Location- and scale-free procedures for distinguishing between distribution tail models

Igor Rodionov[1]

[1]*School of Mathematics, Statistics and Actuarial Science, University of Essex, UK,*
e-mail: igor.rodionov@essex.ac.uk

Abstract: We consider distinguishing between two distribution tail models when tails of one model are lighter (or heavier) than those of the other. Two procedures are proposed: one scale-free and one location- and scale-free, and their asymptotic properties are established. We show the advantage of using these procedures for distinguishing between certain tail models in comparison with the tests proposed in the literature by simulation and apply them to data on daily precipitation in Green Bay, US and Jena, Germany.

1. Introduction

Most methods used in the statistical analysis of distribution tails are provided by extreme value theory and involve the estimation of parameters of one of the general distribution tail models. Applications of this theory are very diverse, from finance and insurance to environment and biology. The peaks-over-threshold (POT) approach based on the Generalized Pareto distribution (GP) model is typically used for estimating distribution tails, whereas the Gumbel block maxima method and the approach based on the extremal index are usually applied for estimating the maxima of time series (see, e.g., [1] or [2]). However, whereas financial and insurance data are usually heavy-tailed and are well approximated by the GP model, the tails of environmental and biological data may be lighter.

The GP model and generalized extreme value (GEV) distributions have been widely applied to fit extremes of environmental data, see, e.g., [3] or [4]. However, [5] and [6] showed that fitting time series maxima using the Weibull model outperforms the classical Gumbel block maxima method on rainfall and flood data, respectively. Moreover, Weibull-type and log-Weibull-type tail models have shown their efficiency in estimating extremes of spectrometric data [7], lifetimes of cancer patients [8] and even financial time series [9]. Surprisingly, there is almost no work on distinguishing between these tail models. The present article is devoted to this problem.

Let $\overline{F}(x) = 1 - F(x)$ denote the tail distribution function of a cdf F, and let $x_F^* := \sup\{y : F(y) < 1\}$ denote its right endpoint. Following [10], we say that $\overline{F}$ and $\overline{G}$ are strictly tail-equivalent, if $x_F^* = x_G^* =: x^*$ and $\overline{F}(x)/\overline{G}(x) \to 1$ as $x \uparrow x^*$ (write $\overline{F} \sim \overline{G}$). We call *the (right) distribution tail* the equivalence class $T(\overline{F})$ of tail distribution functions, i.e., for all $\overline{G}$ the property $\overline{G} \sim \overline{F}$ is equivalent to $\overline{G} \in T(\overline{F})$.

Let $\mathbf{X}_n = (X_1, \ldots, X_n)$ be i.i.d. random variables with cdf F and with $x_F^* = +\infty$. Below we propose a location- and scale-free model selection procedure for distinguishing between two tail models A_0 and A_1 using $\mathbf{X}_n$. Here A_0 and A_1 are families of distribution tails that are "separable" by some cdf F_0; in particular, all tails in A_0 are lighter (or heavier) than those in A_1. We also propose a scale-free procedure for distinguishing A_0 from A_1, that may be more powerful than the first. Both procedures are extensions of those proposed in [11], which are neither location- nor scale-free.

The article is organized as follows. In Section 2, we discuss Weibull-type and log-Weibull-type models for distribution tails. In Section 3.1 we introduce the notion of C-separability, and then scale-free and location- and scale-free model selection procedures for distinguishing between C-separable classes of distribution tails are proposed in Sections 3.2 and 3.3, respectively, and their asymptotic properties are established. These procedures can be used for distinguishing between regularly-varying, Weibull-type and log-Weibull-type tails, though their applicability is not limited to this. The selection of the optimal threshold level for distinguishing between distribution tail models is discussed in Section 3.4. In Section 4, the performance of our methods is examined on simulated and real data sets and compared with methods in the literature. All proofs are postponed to Section 5.

2. Recent models for distribution tails

The classical approach to handling extremes presupposes that the distribution function F of observations belongs to the maximum domain of attraction (MDA) of some extreme value distribution, written $F \in \mathcal{D}(EV_\gamma)$, i.e., for some sequences $a_n > 0$ and b_n,

$$F^n(a_n x + b_n) \to EV_\gamma(x), \quad n \to \infty, \tag{1}$$

where

$$EV_\gamma(x) = \begin{cases} \exp(-(1 + \gamma x)^{-1/\gamma}), \ 1 + \gamma x > 0, & \text{if } \gamma \neq 0, \\ \exp(-\exp(-x)), \ x \in \mathbb{R}, & \text{if } \gamma = 0, \end{cases}$$

is the generalized extreme value distribution and γ is the extreme value index. If $F \in \mathcal{D}(EV_\gamma)$, then as $u \uparrow x_F^*$, the distribution of excesses over a high threshold u

$$F_u(x) = \frac{F(x + u) - F(u)}{1 - F(u)}, \quad 0 \leq x \leq x_F^* - u,$$

can be approximated by the Generalized Pareto distribution $GP_\gamma(x/\sigma(u))$, where

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

according to [12] and [13].

The efficiency of the GP model to fit distribution tails belonging to the Gumbel MDA (that is, if $\gamma = 0$) is sometimes unsatisfactory, so several models for tails within the Gumbel MDA have been proposed in last two decades. First, consider the model developed in [14] (see also the particular case of this model proposed in [15]). Following the description of this model in [16], assume that the distribution tail of random variable X (let $X \geq 1$ a.s. for simplicity) is given by

$$1 - F(x) = \exp(-V^\leftarrow(\ln x)), \ x \geq 1, \tag{2}$$

where $V^\leftarrow(x) := \inf\{y : V(y) \geq x\}$ is the generalized inverse of $V(x) = \ln u_F(1/\exp(x))$ with u_F, the tail quantile function of F, that is $u_F(t) = \overline{F}^\leftarrow(1/t)$. It is also supposed that there exists a positive function a, such that for all $t > 0$ and some real ϑ,

$$\lim_{x\to\infty} \frac{V(tx) - V(x)}{a(x)} = \frac{t^\vartheta - 1}{\vartheta}, \tag{3}$$

where for $\vartheta = 0$ the right-hand side should be understood as $\ln t$. Hence, V is supposed to be of extended regular variation with index ϑ, see p. 371 in [1]. If $\vartheta > 0$ then ϑ governs the tail behavior of the log-Weibull-type distribution, i.e., F is such that

$$\lim_{x\to\infty} \frac{\ln(1 - F(e^{tx}))}{\ln(1 - F(e^x))} = t^{1/\vartheta}, \quad t > 0. \tag{4}$$

Representatives of this class are the lognormal distribution and distributions with cdf $F(x) = 1 - \exp(-(\ln x)^{1/\vartheta})$, for $x > 1$. If $0 < \vartheta < 1$, then F belongs to the Gumbel MDA and if $\vartheta = 1$, then F belongs to the Fréchet MDA, i.e., if $\gamma > 0$ (see Theorem 1 in [16]). The estimation of ϑ, extreme quantiles and probabilities of rare events for this model are considered in [14], [16], [17], and [18].

Another class of distributions from the Gumbel MDA is the class of Weibull-type distributions [19], [20]. We say that F is of Weibull type if there exists $\theta > 0$ such that

$$\lim_{x\to\infty} \frac{\ln(1 - F(tx))}{\ln(1 - F(x))} = t^{1/\theta}, \quad t > 0. \tag{5}$$

Then θ is called the Weibull tail index. Representatives of this class include the normal, exponential, gamma and Weibull distributions. The model (2) does not provide a good description for Weibull-type distributions, since in (3) the parameter ϑ is equal to 0 for all distributions from this class as well as for distributions with lighter tails. The problems of estimation of the Weibull tail index and extreme quantiles for Weibull-type distributions are investigated by [15], [19], [21], [22], and [23], among others.

The literature on distinguishing between the Gumbel and Fréchet MDA is quite extensive [24], [25], but there is an almost complete lack of work on distinguishing between distribution tails within the Gumbel MDA. Apart from [11], we mention [20], where the goodness-of-fit test for Weibull-type behavior was proposed, and two author's papers [26], dealing with distinguishing between Weibull-type distribution tails, and rather theoretical [27], where the close hypotheses about distribution tails within the Gumbel MDA are considered. Note that two latter papers are based on quite different principle (the likelihood ratio method) and much less general than this article. The tests proposed in these four articles are not invariant with respect to scale and location changes, though the test of [20] is scale-free. However, poor estimation or incorrect setting of the location and scale parameters can lead to serious errors in conclusions (see, e.g., Figure 6), which motivates developing location- and scale-free alternatives.

3. Asymptotic results

Hereinafter we assume that all considered distributions are continuous and their right endpoints are infinite. Our methods can be extended to finite right endpoints, but this is beyond the scope of our paper. The model selection procedures introduced in Sections 3.2 and 3.3 are generalizations of those of [11], which are neither location- nor scale-free.

Hereafter for model selection procedures considered below, we call the choice of an alternative class for distributions from a null class the type I error. Also, we call the probability of selecting an alternative class if a distribution belongs to it the power of model selection procedure.

3.1. C-condition and C-separability. Examples

Denote the tail quantile function of F by $u_F(t)$, that is $u_F(t) = \inf\{x : F(x) = 1 - 1/t\}$. Let us introduce the conditions that we use to establish the asymptotic properties of the model selection procedures proposed in this work.

Definition 3.1. Two cdfs H and G are said to satisfy the C-condition $C_\delta(H, G)$, if for some $\delta \geq 0$ there exists t_0 such that for all $t > t_0$ and $c > 1$,

$$\frac{u_H(c^{1+\delta}t)}{u_G(ct)} \leq \frac{u_H(t)}{u_G(t)}. \tag{6}$$

If H and G satisfy the C_0-condition $C_0(H, G)$, then starting from some t_0, the ratio $u_H(t)/u_G(t)$ does not increase. Notice also that $C_\delta(H, G)$ is weaker than $C_{\delta'}(H, G)$ for $\delta' > \delta \geq 0$.

It is easy to see that $C_\delta(H, G)$ is invariant with respect to changes of the scale parameters of G and H. However, these conditions may be violated if $\sigma_1 u_H(t) + a_1$ and $\sigma_2 u_H(t) + a_2$ are substituted for $u_H(t)$ and $u_G(t)$, respectively, but this does not prevent us from proving the results below.

The condition $C_\delta(H, G)$ implies the following relation, which we use in proving the main results.

Proposition 3.2. *Assume that cdfs H and G satisfy $C_\delta(H,G)$ for some $\delta \geq 0$ and t_0. Then for all $t > t_0$ and $c > 1$,*

$$\frac{\left(1 - H(cu_H(t))\right)^{1-\varepsilon}}{(1 - H(u_H(t)))^{1-\varepsilon}} \leq \frac{1 - G(cu_G(t))}{1 - G(u_G(t))}, \tag{7}$$

with $\varepsilon = 1 - (1+\delta)^{-1}$.

Now we introduce the notion of C-separability.

Definition 3.3. Let A_0 and A_1 be two classes of distribution tails such that the tails belonging to A_0 are lighter than those in A_1. We call A_0 and A_1 C-**separable** (from the right) if there exists a cdf F_0 such that for all $G \in A_0$ the condition $C_0(G, F_0)$ holds for some $t_0 = t_0(G)$, and for all $H \in A_1$ the condition $C_\delta(F_0, H)$ holds for some $\delta > 0$ and $t_0 = t_0(H)$.

If classes A_0 and A_1 are C-separable from the right by cdf F_0, we call them $C(F_0)$-separable from the right.

If the distribution tails in A_0 are heavier then those in A_1, then we say that these two classes are $C(F_0)$-separable (from the left), if for some cdf F_0 and all $G \in A_0$ the condition $C_0(F_0, G)$ holds for some $t_0 = t_0(G)$, and for all $H \in A_1$ the condition $C_\delta(H, F_0)$ holds for some $\delta > 0$ and $t_0 = t_0(H)$.

C-separability is satisfied for a wide range of classes of distribution tails, as is now discussed.

Example 3.4. Let A_0 and A_1 be the classes of Weibull-type and log-Weibull-type distribution tails, respectively (see Section 2), and suppose that all such functions with tails in A_0 and A_1 are eventually differentiable. To distinguish between these two classes we recommend selecting the cdf

$$F_0(x) = (1 - \exp\{-\exp\{b(\ln x)^{1/2}\}\})I(x > 1) \tag{8}$$

for some constant $b > 0$. The proof of $C(F_0)$-separability of these classes from the right under some technical conditions is given in Section 5. The proof of their $C(F_0)$-separability from the left is similar. Finite-sample properties of procedures for distinguishing A_0 from A_1 are compared in Section 4.3.

Example 3.5. Let A_0 be the class of log-Weibull-type distribution tails with $\vartheta < 1$, and let A_1 be the class of regularly-varying distribution tails. Recall that the Fréchet MDA consists of only the distributions with regularly-varying tails, that is $F \in \mathcal{D}(EV_\gamma)$, $\gamma > 0$, for such distribution functions. Assume again that all distribution functions with tails in the classes A_0 and A_1 are eventually differentiable. To distinguish between these two classes we recommend selecting the cdf

$$F_0(x) = 1 - \exp(-\exp(b\sqrt{\ln \ln x})\ln x), \quad x > e, \tag{9}$$

for some constant $b > 0$. Indeed, the tails of log-Weibull-type distributions with $\vartheta < 1$ are lighter than that of F_0, and the condition $C_0(G, F_0)$ for $G \in A_0$ follows from (4) and the properties of regularly-varying functions. On the other hand,

the condition $C_\delta(F_0, H)$ for $H \in \mathcal{D}(EV_\gamma), \gamma > 0$, and $\delta > 0$ follows from (9) and the definition of regular variation. We do not provide the detailed proofs of these facts, because they are technical and in many respects similar to the proofs of corresponding facts in Example 1. $C(F_0)$-separability from the right holds in this case as well. Numerical comparisons of performance of procedures and tests for distinguishing between these two classes are provided in Section 4.4.

Example 3.6. Let the class A_0 be such that $u_G(t) = \int_1^t f(x)/x \, dx$ for all $G \in A_0$, where f is positive and $f(x) \to \infty$ as $x \to \infty$. It is clear that G is heavy-tailed. Next, let A_1 be the class of the tails of absolutely continuous Weibull-type distributions with $\theta < 1$. Then the classes A_0 and A_1 are $C(F_0)$-separable from the left by F_0, the cdf of the standard exponential distribution. We omit the proof of this fact, since it is similar to the proof of Example 1.

Example 3.7. Consider a one-parameter family of distributions $F \in \{G_\theta, \theta \in \Theta\}$ such that for all $\theta_1, \theta_2 \in \Theta$, $\theta_1 < \theta_2$, the cdfs G_{θ_1} and G_{θ_2} satisfy the condition $C_\delta(G_{\theta_1}, G_{\theta_2})$ for $\delta > 0$. Then it is easy to see that the classes $A_0 = \{G_\theta, \ \theta < \theta_0\}$ and $A_1 = \{G_\theta, \ \theta > \theta_0\}$ are $C(G_{\theta_0})$-separable from both the right and the left.

