{d}x,\\
I_3 &= \int_1^\infty \big((1+1/\gamma_0)\log\big(1+\gamma_0\widetilde{U}_m(x)\big)+1\big)^p \exp\big(-1/(2x)\big)x^{-2}\,\mathrm{d}x.
\end{aligned}$$

The integral I_1 can be computed explicitly and, for $m \geq 2$,

$$I_1 \leq 2\alpha_n^p \exp\big(-(m-1)/m_0\big). \tag{16}$$

To estimate I_2 and I_3, we need upper and lower bounds for $\widetilde{U}_m(x)$ and we have to distinguish between the three cases $\gamma_0 > 0$, $\gamma_0 \in (-1,0)$ and $\gamma_0 = 0$.

Case $\gamma_0 > 0$: According to Theorem 1, the function U is regularly varying at infinity with index $\gamma_0 > 0$ and

$$1+\gamma_0\widetilde{U}_m(x) = 1+\gamma_0\frac{U(mx)-U(m)}{a_m} = \frac{U(mx)}{U(m)}.$$

We use then Potter's bounds (see, e.g., Proposition B.1.9 in [5]): for all $\varepsilon > 0$, there exists $m_0 \geq 1$ such that for $m \geq m_0$ and $mx \geq m_0$

$$(1-\varepsilon)x^{\gamma_0}\min\big(x^\varepsilon, x^{-\varepsilon}\big) \leq \frac{U(mx)}{U(m)} \leq (1+\varepsilon)x^{\gamma_0}\max\big(x^\varepsilon, x^{-\varepsilon}\big). \tag{17}$$

We fix $\varepsilon \in (0,\gamma_0)$ and choose m_0 accordingly. Using the lower Potter's bound to bound I_2 and the upper Potter's bound to bound I_3, we get

$$\begin{aligned}
I_2 &\leq \int_{m_0/m}^{1} c^p\big((1-\varepsilon)x^{\gamma_0+\varepsilon}\big)^{-p/\gamma_0}\exp\big(-1/(2x)\big)x^{-2}\,\mathrm{d}x\\
&\leq c^p(1-\varepsilon)^{-p/\gamma_0}\int_0^1 x^{-2-p-p\varepsilon/\gamma_0}\exp\big(-1/(2x)\big)\,\mathrm{d}x
\end{aligned}$$

and

$$I_3 \leq \int_1^\infty \big((1+1/\gamma_0)\log\big((1+\varepsilon)x^{\gamma_0+\varepsilon}\big)+1\big)^p\exp\big(-1/(2x)\big)x^{-2}\,\mathrm{d}x.$$

These integrals are finite and this implies that I_2 and I_3 are uniformly bounded for $m \geq m_0$. From equations (15) and (16), we obtain

$$\mathbb{E}\big[\big(|Y_m| \wedge \alpha_n\big)^p\big] \leq 4\alpha_n^p \exp\big(-m/(2m_0)\big) + C,$$

for some constant $C > 0$. Finally, we set $\alpha_n^p \exp(-(m-1)/m_0) = 1$, i.e., $\alpha_n = \exp((m-1)/(pm_0))$. Equation (14) is clearly satisfied and

$$n\alpha_n^{-p} = \exp\big[-\big((m-1)/(m_0 \log n) - 1\big)\log n\big]. \tag{18}$$

We check easily that the condition $\lim_{n\to+\infty} \frac{m(n)}{\log n} = +\infty$ implies equation (13).
Case $\gamma_0 < 0$: It follows from Theorem 1 that the function $t \mapsto U(\infty) - U(t)$ is regularly varying at infinity with index $\gamma_0 < 0$ and that

$$1 + \gamma_0 \widetilde{U}_m(x) = 1 + \gamma_0 \frac{U(mx) - U(m)}{a_m} = \frac{U(\infty) - U(mx)}{U(\infty) - U(m)}.$$

Then, the Potter's bounds become: for all $\varepsilon > 0$, there exists $m_0 \geq 1$ such that for $m \geq m_0$ and $mx \geq m_0$

$$(1-\varepsilon)x^{\gamma_0} \min\big(x^{\varepsilon}, x^{-\varepsilon}\big) \leq \frac{U(\infty) - U(mx)}{U(\infty) - U(m)} \leq (1+\varepsilon)x^{\gamma_0} \max\big(x^{\varepsilon}, x^{-\varepsilon}\big).$$

Using this, the proof is completed in the same way as in the case $\gamma_0 > 0$ with straightforward modifications.
Case $\gamma_0 = 0$: In this case, Theorem B.2.18 in [5] implies that for all $\varepsilon > 0$, there exists $m_0 \geq 1$ such that for $m \geq m_0$ and $mx \geq m_0$,

$$\left| \frac{U(mx) - U(m)}{a_m} - \log x \right| \leq \varepsilon \max\big(x^{\varepsilon}, x^{-\varepsilon}\big).$$

Equivalently, for $m \geq m_0$ and $mx \geq m_0$,

$$\log x - \varepsilon \max\big(x^{\varepsilon}, x^{-\varepsilon}\big) \leq \widetilde{U}_m(x) \leq \log x + \varepsilon \max\big(x^{\varepsilon}, x^{-\varepsilon}\big).$$

