

## STRONG CONVERGENCE OF PEAKS OVER A THRESHOLD

SIMONE A. PADOAN,\* *Bocconi University*  
STEFANO RIZZELLI,\*\* *Catholic University*

### Abstract

Extreme value theory plays an important role in providing approximation results for the extremes of a sequence of independent random variables when their distribution is unknown. An important one is given by the generalised Pareto distribution  $H_\gamma(x)$  as an approximation of the distribution  $F_t(s(t)x)$  of the excesses over a threshold  $t$ , where  $s(t)$  is a suitable norming function. We study the rate of convergence of  $F_t(s(t) \cdot)$  to  $H_\gamma$  in variational and Hellinger distances and translate it into that regarding the Kullback–Leibler divergence between the respective densities.

*Keywords:* Convergence rate; exceedances; extreme quantile; generalised Pareto; tail index

2020 Mathematics Subject Classification: Primary 60G70  
Secondary 60F15; 60B10

### 1. Introduction

Extreme value theory (EVT) develops probabilistic models and methods for describing the random behaviour of extreme observations that occur rarely. These theoretical foundations are very important for studying practical problems in environmental, climate, insurance, and financial fields (e.g. [5, 6, 10]), to name a few.

In the univariate setting, the most popular approaches for statistical analysis are the so-called block maxima (BM) and peaks over threshold (POT) (see, e.g., [3] for a review). Let  $X_1, \dots, X_n$  be independent and identically distributed (i.i.d.) random variables according to a common distribution  $F$ . The first approach concerns the modelling of  $k$  sample maxima derived over blocks of a certain size  $m$ , i.e.  $M_{m,i} = \max(X_{(i-1)m+1}, \dots, X_{im})$ ,  $i \in \{1, \dots, k\}$ . In this case, under some regularity conditions (e.g. [4, Chapter 1]), the weak limit theory establishes that  $F^m(a_mx + b_m)$  converges pointwise to  $G_\gamma(x)$  as  $m \rightarrow \infty$  for every continuity point  $x$  of  $G_\gamma$ , where  $G_\gamma$  is the generalised extreme value (GEV) distribution,  $a_m > 0$  and  $b_m$  are suitable norming constants for each  $m = 1, 2, \dots$ , and  $\gamma \in \mathbb{R}$  is the so-called tail index, which describes the tail heaviness of  $F$  (e.g. [4, Chapter 1]). The second method concerns the modelling of  $k$  random variables out of the  $n$  available that exceed a high threshold  $t$ , or, equivalently, of  $k$  threshold excesses  $Y_j$ ,  $j = 1, \dots, k$ , which are i.i.d. copies of  $Y = X - t \mid X > t$ . In this context, the generalised Pareto (GP) distribution, say  $H_\gamma$ , appears as the weak limit law of appropriately normalised high threshold exceedances, i.e.  $F_t(s(t)x)$  converges pointwise to  $H_\gamma(x)$  as

---

Received 26 December 2022; accepted 22 June 2023.

\* Postal address: Department of Decision Sciences, via Roentgen 1, 20136, Milan, Italy. Email: [simone.padoan@unibocconi.it](mailto:simone.padoan@unibocconi.it)

\*\* Postal address: Department of Statistical Science, via Lanzone 18, 20123, Milan, Italy. Email: [stefano.rizzelli@unicatt.it](mailto:stefano.rizzelli@unicatt.it)

© The Author(s), 2023. Published by Cambridge University Press on behalf of Applied Probability Trust.

$t \rightarrow x^*$  for all the continuity points  $x$  of  $H_\gamma(x)$ , where  $F_t(x) = \mathbb{P}(Y \leq x)$  and  $s(t) > 0$  is a suitable scaling function for any  $t \leq x^*$ , with  $x^* = \sup\{x : F(x) < \infty\}$ . This result motivates the POT approach, which was introduced decades ago by the seminal paper [1]. Since then, few other convergence results have emerged. For instance, the uniform convergence of  $F_t(s(t) \cdot)$  to  $H_\gamma$  and the corresponding convergence rate have been derived by [11, 12], respectively. Similar results but in Wasserstein distance have recently been established by [2]. As for the GEV distribution, more results are available. In particular, there are sufficient conditions to ensure, in addition to weak convergence, that  $F^m(a_m \cdot + b_m)$  converges to  $G_\gamma$ , for example, uniformly and in variational distance, and the density of  $F^m(a_m \cdot + b_m)$  converges pointwise, locally uniformly, and uniformly to that of  $G_\gamma$  (e.g. [7, Chapter 2]; [13, Chapter 2]).