3.2. Scale-free model selection

Assume that classes A_0 and A_1 are $C(F_0)$-separable from the right by some cdf F_0. In this section, we discuss the following procedure for distinguishing A_0 from A_1:

$$\text{select} \begin{cases} A_0, & \text{if } \tilde{R}_{k,n} \leq 1 + \frac{u_{1-\alpha}}{\sqrt{k}}, \\ A_1, & \text{otherwise,} \end{cases} \tag{10}$$

where $u_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the standard normal distribution,

$$\tilde{R}_{k,n} = \ln \frac{k}{n} - \frac{1}{k} \sum_{i=n-k+1}^{n} \ln \left(1 - F_0(u_0(n/k)X_{(i)}/X_{(n-k)})\right), \tag{11}$$

and $u_0(t) = u_{F_0}(t)$. Here α is the asymptotic upper bound on the probabilities of selecting A_1 for distribution tails in A_0 obtained when applying (10) to distinguish between classes A_0 and A_1; it can be considered as analogous to a significance level.

The statistic $\tilde{R}_{k,n}$ is obtained by replacing the order statistics $\{X_{(i)}\}_{i=n-k}^{n}$ with their scale-free counterparts $\{u_0(n/k)X_{(i)}/X_{(n-k)}\}_{i=n-k}^{n}$ from the statistic

$$R_{k,n} = \ln(1 - F_0(X_{(n-k)})) - \frac{1}{k} \sum_{i=n-k+1}^{n} \ln(1 - F_0(X_{(i)})) \tag{12}$$

introduced in [11] and inspired by the Hill estimator [29] of the positive extreme value index γ. The statistic $R_{k,n}$ has fruitful distributional properties allowing to propose a procedure similar to (10). Specifically, if $F = F_0$, the distribution

of $R_{k,n}$ is known, whereas if the right tail of F is heavier or lighter than that of F_0 then $R_{k,n}$ tends in probability to $+\infty$ or $-\infty$, respectively, as $k = k(n) \to \infty, k/n \to 0, n \to \infty$. However, the result of applying this procedure to finite samples might depend strongly on changes in the scale parameters of F and F_0, which is a significant drawback of the method and motivates introducing the scale-free statistics $\tilde{R}_{k,n}$.

Unlike $R_{k,n}$, the distribution of $\tilde{R}_{k,n}$ does not depend on the scale parameters of F_0 and F, so the procedure (10) is scale-free. However, $\tilde{R}_{k,n}$ preserves distributional properties similar to those of $R_{k,n}$, see Theorem 3.9 and Theorem 3.10 below. If for some cdf F_0 classes A_0 and A_1 are $C(F_0)$-separable from the left, then we use the following procedure to distinguish A_0 from A_1:

$$\text{select} \begin{cases} A_0, & \text{if } \tilde{R}_{k,n} \geq 1 + \frac{u_\alpha}{\sqrt{k}}, \\ A_1, & \text{otherwise}, \end{cases}$$

where α and u_α have the same meaning as above.

Now let us discuss the asymptotic properties of the procedure (10). To formulate them, we need the following condition, which is classical for extreme value theory, see, e.g., formula (1.1.30) in [1].

Condition 3.8. We say that a cdf F satisfies the von Mises condition if for some $\gamma \in \mathbb{R}$

$$\lim_{x \uparrow x_F^*} \frac{(1 - F(x))F''(x)}{(F'(x))^2} = -\gamma - 1. \tag{13}$$

It is known (see, e.g., Theorem 1.1.8 ibid.) that the von Mises condition is sufficient for F to belong to $\mathcal{D}(EV_\gamma)$. Hereinafter we call a natural sequence $k = k(n)$ satisfying

$$k \to \infty, k/n \to 0 \text{ as } n \to \infty, \tag{14}$$

an intermediate sequence, which is also classical for statistics of extremes.

The first result gives the behavior of the statistic $\tilde{R}_{k,n}$ if the cdf F of a sample $X_1, \ldots, X_n$ coincides with F_0.

Theorem 3.9. *Let* $X_1, \ldots, X_n$ *be i.i.d. random variables with a cdf F_0 satisfying Condition 3.8. If a sequence $k = k(n)$ is intermediate, then*

$$\sqrt{k}(\tilde{R}_{k,n} - 1) \xrightarrow{d} N(0, 1), \qquad n \to \infty.$$

Two further results together with Theorem 3.9 imply that the upper bound over the type I error probabilities of the procedure (10) for distribution tails from A_0 is at most α asymptotically, and the power of (10) tends to 1 as $n \to \infty$.

Theorem 3.10. *Let* $X_1, \ldots, X_n$ *be i.i.d. random variables with a cdf F_1. Assume that F_0 satisfies Condition 3.8 and a sequence $k = k(n)$ is intermediate. If the condition $C_\delta(F_0, F_1)$ holds for some $\delta > 0$, then*

$$\sqrt{k}(\tilde{R}_{k,n} - 1) \xrightarrow{P} +\infty, \qquad n \to \infty,$$

and if $C_\delta(F_1, F_0)$ holds for some $\delta > 0$, then

$$\sqrt{k}(\tilde{R}_{k,n} - 1) \xrightarrow{P} -\infty, \qquad n \to \infty.$$

Corollary 3.11. *Under the assumptions of Theorem 3.10, let $X_1^0, \ldots, X_n^0$ be i.i.d. random variables with the cdf F_0. Denote the values of the statistic (11) built by samples $(X_1^0, \ldots, X_n^0)$ and $(X_1, \ldots, X_n)$ by $\tilde{R}_{k,n}^0$ and $\tilde{R}_{k,n}^1$, respectively. If $C_0(F_0, F_1)$ holds, then for all $x > 0$ eventually*

$$P(\tilde{R}_{k,n}^0 > x) \le P(\tilde{R}_{k,n}^1 > x),$$

and if $C_0(F_1, F_0)$ holds, then for all $x > 0$ eventually

$$P(\tilde{R}_{k,n}^0 > x) \ge P(\tilde{R}_{k,n}^1 > x).$$

Remark 3.12. Although the procedure (10) possesses the asymptotic properties mentioned above for all F_0 satisfying Condition 3.8 and such that classes A_0 and A_1 are $C(F_0)$-separable, the choice of F_0 matters for finite n. However, the problem of optimal selection of F_0 is beyond the scope of this paper. This remark is related also to the procedure proposed in the next section. However, some practical recommendations can be given; see Section 4.1.

Remark 3.13. It follows from the results of this section that if classes A_0 and A_1 are $C(F_0)$-separable and F_0 does not belong to A_0, then the type I error probabilities of (10) when applying it to distinguish A_0 from A_1 are asymptotically lower than α. Thus, if A_0 has a natural "left bound" $\tilde{F}$, i.e., for all $G \in A_0$ the condition $C_0(G, \tilde{F})$ holds, but $C_\delta(G, \tilde{F})$ does not necessary hold for any $\delta > 0$, and then we recommend selecting $F_0 = \tilde{F}$ to maximize the power of the procedure.

3.3. Location- and scale-free model selection

The procedure proposed in the previous section and the tests mentioned at the end of Section 2 are not invariant with respect to the location parameter, which can restrict their application in some situations; see, e.g., Fig. 6. In this section, we propose a procedure for distinguishing between separable classes of distribution tails that is invariant with respect to both scale and location parameters. Let us consider the statistic

$$\widehat{R}_{k,n} = \ln \frac{k}{n} - \tag{15}$$
$$\frac{1}{k} \sum_{i=n-k+1}^{n} \ln \left[1 - F_0 \left(u_0(n/k) + \frac{X_{(i)} - X_{(n-k)}}{X_{(n-k)} - X_{(n-2k)}} \left(u_0(n/k) - u_0(n/(2k)) \right) \right) \right],$$

which can be obtained from $R_{k,n}$ (12) by replacing the order statistics $\{X_{(i)}\}_{i=n-k}^{n}$ with their location- and scale-free counterparts

$$u_0(n/k) + \frac{X_{(i)} - X_{(n-k)}}{X_{(n-k)} - X_{(n-2k)}} \left(u_0(n/k) - u_0(n/(2k)) \right), \quad i = n - k, \ldots, n.$$

The asymptotic properties of $\widehat{R}_{k,n}$ are stated below.

Theorem 3.14. *Under the conditions of Theorem 3.9,*

$$\sqrt{k}(\widehat{R}_{k,n} - 1) \xrightarrow{d} N\Big(0, 1 + \frac{\gamma^2}{2(\gamma+1)^2(2^\gamma - 1)^2}\Big), \qquad n \to \infty,$$

where the ratio on the right-hand side should be understood as $1/(2(\ln 2)^2)$ for $\gamma = 0$.

Theorem 3.15. *Assume the conditions of Theorem 3.10. If the condition $C_\delta(F_0, F_1)$ holds for some $\delta > 0$ then*

$$\sqrt{k}(\widehat{R}_{k,n} - 1) \xrightarrow{P} +\infty, \qquad n \to \infty,$$

and if $C_\delta(F_1, F_0)$ holds for some $\delta > 0$ then

$$\sqrt{k}(\widehat{R}_{k,n} - 1) \xrightarrow{P} -\infty, \qquad n \to \infty.$$

Corollary 3.16. *Assume the conditions of Theorem 3.10. Let $X_1^0, \ldots, X_n^0$ be i.i.d. random variables with the cdf F_0. Denote the values of statistic (15) built by samples $(X_1^0, \ldots, X_n^0)$ and $(X_1, \ldots, X_n)$ by $\widehat{R}_{k,n}^0$ and $\widehat{R}_{k,n}^1$, respectively. If $C_0(F_0, F_1)$ holds then for all $x > 0$ eventually*

$$P(\widehat{R}_{k,n}^0 > x) \leq P(\widehat{R}_{k,n}^1 > x),$$

and if $C_0(F_1, F_0)$ holds then for all $x > 0$ eventually

$$P(\widehat{R}_{k,n}^0 > x) \geq P(\widehat{R}_{k,n}^1 > x).$$

Denote for convenience

$$\sigma^2(\gamma) = 1 + \frac{\gamma^2}{2(\gamma+1)^2(2^\gamma - 1)^2}, \quad \gamma > 0,$$

and set $\sigma^2(0) = 1 + 1/(2(\ln 2)^2)$ as its limit as $\gamma \to 0$. As in the previous section, consider two classes A_0 and A_1 of distribution tails such that tails in A_0 are lighter than those in A_1, and assume that these classes are $C(F_0)$-separable from the right, where $F_0 \in \mathcal{D}(EV_\gamma)$. Let us propose the following procedure to distinguish A_0 from A_1:

$$\text{select} \begin{cases} A_0, & \text{if } \widehat{R}_{k,n} \leq 1 + \frac{\sigma(\gamma)u_{1-\alpha}}{\sqrt{k}}, \\ A_1, & \text{otherwise.} \end{cases} \tag{16}$$

This procedure is clearly location- and scale-free by the definition of $\widehat{R}_{k,n}$. Moreover, Theorem 3.14, Theorem 3.15, and Corollary 3.16 imply that the upper bound over the type I error probabilities of procedure (16) is at most α asymptotically, and its power tends to 1 as $n \to \infty$. The procedure for distinguishing A_0 from A_1 in case of their $C(F_0)$-separability from the left can be proposed similarly to the previous section.

Remark 3.17. Assume that A_0 and A_1 are $C(F_0)$-separable from the right. Formally speaking, the procedures (10) and (16) distinguish not the classes A_0 and A_1, but the wider classes $\hat{A}_0$ and $\hat{A}_1$, where $\hat{A}_0$ consists of all (continuous) cdfs G satisfying condition $C_0(G, F_0)$, and $\hat{A}_1$ consists of all (continuous) cdfs H satisfying $C(F_0, H)$.

Remark 3.18. Distributions belonging to A_0 or A_1 (apart from the cdf F_0 in case $F_0 \in A_0$) do not have to satisfy the assumptions of the extreme value theorem (1).

3.4. The choice of k

An important problem for practitioners is the optimal choice of k. It should not be too small, since the more observations we use, the more accurate our conclusions are, but it should not be too large because, roughly speaking, the tail may "end" earlier and our conclusions about tail behavior in this case may turn out to be incorrect.

There are many papers devoted to the optimal selection of k in more general context. For instance, we refer to [30] for a review of such methods for tail index estimation, see also Section 5.4 in [31], and to [32] for those for extremal index estimation. These methods are most often based on either empirical principles, like choosing the optimal value of estimator from some interval of stability of its values, or on optimization of some metrics (e.g., MSE), often using the bootstrap.

However, the methods of choosing k proposed in the framework of statistics of extremes are unsuitable when distinguishing between distribution tail models. Indeed, statistics of such procedures and tests quite often have a steady increasing or decreasing trend as k increases, so the stability interval method cannot be used, and almost never tends to a certain finite value, implying that the method of optimizing some metric in the form used for parameter estimation cannot be applied either. To the best of our knowledge, so far in papers devoted to testing hypotheses about distribution tails this question has been omitted. Moreover, the most comprehensive review to date of methods for testing hypotheses about distribution tails [24] suggests that the optimal selection of k is an open problem.

Nevertheless, judging by our observations, some regularities can be distinguished. As a rule, a test for a hypothesis about the distribution tail (or a model selection procedure in our case) can be represented as follows

$$\text{if } S(k) > u, \text{ then reject } H_0 \text{ (select } A_1\text{)},$$

where the statistic $S(k)$ is built from the k largest order statistics of a sample, and u is a constant independent of k. For instance, all the procedures considered in this article can be represented in such a way. Moreover, it is often possible to ensure that the limit distribution of $S(k)$ does not depend on k under the null hypothesis (or for some "boundary" distributions, like in this article, see Theorems 3.9 and 3.14), where k is an intermediate sequence, i.e., it satisfies (14). Let

us consider such a case. Then three types of $S(k)$ behavior are most common as k increases, and the type of this behavior does not change on different samples from the same distribution. Below we provide recommendations for testing the null hypothesis (or selecting the distribution tail model) for these three types of behavior:

1. $S(k)$ has a well expressed increasing trend, and its values exceed u almost immediately. In this case, H_0 should be rejected (the class A_1 should be selected);
2. $S(k)$ has a well expressed decreasing trend, and its values become smaller than u almost immediately. In this case, H_0 should not be rejected (the class A_0 should be selected);
3. the values of $S(k)$ oscillate around u for some interval of small values of k. As a rule, there is no stability of the $S(k)$ values in this case. If there is no stable interval of these values below or above level u, then no decision is possible.

Of course, these recommendations are empirical, and the problem of selection of k while testing hypotheses about distribution tails deserves a separate study. Possible solutions to this problem may be the use of multiple hypothesis testing or sequential analysis.

4. Simulation study

In this section, we examine the efficiency of the procedures introduced by the article in comparison with some tests proposed in the literature. Before this, we provide some practical recommendations on selecting the parameters of F_0 for better performance of the procedures. We also apply our methods to data from the Daily Global Historical Climatology Network (GHCNd) [33].

The role of the parameter α below is the following: it denotes the significance level for all tests considered in this section and corresponds to the asymptotic upper bound over the empirical type I error probabilities of the model selection procedures (10) and (16). We set its value equal to 0.05. The number of replicates used to calculate empirical characteristics below is always equal to 2000.

4.1. Practical recommendations on selecting F_0

When discussing the selection of F_0 in Remarks 3.12 and 3.13, we noted that we do not provide a procedure for the optimal choice of F_0 (and, consequently, its parameters). Nevertheless, some practical recommendations can be given.

First note that the statistic used in the location- and scale-free procedure does not depend on the scale and location parameters of F_0, and the statistic used in the scale-free procedure does not depend on the scale parameter of F_0, so these parameters need not be chosen.

Next, assume that classes A_0 and A_1 are $C(F_0)$-separable from the right. If A_0 has a natural "left bound" (see Remark 3.13 for definition and, e.g., Exam-

ples [3.6], [3.7]), then F_0 should be chosen equal to this bound. In this case the scale and location parameters need not be chosen as well.

