A test for comparing tail indices for heavy-tailed distributions via empirical likelihood

Julien Worms, Rym Worms

A test for comparing tail indices for heavy-tailed distributions via empirical likelihood

Julien Worms(1) & Rym Worms (2)

(1) Université de Versailles-Saint-Quentin-En-Yvelines
Laboratoire de Mathématiques de Versailles (CNRS UMR 8100),
UFR de Sciences, Bât. Fermat,
45 av. des Etats-Unis, 78035 Versailles Cedex,
e-mail : julien.worms@uvsq.fr

(2) Université Paris-Est-Créteil
Laboratoire d'Analyse et de Mathématiques Appliquées (CNRS UMR 8050),
61 av. du Général de Gaulle, 94010 Créteil cedex,
e-mail : rym.worms@u-pec.fr

AMS Classification. Primary 62G32 ; Secondary 62G10.

Keywords and phrases. Extreme values, Tail index, Comparison Test, Hill Estimator, Empirical Likelihood.

Abstract

In this work, the problem of testing whether different (≥ 2) independent samples, with (possibly) different heavy tailed distributions, share the same extreme value index, is addressed. The test statistic proposed is inspired by the empirical likelihood methodology and consists in an ANOVA-like confrontation of Hill estimators. Asymptotic validity of this simple procedure is proved and efficiency, in terms of empirical type I error and power, is investigated through simulations under a variety of situations. Surprisingly, this topic had hardly been addressed before, and only in the two sample case, though it can prove useful in applications.

1. Introduction

In the topic of univariate extreme value analysis, the interest lies generally in the study of a single sample and inference in the tail of its underlying distribution. The main parameter describing the tail behavior of a continuous distribution function F is its extreme value index. In many fields of applications, distributions of interest are those exhibiting a heavy tail phenomenon. In this case the extreme value index appears as the positive number γ ($1/\gamma$ is then called the tail index) such that the survival function $1-F$ is regularly varying with order $-1/\gamma$, which means that $\bar{F}(x)$ grossly behaves like $x^{-1/\gamma}$ for large x (precise definition of regular variation is given below), *i.e.* tail decreases to 0 at a polynomial rate. Therefore, the greater the value of γ is, the greater is the chance that samples drawn from F exhibits extreme values.

Most of the literature on univariate extreme value statistics focuses on methods for estimating as accurately as possible the extreme value index, which appears as the natural measure of heaviness of the tail of the underlying distribution. This work is devoted to a different topic, namely testing whether different distributions could share the same value of the tail index, in the heavy tail framework, on the basis of the observation of independent samples of these distributions. More precisely, we suppose that we observe K independent i.i.d. samples $(X_i^{(1)})_{i\leq n_1}, \ldots, (X_i^{(K)})_{i\leq n_K}$, respectively coming from heavy-tailed distribution functions $F_1, \ldots, F_K$, *i.e.* such that for each $1 \leq j \leq K$, F_j satisfies

$$\frac{1-F_j(tx)}{1-F_j(t)} \longrightarrow x^{-1/\gamma_{0j}}, \text{ as } t \to +\infty, \tag{1}$$

for all $x > 0$, where the parameter $\gamma_{0j} > 0$ is the extreme value index of the j-th sample.

The aim of this paper is to provide a simple yet effective test procedure for comparison of $\gamma_{01}, \ldots, \gamma_{0K}$, for instance by testing the equality hypothesis

$$H_0 : \gamma_{01} = \ldots = \gamma_{0K}$$

This problem has not been so much addressed in the literature : it nonetheless has practical applications in the usual fields where heavy tail phenomena occur, such as insurance, finance, or teletraffic data analysis. For instance, the data studied in Bottolo et al (2003) consists of insurance claims having different origins, and it is of interest to decide whether the claim type can have an effect on the tail behavior of the claim distribution, and if so, which types can be found to be equivalent in this sense. In Mougeot and Tribouley (2010), the authors consider various financial data of different firms, and address the problem of comparing (pairwisely) their associated financial risks through their corresponding tail indices. For this purpose, they proposed a data-driven procedure of comparison of (only) two positive tail indices : in Section 3, we will compare their results to those corresponding to the test statistic we propose, especially in terms of coverage accuracy and power, through a simulation study.

The organization of the paper is classical : in Section 2, the methodology, assumptions, and results are stated, and Section 3 is devoted to a simulation study which shows satisfactory power and coverage accuracy of our method. The proofs are delegated to the Appendix.