The main contribution of this article is to provide new convergence results that can be useful in practical problems for the POT approach. Motivated by the utility in the statistical field of assessing the asymptotic accuracy of estimation procedures, we study stronger forms of convergence than the pointwise one, as  $\lim_{t \rightarrow x^*} \mathcal{D}(F_t(s(t) \cdot); H_\gamma) = 0$ , where  $\mathcal{D}(\cdot; \cdot)$  is either the variational distance, the Hellinger distance, or the Kullback–Leibler divergence. In particular, we provide upper bounds for the rate of convergence to zero of  $\mathcal{D}(F_t(s(t) \cdot); H_\gamma)$  in the case that  $\mathcal{D}(\cdot; \cdot)$  is the variational and Hellinger distance, and further translate them into bounds on Kullback–Leibler divergence between the densities of  $F_t(s(t) \cdot)$  and  $H_\gamma$ , respectively. We also pinpoint cases where recentering of exceedances is necessary to reach the optimum rate, namely where  $F_t(s(t) \cdot + c(t))$  has to be considered in place of  $F_t(s(t) \cdot)$ , for a suitable real valued function  $c(t)$ .

The article is organised as follows. Section 2 provides a brief summary of the probabilistic context on which our results are based. Section 3 provides our new results on strong convergence to a GP distribution. Section 4 provides the proofs of the main results.

## 2. Background

Let  $X$  be a random variable with a distribution function  $F$  that is in the domain of attraction of the GEV distribution  $G_\gamma$ , denoted as  $F \in \mathcal{D}(G_\gamma)$ . This means that there are norming constants  $a_m > 0$  and  $b_m \in \mathbb{R}$  for  $m = 1, 2, \dots$  such that

$$\lim_{m \rightarrow \infty} F^m(a_m x + b_m) = \exp(-(1 + \gamma x)^{-1/\gamma}) =: G_\gamma(x), \quad (2.1)$$

for all  $x \in \mathbb{R}$  such that  $1 + \gamma x > 0$ , where  $\gamma \in \mathbb{R}$ , and this is true if and only if there is a scaling function  $s(t) > 0$  with  $t < x^*$  such that

$$\lim_{t \rightarrow x^*} F_t(s(t)x) = 1 - (1 + \gamma x)^{-1/\gamma} =: H_\gamma(x), \quad (2.2)$$

(e.g. [4, Theorem 1.1.6]). The densities of  $H_\gamma$  and  $G_\gamma$  are  $h_\gamma(x) = (1 + \gamma x)^{-(1/\gamma+1)}$  and  $g_\gamma(x) = G_\gamma(x)h_\gamma(x)$ , respectively. Let  $U(v) := F^{\leftarrow}(1 - 1/v)$  for  $v \geq 1$ , where  $F^{\leftarrow}$  is the left-continuous inverse function of  $F$  and  $G^{\leftarrow}(\exp(-1/x)) = (x^\gamma - 1)/\gamma$ . Then, we recall that the first-order condition in (2.1) is equivalent to the limit result

$$\lim_{v \rightarrow \infty} \frac{U(vx) - U(v)}{a(v)} = \frac{x^\gamma - 1}{\gamma}, \quad (2.3)$$

for all  $x > 0$ , where  $a(v) > 0$  is a suitable scaling function. In particular, we can set  $s(t) = a(1/(1 - F(t)))$ ; see [4, Chapter 1] for possible selections of the function  $a$ .

A stronger convergence form than that in formula (2.2) is the uniform one, i.e.

$$\sup_{x \in [0, (x^* - t)/s(t)]} |F_t(s(t)x) - H_\gamma(x)| \rightarrow 0, \quad t \rightarrow x^*.$$

In the case of distributions  $F$  with finite end-point  $x^*$ , the following slightly more general form of convergence is also of interest

$$\sup_{x \in [0, (x^* - t - c(t))/s(t)]} |F_t(s(t)x + c(t)) - H_\gamma(x)| \rightarrow 0, \quad t \rightarrow x^*,$$

for a centering function  $c(t)$  satisfying  $c(t)/s(t) \rightarrow 0$  as  $t \rightarrow x^*$ . To establish the speed at which  $F_t(s(t)x)$  or  $F_t(s(t)x + c(t))$  converges uniformly to  $H_\gamma(x)$ , [12] relied on a specific formulation of the well-known second-order condition. In its general form, the second-order condition requires the existence of a positive function  $a$  and a positive or negative function  $A$ , named the rate function, such that  $\lim_{v \rightarrow \infty} |A(v)| = 0$  and

$$\lim_{v \rightarrow \infty} \frac{(U(vx) - U(v))/a(v) - (x^\gamma - 1)/\gamma}{A(v)} = D(x), \quad x > 0,$$

where  $D$  is a non-null function which is not a multiple of  $(x^\gamma - 1)/\gamma$  [4, Definition 2.3.1]. The rate function  $A$  is necessarily regularly varying at infinity with index  $\rho \leq 0$ , named the second-order parameter [4, Theorem 2.3.3]. In the following, we use the same specific form of second-order condition of [12] to obtain decay rates for stronger metrics than uniform distance between distribution functions.

### 3. Strong results for POT

In this section we discuss strong forms of convergence for the distribution of renormalised exceedances over a threshold. First, in Section 3.1, we discuss convergence to a GP distribution in variational and Hellinger distance, drawing a connection with known results for density convergence of normalized maxima. In Section 3.2 we quantify the speed of convergence in variational and Hellinger distance. Moreover, we show how these can be used to also bound Kullback–Leibler divergences. Throughout, for a twice-differentiable function  $W(x)$  on  $\mathbb{R}$ , we denote by  $W'(x) = (\partial/\partial x)W(x)$  and  $W''(x) = (\partial^2/\partial x^2)W(x)$  the first- and second-order derivatives, respectively.