If A_0 has no natural "left bound", like in Examples [3.4] and [3.5], it is enough to vary only one parameter to derive a good quality procedure. We suggest choosing some "basic" distribution in A_0 quite close to A_1 in terms of tail heaviness (for example, in Section [4.3] we choose the Weibull distribution with cdf $F(x) = 1 - \exp(-x^a), x > 0$, and $a = 0.5$) and selecting the shape parameter of F_0 in such a way that the average empirical type I error probability would be close to α. For this, the preliminary simulation study should be carried out, see Section [4.2].

Finally, we note that using the location- and scale-free procedure is preferable when distinguishing between distributions from the Gumbel MDA, in which the location parameter can strongly affect the tail behavior for small samples (see Fig. [6]). Moreover, given fixed classes A_0 and A_1, for different sample sizes the same separating cdf F_0 can be used.

4.2. Impact of F_0 choice on procedure performance

In this section, we discuss the impact of the choice of separating cdf F_0 on the performance of the procedures proposed by this article.

First, let us give the details of the preliminary simulation procedure mentioned in Section [4.1] whose purpose is to select the optimal value of the shape parameter given a fixed choice of F_0 family. We focus on the performance of the procedure [16] for distinguishing Weibull-type from log-Weibull-type tails, which is discussed in detail in Section [4.3]. We select F_0 given by [8] and show how the analogs of the type I error probabilities and the power of this procedure change depending on the choice of the parameter b. Two distributions are selected:

- the Weibull distribution $\mathcal{W}(0.5, 1)$ with cdf

$$F(x) = 1 - \exp(-x^{0.5}), x > 0, \tag{17}$$

  whose tail is of Weibull type;
- the standard lognormal distribution $\mathcal{LN}(0, 1)$ with density

$$p(x) = \frac{1}{\sqrt{2\pi}x} \exp\big(- (\ln x)^2/2 \big),$$

  whose tail is of log-Weibull type.

Recall that it is the distribution [17] that we propose as "basic" for distinguishing between Weibull-type and log-Weibull-type tails according to the recommendations given in Section [4.1].

For each distribution, we generated 5000 samples of size $n = 2500$ and computed the rejection rates of the procedure [16] for k from 5 up to 500 in steps of 5. In Figure [1] we plot the empirical I type error probabilities (left panel) and
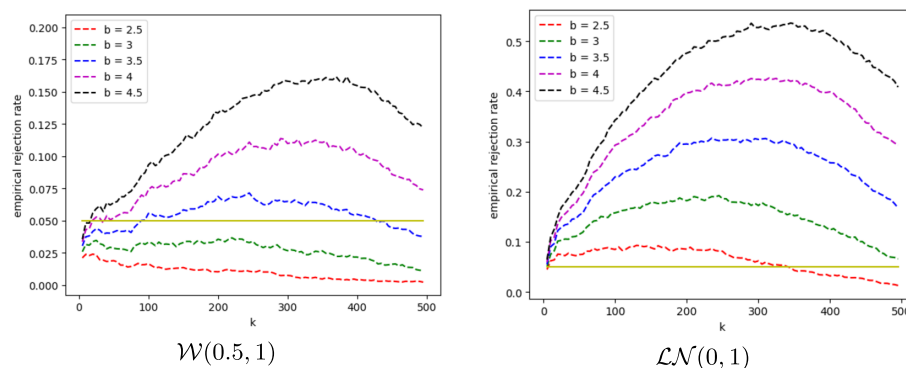
FIG 1. *Empirical type I error probabilities (left) and empirical power (right) of the procedure* (16) *for distinguishing Weibull-type from log-Weibull type tails with F_0 given by* (8) *and 5 different values of its shape parameter b, the level $\alpha = 0.05$ is indicated by a yellow solid line, $n = 2500$.*

empirical power (right panel) of this procedure for $b = 2.5, 3, 3.5, 4$, and 4.5, the shape parameter of F_0 given by (8).

According to Figure 1 and the recommendations given in Section 4.1, the best choice for the parameter b would be 3.5 because the average empirical type I error probability of the procedure (16) on the "basic" distribution $\mathcal{W}(0.5, 1)$ is closer to 0.05 than those for other values of b. Of course, the choice of b strongly affects the power of the procedure, see the right panel of Figure 1: the smaller b the smaller the empirical power.

It is also worth to analyse how much the performance of the procedure changes if we vary the form of the separating cdf F_0. Let us consider distinguishing regularly-varying from Weibull-type distribution tails, in this case it is reasonable to select F_0 of log-Weibull type. Indeed, every log-Weibull-type distribution is separating for these two classes. So, we select four log-Weibull-type distributions as candidates for F_0:

- the log-Weibull distribution $\mathcal{LW}(\lambda, 1)$ with cdf

$$F(x) = 1 - \exp(-(\ln x)^{\lambda}), \quad x > 1, \ \lambda > 0,$$

  and $\lambda = 1.5, 2$ and 3;
- the standard lognormal distribution $\mathcal{LN}(0, 1)$;

and analyse the performance of the procedure (16) on two distributions, where the first is regularly-varying and the second is of Weibull type:

- the generalized Pareto distribution $\mathcal{GP}(0.25, 1)$ with cdf

$$F(x) = 1 - (1 + 0.25x)^{-4}, \quad x > 0;$$

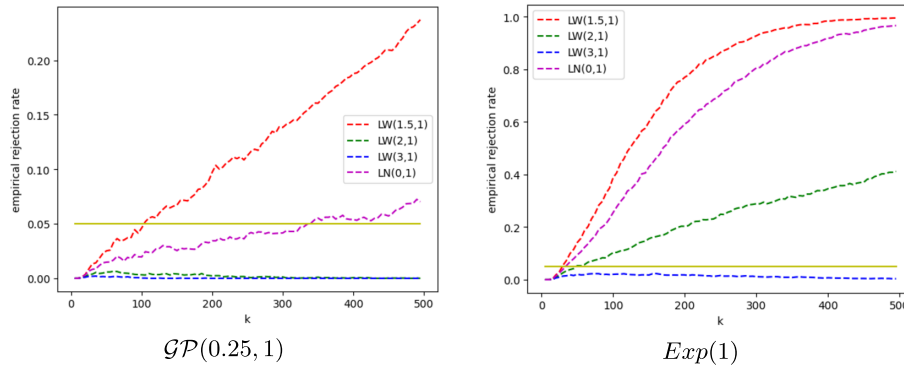- the standard exponential distribution $Exp(1)$.

$\mathcal{GP}(0.25, 1)$ $Exp(1)$

Fɪɢ 2. *Empirical type I error probabilities (left) and empirical power (right) of the procedure* (16) *for distinguishing regularly-varying from Weibull-type tails with four F_0 listed above, the level $\alpha = 0.05$ is indicated by a yellow solid line, $n = 2500$.*

The $\mathcal{GP}(0.25, 1)$ distribution can be considered as "basic" in this example. For each distribution, we generate 5000 samples of size $n = 2500$, compute the rejection rates of the procedure (16) with four candidate F_0 listed above, and plot these rates in Figure 2.

Despite $\mathcal{LW}(1.5, 1)$ is the closest among considered distributions to the class of distributions with regularly-varying tails in terms of tail heaviness, it is not suitable to be a separating cdf for distinguishing between regularly-varying and Weibull-type tails because of high type I error probability. On the other hand, selecting $\mathcal{LW}(3, 1)$ as F_0 for this problem has no sense due to poor performance of the procedure on the $Exp(1)$ distribution. Due to the recommendations given in Section 4.1, it is reasonable to select the $\mathcal{LN}(0, 1)$ distribution as F_0 for this problem (or consider for this purpose the $\mathcal{LN}(0, b)$ distribution with the shape parameter b selected by the procedure described above in this section). It is worth noting the rather different performance of the procedures with $F_0 = \mathcal{LN}(0, 1)$ and $F_0 = \mathcal{LW}(2, 1)$ despite the similar behavior of their tails. This effect can be explained by the different shapes of these distributions which influences the different finite sample behavior of the procedure.

4.3. Distinguishing Weibull-type from log-Weibull-type distribution tails

As mentioned in Section 2, to the best of our knowledge, only one existing publication concerns distinguishing Weibull-type from log-Weibull-type tails, namely [20]. They introduced the Jackson- and Lewis-type goodness-of-fit tests for Weibull-type tail behavior.

We compare the procedures (10) and (16) adapted for distinguishing between the classes $\mathcal{W}$ and $\mathcal{LW}$, i.e., the classes of Weibull-type and log-Weibull-type distribution tails, respectively, with the Jackson- and Lewis-type tests from [20].

According to Remark 3.17 we can consider the distributions with tails heavier than those of $\mathcal{LW}$, in particular, regularly-varying distributions, to be included in $\mathcal{LW}$. As a separating cdf for procedures (10) and (16), we select F_0 given by (8) with b equal to 1.8 and 3.5, respectively, according to the procedure described in Section 4.2. We examine the performance of the procedures and tests listed above on the following set of distributions, which mostly coincides with those used by [20]. First, we list Weibull-type distributions:

- The Weibull distribution $\mathcal{W}(a, \lambda)$

$$F(x) = 1 - \exp(-\lambda x^a), \quad x > 0, \ a, \lambda > 0,$$

  with $(a, \lambda) = (0.5, 1)$.
- The standard normal distribution $\mathcal{N}(0, 1)$.
- The gamma distribution $\Gamma(a, \lambda)$ with density

$$p(x) = \frac{\lambda^a x^{a-1}}{\Gamma(a)} e^{-\lambda x}, \quad x > 0, \ \alpha, \lambda > 0,$$

  with $(a, \lambda) = (0.25, 1), \ (4, 1)$.
- The modified standard exponential distribution $\mathcal{ME}(1)$, defined as the distribution of the random variable $Y = X \ln X$, where $X \sim Exp(1)$.
- The extended Weibull model $\mathcal{EW}(a, \beta)$

$$F(x) = 1 - r(x)e^{-x^a}, \quad x > 0,$$

  where $a > 0$ and $r(x)$ is regularly varying at infinity with index $\beta \in \mathbb{R}$. We set $r(x) = (x+1)^{-1}$ and $a = 2$.

Now we list the log-Weibull-type distributions and distributions with heavier tails which we use in this section:

- The lognormal distribution $\mathcal{LN}(\mu, \sigma)$ with density

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}x} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), \quad x > 0, \ \mu \in \mathbb{R}, \ \sigma^2 > 0,$$

  with $\mu = 0$ and $\sigma^2 = 1$.
- The log-Weibull distribution $\mathcal{LW}(\lambda, c)$ with cdf

$$F(x) = 1 - \exp(-(\ln(x/c))^\lambda), \quad x > c, \ \lambda, c > 0,$$

  and $(\lambda, c) = (1.5, 1)$.
- The generalized Pareto distribution $\mathcal{GP}(\gamma, \sigma)$ with cdf

$$F(x) = 1 - (1 + \gamma x/\sigma)^{-1/\gamma}, \quad x > 0, \ \gamma, \sigma > 0,$$

  with the cases $(\gamma, \sigma) = (0.25, 1), (0.5, 1), (1, 1)$, denoted by P_1, P_2 and P_3, respectively.
- The t-distribution $\mathcal{T}(n)$ with $n = 3$ degrees of freedom.
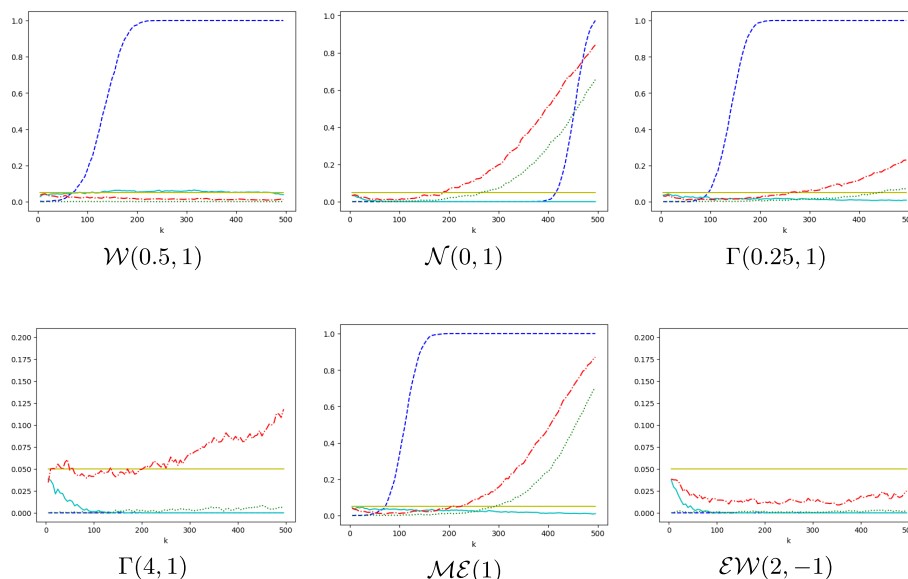
FIG 3. *Empirical type I error probabilities of the procedures* (10) *(blue dashed line),* (16) *(cyan solid line), Jackson-type (green dotted line) and Lewis-type (red dash-dotted line) tests for various distributions, the level* $\alpha = 0.05$ *is indicated by a yellow solid line,* $n = 2500$.

For each distribution, we generated samples of size $n = 2500$ and computed the rejection rates of all procedures and tests for k from 5 up to 500 in steps of 5.

In Figure 3 for the six distributions from $\mathcal{W}$ we plot the empirical type I error probabilities of the scale-free procedure (10), location- and scale-free procedure (16), Jackson- and Lewis-type tests, proposed by [20]. Since $\mathcal{W}$ is the null class, we expect to see empirical type I error probabilities below level α for distributions from $\mathcal{W}$; but only procedure (16) has these probabilities below 0.05 (for $\mathcal{W}(0.5, 1)$, reasonably close to 0.05) for all $k \leq 500$ and all distributions from $\mathcal{W}$. However, the Jackson- and Lewis-type tests provide reasonable empirical type I error probabilities as long as k is not too high, for instance, if $k \leq 200$. We focus on this interval of k values.

In Figure 4 for six distributions from $\mathcal{LW}$ we plot empirical power of the procedures and tests mentioned above. Procedure (16) is more powerful than the Jackson- and Lewis-type tests and procedure (10) is more powerful than the Jackson-type test on distributions from the class $\mathcal{LW}$ when $k \leq 200$.