Using the lower bound to estimate I_2 and the upper bound to estimate I_3, we obtain

$$\begin{aligned} I_2 &= \int_{m_0/m}^1 c^p \exp\big(-p\widetilde{U}_m(x)\big) \exp\big(-1/(2x)\big) x^{-2}\, \mathrm{d}x \\ &\leq c^p \int_0^1 \exp\big(-p\log x + p\varepsilon x^{-\varepsilon} - 1/(2x)\big) x^{-2}\, \mathrm{d}x \end{aligned}$$

and

$$\begin{aligned} I_3 &= \int_1^{\infty} \big(\widetilde{U}_m(x) + 1\big)^p \exp\big(-1/(2x)\big) x^{-2}\, \mathrm{d}x \\ &\leq \int_1^{\infty} \big(\log x + \varepsilon x^{\varepsilon} + 1\big)^p \exp\big(-1/(2x)\big) x^{-2}\, \mathrm{d}x. \end{aligned}$$

For $\varepsilon \in (0, 1/p)$, the integrals appearing in the upper bounds are finite and independent of m. This shows that I_2 and I_3 are uniformly bounded for $m \geq m_0$. The proof is then completed as in the case $\gamma_0 > 0$. $\square$

Remark 5. In view of equation (18), one can easily check that the conclusion of the proof requires only the weaker condition

$$\liminf_{n\to+\infty} \frac{m(n)}{\log n} > 2m_0$$

with m_0 given by Potter's bounds (17). Furthermore, if $m_0 = 1$, the integral I_1 vanishes while I_2 and I_3 remain bounded and one can check that the conclusion of the proof requires only

$$\lim_{n\to+\infty} m(n) = +\infty.$$

This last case occurs for instance with the Pareto distribution function

$$F(t) = 1 - \left(\frac{t}{t_0}\right)^{-1/\gamma_0}, \qquad t \geq t_0.$$

Then $F \in D(G_{\gamma_0})$ and $U(x) = t_0 x^{\gamma_0}$, $x \geq 1$, so that the Potter's bounds (17) is trivially satisfied with $m_0 = 1$. The same comments hold true in the cases $\gamma_0 < 0$ or $\gamma_0 = 0$.

Acknowledgements

Clément Dombry is grateful to Laurens de Haan for suggesting the research problem in the first place and for useful comments that improved greatly an early version of the manuscript. Two anonymous referees and an associate editor are also acknowledged for thorough reading and constructive suggestions from which the paper has benefited a lot.

References

[1] Balkema, A.A. and de Haan, L. (1974). Residual life time at great age. *Ann. Probab.* **2** 792–804. MR0359049

[2] Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2004). *Statistics of Extremes. Theory and Applications. Wiley Series in Probability and Statistics*. Chichester: Wiley. MR2108013

[3] Billingsley, P. (1995). *Probability and Measure*, 3rd ed. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR1324786

[4] de Haan, L. (1984). Slow variation and characterization of domains of attraction. In *Statistical Extremes and Applications* (*Vimeiro*, 1983). *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.* **131** 31–48. Dordrecht: Reidel. MR0784812

[5] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory. An Introduction. Springer Series in Operations Research and Financial Engineering*. New York: Springer. MR2234156

[6] Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* **17** 1833–1855. MR1026315

[7] Drees, H., Ferreira, A. and de Haan, L. (2004). On maximum likelihood estimation of the extreme value index. *Ann. Appl. Probab.* **14** 1179–1201. MR2071420
[8] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events. For Insurance and Finance. Applications of Mathematics* (*New York*) **33**. Berlin: Springer. MR1458613
[9] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163–1174. MR0378204
[10] Hosking, J.R.M. (1985). Algorithm AS 215: Maximum likelihood estimation of the parameters of the generalized extreme value distribution. *Appl. Stat.* **34** 301–310.
[11] Hosking, J.R.M., Wallis, J.R. and Wood, E.F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* **27** 251–261. MR0797563
[12] Macleod, A.J. (1989). AS R76 – A remark on algorithm AS 215: Maximum likelihood estimation of the parameters of the generalized extreme value distribution. *Appl. Stat.* **38** 198–199.
[13] Petrov, V.V. (1995). *Limit Theorems of Probability Theory. Sequences of Independent Random Variables. Oxford Studies in Probability* **4**. New York: The Clarendon Press Oxford Univ. Press. MR1353441
[14] Pickands, J. III (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3** 119–131. MR0423667
[15] Prescott, P. and Walden, A.T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika* **67** 723–724. MR0601119
[16] Prescott, P. and Walden, A.T. (1983). Maximum likelihood estimation of the parameters of the three-parameter extreme-value distribution from censored samples. *J. Stat. Comput. Simul.* **16** 241–250.
[17] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. New York: Wiley. MR0838963
[18] Smith, R.L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72** 67–90. MR0790201
[19] Smith, R.L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15** 1174–1207. MR0902252
[20] van der Vaart, A.W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247
[21] Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Stat.* **20** 595–601. MR0032169
[22] Zhou, C. (2009). Existence and consistency of the maximum likelihood estimator for the extreme value index. *J. Multivariate Anal.* **100** 794–815. MR2478199

Received January 2013 and revised September 2013