#### 3.1. Strong convergence under classical assumptions

Let the distribution function  $F$  be twice differentiable. We write  $f = F'$ ,  $g_m = (F^m(a_m \cdot + b_m))'$ , and  $f_t = F'_t$ . Under the classical von Mises-type conditions

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{xf(x)}{1 - F(x)} &= \frac{1}{\gamma}, \quad \gamma > 0, \\ \lim_{x \rightarrow x^*} \frac{(x^* - x)f(x)}{1 - F(x)} &= -\frac{1}{\gamma}, \quad \gamma < 0, \\ \lim_{x \rightarrow x^*} \frac{f(x) \int_x^{x^*} (1 - F(v)) \, dv}{(1 - F(x))^2} &= 0, \quad \gamma = 0, \end{aligned} \tag{3.1}$$

we know that the first-order condition in (2.3) is satisfied, and that

$$\lim_{v \rightarrow \infty} va(v)f(a(v)x + U(v)) = (1 + \gamma x)^{-1/\gamma - 1} \tag{3.2}$$

locally uniformly for  $(1 + \gamma x) > 0$ . Since the equality  $g_m(x) = F^{m-1}(a_mx + b_m)h_m(x)$  holds with  $b_m = U(m)$ ,  $a_m = a(m)$ , and  $h_m(x) = ma_m f(a_mx + b_m)$ , and since  $F^{m-1}(a_mx + b_m)$  converges to  $G_\gamma(x)$  locally uniformly as  $m \rightarrow \infty$ , the convergence result in (3.2) thus implies that  $g_m(x)$  converges to  $g_\gamma(x)$  locally uniformly [13, Section 2.2].

On the other hand, the density pertaining to  $F_t(s(t)x)$  is

$$l_t(x) := f_t(s(t)x)s(t) = \frac{s(t)f(s(t)x + t)}{1 - F(t)}$$

and, setting  $v = 1/(1 - F(t))$ , we have  $a(v) = s(t)$  and  $v \rightarrow \infty$  as  $t \rightarrow x^*$ . Therefore, a further implication of the convergence result in (3.2) is that  $l_t(x)$  converges to  $h_\gamma(x)$  locally uniformly for  $x > 0$  if  $\gamma \geq 0$ , or for  $x \in (0, -1/\gamma)$  if  $\gamma < 0$ .

In turn, by Scheffe's lemma we have  $\lim_{t \rightarrow x^*} \mathcal{V}(\mathcal{P}_t, \mathcal{P}) = 0$ , where  $\mathcal{V}(\mathcal{P}_t; \mathcal{P}) = \sup_{B \in \mathbb{B}} |\mathcal{P}_t(B) - \mathcal{P}(B)|$  is the total variation distance between the probability measures  $\mathcal{P}_t(B) := \mathbb{P}((X - t)/s(t) \in B \mid X > t)$  and  $\mathcal{P}(B) := \mathbb{P}(Z \in B)$ , and where  $Z$  is a random variable with distribution  $H_\gamma$  and  $B$  is a set in the Borel  $\sigma$ -field of  $\mathbb{R}$ , denoted by  $\mathbb{B}$ . Let  $\mathcal{H}^2(l_t; h_\gamma) := \int [\sqrt{l_t(x)} - \sqrt{h_\gamma(x)}]^2 dx$  be the square of the Hellinger distance. It is well known that the Hellinger and total variation distances are related as

$$\mathcal{H}^2(l_t; h_\gamma) \leq 2\mathcal{V}(\mathcal{P}_t; \mathcal{P}) \leq 2\mathcal{H}(l_t; h_\gamma) \quad (3.3)$$

(see, e.g., [9, Appendix B]). Therefore, the conditions in (3.1) ultimately entail that the Hellinger distance between the density of rescaled peaks over a threshold  $l_t$  and the GP density  $h_\gamma$  also converges to zero as  $t \rightarrow x^*$ . In the next subsection we introduce a stronger assumption, allowing us to also quantify the speed of such convergence.

### 3.2. Convergence rates

As in [12] we rely on the following assumption, in order to derive the convergence rate for the variational and Hellinger distances.

**Condition 3.1.** Assume that  $F$  is twice differentiable. Moreover, assume that there exists  $\rho \leq 0$  such that

$$A(v) := \frac{vU''(v)}{U'(v)} + 1 - \gamma$$

defines a function of constant sign near infinity, whose absolute value  $|A(v)|$  is regularly varying as  $v \rightarrow \infty$  with index of variation  $\rho$ .

When Condition 3.1 holds then the classical von Mises conditions in (3.1) are also satisfied for the cases where  $\gamma$  is positive, negative, or equal to zero, respectively. Furthermore, Condition 3.1 implies that an appropriate scaling function for the exceedances of a high threshold  $t < x^*$ , which complies with the equivalent first-order condition (2.2), is defined as  $s(t) = (1 - F(t))/f(t)$ . With such a choice of the scaling function  $s$ , we establish the following results.