Analyzing Figs. 3–4, one may conclude that procedure (16) is always preferable to procedure (10). The reason to use (10) instead of (16), in particular, can be a small sample; see Fig. 5. It can also be reasonable to use (10) for distinguishing between quite heavy (in particular, regularly-varying) distribution tails. Fig. 5 shows empirical power of all tests and procedures considered in this section for distributions $\mathcal{LN}(0, 1)$, $\mathcal{GP}(0.25, 1)$ and $\mathcal{T}(3)$ and $n = 1000$. The empirical power of (10) is substantially larger than the empirical power of other tests/procedures for moderate k.
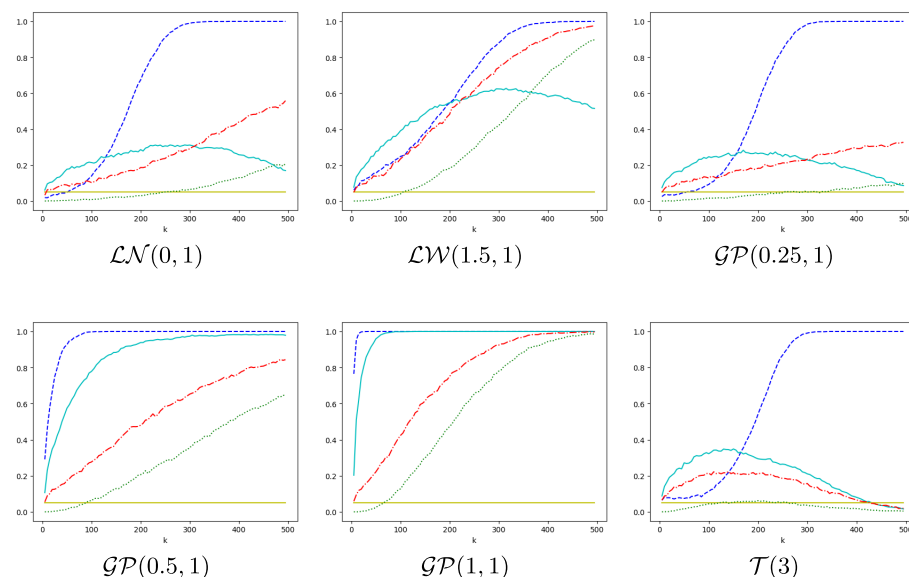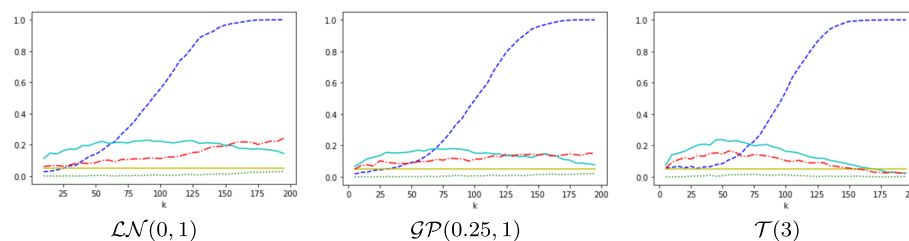
FIG 4. *Empirical power of the procedures* (10) *(blue dashed line),* (16) *(cyan solid line), Jackson-type (green dotted line) and Lewis-type (red dash-dotted line) tests for distributions from* $\mathcal{W}$*, the level* $\alpha = 0.05$ *is indicated by a yellow solid line,* $n = 2500$.

FIG 5. *Empirical power of the procedures* (10) *(blue dashed line),* (16) *(cyan solid line), the Jackson-type (green dotted line) and Lewis-type (red dash-dotted line) tests for distributions from* $\mathcal{LW}$*, the level* $\alpha = 0.05$ *is indicated by a yellow solid line,* $n = 1000$.

However, the Jackson- and Lewis-type tests and the procedure (10) are quite sensitive to the location parameter. In Fig. 6, we provide the empirical type I error probabilities of all the tests and procedures considered in this section for the exponential density $p(x) = \exp(-(x-a))I(x > a)$ and values $a = -1.5, 0, 1.5$ of the location parameter, with $n = 2500$. For $a = 0$, the empirical type I error probabilities of all the tests and procedures are below α for all $k \le 500$ ($k \le 400$ for the test (10)) whereas for $a = -1.5$ the empirical type I error probabilities of all the tests/procedures except (16) increase substantially. Moreover, for $a = 1.5$ these probabilities for Lewis-type test are higher 0.05. Since even small changes of the location parameter can cause large changes in the properties of
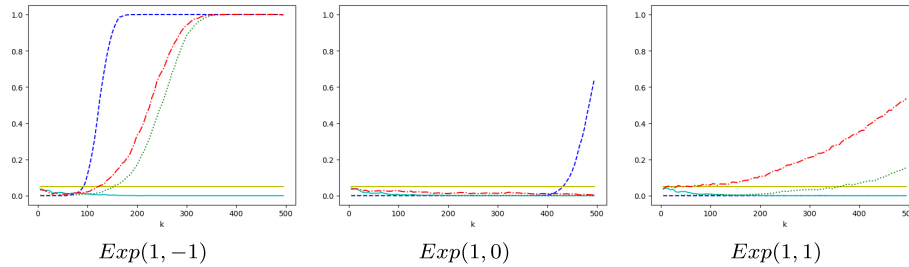
$Exp(1,-1)$ $Exp(1,0)$ $Exp(1,1)$

FIG 6. *Empirical power of the procedures* (10) *(blue dashed line),* (16) *(cyan solid line), the Jackson-type (green dotted line) and Lewis-type (red dash-dotted line) tests for exponential distributions with various values of the location parameter, the level $\alpha = 0.05$ is indicated by a yellow solid line, $n = 2500$.*

the Jackson- and Lewis-type tests as well as in (10), they should be applied with great caution in practice when information about the location parameter is lacking. Moreover, these tests cannot be applied to samples containing only negative observations, an additional difficulty when using them. Summarizing the above, we can conclude that the use of the procedure (16) as compared to the other three procedures considered above is more beneficial for distinguishing between the classes $\mathcal{W}$ and $\mathcal{LW}$.

Remark 4.1. The procedure (16) seems to be quite conservative in the majority of situations (however, the same can be said about the Jackson-type test), the empirical type I error probabilities being very far from the nominal level 0.05. This phenomenon can be explained by the analogy with the uniformly most powerful test for testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ if a parametric family of distributions $\mathcal{P} = \{P_\theta, \theta \in \Theta \subset \mathbb{R}\}$ possesses the monotone likelihood ratio property. This test is in fact conservative for every distribution from the null hypothesis except P_{θ_0} which acts as a "boundary" for the null hypothesis, or as a "separating distribution" (compare with F_0 in our setting) for classes $\mathcal{P}_0 = \{P_\theta, \theta < \theta_0\}$ and $\mathcal{P}_1 = \{P_\theta, \theta > \theta_0\}$. The C_0- and C-conditions are somewhat analogs of the monotone likelihood ratio property for the procedures (10) and (16), which are eventually conservative according to Corollary 3.11 and Corollary 3.16, respectively.

4.4. *Distinguishing log-Weibull-type from regularly-varying distribution tails*

The range of tests that can be used for the problem of this section is quite wide. Here we consider the procedures (10) and (16) adapted for distinguishing between the classes $\mathcal{LW}$ and $\mathcal{RV}$, where $\mathcal{RV}$ is the class of distributions with regularly-varying tails, which coincides with the Fréchet MDA $\mathcal{D}(EV_\gamma)$ for $\gamma > 0$. We compare their efficiency with that of some methods for testing the hypothesis $\tilde{H}_0 : F \in \mathcal{D}(EV_0)$, i.e., the hypothesis of belonging to the Gumbel MDA, against

the alternative $H_1 : F \in \mathcal{D}(EV_\gamma)$, $\gamma > 0$; see [25] for an overview of such methods. Namely, we use

1. the test of [34], see also [35], based on the Shapiro–Wilk type statistic

$$W_n(k) = \frac{k\left(\frac{1}{k}\sum_{i=0}^{k-1} X_{(n-i)} - X_{(n-k+1)}\right)^2}{(k-1)\sum_{j=0}^{k-1}\left(\frac{1}{k}\sum_{i=0}^{k-1} X_{(n-i)} - X_{(n-j)}\right)^2};$$

2. the test proposed in [35] and based on the Greenwood-type statistic $G_n(k)$, a modification of the previous test;
3. the ratio test proposed in [36] and based on the statistic

$$R_n(k) = \frac{X_{(n)} - X_{(n-k)}}{\frac{1}{k}\sum_{i=0}^{k-1} X_{(n-i)} - X_{(n-k)}};$$

4. and the likelihood ratio test, see, e.g., [36].

Numerical comparisons of these tests can be found in [35] and [36], see also [25].

As a separating cdf for (10) and (16) we select F_0 given by (9) with b equal to 0.6 and 1.1, respectively. We compare the efficiency of the tests and procedures listed above on the following distributions (many of which were introduced in the previous section) belonging to the Gumbel MDA:

- the standard exponential distribution $Exp(1)$;
- the Weibull distribution $\mathcal{W}(0.5, 1)$;
- the gamma distribution $\Gamma(0.25, 1)$;
- the modified standard exponential distribution $\mathcal{ME}(1)$;
- the standard lognormal distribution $\mathcal{LN}(0, 1)$;
- the log-Weibull distribution $\mathcal{LW}(\lambda, c)$ with $(\lambda, c) = (1.5, 1), (3, 1)$;

and on others belonging to the Fréchet MDA:

- the generalized Pareto distribution $\mathcal{GP}(\alpha, \sigma)$ with $(\gamma, \sigma) = (1/3, 1)$, $(0.5, 1), (1, 1)$; respectively.
- the standard Cauchy distribution $\mathcal{C}(1)$;
- the t-distribution with 3 degrees of freedom $\mathcal{T}(3)$;
- the Burr distribution $Burr(c, d)$ with cdf

$$F(x) = 1 - \frac{1}{(1 + x^{-c})^d}, \quad x > 0,$$

with $(c, d) = (2, 2), (3, 2)$.

For each distribution we generated samples of size $n = 2500$ and computed the empirical rejection rates of all tests and procedures for k from 5 up to 500. The empirical properties of the Hasofer–Wang, Greenwood-type and likelihood ratio tests are quite similar, so we do not show the results for the last two.

In Figs. 7–8 we plot the empirical type I error probabilities for seven distributions from the Gumbel MDA and empirical power for seven distributions
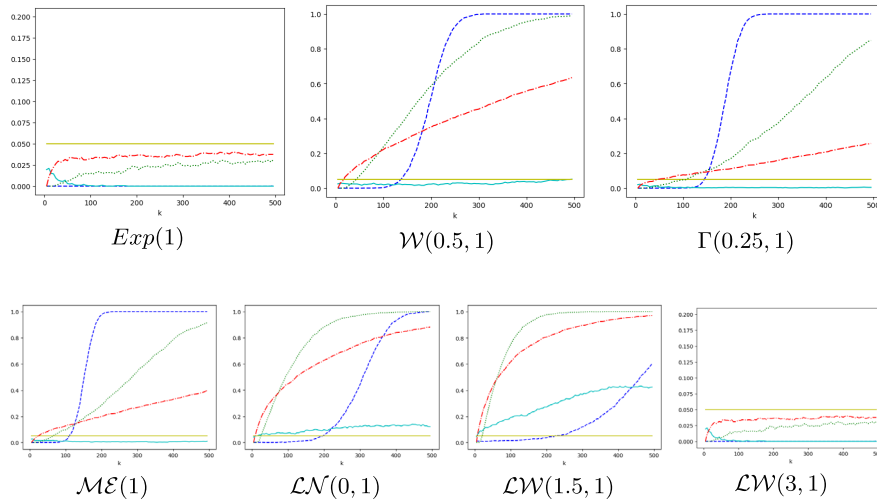
$Exp(1)$ $\mathcal{W}(0.5,1)$ $\Gamma(0.25,1)$

$\mathcal{ME}(1)$ $\mathcal{LN}(0,1)$ $\mathcal{LW}(1.5,1)$ $\mathcal{LW}(3,1)$

Fig 7. *Empirical type I error probabilities of the procedures* (10) *(blue dashed line),* (16) *(cyan solid line), the Hasofer–Wang (green dotted line) and ratio (red dash-dotted line) for various distributions, the level $\alpha = 0.05$ is indicated by a yellow solid line, $n = 2500$.*

from the Fréchet MDA of the scale-free procedure (10), location- and scale-free procedure (16), Hasofer–Wang and ratio tests.

For all the distributions from the Gumbel MDA, except for $Exp(1)$ and $\mathcal{LW}(3,1)$, the empirical type I error probabilities of the Hasofer–Wang and ratio tests are greater than α for almost all k values. On the other hand, the empirical type I error probabilities of procedure (16) exceed α only for $\mathcal{LN}(0,1)$ and $\mathcal{LW}(1.5,1)$. However, even for these two distributions, these probabilities are much less than those of the Hasofer–Wang and ratio tests. Procedure (10) is conservative for all $k < 100$ on all the distributions considered from the class $\mathcal{LW}$. But one should not forget that this procedure is the only one in our comparison that is not location-free. On the other hand, all the tests and procedures considered show quite good efficiency on the distributions from the class $\mathcal{RV}$, except poor performance of (10) and (16) on the distributions $\mathcal{T}(3)$ and $Burr(3,2)$.

Summarizing the results of this section, we can conclude that the procedure (16) outperforms other tests/procedures analyzed in this section in terms of its empirical properties on the distributions from class $\mathcal{LW}$ but might have less power on those of $\mathcal{RV}$, like on $\mathcal{GP}(1/3,1)$, $\mathcal{T}(3)$, and $Burr(3,2)$. However, the high type I error probabilities of Hasofer–Wang and ratio tests do not allow us to consider them adequate for testing the hypothesis $H_0 : F \in \mathcal{LW}$ and even $\tilde{H}_0 : F \in \mathcal{D}(EV_0)$. This means that tests/procedures for distinguishing between $\mathcal{D}(EV_0)$ and $\mathcal{D}(EV_\gamma)$, $\gamma > 0$, based on tail heaviness, might be more efficient than tests based on the properties of the GEV and GP models.
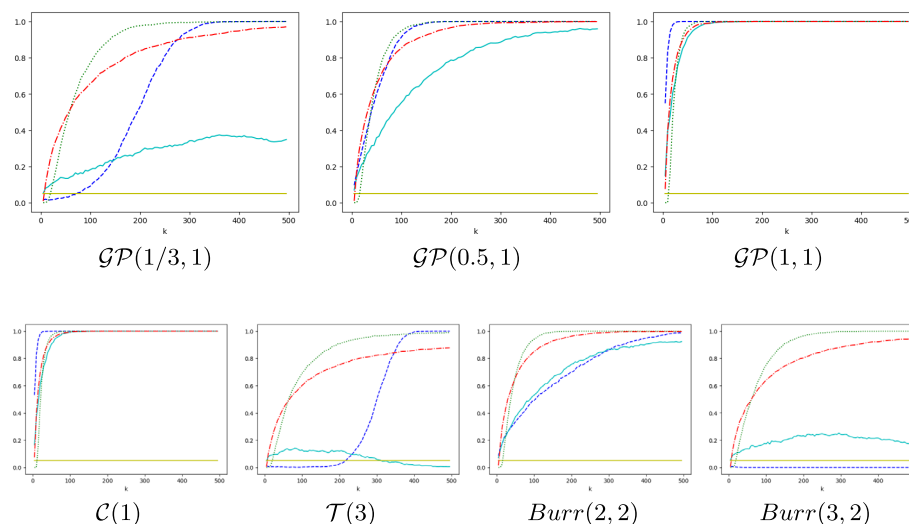
FIG 8. *Empirical power of the procedures* (10) *(blue dashed line),* (16) *(cyan solid line), the Hasofer–Wang (green dotted line) and ratio (red dash-dotted line) tests for various distributions, the significance level $\alpha = 0.05$ is indicated by a yellow solid line, $n = 2500$.*

4.5. Real data example

In this section, the Jackson- and Lewis-type tests from [20] and the procedure (16) for distinguishing $\mathcal{W}$ from $\mathcal{LW}$ (see Section 4.3) are compared on daily precipitation data collected in Green Bay, US and Jena, Germany; the total number of observations in both datasets is 3630. The data was taken from the Daily Global Historical Climatology Network (GHCNd), an integrated database of daily climate summaries from land surface stations across the globe, [33]. The two datasets cover the period 1901–2021, but for stationarity reasons, only days in April were considered, leading to 1331 and 1693 non-zero observations in Green Bay and Jena, respectively. According to [4] (see paragraph [38] there) which analyzes the GHCNd data as we do, there is no evidence of monotonic trend in the data. Moreover, to check for possible clustering of extremes, we estimate the extremal index of these data using the intervals estimator [37]. To apply it, we consider a set of thresholds equal to the empirical quantiles of levels from 0.9 to 0.99. For both datasets and all selected thresholds, the values of the intervals estimator lie above the 0.9 level, which confirms that the extremes of both datasets can be considered as asymptotically independent and the application of the Jackson- and Lewis-type tests as well as the procedure (16) to these data is correct.