2. Methodology and statement of the results

We suppose that we observe K independent samples $(X_i^{(1)})_{i\leq n_1}, \ldots, (X_i^{(K)})_{i\leq n_K}$ satisfying (1), as presented previously. In order to derive our asymptotic result, we need a slightly stronger condition, which specifies the rate of convergence in (1). Denoting by U_j the inverse function of $1/(1-F_j)$ $(j = 1, \ldots, K)$, we suppose that there exists a function A_j tending to 0 at infinity such that

$$\frac{U_j(tx)/U_j(t) - x^{\gamma_{0j}}}{A_j(t)} \longrightarrow x^{\gamma_{0j}} \frac{x^{\rho_j}-1}{\rho_j}, \text{ as } t \to +\infty, \tag{2}$$

for all $x > 0$, where $\rho_j < 0$. This so-called second order condition is classical in the extreme value theory framework, and is known to hold for most commonly encountered

heavy-tailed distributions.

The method we propose for testing the null hypothesis $H_0 : \gamma_{01} = \ldots = \gamma_{0K}$, is inspired by a version of the classical ANOVA test based on the empirical likelihood methodology, proposed by Owen (1991). The starting point is the famous Hill estimator for γ_{0j} $(j = 1 \ldots K)$ which is defined as

$$\hat{\gamma}_j := \frac{1}{k_j} \sum_{i=1}^{k_j} Y_i^{(j)},$$

where, denoting by $X_{(1)}^{(j)} \leq \ldots \leq X_{(n_j)}^{(j)}$ the order statistics associated to the j-th sample, the log spacings $Y_i^{(j)}$ are defined by

$$Y_i^{(j)} := i \log \frac{X_{(n_j-i+1)}^{(j)}}{X_{(n_j-i)}^{(j)}} \qquad (i = 1, \ldots, n_j)$$

and k_j, the sample fraction of observations to keep from $(X_i^{(j)})_{i=1\ldots n_j}$, satisfies

$$k_j \to +\infty, \;\; k_j/n_j \to 0, \text{ and } \sqrt{k_j} A_j(n_j/k_j) \to 0, \;\; \text{as } n_j \to +\infty. \tag{3}$$

Let $n = \sum_{j=1}^K n_j$ and $k = \sum_{j=1}^K k_j$. Note that we do not use the index n in the definition of the k_js (nor in that of other quantities depending on n) in order to lighten the notations.

We define, for any $\gamma = (\gamma_j)_{1 \leq j \leq K}$, the empirical likelihood ratio

$$ELR(\gamma) \;:=\; \sup_{(p_{ij})} \left\{ \prod_{j=1}^K \prod_{i=1}^{k_j} (k p_{ij}) \;;\; p_{ij} \geq 0, \; \sum_{i=1}^{k_j} p_{ij} = 1, \; \sum_{i=1}^{k_j} p_{ij} (Y_i^{(j)} - \gamma_j) = 0 \right\}.$$

Clearly, this function is maximum at $\hat{\gamma} := (\hat{\gamma}_j)_{1 \leq j \leq K}$ and $ELR(\hat{\gamma}) = 1$. Thus, if we note

$$l(\gamma) \;:=\; -2 \log ELR(\gamma),$$

testing H_0 can be based on the statistic

$$\inf_{\gamma \in \mathcal{A}} l(\gamma) = -2 \log \frac{\sup_{\gamma \in \mathcal{A}} ELR(\gamma)}{\sup_{\gamma} ELR(\gamma)},$$

where $\mathcal{A} = \{\gamma \in]0, +\infty[^K, \; \gamma_1 = \ldots = \gamma_K\}$.

The first part of the following theorem yields the asymptotic distribution of this statistic and thus provides a critical region for this empirical likelihood ratio test of H_0, with prescribed asymptotic level. The second part shows that the method extends to more general linear hypotheses on the extreme value indices.

Theorem 1. *Under assumptions* (1)-(3)*, if* $\min_{j \leq K} k_j/k$ *is bounded away from* 0 *then, under* H_0*,*

$$\inf_{\gamma \in \mathcal{A}} l(\gamma) \xrightarrow{d} \chi^2(K-1), \quad \text{as } n \to +\infty.$$

Moreover, if H_0' *: "*$C\gamma_0 = 0$*", where* $\gamma_0 = (\gamma_{0j})_{1 \leq j \leq K}$ *and* C *is a* $d \times K$ *matrix with full*

rank $1 \leq d < K$, *then under* H_0',

$$\inf_{\gamma \in \mathcal{A}'} l(\gamma) \xrightarrow{d} \chi^2(d), \quad as\ n \to +\infty$$

where $\mathcal{A}' = \{\gamma \in]0, +\infty[^K \ ; \ C\gamma = 0\}$.