**Theorem 3.1.** Assume Condition 3.1 is satisfied. Then, there exist constants  $c > 0$ ,  $\alpha_j > 0$  with  $j = 1, 2$ ,  $K > 0$ , and  $t_0 < x^*$ , depending on  $\gamma$ , such that

$$\frac{\mathcal{H}^2(\tilde{l}_t; h_\gamma)}{K|A(v)|^2} \leq S(v)$$

for all  $t \geq t_0$ , where  $v = 1/(1 - F(t))$ ,

$$\tilde{l}_t = \begin{cases} l_t, & \gamma \geq 0, \\ l_t(\cdot + (x^* - t)/s(t) + 1/\gamma), & \gamma < 0, \end{cases}$$

and  $S(v) = 1 - |A(v)|^{\alpha_1} + 4 \exp(c|A(v)|^{\alpha_2})$ .

Note that  $\tilde{l}_t$  is the density of  $F_t(s(t) \cdot + c(t))$ , with centring function

$$c(t) := \begin{cases} 0, & \gamma \geq 0, \\ x^* - t + \gamma^{-1}s(t), & \gamma < 0, \end{cases} \tag{3.4}$$

for  $t < x^*$ . Given the relationship between the total variation and Hellinger distances in (3.3), with obvious adaptations when a non-null recentring is considered, the following result is a direct consequence of Theorem 3.1.

**Corollary 3.1.** *Under the assumptions of Theorem 3.1, for all  $t \geq t_0$ ,  $\mathcal{V}(\tilde{\mathcal{P}}_t; \mathcal{P}) \leq |A(v)|\sqrt{KS(v)}$ , with  $\tilde{\mathcal{P}}_t$  the probability measure pertaining to  $\tilde{l}_t$ .*

Theorem 3.1 implies that when  $\gamma \geq 0$  the Hellinger and variational distances of the probability density and measure of rescaled exceedances from their GP distribution counterparts are bounded from above by  $C_1|A(v)|$ , for a positive constant  $C_1$ , as the threshold  $t$  approaches the end-point  $x^*$ . Since, for a fixed  $x \in \cap_{t \geq t_0}(0, (x^* - t)/s(t))$ ,  $|F_t(s(t)x) - H_\gamma(x)| \leq \mathcal{V}(\mathcal{P}_t; \mathcal{P})$ , and since [12, Theorem 2(i)] implies that  $|F_t(s(t)x) - H_\gamma(x)|/|A(v)|$  converges to a positive constant, there also exists  $C_0 > 0$  such that, for all large  $t$ ,  $C_0|A(v)|$  is a lower bound for the variational and Hellinger distances. Therefore, since  $C_0|A(v)| \leq \mathcal{V}(\mathcal{P}_t; \mathcal{P}) \leq \mathcal{H}(l_t; h_\gamma) \leq C_1|A(v)|$ , the decay rate of the variational and Hellinger distances is precisely  $|A(v)|$  as  $t \rightarrow x^*$ . When  $\gamma < 0$ , analogous considerations apply to  $\tilde{l}_t$  and  $\tilde{\mathcal{P}}_t$ . With the following results, we give precise indications of when a recentred version of  $l_t$  is necessary to achieve the optimal rate.

**Proposition 3.1.** *Under the assumptions of Theorem 3.1, when  $\gamma < 0$  there are constants  $c_j$ ,  $j = 1, 2$ , and  $t_1 < x^*$ , depending on  $\gamma$ , such that, for all  $t > t_1$ ,*

$$c_1|A(v)|^{-1/2\gamma} < \mathcal{H}(h_\gamma; h_\gamma(\cdot - \mu_t)) < c_2|A(v)|^{\min(1, -1/2\gamma)},$$

where  $\mu_t := c(t)/s(t)$  and  $c(t)$  is as in the second line of (3.4).

**Corollary 3.2.** *Under the assumptions of Theorem 3.1:*

- (i) *when  $-1/2 \leq \gamma < 0$ , there are constants  $c_3 > 0$  and  $t_2 < x^*$ , depending on  $\gamma$ , such that, for all  $t > t_2$ ,  $\mathcal{H}(l_t; h_\gamma) \leq c_3|A(v)|$ ;*
- (ii) *when  $\gamma < -1/2$ , there are constants  $c_4 > 0$  and  $t_3 < x^*$ , depending on  $\gamma$ , such that, for all  $t > t_3$ ,  $\mathcal{H}(l_t; h_\gamma) \geq c_4|A(v)|^{-1/2\gamma}$ .*