We focus on the 150 largest observations of these samples. The p-values of the Jackson- and Lewis-type tests and analogs of p-values of the procedure (16) (calculated as $p = 1 - \Phi\big(\sqrt{k}(\widehat{R}_{k,n} - 1)/\sigma(0)\big)$, where Φ is the standard normal cdf) are shown in Figure 9 for $k = 10, \ldots, 150$. Their quite unstable behavior

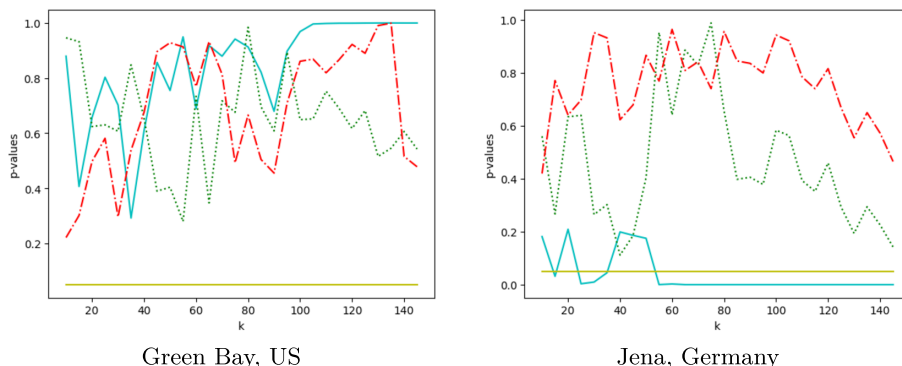Green Bay, US   Jena, Germany

FIG 9. *P-values of the procedure* (16) *(cyan solid line), the Jackson-type (green dotted line) and Lewis-type (red dash-dotted line) tests for daily precipitation data in Green Bay, US (left panel) and Jena, Germany (right panel); the level* $\alpha = 0.05$ *is indicated by a yellow solid line,* $k = 10, \ldots, 150$.

is because there are many identical observations (the total number of unique non-zero observations is 127 in Green Bay and 172 in Jena; this is explained by the fact that the amount of precipitation in the GHCNd database is given with an accuracy of up to one millimeter).

The left panel of Figure 9 suggests that the class $\mathcal{W}$ should be selected as an appropriate tail model for data collected in Green Bay, though traditionally extremes of precipitation data are modelled by extreme value distributions with positive shape parameter, see [3]. Indeed, p-values of the Jackson- and Lewis-type tests and its analogs of the procedure (16) are higher than the level $\alpha = 0.05$ for all k values shown.

This conclusion is confirmed by the quantile-quantile plots in Figure 10. The left panel is a Hill plot, or Pareto quantile plot ([2], p.102), showing the fit of the largest sample observations by the Pareto distribution (and as a consequence by any regularly-varying distribution), and containing points

$$\left\{ \left(\ln\left(\frac{n+1}{j}\right), \ln(X_{(n-j+1)}) \right); \quad j = 1, \ldots, 150 \right\}.$$

The right panel is a Weibull plot [38], since it is exactly a quantile-quantile plot for high quantiles of the Weibull distribution $\mathcal{W}(a, \lambda)$, see Section 4.3 for definition. It is defined by points

$$\left\{ \left(\ln\left(\ln\left(\frac{n+1}{j}\right) \right), \ln(X_{(n-j+1)}) \right); \quad j = 1, \ldots, 150 \right\}$$

and is inspired by approximation (6) in [39]. The central panel is a log-Weibull plot also introduced in [38], a quantile-quantile plot for high quantiles of the log-Weibull distribution $\mathcal{LW}(\lambda, c)$, containing points

$$\left\{ \left(\ln\left(\ln\left(\frac{n+1}{j}\right) \right), \ln\left(\ln(X_{(n-j+1)}) \right) \right); \quad j = 1, \ldots, 150 \right\}$$
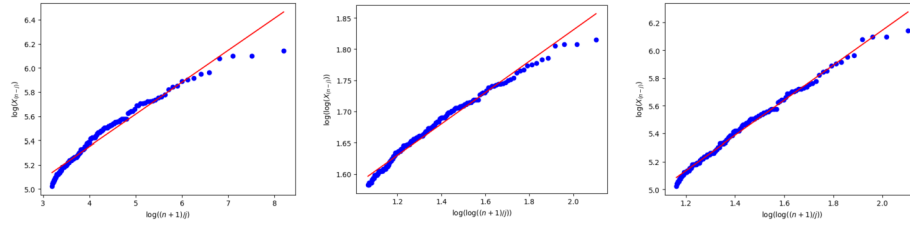
FIG 10. *Hill (left), log-Weibull (center) and Weibull (right) plots of the largest $k = 150$ daily precipitation observations in Green Bay, US, in April during 1901–2021, $n = 1331$, the red solid lines correspond to the best linear fits to these data.*
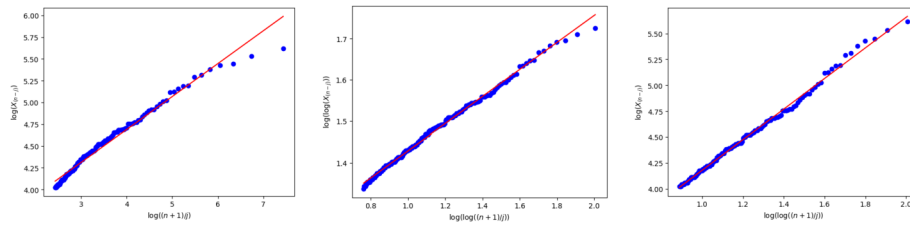

FIG 11. *Hill (left), log-Weibull (center) and Weibull (right) plots of the largest $k = 150$ daily precipitation observations in Jena, Germany, in April during 1901–2021, $n = 1693$, the red solid lines correspond to the best linear fits to these data.*

The quite good fit shown by the Weibull plot can be interpreted as an empirical validation of the Weibull-type behavior of the tail, whereas the Hill and log-Weibull plots are clearly concave.

The right panel of Figure 9 is not so univocal as its left one. The Jackson- and Lewis-type tests suggest that the class $\mathcal{W}$ should be selected as a tail model for data collected in Jena. However, analogs of p-values of the procedure (16) occur both below and above the level $\alpha = 0.05$; moreover, there is a stable interval of $\widehat{R}_{k,n}$ values below α for k from 60 to 150, which suggests that the class $\mathcal{LW}$ (or, maybe, a model with heavier tails) should be selected as a tail model for this data according to Section 3.4.

The quantile-quantile plots in Figure 11 also do not allow us to draw a clear conclusion about which of the three models is the best. Whereas the Hill plot looks a bit concave, the log-Weibull and Weibull plots show a good fit, that is both the log-Weibull and Weibull tail models look to be well suited to describe the precipitation data collected in Jena.

5. Proofs

5.1. Proof of Proposition 3.2

The continuity of G and infinity of its right endpoint imply that for all $c > 1$ there exists some $\tilde{c} > 1, \tilde{c} = \tilde{c}(t)$, such that $c = u_G(\tilde{c}t)/u_G(t)$. From $C_\delta(H, G)$ it

follows that

$$u_H(t) \geq u_G(t)\frac{u_H\big(\tilde{c}^{1+\delta}t\big)}{u_G(\tilde{c}t)}.$$

Using this relation, we derive

$$\frac{\big(1 - H(cu_H(t))\big)^{1-\varepsilon}}{1 - G(cu_G(t))} = \frac{\left(1 - H\Big(\frac{u_G(\tilde{c}t)}{u_G(t)}u_H(t)\Big)\right)^{1-\varepsilon}}{1 - G(u_G(\tilde{c}t))}$$

$$\leq \frac{\left(1 - H\Big(\frac{u_G(\tilde{c}t)}{u_G(t)}\frac{u_H(\tilde{c}^{1+\delta}t)}{u_G(\tilde{c}t)}u_G(t)\Big)\right)^{1-\varepsilon}}{1 - G(u_G(\tilde{c}t))}$$

$$= \frac{\big(1 - H(u_H(\tilde{c}^{1+\delta}t))\big)^{1-\varepsilon}}{1 - G(u_G(\tilde{c}t))} = \frac{(\tilde{c}^{1+\delta}t)^{-(1-\varepsilon)}}{(\tilde{c}t)^{-1}} = t^{\varepsilon}$$

$$= \frac{(1 - H(u_H(t)))^{1-\varepsilon}}{1 - G(u_G(t))},$$

whence the proposition follows.

5.2. Proof of Example 3.4

First let us show that for all $H \in A_1$ the condition $C_\delta(F_0, H)$ holds for some $\delta > 0$. We have $u_{F_0}(t) = \exp\big((\ln\ln t/b)^2\big)$. By its definition (4), the function $\ln(1 - H(e^t))$ is regularly varying at infinity with index $1/\vartheta$, $\vartheta > 0$, thus $1 - H(t) = \exp\big(-(\ln t)^{1/\vartheta}\ell(\ln t)\big), t > 1$, for some slowly varying at infinity function $\ell(t)$. Therefore, by properties of slowly varying functions (see, e.g., [40]), there is a slowly varying function $\ell^\sharp(t)$ such that $u_H(t) = \exp\big((\ln t)^\vartheta \ell^\sharp(\ln t)\big)$.

So, to check the condition $C_\delta(F_0, H)$, we show that for some $\delta > 0$ there is a t_0 such that for all $t > t_0$ the ratio

$$u_{F_0}(c^{1+\delta}t)/u_H(ct) = \exp\big((\ln\ln(c^{1+\delta}t)/b)^2 - (\ln(ct))^\vartheta \ell^\sharp(\ln(ct))\big) = \exp(f(c,t))$$

does not increase with respect to $c > 1$. For this purpose it is enough to show that the derivative with respect to c of the function in the exponent on the right-hand side of the latter relation is negative for all $t > t_0$ and $c > 1$. Below we assume that the derivative of $\ell^\sharp$ is eventually monotone. If the slowly varying function $\ell(t)$ is differentiable and its derivative is eventually monotone, then

$$\lim_{t\to\infty} t\ell'(t)/\ell(t) = 0, \tag{18}$$

see, e.g., Theorem 1.7.2, [40]. Thus, for large t_0 we have

$$\frac{\partial f(c,t)}{\partial c} = \frac{1}{c}\frac{2(1+\delta)}{b^2}\frac{\ln\ln(c^{1+\delta}t)}{\ln(c^{1+\delta}t)} - \frac{1}{c}\vartheta(\ln(ct))^{\vartheta-1}\ell^\sharp(\ln(ct))(1+o(1))$$

$$< t\left(\frac{1}{ct}\frac{2(1+\delta)}{b^2}\frac{\ln\ln(ct)}{\ln(ct)} - \frac{1}{ct}\vartheta(\ln(ct))^{\vartheta-1}\ell^\sharp(\ln(ct))(1+o(1))\right)$$

$$= t\,\tilde{f}'(ct), \tag{19}$$

with

$$\tilde{f}(x) = (1+\delta)((\ln\ln x)/b)^2 - (\ln x)^\vartheta \ell^\sharp(\ln x).$$

It is easy to see that there exists x_0 such that for all $x > x_0$ the derivative of $\tilde{f}(x)$ is negative. Selecting $t_0 = x_0$, we finally derive by (19) that $\partial f(c,t)/\partial c < 0$ for all $t > t_0$ and $c > 1$.

Checking the condition $C_0(G, F_0)$ for $G \in A_0$, where A_0 is the class of tails of Weibull-type distributions, is much easier. Indeed, using properties of slowly varying functions and the definition of Weibull-type distributions (see formula (4)), we obtain $u_G(t) = (\ln t)^\theta \ell^\sharp(\ln t)$ for $\theta > 0$ and some slowly varying $\ell^\sharp(t)$. Then we need to show that the function

$$u_G(t)/u_{F_0}(t) = \exp\left(\theta \ln\ln t + \ln\ell^\sharp(\ln t) - ((\ln\ln t)/b)^2\right)$$

does not increase eventually. Assuming again that the derivative of $\ell^\sharp$ is eventually monotone, we can show this by proving that the derivative of the function in the exponent on the right-hand side of the latter relation is eventually negative. This follows immediately by using (18).

5.3. Proof of Theorem 3.9

We have

$$\sqrt{k}(\tilde{R}_{k,n} - 1) = \sqrt{k}(\tilde{R}_{k,n} - R_{k,n}) + \sqrt{k}(R_{k,n} - 1),$$

where $R_{k,n}$ is defined by (12).

First, let us consider the asymptotic behavior of $R_{k,n}$. We have, $-\ln(1 - F_0(X_{(n-i)})) \stackrel{d}{=} E_{(n-i)}$, $i = 0, \ldots, n-1$, where $E_{(1)} \leq \cdots \leq E_{(n)}$ denote the order statistics of independent standard exponential random variables $E_1, \ldots, E_n$. Then, by the Rényi representation [41]

$$R_{k,n} \stackrel{d}{=} \frac{1}{k}\sum_{i=1}^{k}\left(E_{(n-k+i)} - E_{(n-k)}\right) \stackrel{d}{=} \frac{1}{k}\sum_{i=1}^{k}E_{(i)}^* = \frac{1}{k}\sum_{i=1}^{k}E_i^*, \qquad (20)$$

where $\{E_{(i)}^*\}_{i=1}^k$ are the order statistics of independent standard exponential random variables $\{E_i^*\}_{i=1}^k$. Thus by the central limit theorem,

$$\sqrt{k}(R_{k,n} - 1) \stackrel{d}{\to} N(0, 1), \qquad (21)$$

see also the proof of Theorem 1 in [11].

So to complete the proof, it is enough to show that

$$\sqrt{k}(\tilde{R}_{k,n} - R_{k,n}) \stackrel{P}{\to} 0.$$

To prove this, we use the Chebyshev inequality for $\sqrt{k}(\tilde{R}_{k,n} - R_{k,n})$ given $X_{(n-k)} = q$ and the law of total probability. For this purpose, we evaluate the asymptotic behaviour of $\mathbb{E}(\tilde{R}_{k,n} - R_{k,n} \mid X_{(n-k)})$ and $\mathrm{Var}(\tilde{R}_{k,n} - R_{k,n} \mid X_{(n-k)})$.