In order to prove Theorem 1, we have to introduce some important quantities and provide their asymptotic behavior in Lemma 1 stated below : for any $\gamma = (\gamma_j)_{1 \leq j \leq K}$ with positive components,

$$\begin{aligned} &G_n(\gamma) := (\hat{\gamma}_j - \gamma_j)_{j=1\ldots K} && S_j^2(\gamma) := \tfrac{1}{k_j} \textstyle\sum_{i=1}^{k_j} (Y_i^{(j)} - \gamma_j)^2 \\ &\widetilde{G}_n(\gamma) := (\sqrt{k_j}(\hat{\gamma}_j - \gamma_j))_{j=1\ldots K} && B_n(\gamma) := diag(S_1^2(\gamma), \ldots, S_K^2(\gamma)) \\ &M_n(\gamma) := \max_{j \leq K} \max_{i \leq k_j} |Y_i^{(j)} - \gamma_j| \end{aligned}$$

Note that the proof of Theorem 1 will make it clear that the statistic $\inf_{\gamma \in \mathcal{A}} l(\gamma)$, seemingly lengthy to compute, is very close to the quadratic quantity

$$\sum_{j=1}^{K} k_j (\hat{\gamma}_j - \tilde{\gamma})^2 / S_j(\hat{\gamma}) \ \text{ where } \ \tilde{\gamma} = \sum_{j=1}^{K} \delta_j^2 \hat{\gamma}_j \Big/ \sum_{j=1}^{K} \delta_j^2 \ \text{ and } \ \delta_j^2 = k_j / S_j^2(\hat{\gamma}). \tag{4}$$

Lemma 1. *Under the conditions of Theorem 1, if* $B := diag(\gamma_{01}^2, \ldots, \gamma_{0K}^2)$, *then, as* $n \to +\infty$, *we have*

$$\begin{aligned} &G_n(\gamma_0) \xrightarrow{\mathbb{P}} 0, \quad \widetilde{G}_n(\gamma_0) \xrightarrow{d} N(0, B), \\ &B_n(\gamma_0) \xrightarrow{\mathbb{P}} B, \quad M_n(\gamma_0) = o_{\mathbb{P}}(\sqrt{k}). \end{aligned}$$

The results contained in this Lemma are due to the asymptotic normality of the Hill estimator (see de Haan and Peng (1998) for example) and to equations (8)-(9) of J.C. Lu and L. Peng (2002).

Proof of Theorem 1

- First step : we shall prove in Subsection 5.1 that for any $C > 0$,

$$l(\gamma) = Q(\gamma) + o_{\mathbb{P}}(1), \quad \text{uniformly for } \gamma \in \mathcal{B}_{n,C}, \tag{5}$$

  where

$$Q(\gamma) := \widetilde{G}_n^t(\gamma)(B_n(\hat{\gamma}))^{-1}\widetilde{G}_n(\gamma) = \sum_{j=1}^{K} k_j (\hat{\gamma}_j - \gamma_j)^2 / S_j(\hat{\gamma}).$$

  and

$$\mathcal{B}_{n,C} := \{\gamma; \ ||\gamma - \gamma_0|| \leq C k^{-1/2}\}$$

- Second step : we shall prove in Subsection 5.2 that

$$\inf_{\gamma \in \mathcal{A}} Q(\gamma) \ = Q(\bar{\gamma}) \ \xrightarrow{d} \chi^2(K-1), \text{ as } n \to +\infty. \tag{6}$$

  where $\bar{\gamma} = (\tilde{\gamma}, \ldots, \tilde{\gamma})$ and $\tilde{\gamma}$ is defined in Equation (4). In view of (5), it thus remains to prove that

$$\lim_{C \to +\infty} \lim_{n \to +\infty} \mathbb{P}[\, \bar{\gamma} \notin \mathcal{B}_{n,C}\,] \ = \ 0,$$

since this implies the contiguity of $l(\bar{\gamma})$ and $Q(\bar{\gamma})$. Using the assumption that $\min_{j\leq K} k_j/k$ is bounded away from 0, we can find a positive constant c (not depending on C) such that

$$\begin{aligned}\mathbb{P}\left[\ \bar{\gamma} \notin \mathcal{B}_{n,C}\ \right] &\leq \mathbb{P}\left[\ |\tilde{\gamma} - \gamma_{01}| > K^{-1/2}Ck^{-1/2}\ \right] \\ &\leq \mathbb{P}\left[\ \sum_{j=1}^{K} |\hat{\gamma}_j - \gamma_{0j}| > K^{-1/2}Ck^{-1/2}\ \right] \\ &\leq \sum_{j=1}^{K} \mathbb{P}\left[\ \sqrt{k_j}|\hat{\gamma}_j - \gamma_{0j}| > cC\ \right]\end{aligned}$$

By the asymptotic normality of the Hill estimators of the K extreme value indices (stated in Lemma 1), the right-hand side of the last inequation converges (as $n\to \infty$) to a quantity which vanishes as C goes to $+\infty$.