According to Corollary 3.2(ii), the density  $l_t$  of rescaled exceedances  $Y/s(t)$  does not achieve the optimal convergence rate  $|A(V)|$  whenever  $\gamma < -1/2$ , in which case the rate is only of order  $|A(v)|^{-1/2\gamma}$ . In simple terms, this is due to the fact that, when  $\gamma$  is negative, the supports of  $l_t$  and  $h_\gamma$  can be different and the approximation error is affected by the amount of probability mass in the unshared region of points. Indeed, we recall that the

end-point  $(x^* - t)/s(t)$  of  $l_t$  converges to  $-1/\gamma$  as  $t$  approaches  $x^*$  at rate  $A(v)$  (see, e.g., [4, Lemma 4.5.4]). Nevertheless, for  $t < x^*$  it can be that  $(x^* - t)/s(t) > -1/\gamma$  or  $(x^* - t)/s(t) < -1/\gamma$ . In turn, when  $\gamma$  is smaller than  $-1/2$ , the approximation error due to support mismatch has a dominant effect. However, if scaled exceedances are shifted by subtracting the quantity  $\mu_t$ , in this case the upper end-point of the density  $\tilde{l}_t$  is the same of that of  $h_\gamma$ , hence no support mismatch occurs and the optimal convergence rate is also achieved in the case where  $\gamma < -1/2$ .

A further implication of Theorem 3.1 concerns the speed of convergence to zero of the Kullback–Leibler divergence  $\mathcal{K}(\tilde{l}_t; h_\gamma) := \int \ln\{\tilde{l}_t(x)/h_\gamma(x)\}\tilde{l}_t(x) dx$ , and the divergences of higher order  $p \geq 2$ ,  $\mathcal{D}_p(\tilde{l}_t; h_\gamma) := \int |\ln\{\tilde{l}_t(x)/h_\gamma(x)\}|^p \tilde{l}_t(x) dx$ . Using the uniform bound on density ratio provided in Lemma 4.7 we are able to translate the upper bounds on the squared Hellinger distance  $\mathcal{H}^2(\tilde{l}_t; h_\gamma)$  into upper bounds on the Kullback–Leibler divergence  $\mathcal{K}(\tilde{l}_t; h_\gamma)$  and higher-order divergences  $\mathcal{D}_p(\tilde{l}_t; h_\gamma)$ .

**Corollary 3.3.** *Under the assumptions of Theorem 3.1, in particular with  $\rho < 0$  and  $\gamma \neq 0$ , there exist constants  $M > 0$  and  $t_4 < x^*$ , depending on  $\gamma$ , such that, for all  $t \geq t_4$ ,*

- (i)  $\mathcal{K}(\tilde{l}_t; h_\gamma) \leq 2MKS(v)|A(v)|^2$ ;
- (ii)  $\mathcal{D}_p(\tilde{l}_t; h_\gamma) \leq 2p!KS(v)|A(v)|^2$ , with  $p \geq 2$ .

To extend the general results in Lemma 4.7 and Corollary 3.3 to the case of  $\gamma = 0$  seems to be technically over-complicated. Nevertheless, there are specific examples where the properties listed in such lemmas are satisfied, such as the following.

**Example 3.1.** Let  $F(x) = \exp(-\exp(-x))$ ,  $x \in \mathbb{R}$ , be the Gumbel distribution function. In this case, Condition 3.1 is satisfied with  $\gamma = 0$  and  $\rho = -1$ , so that Theorem 3.1 applies to this example, and, for an arbitrarily small  $\varepsilon > 0$ , we have  $l_t(x)/h_0(x) \leq \exp(\exp(-t)) < 1 + \varepsilon$  for all  $x > 0$  and suitably large  $t$ . Hence, the bounded density ratio property is satisfied and it is still possible to conclude that  $\mathcal{D}_p(l_t; h_0)/|A(v)|^2$  and  $\mathcal{K}(l_t; h_0)/|A(v)|^2$  can be bounded from above as in Corollary 3.3.

## 4. Proofs

### 4.1. Additional notation

For  $y > 0$  we write  $T(y) = U(e^y)$  and, for  $t < x^*$ , we define the functions

$$p_t(y) = \begin{cases} (T(y + T^{-1}(t)) - t)/s(t) - (e^{\gamma y} - 1)/\gamma, & \gamma > 0, \\ (T(y + T^{-1}(t)) - t)/s(t) - y, & \gamma = 0, \\ (T(y + T^{-1}(t)) - x^* - \gamma^{-1}s(t))/s(t) - (e^{\gamma y} - 1)/\gamma, & \gamma < 0, \end{cases}$$

with  $s(t) = (1 - F(t))/f(t)$ , and

$$q_t(y) = \begin{cases} (1/\gamma) \ln [1 + \gamma e^{-\gamma y} p_t(y)], & \gamma \neq 0, \\ p_t(y), & \gamma = 0. \end{cases}$$

Moreover, for  $t < x^*$  we set

$$\tilde{x}_t^* = \begin{cases} (x^* - t)/s(t), & \gamma \geq 0, \\ -1/\gamma, & \gamma < 0. \end{cases}$$

Furthermore, for  $x \in (0, x^* - t)$ , we let  $\phi_t(x) = T^{-1}(x + t) - T^{-1}(t)$ . Finally, for  $x \in \mathbb{R}$ ,  $\gamma \in \mathbb{R}$ , and  $\rho \leq 0$ , we set

$$I_{\gamma, \rho}(x) = \begin{cases} \int_0^x e^{\gamma s} \int_0^s e^{\rho z} dz ds, & \gamma \geq 0, \\ -\int_x^\infty e^{\gamma s} \int_0^s e^{\rho z} dz ds, & \gamma < 0. \end{cases}$$

### 4.2. Auxiliary results

In this section we provide some results which are auxiliary to the proofs of the main ones, presented in Section 3. Throughout, for Lemmas 4.1–4.6, Condition 3.1 is implicitly assumed to hold. The proofs are provided in the Supplementary Material. In particular, Lemmas 4.1 and 4.2 are used directly in the proof of our main result, Theorem 3.1.