Explicit form of $\mathbb{E}(\tilde{R}_{k,n} - R_{k,n} \mid X_{(n-k)} = q)$**.**

We have

$$
\tilde{R}_{k,n} - R_{k,n} = \left(\ln \frac{k}{n} - \ln(1 - F_0(X_{(n-k)})) \right)
$$
$$
- \frac{1}{k} \sum_{i=n-k+1}^{n} \ln \frac{1 - F_0(u_0(n/k)X_{(i)}/X_{(n-k)})}{1 - F_0(X_{(i)})}.
$$

Consider the conditional distribution of $\tilde{R}_{k,n} - R_{k,n}$ given $X_{(n-k)} = q$. By Lemma 3.4.1 in [1], the joint conditional distribution of the set of order statistics $\{X_{(i)}\}_{i=n-k+1}^{n}$ given $X_{(n-k)} = q$ coincides with the (unconditional) joint distribution of the set of order statistics $\{X_{(i)}^*\}_{i=1}^{k}$ of i.i.d. random variables $\{X_i^*\}_{i=1}^{k}$ with common cdf

$$
F_q(x) = P(X \le x \mid X > q) = \frac{F_0(x) - F_0(q)}{1 - F_0(q)}, \quad x > q.
$$

Thus, given $X_{(n-k)} = q$,

$$
\tilde{R}_{k,n} - R_{k,n} \stackrel{d}{=} \left(\ln \frac{k}{n} - \ln(1 - F_0(q)) \right) - \frac{1}{k} \sum_{i=1}^{k} \ln \frac{1 - F_0(u_0(n/k)X_i^*/q)}{1 - F_0(X_i^*)}. \quad (22)
$$

Hereafter we write u_0 instead of $u_0(n/k)$ for convenience. For $Y^* := \ln(1 - F_0(u_0 X_1^*/q)) - \ln(1 - F_0(X_1^*))$, after direct, but tedious calculations, we derive

$$
\mathbb{E}Y^* = \ln \frac{1 - F_0(u_0)}{1 - F_0(q)} - (f(q) - 1), \quad (23)
$$

where

$$
f(q) = f(q; n, k) = \int_q^\infty \frac{u_0}{q} \frac{1 - F_0(x)}{1 - F_0(q)} \frac{p_0(u_0 x/q)}{1 - F_0(u_0 x/q)} dx, \quad (24)
$$

and $p_0(t)$ denotes the pdf of F_0. This and (22) imply that

$$
\mathbb{E}(\tilde{R}_{k,n} - R_{k,n} \mid X_{(n-k)} = q) = f(q) - f(u_0) = f(q) - 1.
$$

Asymptotics of $\sqrt{k}\,\mathbb{E}(\tilde{R}_{k,n} - R_{k,n} \mid X_{(n-k)})$**.**

Next, we show that

$$
\sqrt{k}\,\mathbb{E}(\tilde{R}_{k,n} - R_{k,n} \mid X_{(n-k)}) = \sqrt{k}(f(X_{(n-k)}) - f(u_0)) \stackrel{P}{\to} 0. \quad (25)
$$

As long as $F_0(u_0(x)) = 1 - 1/x$, we have $p_0(u_0(x))u_0'(x) = x^{-2}$ and $(xu_0'(x))^{-1} = xp_0(u_0(x))$. Lemma 2.2.1 in [1] implies that

$$
\frac{n}{\sqrt{k}} p_0(u_0)(X_{(n-k)} - u_0) \stackrel{d}{\to} N(0,1). \quad (26)
$$

Applying the delta method to (26), we obtain

$$\frac{n}{\sqrt{k}}p_0(u_0)\frac{f(X_{(n-k)}) - f(u_0)}{f'(u_0)} \xrightarrow{d} N(0,1). \tag{27}$$

This is not the classical delta method, since u_0 is not a constant and so (27) should be established. By the mean value theorem, we have $f(X_{(n-k)}) - f(u_0) = f'(c)(X_{(n-k)} - u_0)$ a.s., where c is some (random) value between $X_{(n-k)}$ and u_0. Thus it is enough to show that $f'(c)/f'(u_0) \xrightarrow{P} 1$. The proof uses the explicit form of f' provided below and is based on the multiple application of Condition 3.8 and (26); it is very technical and is not of special interest. Thus here and in similar situations below we omit such technical details.

Next, calculating directly $f'(q)$ and substituting $q = u_0$, we derive

$$f'(u_0) = \frac{n}{k}\left(p_0(u_0) - \int_{u_0}^{\infty}\frac{x}{u_0}\frac{p_0^2(x)}{1 - F_0(x)}dx\right).$$

For absolutely continuous distributions with infinite right endpoint belonging to $\mathcal{D}(EV_\gamma)$,

$$\frac{1 - F_0(u)}{up_0(u)} \to \gamma, \quad u \to \infty, \tag{28}$$

see, e.g., Theorem 1.1.11 in [1], for $\gamma > 0$ and p. 18 ibid. for $\gamma = 0$. Using the latter, L'Hôpital's rule and Condition 3.8, we get

$$\begin{aligned}
\lim_{u \to \infty}\frac{up_0(u)}{\int_u^{\infty}x\frac{p_0^2(x)}{1-F_0(x)}dx} &= -\lim_{u \to \infty}\frac{up_0'(u) + p_0(u)}{up_0^2(u)/(1 - F_0(u))} \\
&= -\lim_{u \to \infty}\frac{p_0'(u)(1 - F_0(u))}{(p_0(u))^2} - \lim_{u \to \infty}\frac{1 - F_0(u)}{up_0(u)} \\
&= 1,
\end{aligned} \tag{29}$$

whence

$$\frac{k}{np_0(u_0)}f'(u_0) \xrightarrow{P} 0.$$

Combining the latter and (27), we derive (25).

Asymptotics of $\mathrm{Var}(\tilde{R}_{k,n} - R_{k,n} \mid X_{(n-k)})$**.**

First, we have,

$$\begin{aligned}
\mathrm{Var}(\tilde{R}_{k,n} - R_{k,n} \mid X_{(n-k)} = q) &= \frac{1}{k}\mathrm{Var}\left(\ln\frac{k}{n} - \ln(1 - F_0(q)) - Y^*\right) \\
&= \frac{1}{k}\mathrm{Var}(Y^*).
\end{aligned}$$

Integrating by parts, we derive for the second moment of Y^*

$$\mathbb{E}(Y^*)^2 = \int_q^{\infty}\left(\ln\frac{1 - F_0(u_0x/q)}{1 - F_0(x)}\right)^2 d\left(\frac{F_0(x) - F_0(q)}{1 - F_0(q)}\right) = \left(\ln\frac{1 - F_0(q)}{1 - F_0(u_0)}\right)^2$$

$$+2\int_q^\infty \ln\frac{1-F_0(u_0x/q)}{1-F_0(x)}\left(\frac{p_0(x)}{1-F_0(x)}-\frac{u_0}{q}\frac{p_0(u_0x/q)}{1-F_0(u_0x/q)}\right)\frac{1-F_0(x)}{1-F_0(q)}dx$$

$$=:\left(\ln\frac{1-F_0(q)}{1-F_0(u_0)}\right)^2+2h(q).$$

From the latter and (23),

$$\mathbb{E}(Y^*)^2-(\mathbb{E}Y^*)^2=2h(q)-2(f(q)-1)\ln\frac{1-F_0(q)}{1-F_0(u_0)}-(f(q)-1)^2=:V(q).$$

Fix δ: $0<\delta<1/2$. Let us show that

$$k^{2\delta}V(X_{(n-k)})\xrightarrow{P}0. \tag{30}$$

Note that $V(X_{(n-k)})=k\mathrm{Var}(\tilde{R}_{k,n}-R_{k,n}\mid X_{(n-k)})$. Let $E_1,\ldots,E_n$ be i.i.d. standard exponential variables and $E_{(1)}\le\cdots\le E_{(n)}$ be their order statistics. By continuity of F_0, we get $-\ln(1-F_0(X_{(n-k)}))\overset{d}{=}E_{(n-k)}$, thus

$$-\sqrt{k}\ln\left(\frac{1-F_0(X_{(n-k)})}{1-F_0(u_0)}\right)\overset{d}{=}\sqrt{k}\big(E_{(n-k)}-\ln(n/k)\big)\xrightarrow{d}N(0,1),$$

by Lemma 2.2.1 in [1]. From the latter, (25) and the relation $f(u_0)=1$, it follows that

$$k^{2\delta}\left(2(f(X_{(n-k)})-1)\ln\frac{1-F_0(X_{(n-k)})}{1-F_0(u_0)}+(f(X_{(n-k)})-1)^2\right)\xrightarrow{P}0.$$

To prove (30) it remains to show that $k^{2\delta}h(X_{(n-k)})\xrightarrow{P}0$. For this purpose we apply the delta method again. In this situation we cannot use the version of the delta method which we applied above since $h'(u_0)=0$, so we aim at finding the second derivative of $h(q)$ and substituting $q=u_0$. After tedious calculations we obtain

$$h''(u_0)=\frac{1}{u_0^2(1-F_0(u_0))}\int_{u_0}^\infty\frac{x^2p_0^3(x)}{(1-F_0(x))^2}dx.$$

Note also that $h(u_0)=0$. From (26) and the relation $1-F_0(u_0)=k/n$ we obtain

$$\frac{n^2}{k}\frac{p_0^2(u_0)(h(X_{(n-k)})-h(u_0))}{2h''(u_0)}=\frac{np_0^2(u_0)h(X_{(n-k)})}{\int_{u_0}^\infty\frac{2x^2}{u_0^2}\frac{p_0^3(x)}{(1-F_0(x))^2}dx}\xrightarrow{d}\xi^2, \tag{31}$$

where $\xi\sim N(0,1)$ and the asymptotics of the integral in the denominator follows from the L'Hôpital rule, Condition 3.8 and (28),

$$\lim_{u\to\infty}\frac{u^2p_0^2(u)/(1-F_0(u))}{\int_u^\infty 2x^2\frac{p_0^3(x)}{(1-F_0(x))^2}dx}=\frac{1}{2}.$$

Substituting the latter into (31) and using the Slutsky theorem, we derive

$$kh(X_{(n-k)}) \xrightarrow{d} \xi^2.$$

Therefore, for $\delta \in (0, 1/2)$ it holds $k^{2\delta}h(X_{(n-k)}) \xrightarrow{P} 0$, that proves (30).

Application of the Chebyshev inequality.

To complete the proof, we must show that

$$T_{k,n} := | \sqrt{k}(\tilde{R}_{k,n} - R_{k,n}) - \sqrt{k}(f(X_{(n-k)}) - f(u_0)) | \xrightarrow{P} 0, \qquad (32)$$

that, together with (25), implies the theorem. By the Chebyshev inequality, for every $\varepsilon > 0$ we get

$$P\big(T_{k,n} > \varepsilon \mid X_{(n-k)} = q\big) \leq \frac{k\mathrm{Var}(\tilde{R}_{k,n} - R_{k,n} \mid X_{(n-k)} = q)}{\varepsilon^2} = \frac{V(q)}{\varepsilon^2}.$$

By the latter and the law of total probability, we finally derive

$$\begin{aligned}
P\left(T_{k,n} > \varepsilon\right) &= P\left(T_{k,n} > \varepsilon,\ k^\delta\sqrt{V(X_{(n-k)})} \leq \varepsilon\right) \\
&\quad + P\left(T_{k,n} > \varepsilon,\ k^\delta\sqrt{V(X_{(n-k)})} > \varepsilon\right) \\
&\leq P\left(T_{k,n} > k^\delta\sqrt{V(X_{(n-k)})}\right) + P\left(k^\delta\sqrt{V(X_{(n-k)})} > \varepsilon\right) \\
&= \int_{\mathbb{R}} P\left(T_{k,n} > k^\delta\sqrt{V(X_{(n-k)})} \mid X_{(n-k)} = q\right) p_{X_{(n-k)}}(q)dq \\
&\quad + P\left(k^\delta\sqrt{V(X_{(n-k)})} > \varepsilon\right) \\
&\leq \int_{\mathbb{R}} k^{-2\delta} p_{X_{(n-k)}}(q)dq + P\left(k^\delta\sqrt{V(X_{(n-k)})} > \varepsilon\right) \\
&= k^{-2\delta} + P\left(k^\delta\sqrt{V(X_{(n-k)})} > \varepsilon\right),
\end{aligned}$$

where $p_{X_{(n-k)}}(q)$ is the pdf of $X_{(n-k)}$. The result follows from (30).

5.4. Proof of Theorem 3.10

To prove Theorem 3.10, we need the following result.

Lemma 5.1. *Let* $X_1^0, \ldots, X_n^0$ *be i.i.d. random variables with common cdf* F_0. *Then under the conditions of Theorem 3.10*

$$\frac{1 - F_0(cX_{(n-k)}^0)}{1 - F_0(X_{(n-k)}^0)} - \frac{1 - F_0(cu_0)}{1 - F_0(u_0)} = O_P(1/\sqrt{k}) \qquad (33)$$

uniformly in $c > 1$.

Proof. Applying the delta method for the function $f(x) = (1 - F_0(cx))/(1 - F_0(x))$ to the relation (26), we derive

$$\frac{n}{\sqrt{k}} p_0(u_0) \frac{f(X^0_{(n-k)}) - f(u_0)}{f'(u_0)} \xrightarrow{d} N(0, 1).$$

Thus to prove (33) it is enough to show that

$$\frac{\sqrt{k}}{n} \frac{f'(u_0)}{p_0(u_0)} = O(1/\sqrt{k}) \tag{34}$$

uniformly in $c > 1$. As long as

$$f'(u_0) = \left(\frac{1 - F_0(cx)}{1 - F_0(x)} \right)'_{x=u_0} = \frac{p_0(u_0)(1 - F_0(cu_0)) - cp_0(cu_0)(1 - F_0(u_0))}{(1 - F_0(u_0))^2},$$

then using $1 - F_0(u_0) = k/n$ we can simplify the relation (34) to

$$\frac{1 - F_0(cu_0)}{1 - F_0(u_0)} - c\frac{p_0(cu_0)}{p_0(u_0)} = O(1) \tag{35}$$

uniformly in $c > 1$. Clear, $\overline{F}_0(cu_0)/\overline{F}_0(u_0) \leq 1$ for all $c > 1$. If $F_0 \in \mathcal{D}(G_\gamma)$ with $\gamma > 0$, boundness from above of the ratio $cp_0(cu_0)/p_0(u_0)$ follows from Potter's inequality, see, e.g., Proposition B.1.9 in [1]. If $\gamma = 0$, then the required fact follows from formula (1.1.33) ibid., that holds under Condition 3.8, relation $p_0(u_0(t))u'_0(t) = t^{-2}$ and Potter's inequality again. Hence (35) and (33) hold. □

Let us show that under the condition $C_\delta(F_0, F_1)$, $\delta > 0$,

$$\sqrt{k}(\tilde{R}_{k,n} - 1) \gg \sqrt{k} \left(\frac{1}{k} \sum_{i=1}^{k} E_i - 1 \right)$$

for some i.i.d. random variables $\{E_i\}_{i \geq 1}$ with mean more than 1, and the result follows from the central limit theorem. Here $X \gg Y$ means that X is stochastically larger than Y. The proof remains almost the same if $C_\delta(F_1, F_0)$ holds with $\delta > 0$ instead of $C_\delta(F_0, F_1)$.