**Lemma 4.1.** *For every  $\varepsilon > 0$  and every  $\alpha > 0$ , there exist  $x_1 < x^*$  and  $\kappa_1 > 0$  (depending on  $\gamma$ ) such that, for all  $t \geq x_1$  and  $y \in (0, -\alpha \ln |A(e^{T^{-1}(t)})|)$ ,*

$$\exp\{-\kappa_1 |A(e^{T^{-1}(t)})| e^{2\varepsilon y}\} < e^{-q_t(y)} < \exp\{\kappa_1 |A(e^{T^{-1}(t)})| e^{2\varepsilon y}\}.$$

**Lemma 4.2.** *For every  $\varepsilon > 0$  and every  $\alpha > 0$ , there exist  $x_2 < x^*$  and  $\kappa_2 > 0$  (depending on  $\gamma$ ) such that, for all  $t \geq x_2$  and  $y \in (0, -\alpha \ln |A(e^{T^{-1}(t)})|)$ ,*

$$\exp\{-\kappa_2 |A(e^{T^{-1}(t)})| e^{2\varepsilon y}\} < 1 + q'_t(y) < \exp\{\kappa_2 |A(e^{T^{-1}(t)})| e^{2\varepsilon y}\}.$$

**Lemma 4.3.** *If  $\gamma > 0$  and  $\rho < 0$ , there exists a regularly varying function  $\mathcal{R}$  with negative index  $\varrho$  such that, defining the function*

$$\eta(t) := \frac{(1 + \gamma t)f(t)}{1 - F(t)} - 1,$$

as  $v \rightarrow \infty$ ,  $\eta(U(v)) = O(\mathcal{R}(v))$ .

**Lemma 4.4.** *If  $\gamma > 0$  and  $\rho < 0$ , there exists  $x_3 \in (0, \infty)$  and  $\delta > 0$  such that, for all  $x \geq x_3$ ,  $f(x) = h_\gamma(x)[1 + O(\{1 - H_\gamma(x)\}^\delta)]$ .*

**Lemma 4.5.** *If  $\gamma < 0$  and  $\rho < 0$ , there exists a regularly varying function  $\tilde{\mathcal{R}}$  with negative index  $\tilde{\varrho} = (-1) \vee (-\rho/\gamma)$  such that, defining the function*

$$\tilde{\eta}(y) := \frac{(1 - \gamma y)f(x^* - 1/y)}{[1 - F(x^* - 1/y)]y^2} - 1,$$

as  $y \rightarrow \infty$ ,  $\tilde{\eta}(y) = O(\tilde{\mathcal{R}}(y))$ .

**Lemma 4.6.** *If  $\gamma < 0$  and  $\rho < 0$ , there exist  $\tilde{\delta} > 0$  such that, as  $y \rightarrow \infty$ ,*

$$\frac{f(x^* - 1/y)}{y^2} = (1 - \gamma y)^{1/\gamma - 1} [1 + O(\{1 - H_{-\gamma}(y)\}^{\tilde{\delta}})].$$

Finally, in order to exploit Theorem 3.1 to give bounds on Kullback–Leibler and higher-order divergences, we introduce by the next lemma a uniform bound on density ratios.

**Lemma 4.7.** *Under the assumptions of Theorem 3.1, if  $\rho < 0$  and  $\gamma \neq 0$ , then there exist a  $t^* < x^*$  and a constant  $M \in (0, \infty)$  such that*

$$\sup_{t \geq t^*} \sup_{0 < x < \tilde{x}_t^*} \frac{\tilde{l}_t(x)}{h_\gamma(x)} < M.$$

### 4.3. Proof of Theorem 3.1

*Proof of Theorem 3.1.* For every  $x_t > 0$ ,

$$\begin{aligned} & \mathcal{H}^2(l_t; h_\gamma) \\ &= \int_0^{x_t} + \int_{x_t}^\infty \left[ \sqrt{f_t(x)} - \sqrt{h_\gamma((x - c(t))/s(t))/s(t)} \right]^2 dx \\ &\leq \int_0^{\phi_t(x_t)} e^{-y} \left[ 1 - \sqrt{e^{-q_t(y)}(1 + q_t'(y))} \right]^2 dy + \left[ \sqrt{1 - F_t(x_t)} + \sqrt{1 - H_\gamma\left(\frac{x_t - c(t)}{s(t)}\right)} \right]^2 \\ &=: \mathcal{I}_1(t) + \mathcal{I}_2(t). \end{aligned}$$

Let  $x_t$  be such that  $\phi_t(x_t) = -\alpha \ln |A(e^{T^{-1}(t)})|$  for a positive constant  $\alpha$  to be specified later. Then, by Lemmas 4.1 and 4.2, for a suitably small  $\varepsilon > 0$  there exists  $\kappa_3 > 0$  such that, for all sufficiently large  $t$ ,

$$\mathcal{I}_1(t) \leq \int_0^{-\alpha \ln |A(e^{T^{-1}(t)})|} \kappa_3 |A(e^{T^{-1}(t)})|^2 e^{(4\varepsilon - 1)y} dy \leq \kappa_3 |A(e^{T^{-1}(t)})|^2 [1 - |A(e^{T^{-1}(t)})|^{\alpha_1}],$$

where  $\alpha_1 := \alpha(1 - 4\varepsilon)$  is positive. Moreover, on one hand we have the identity  $1 - F_t(x_t) = |A(e^{T^{-1}(t)})|^\alpha$ . On the other hand, for some constant  $\kappa_5 > 0$  we have the inequality

$$\begin{aligned} 1 - H_\gamma\left(\frac{x_t - c(t)}{s(t)}\right) &= |A(e^{T^{-1}(t)})|^\alpha \exp\{-q_t(-\alpha \ln |A(e^{T^{-1}(t)})|)\} \\ &\leq |A(e^{T^{-1}(t)})|^\alpha \exp\{\kappa_5 |A(e^{T^{-1}(t)})|^{1-2\varepsilon\alpha}\}. \end{aligned}$$

Consequently,

$$\mathcal{I}_2(t) \leq |A(e^{T^{-1}(t)})|^\alpha \left[ 1 + \exp\left\{\frac{\kappa_5}{2} |A(e^{T^{-1}(t)})|^{1-2\varepsilon\alpha}\right\} \right].$$

Now, we can choose  $\alpha > 2$  and  $\varepsilon$  small enough that  $|A(e^{T^{-1}(t)})|^\alpha < |A(e^{T^{-1}(t)})|^2$  and  $\alpha_2 := 1 - 2\varepsilon\alpha > 0$ . The conclusion then follows noting that  $T^{-1}(t) = -\ln(1 - F(t))$  and, in turn,  $|A(e^{T^{-1}(t)})| = |A(v)|$ .  $\square$



### 4.4. Proof of Proposition 3.1

*Proof of Proposition 3.1.* Assume first that, for all large  $t$ ,  $\mu_t = (x^* - t)/s(t) + 1/\gamma$  is positive. In this case, we have the following identities:

$$\begin{aligned} \mathcal{H}^2(h_\gamma; h_\gamma(\cdot - \mu_t)) &= \int_0^{-1/\gamma} \left[ \sqrt{h_\gamma(x)} - \sqrt{h_\gamma(x - \mu_t)} \right]^2 dx + 1 - H_\gamma\left(-\frac{1}{\gamma} - \mu_t\right) \\ &= \int_0^\infty e^{-s} \left[ \sqrt{\frac{\exp\{-(1/\gamma) \ln(1 - \gamma e^{-\gamma s} \mu_t)\}}{1 - \gamma e^{-\gamma s} \mu_t}} - 1 \right]^2 ds + (-\gamma \mu_t)^{-1/\gamma}. \end{aligned}$$

Concerning the first term on the right-hand side, for all  $s > 0$  as  $t \rightarrow x^*$  we have  $0 \leq \ln(1 - \gamma e^{-\gamma s} \mu_t) \leq \ln(1 - \gamma \mu_t) = O(\mu_t)$  and

$$1 \geq \frac{1}{1 - \gamma e^{-\gamma s} \mu_t} \geq \frac{1}{1 - \gamma \mu_t} = 1 + O(\mu_t),$$

where, for a positive constant  $\tau$ ,  $\mu_t$  satisfies  $\mu_t/|A(v)| = (1 + o(1))\tau$  (see, e.g., [4, Lemma 4.5.4]). Then, as  $t \rightarrow x^*$ ,

$$\int_0^\infty e^{-s} \left[ \sqrt{\frac{\exp\{-(1/\gamma) \ln(1 - \gamma e^{-\gamma s} \mu_t)\}}{1 - \gamma e^{-\gamma s} \mu_t}} - 1 \right]^2 ds = O(\mu_t^2) = O(|A(v)|^2).$$