Let $X_1, \ldots, X_n$ be i.i.d. random variables with cdf F_1. Consider the asymptotic behavior of the statistic $\tilde{R}_{k,n}$ as $n \to \infty$. Denote $Y_i = \ln(k/n) - \ln(1 - F_0(u_0 X_i^*/q))$, $i = 1, \ldots, k$, where $\{X_i^*\}_{i=1}^{k}$ are i.i.d. random variables defined in the same way as in Theorem 2 with common cdf

$$F_q^1(x) = \frac{F_1(x) - F_1(q)}{1 - F_1(q)}, \quad q < x.$$

By Lemma 3.4.1 in [1], the joint distribution of the order statistics $\{Y_{(i)}\}_{i=1}^{k}$ of $\{Y_j\}_{j=1}^{k}$ coincides with the conditional joint distribution of the set of random variables $\{Z_j\}_{j=1}^{k}$ given $X_{(n-k)} = q$, with

$$Z_j = \ln(k/n) - \ln(1 - F_0(u_0 X_{(n-j+1)}/X_{(n-k)})), \quad j = 1, ..., k.$$

Clearly, $\tilde{R}_{k,n} = \frac{1}{k}\sum_{j=1}^{k} Z_j$. Thus, the conditional distribution of $\tilde{R}_{k,n}$ given $X_{(n-k)} = q$ coincides with the distribution of $\frac{1}{k}\sum_{i=1}^{k} Y_i$. Therefore,

$$
\begin{aligned}
P(Y_1 \le x) &= P(\ln(1 - F_0(u_0 X_1^*/q)) \ge \ln(k/n) - x) \\
&= P\big(X_1^* \le q u_0^{-1} F_0^{\leftarrow}(1 - ke^{-x}/n)\big) \\
&= \frac{F_1\big(q u_0^{-1} F_0^{\leftarrow}(1 - ke^{-x}/n)\big) - F_1(q)}{1 - F_1(q)}.
\end{aligned}
\tag{36}
$$

By assumption, the cdfs F_1 and F_0 satisfy either the condition $C_\delta(F_0, F_1)$ or $C_\delta(F_1, F_0)$ with $\delta > 0$ by assumptions. Assume that $C_\delta(F_0, F_1)$ holds with some $\delta > 0$ and t_0, if $C_\delta(F_1, F_0)$ is satisfied the proof is similar. If $x^* = +\infty$, then $X_{(n-k)} \to +\infty$ almost surely, so we can deal only with the case $n/k > t_0$. Denote the quantile function of F_1 by $u_1(t)$ and write u_1 instead of $u_1(n/k)$. By Proposition 1,

$$
\frac{1 - F_1(cu_1)}{1 - F_1(u_1)} \ge \frac{(1 - F_0(cu_0))^{1-\varepsilon}}{(1 - F_0(u_0))^{1-\varepsilon}}, \quad c > 1,
\tag{37}
$$

for some $\varepsilon > 0$.

Denote $r = u_1^{\leftarrow}(q)$. Note also that if $X_{(n-k)} = u_1(\xi)$ for some random ξ, then $u_0(\xi) \stackrel{d}{=} X_{(n-k)}^0$ with $\{X_i^0\}_{i=1}^n$ introduced in Lemma 5.1. Let us rewrite (36), using (37) with $c = u_0^{-1} F_0^{\leftarrow}(1 - ke^{-x}/n)$ and (33). Uniformly in $x > 0$, we have

$$
\begin{aligned}
P(Y_1 \le x) &= 1 - \frac{1 - F_1\big(u_1(r)u_0^{-1}F_0^{\leftarrow}(1 - ke^{-x}/n)\big)}{1 - F_1(u_1(r))} \\
&\le 1 - \frac{\big(1 - F_0(u_0(r)u_0^{-1}F_0^{\leftarrow}(1 - ke^{-x}/n))\big)^{1-\varepsilon}}{\big(1 - F_0(u_0(r))\big)^{1-\varepsilon}} \\
&= 1 - \left(\frac{1 - F_0(F_0^{\leftarrow}(1 - ke^{-x}/n))}{1 - F_0(u_0)} + O(1/\sqrt{k})\right)^{1-\varepsilon} \\
&= 1 - \left(e^{-x} + O(1/\sqrt{k})\right)^{1-\varepsilon} = 1 - e^{-(1-\varepsilon)x} + O(1/\sqrt{k}).
\end{aligned}
$$

Since $O(1/\sqrt{k})$ vanishes, for k large enough there exists a cdf $G(x)$ such that

$$
G(x) \ge \min(1 - e^{-(1-\varepsilon)x} + O(1/\sqrt{k}), 1)
$$

with mean $(1-\varepsilon_1)^{-1}$, $\varepsilon_1 \in (0, \varepsilon)$, and variance $\sigma^2 > 0$. Hence, Y_1 is stochastically larger than a random variable E with cdf G. Next, let $E_1, \ldots, E_k$ be i.i.d. random variables with common cdf G, then

$$
\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k} Y_i - 1\right) \gg \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k} E_i - 1\right).
\tag{38}
$$

Provided (38) holds for all $q > x_0$,

$$
\sqrt{k}(\tilde{R}_{k,n} - 1) \gg \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k} E_i - 1\right).
\tag{39}
$$

By the central limit theorem,

$$\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k}E_i - \frac{1}{1-\varepsilon_1}\right) \xrightarrow{d} N(0,\sigma^2),$$

so

$$\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k}E_i - 1\right) \xrightarrow{P} +\infty,$$

which with (39) completes the proof of Theorem 3.10.

5.5. *Proof of Corollary 3.11*

Let us assume that $C_0(F_0, F_1)$ holds; the proof under $C_0(F_1, F_0)$ is similar. Condition $C_0(F_0, F_1)$ implies that

$$\frac{u_0(ct)}{u_0(t)} \leq \frac{u_1(ct)}{u_1(t)} \tag{40}$$

for all $c \geq 1$ and $t \geq t_0$. As mentioned in the proof of Theorem 3.10, under the assumptions of Corollary 3.11, $X_1^0 \stackrel{d}{=} u_0(u_1^{\leftarrow}(X_1)) =: V_1$. Hence substituting $t = u_1^{\leftarrow}(X_{(n-k)})$ and $c = u_1^{\leftarrow}(X_{(i)})/u_1^{\leftarrow}(X_{(n-k)})$ in (40), we get for all $i \in \{n-k+1, \ldots, n\}$ and n large enough

$$\frac{V_{(i)}}{V_{(n-k)}} = \frac{u_0(u_1^{\leftarrow}(X_{(i)}))}{u_0(u_1^{\leftarrow}(X_{(n-k)}))} \leq \frac{u_1(u_1^{\leftarrow}(X_{(i)}))}{u_1(u_1^{\leftarrow}(X_{(n-k)}))} = \frac{X_{(i)}}{X_{(n-k)}}, \tag{41}$$

where $V_{(1)} \leq \cdots \leq V_{(n)}$ are the order statistics of i.i.d. random variables $\{V_i\}_{i=1}^n$, $V_i = u_0(u_1^{\leftarrow}(X_i))$ for $i \in \{1, \ldots, n\}$. Since $(V_{(1)}, \ldots, V_{(n)}) \stackrel{d}{=} (X_{(1)}^0, \ldots, X_{(n)}^0)$, it follows from (41) that

$$\tilde{R}_{k,n}^0 \stackrel{d}{=} \ln\frac{k}{n} - \frac{1}{k}\sum_{i=n-k+1}^{n} \ln(1 - F_0(u_0 V_{(i)}/V_{(n-k)}))$$

$$\leq \ln\frac{k}{n} - \frac{1}{k}\sum_{i=n-k+1}^{n} \ln(1 - F_0(u_0 X_{(i)}/X_{(n-k)})) = \tilde{R}_{k,n}.$$

The result follows.

5.6. *Proof of Theorem 3.14*

To prove Theorem 3.14, we need the following result.

Lemma 5.2. *Let $E_1, \ldots, E_n$ be i.i.d. standard exponential with $E_{(1)} \leq \cdots \leq E_{(n)}$, their order statistics. Then*

$$\frac{n}{\sqrt{k}}\left(\begin{pmatrix}E_{(k)}\\E_{(2k)}\end{pmatrix} + \begin{pmatrix}\ln(1 - k/n)\\\ln(1 - 2k/n)\end{pmatrix}\right) \xrightarrow{d} N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 1\\1 & 2\end{pmatrix}\right), \tag{42}$$

if $k \to \infty$, $k/n \to 0$ as $n \to \infty$.

Proof. The result of the lemma is almost the direct consequence of the following two-dimensional extension of Smirnov's lemma [42], see also Lemma 2.2.3 in [1]. Let $U_{(1)} \leq \cdots \leq U_{(n)}$ be order statistics from a standard uniform sample. Then, if $k \to \infty$, $k/n \to 0$ as $n \to \infty$,

$$\begin{pmatrix} \frac{U_{(n-k)} - b_{n,k}}{a_{n,k}} \\ \frac{U_{(n-2k)} - b_{n,2k}}{a_{n,2k}} \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1/\sqrt{2} \\ 1/\sqrt{2} & 1 \end{pmatrix} \right), \tag{43}$$

where for $t \in \mathbb{N}$, $t < n$, $b_{n,t} = (n-t)/(n-1)$ and $a_{n,t} = \sqrt{b_{n,t}(1-b_{n,t})/(n-1)}$. The proof of (43) is just a repetition of that of Lemma 2.2.3 in [1]. Firstly, the pdf of the random vector $(U_{(n-k)}, U_{(n-2k)})$ is

$$\frac{n!}{((k-1)!)^2(n-2k)!}(1-x)^{k-1}(x-y)^{k-1}y^{n-2k},$$

thus the pdf of the vector $\big((U_{(n-k)} - b_{n,k})/a_{n,k}, \ (U_{(n-2k)} - b_{n,2k})/a_{n,2k}\big)$ is

$$\frac{n!}{((k-1)!)^2(n-2k)!} a_{n,k} a_{n,2k}(1-b_{n,k})^{k-1}(b_{n,k} - b_{n,2k})^{k-1}(b_{n,2k})^{n-2k}$$

$$\times \left(1 - \frac{a_{n,k}x}{1-b_{n,k}}\right)^{k-1}\left(1 + \frac{a_{n,k}x - a_{n,2k}y}{b_{n,k} - b_{n,2k}}\right)^{k-1}\left(1 + \frac{a_{n,2k}x}{b_{n,2k}}\right)^{n-2k}.$$

By the Stirling formula one can conclude that the expression on the first line of the latter formula tends to $(\sqrt{2}\pi)^{-1}$, whereas the expression on the second line tends to

$$\exp(-x^2 + \sqrt{2}xy - y^2).$$

Thus, (43) follows from the fact that pointwise convergence of the sequence of pdfs implies weak convergence of the probability distributions (Scheffe's theorem).

Next, noticing that $a_{n,k} \sim a_{n,2k}/\sqrt{2} \sim \sqrt{k}/n$ and using the properties of multivariate normal distribution, we have

$$\frac{n}{\sqrt{k}}\left(\begin{pmatrix} U_{(n-k)} \\ U_{(n-2k)} \end{pmatrix} - \begin{pmatrix} \frac{n-k}{n-1} \\ \frac{n-2k}{n-1} \end{pmatrix} \right) \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \right). \tag{44}$$

Notice that $(E_{(k)}, E_{(2k)}) \stackrel{d}{=} (-\ln(U_{(n-k)}), -\ln(U_{(n-2k)}))$ and denote $f(x) = -\ln x$. We derive by the mean value theorem

$$\frac{n}{\sqrt{k}}\left(\begin{pmatrix} E_{(k)} \\ E_{(2k)} \end{pmatrix} + \begin{pmatrix} \ln\left(\frac{n-k}{n-1}\right) \\ \ln\left(\frac{n-2k}{n-1}\right) \end{pmatrix} \right) \stackrel{d}{=} \frac{n}{\sqrt{k}}\begin{pmatrix} f(U_{(n-k)}) - f\left(\frac{n-k}{n-1}\right) \\ f(U_{(n-2k)}) - f\left(\frac{n-2k}{n-1}\right) \end{pmatrix}$$

$$= \frac{n}{\sqrt{k}}\begin{pmatrix} f'(c_1)\big(U_{(n-k)} - \frac{n-k}{n-1}\big) \\ f'(c_2)\big(U_{(n-2k)} - \frac{n-2k}{n-1}\big) \end{pmatrix}$$

where c_i, $i = 1, 2$, are some (random) values between $U_{(n-ik)}$ and $(n-ik)/(n-1)$, $i = 1, 2$, respectively. But for $i = 1, 2$ $f'(c_i) = -1/c_i$, $U_{(n-ik)} \to 1$ in probability and $(n-ik)/(n-1) \to 1$, thus $f'(c_i) \to -1$ in probability. This and the Slutsky theorem imply (42). $\square$

The schemes of the proofs of Theorems 3.9 and 3.14 coincide. However, the proof of the latter is more difficult technically since the statistic $\widehat{R}_{k,n}$ depends on $X_{(n-k)}$ and $X_{(n-2k)}$ together. Moreover, if $\gamma > 0$ then $\sqrt{k}(\widehat{R}_{k,n} - R_{k,n})$ no longer tends to 0 in probability.

Explicit form of $\mathbb{E}(\widehat{R}_{k,n} - R_{k,n} \mid X_{(n-k)} = q, X_{(n-2k)} = \hat{q})$**.**

Let us write $\sqrt{k}(\widehat{R}_{k,n} - R_{k,n})$ in explicit form

$$
\sqrt{k}(\widehat{R}_{k,n} - R_{k,n}) = \sqrt{k}\left(\ln\frac{k}{n} - \ln(1 - F_0(X_{(n-k)}))\right)
$$
$$
- \frac{1}{\sqrt{k}}\sum_{i=n-k+1}^{n}\left[\ln\left(1 - F_0\left(u_0 + \frac{X_{(i)} - X_{(n-k)}}{X_{(n-k)} - X_{(n-2k)}}(u_0 - \hat{u}_0)\right)\right)\right.
$$
$$
\left. - \ln(1 - F_0(X_{(i)}))\right],
$$

where $\hat{u}_0 = u_0(n/(2k))$, and consider its conditional distribution given $X_{(n-k)} = q$ and $X_{(n-2k)} = \hat{q}$. Using argument similar to that from the proof of Theorem 3.9 and inheriting the notation from there, we get

$$
\sqrt{k}(\widehat{R}_{k,n} - R_{k,n}) \mid X_{(n-k)} = q, X_{(n-2k)} = \hat{q}
$$
$$
\stackrel{d}{=} \sqrt{k}\left(\ln\frac{k}{n} - \ln(1 - F_0(q))\right)
$$
$$
- \frac{1}{\sqrt{k}}\sum_{i=1}^{k}\left[\ln\left(1 - F_0\left(u_0 + \frac{X_i^* - q}{q - \hat{q}}(u_0 - \hat{u}_0)\right)\right)\right. \tag{45}
$$
$$
\left. - \ln(1 - F_0(X_i^*))\right].
$$

Hereinafter we write $\overline{F}(x)$ instead of $1 - F(x)$. Write

$$
Y^* = \ln\overline{F}_0\left(u_0 + \frac{X_1^* - q}{q - \hat{q}}(u_0 - \hat{u}_0)\right) - \ln\overline{F}_0(X_1^*)
$$

and find its mean. After integrating by parts, we obtain

$$
\mathbb{E}Y^* = \ln\frac{\overline{F}_0(u_0)}{\overline{F}_0(q)}
$$
$$
- \int_q^\infty \frac{\overline{F}_0(x)}{\overline{F}_0(q)}\left(\frac{u_0 - \hat{u}_0}{q - \hat{q}}\frac{p_0(u_0 + (u_0 - \hat{u}_0)(x - q)/(q - \hat{q}))}{\overline{F}_0(u_0 + (u_0 - \hat{u}_0)(x - q)/(q - \hat{q}))} - \frac{p_0(x)}{\overline{F}_0(x)}\right)dx
$$
$$
=: \ln\frac{\overline{F}_0(u_0)}{\overline{F}_0(q)} + 1 - f(q, \hat{q}).
$$

from which and (45) we get

$$\mathbb{E}(\widehat{R}_{k,n} - R_{k,n} \mid X_{(n-k)} = q, X_{(n-2k)} = \hat{q}) = f(q, \hat{q}) - 1.$$

Evaluating the asymptotics of $\sqrt{k}\mathbb{E}(\tilde{R}_{k,n} - R_{k,n} \mid X_{(n-k)}, X_{(n-2k)})$.

Now find the asymptotics of

$$\sqrt{k}(f(X_{(n-k)}, X_{(n-2k)}) - 1) = \sqrt{k}\mathbb{E}(\tilde{R}_{k,n} - R_{k,n} \mid X_{(n-k)}, X_{(n-2k)}).$$

In contrast to the proof of Theorem 3.9, we need the multivariate delta method. First of all, observe that

$$(X_{(n-k)}, X_{(n-2k)}) \stackrel{d}{=} \left(u_0\Big(\frac{1}{1 - \exp(-E_{(k)})}\Big), u_0\Big(\frac{1}{1 - \exp(-E_{(2k)})}\Big)\right),$$

where the order statistics $E_{(k)}$ and $E_{(2k)}$ were introduced in Lemma 5.2. Applying the multivariate delta method for the function

$$g(x, y) = f\left(u_0\Big(\frac{1}{1 - e^{-x}}\Big), u_0\Big(\frac{1}{1 - e^{-y}}\Big)\right)$$

to the relation (42), we get

$$\frac{\sqrt{k}\big(f(X_{(n-k)}, X_{(n-2k)}) - f(u_0, \hat{u}_0)\big)}{\frac{k}{n}\sqrt{D}} \stackrel{d}{\to} N(0, 1), \tag{46}$$

where

$$D = \nabla g \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} (\nabla g)^T$$

and ∇g is the gradient of $g(x, y)$ at the point $(-\ln(1 - k/n), -\ln(1 - 2k/n))$. The correctness of applying the multivariate delta method here is established using the same argumentation as of (27) (in fact, (46) can be derived from the two-dimensional version of (26)), is very technical and thus omitted.

Observe that $f(u_0, \hat{u}_0) = 1$. Thus the asymptotics of $\sqrt{k}\mathbb{E}(\tilde{R}_{k,n} - R_{k,n} \mid X_{(n-k)}, X_{(n-2k)})$ strongly depend on those of $\frac{k}{n}\sqrt{D}$. One can show that

$$\frac{k}{n}\sqrt{D} \to \frac{1}{\sqrt{2}(2^\gamma - 1)}\frac{\gamma}{\gamma + 1}, \tag{47}$$

where for $\gamma = 0$ the right-hand side should be understood as $(\sqrt{2}\ln 2)^{-1}$. The proof of (47) is technical and includes multiple use of the properties of regularly-varying functions and L'Hôpital's rule; it is given in the Supplementary material [28].

Summarizing the above, we proved that under the assumptions of the theorem,

$$\sqrt{k}\big(f(X_{(n-k)}, X_{(n-2k)}) - 1\big) \stackrel{d}{\to} N\Big(0, 1 + \frac{\gamma^2}{2(\gamma + 1)^2(2^\gamma - 1)^2}\Big), \tag{48}$$

where for $\gamma = 0$ the ratio on the right-hand side should be understood as $1/(2(\ln 2)^2)$.

Final steps

The proof of the relation

$$\sqrt{k}\big(\widehat{R}_{k,n} - R_{k,n}\big) - \sqrt{k}\big(f(X_{(n-k)}, X_{(n-2k)}) - 1\big) \overset{P}{\to} 0 \qquad (49)$$

does not contain new ideas as compared to the proof of the similar relation (32) and the previous steps of the current proof, so we omit it. Finally, we have

$$\sqrt{k}(\widehat{R}_{k,n} - 1) = \sqrt{k}(R_{k,n} - 1)$$
$$+ \sqrt{k}\Big[\big(\widehat{R}_{k,n} - R_{k,n}\big) - \big(f(X_{(n-k)}, X_{(n-2k)}) - 1\big)\Big]$$
$$+ \sqrt{k}\big(f(X_{(n-k)}, X_{(n-2k)}) - 1\big).$$

Observe that the properties of iid exponential random variables imply that the third summand on the right-hand side of the latter relation is independent of the first. Indeed, recall that by (20)

$$R_{k,n} \overset{d}{=} \frac{1}{k} \sum_{i=0}^{k-1} (E_{(n-i)} - E_{(n-k)})$$

and $-\ln(1 - F_0(X_{(i)})) \overset{d}{=} E_{(i)}$, $i = 1, \ldots, n$, with $\{E_i\}_{i=1}^n$ i.i.d. standard exponential. Next, it follows from the Rényi representation [41] that $E_{(n-k)}$ and $E_{(n-2k)}$ are independent of $\{E_{(n-i)} - E_{(n-k)}\}_{i=0}^{k-1}$. Hence, $\ln(1 - F_0(X_{(n-k)}))$ and $\ln(1 - F_0(X_{(n-2k)}))$ are independent of $R_{k,n}$, and the required statement follows from the fact that

$$(X_{(n-k)}, X_{(n-2k)}) = \Big(\ell(-\ln(1 - F_0(X_{(n-k)}))), \ell(-\ln(1 - F_0(X_{(n-2k)})))\Big)$$

a.s. with $\ell(x) = u_0\big((1 - e^{-x})^{-1}\big)$. This argument together with (21), (48) and (49) completes the proof of Theorem 3.14.

5.7. Proof of Theorem 3.15

The scheme of the proof of Theorem 3.15 is the same as for Theorem 3.10, so it is omitted. Note only, that for the proof of the relation similar to (34), one needs to use the multivariate delta method and Lemma 5.2. Next, instead of the relation (37) used to find the upper bound for $P(Y_1 \le x)$, one should use the relation

$$\frac{\overline{F}_1(u_1(t) + x(u_1(t) - u_1(t/2)))}{\overline{F}_1(u_1(t))} \ge \frac{\big(\overline{F}_0(u_0(t) + x(u_0(t) - u_0(t/2)))\big)^{1-\varepsilon}}{(\overline{F}_0(u_0(t)))^{1-\varepsilon}}$$

for all $x > 0$. The latter follows from the condition $C_\delta(F_0, F_1)$. Indeed,

$$
\begin{aligned}
\frac{\overline{F}_1(u_1(t) + x(u_1(t) - u_1(t/2)))}{\overline{F}_1(u_1(t))} &= \frac{\overline{F}_1\big(u_1(t)(1 + x - xu_1(t/2)/u_1(t))\big)}{\overline{F}_1(u_1(t))} \\
&\geq \frac{\overline{F}_1\big(u_1(t)(1 + x - xu_0(t/2)/u_0(t))\big)}{\overline{F}_1(u_1(t))} \\
&\geq \frac{\big(\overline{F}_0(u_0(t) + x(u_0(t) - u_0(t/2)))\big)^{1-\varepsilon}}{(\overline{F}_0(u_0(t)))^{1-\varepsilon}},
\end{aligned}
$$

where the first and second relations follow from $C_0(F_0, F_1)$ and Proposition 1, respectively.

5.8. Proof of Corollary 3.16

The proof of Corollary 3.16 is like that of Corollary 3.11 except that instead of the relation (41) one should use the inequality

$$
\frac{V_{(i)} - V_{(n-k)}}{V_{(n-k)} - V_{(n-2k)}} \leq \frac{X_{(i)} - X_{(n-k)}}{X_{(n-k)} - X_{(n-2k)}},
$$

that immediately follows from $C_0(F_0, F_1)$ and (41).

Declarations

Conflict of interest statement. The author has no conflict of interest to declare.

Data availability statement. The datasets analysed during the current study are available on the site of Daily Global Historical Climatology Network, https://www.ncei.noaa.gov/data/global-historical-climatology-network-daily/.

Acknowledgments

Supplementary Material

Supplement to "Location- and scale-free procedures for distinguishing between distribution tail models"
(doi: 10.1214/24-EJS2328SUPP; .pdf). The Supplementary material contains a proof of the relation (47) from the proof of Theorem 3.14

References

[1] DE HAAN, L. and FERREIRA, A. (2006) *Extreme Value Theory. An Introduction.* Springer, New York. MR2234156

[2] BEIRLANT, T., GOEGEBEUR, Y., SEGERS, J. AND TEUGELS, J. (2004) *Statistics of Extremes. Theory and Applications.* Wiley, London. MR2108013

[3] KOUTSOYIANNIS, D. (2004) Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records. *Hydrological Sciences Journal* **49** 591–610.

[4] SERINALDI, F. AND KILSBY, C. G. (2014) Rainfall extremes: Toward reconciliation after the battle of distributions. *Water Resources Research* **50** 336–352.

[5] MARANI, M. AND IGNACCOLO, M. (2015) A metastatistical approach to rainfall extremes. *Advances in Water Resources* **79** 121–126.

[6] MINIUSSI, A., MARANI, M. AND VILLARINI, G. (2020) Metastatistical Extreme Value Distribution applied to floods across the continental United States. *Advances in Water Resources* **136** 103498.

[7] GARDES, L. AND GIRARD, S. (2016) On the estimation of the functional Weibull tail-coefficient. *Journal of Multivariate Analysis* **146** 29–45. MR3477647

[8] WORMS, J. AND WORMS, R. (2019) Estimation of extremes for Weibulltail distributions in the presence of random censoring. *Extremes* **22** 667–704. MR4031853

[9] MALEVERGNE, Y., PISARENKO, V. AND SORNETTE, D. (2005) Empirical distributions of stock returns: between the stretched exponential and the power law? *Quantitative Finance* **5** 379–401. MR2239385

[10] RESNICK, S.I. (1971) Tail equivalence and its applications. *Journal of Applied Probability* **8** 136–156. MR0278363

[11] RODIONOV, I.V. (2018) On discrimination between classes of distribution tails. *Problems of Information Transmission* **54**(2) 124–138. arXiv:2202.11619 MR3845495

[12] BALKEMA, A. A. AND DE HAAN, L. (1974) Residual life time at great age. *Annals of Probability* **2** 792–804. MR0359049

[13] PICKANDS, J. III (1975) Statistical inference using extreme order statistics. *Annals of Statistics* **3** 119–131. MR0423667

[14] DE VALK, C. (2016) Approximation of high quantiles from intermediate quantiles. *Extremes* **19** 661–686. MR3558349

[15] EL METHNI, J., GARDES, L., GIRARD, S. AND GUILLOU, A. (2012) Estimation of extreme quantiles from heavy and light tailed distributions. *Journal of Statistical Planning and Inference* **142** 2735–2747. MR2925962

[16] ALBERT, C., DUTFOY, A., GARDES, L. AND GIRARD, S. (2020) An extreme quantile estimator for the log-generalized Weibull-tail model. *Econometrics and Statistics* **13** 137–174. MR4058331

[17] DE VALK, C. (2016) Approximation and estimation of very small probabilities of multivariate extreme events. *Extremes* **19** 687–717. MR3558350

[18] DE VALK, C. AND CAI, J.-J. (2018) A high quantile estimator based on the log-generalized Weibull tail limit. *Econometrics and Statistics* **6** 107–128. MR3797978

[19] GARDES, L., GIRARD, S. AND GUILLOU, A. (2011) Weibull tail-distributions revisited: a new look at some tail estimators. *Journal of Statistical Planning and Inference* **141**(1) 429–444. MR2719507

[20] GOEGEBEUR, J. AND GUILLOU, A. (2010) Goodness-of-fit testing for Weibull-type behavior. *Journal of Statistical Planning and Inference* **140**(6) 1417–1436. MR2592221

[21] BRONIATOWSKI, M. (1993) On the estimation of the Weibull tail coefficient. *Journal of Statistical Planning and Inference* **35** 349–366. MR1229249

[22] BEIRLANT, J., BRONIATOWSKI, M., TEUGELS, J.L. AND VYNCKIER, P. (1995) The mean residual life function at great age: applications to tail estimation. *Journal of Statistical Planning and Inference* **45** 21–48. MR1342084

[23] GARDES, L. AND GIRARD, S. (2006) Comparison of Weibull tail-coefficient estimators. *REVSTAT – Statistical Journal* **4** 163–188. MR2259371

[24] HÜSLER, J. AND PENG, L. (2008) Review of testing issues in extremes: in honor of Professor Laurens de Haan. *Extremes* **11** 99–111. MR2420198

[25] NEVES, C. AND FRAGA ALVES, M.I. (2008) Testing extreme value conditions – an overview and recent approaches. *REVSTAT – Statistical Journal* **6** 83–100. MR2386301

[26] RODIONOV, I.V. (2018) A discrimination test for tails of Weibull-type distributions. *Theory of Probability and its Applications* **63**(2) 327–335. MR3796495

[27] RODIONOV, I.V. (2019) Discrimination of close hypotheses about the distribution tails using highest order statistics. *Theory of Probability and its Applications* **63**(3) 364–380. MR3833092

[28] RODIONOV, I.V. (2024) Supplement to "Location- and scale-free procedures for distinguishing between distribution tail models" https://doi.org/10.1214/24-EJS2328SUPP

[29] HILL, B.M. (1975) A simple general approach to inference about the tail of a distribution. *Annals of Statistics* **3**(5) 1163–1174. MR0378204

[30] DANIELSSON, J., ERGUN, L.M., DE HAAN, L. AND DE VRIES, C. (2016) Tail Index Estimation: Quantile Driven Threshold Selection. Available at SSRN: https://ssrn.com/abstract=2717478 or http://dx.doi.org/10.2139/ssrn.2717478

[31] GOMES, M.I. AND GUILLOU, A. (2015) Extreme value theory and statistics of univariate extremes: A review. *International Statistical Review* **83**(2) 263–292. MR3377081

[32] MARKOVICH, N.M. AND RODIONOV, I.V. (2024) Threshold selection for extremal index estimation. *Journal of Nonparametric Statitics* **36**(3) 527–546. MR4790634

[33] MENNE, M.J., DURRE, I., VOSE, R., GLEASON, B. AND HOUSTON, T. (2012) An overview of the global historical climatology network-daily database. *Journal of Atmospheric and Oceanic Technology* **29** 897–910.

[34] HASOFER, A. AND WANG, J.Z. (1992) A test for extreme value domain of attraction. *Journal of American Statistical Association* **87** 171–177.

[35] NEVES, C. AND FRAGA ALVES, M.I. (2007) Semi-parametric approach to Hasofer–Wang and Greenwood statistics in extremes. *Test* **16** 297–313. MR2393657

[36] NEVES, C., PICEK, J. AND FRAGA ALVES, M.I. (2006) The contribution of the maximum to the sum of excesses for testing max-domains of attraction. *Journal of Statistical Planning and Inference* **136**(4) 1281–1301. MR2253763

[37] FERRO, C.A.T. AND SEGERS, J. (2003) Inference for clusters of extreme values. *Journal of the Royal Statistical Society Series B* **65** 545–556. MR1983763

[38] BEIRLANT, J. AND BLADT, M. (2023) Tail classification using non-linear regression on tail model plots, preprint. DOI: 10.21203/rs.3.rs-3641407/v1

[39] GIRARD, S. (2004) A Hill-type estimator of the Weibull tail-coefficient. *Communications in Statistics – Theory and Methods* **33**(2) 205–234. MR2045313

[40] BINGHAM, N.H., GOLDIE, C.M. AND TEUGELS, J.L. (1987) *Regular Variation.* Cambridge University Press, Cambridge. MR0898871

[41] RÉNYI, A. (1953) On the theory of order statistics. *Acta Mathematica Scient. Hungar.* **4** 191–227. MR0061792

[42] SMIRNOV, N.V. (1949) Limit distributions for the terms of a variational series. In Russian: *Trudy Mat. Inst. Steklov.* **25**. Translation: *Transl. Amer. Math. Soc.* **11** 82–143 (1952). MR0034549