Concerning the second term, as  $t \rightarrow x^*$ ,  $(-\gamma \mu_t)^{-1/\gamma} = (-\gamma)^{-1/\gamma} (1 + o(1))\tau^{-1/\gamma} |A(v)|^{-1/\gamma}$ . The result now follows for the case where  $\mu_t$  is ultimately positive. When it is ultimately negative, simply note that

$$\mathcal{H}^2(h_\gamma; h_\gamma(\cdot - \mu_t)) = \mathcal{H}^2(h_\gamma(\cdot + \mu_t); h_\gamma) = \mathcal{H}^2(h_\gamma(\cdot - (-\mu_t)); h_\gamma), \tag{4.1}$$

and proceed as above, but replacing  $\mu_t$  with  $-\mu_t$ . □

### 4.5. Proof of Corollary 3.2

*Proof of Corollary 3.2.* Observe that, for  $\gamma < 0$ ,  $\mathcal{H}(l_t; h_\gamma) = \mathcal{H}(\tilde{l}_t; h_\gamma(\cdot + \mu_t))$ . Moreover, note that, by the triangular inequality and the first identity in (4.1),

$$\begin{aligned} \mathcal{H}(h_\gamma; h_\gamma(\cdot - \mu_t)) - \mathcal{H}(\tilde{l}_t; h_\gamma(\cdot + \mu_t)) &\leq \mathcal{H}(\tilde{l}_t; h_\gamma(\cdot + \mu_t)) \\ &\leq \mathcal{H}(h_\gamma; h_\gamma(\cdot - \mu_t)) + \mathcal{H}(\tilde{l}_t; h_\gamma(\cdot + \mu_t)). \end{aligned}$$

Whenever  $\gamma \geq -1/2$ , by Theorem 3.1 and Proposition 3.1, the term on the second line is of order  $O(|A(v)|)$  as  $t \rightarrow x^*$ . Instead, if  $\gamma < -1/2$ , by Theorem 3.1 and Proposition 3.1, as  $t \rightarrow x^*$  the term on the left-hand side of the first line satisfies

$$\begin{aligned} \mathcal{H}(h_\gamma; h_\gamma(\cdot - \mu_t)) - \mathcal{H}(\tilde{l}_t; h_\gamma(\cdot + \mu_t)) &\geq c_3 |A(v)|^{-1/2\gamma} - O(|A(v)|) \\ &= c_3 |A(v)|^{-1/2\gamma} (1 + o(1)). \end{aligned}$$

The result now follows. □

#### 4.6. Proof of Corollary 3.3

*Proof of Corollary 3.3.* By [8, Lemma 8.2],

$$\mathcal{H}(\tilde{l}_t; h_\gamma) \leq 2 \left[ \sup_{0 < x < \tilde{x}_t^*} \frac{\tilde{l}_t(x)}{h_\gamma(x)} \right] \mathcal{H}^2(\tilde{l}_t; h_\gamma).$$

Moreover, by [9, Lemma B.3], for  $p \geq 2$ ,

$$\mathcal{D}_p(\tilde{l}_t; h_\gamma) \leq 2p! \left[ \sup_{0 < x < \tilde{x}_t^*} \frac{\tilde{l}_t(x)}{h_\gamma(x)} \right] \mathcal{H}^2(\tilde{l}_t; h_\gamma).$$

The conclusion now follows by combining the above inequalities and applying Theorem 3.1 and Lemma 4.7.  $\square$

#### Acknowledgements

Simone Padoan is supported by the Bocconi Institute for Data Science and Analytics (BIDSA), Italy.

#### Funding information

There are no funding bodies to thank relating to the creation of this article.

#### Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

#### Supplementary material

The supplementary material for this article can be found at <https://doi.org/10.1017/jpr.2023.53>

#### References

- [1] BALKEMA, A. A. AND DE HAAN, L. (1974). Residual life time at great age. *Ann. Prob.* **2**, 792–804.
- [2] BOBBIA, B., DOMBRY, C. AND VARRON, D. (2021). The coupling method in extreme value theory. *Bernoulli* **27**, 1824–1850.
- [3] BÜCHER, B. AND ZHOU, C. (2021). A horse race between the block maxima method and the peak-over-threshold approach. *Statist. Sci.* **36**, 360–378.
- [4] DE HAAN, L. AND FERREIRA, A. (2006). *Extreme Value Theory: An Introduction*. Springer, New York.
- [5] DEY, D. K. AND YAN, J. (2016). *Extreme Value Modeling and Risk Analysis: Methods and Applications*. CRC Press, New York.
- [6] EMBRECHTS, P., KLÜPPELBERG, C. AND MIKOSCH, T. (2013). *Modelling Extremal Events: For Insurance and Finance*. Springer, Berlin.
- [7] FALK, M., HÜSLER, J. AND REISS, R.-D. (2010). *Laws of Small Numbers: Extremes and Rare Events*. Birkhäuser, Basel.
- [8] GHOSAL, S., GHOSH, J. K. AND VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28**, 500–531.
- [9] GHOSAL, S. AND VAN DER VAART, A. W. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press.

- [10] KULIK, R. AND SOULIER, P. (2020). *Heavy-Tailed Time Series*. Springer, New York.
- [11] PICKANDS III, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119–131.
- [12] RAOULT J.-P. AND WORMS, R. (2003). Rate of convergence for the generalized Pareto approximation of the excesses. *Adv. Appl. Prob.* **35**, 1007–1027.
- [13] RESNICK, S. I. (2007). *Extreme Values, Regular Variation, and Point Processes*. Springer, New York.