Remark 1. According to (5) and (6), we can use in practice indifferently $l(\bar{\gamma})$ or $Q(\bar{\gamma})$ as a statistic for our test. As expected and can be seen from assumption (3) and Figure 1, the accuracy of our procedure depends on the choice of the sample fractions k_j, for $j = 1, \ldots, K$, as for the estimation problem. We shall present, in Section 3 below, the method we used for choosing adequate values for these tuning parameters.

3. Simulations

The purpose of this section is to investigate, through extensive simulations, the performance of our test procedure in the case of the comparison of tail indices of two samples X_1 and X_2 (case $K = 2$). In this testing framework, we naturally focus our attention on both the empirical type I error and power function of the procedure.

As was made in Mougeot and Tribouley (2010), we made simulations based on 2000 random samples of size $n = 800$ (for the first sample X_1) and $m = 700$ (for the second sample X_2) generated from the following families of distributions which satisfy the first and second order conditions (1) and (2) :

- The Fréchet distribution (denoted by F) with parameter $\gamma > 0$ given by $F(x) = \exp(-x^{-1/\gamma})\mathbb{I}_{x>0}$, for which $\rho = -1$.
- The Student distribution (denoted by t) with ν degrees of freedom, for which $\gamma = 1/\nu$ and $\rho = -2\gamma$.
- The Burr distribution (denoted by B) with parameters $\gamma > 0$ and $\rho < 0$, given by $F(x) = 1 - (1 + x^{\rho/\gamma})^{-1/\rho}\mathbb{I}_{x>0}$.

In Subsections 3.1 and 3.2, we apply the method proposed by Hall and Welsh (1985) for the estimation of the optimal (in the sense of minimization of the mean square error) sample fractions k_1 and k_2 given by :

$$\hat{k}_1 = n^{-2\hat{\rho}_1/(1-2\hat{\rho}_1)} \ \text{ and } \ \hat{k}_2 = m^{-2\hat{\rho}_2/(1-2\hat{\rho}_2)},$$

where $\hat{\rho}_j$ is the estimator of the second order parameter ρ_j ($j = 1$ or 2). We used the one proposed in Fraga Alves et al (2003) for comparison reasons. Recent alternatives can be found in Worms (2012), Ciuperca and Mercadier (2010), Goegebeur et al (2010) and de Wet et al (2012). Results for the type I error are given in Subsection 3.1 and for the power function in Subsection 3.2 .

In Subsection 3.1, we also investigated the effect of the choice of k_1 and k_2 on the type I accuracy of the test, by performing simulations under a large range of values for the sample fractions k_1 and k_2.

3.1. Type I error estimation

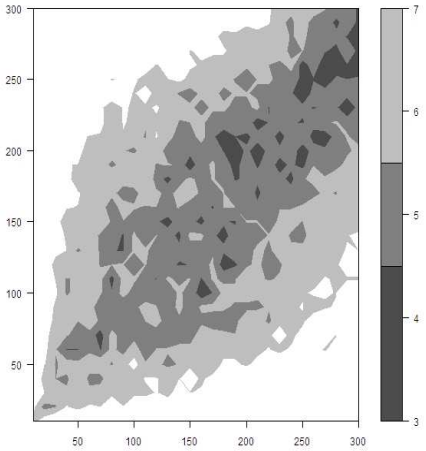
Table 1 provides the estimated type I error $\hat{\alpha}$ at the optimal sample fractions mentioned above, for nominal risks $\alpha = 5\%$ and $\alpha = 10\%$, and for different combinations of the above mentioned distributions, where the second order parameters ρ_1 and ρ_2 can be equal or different. We add to our results those of Mougeot and Tribouley (2010) taken from their Table 2, when using the Hill estimator. In the following table, our results are noted WW and those of Mougeot and Tribouley (2010), MT.

Table 1. Type I error $\hat{\alpha}$

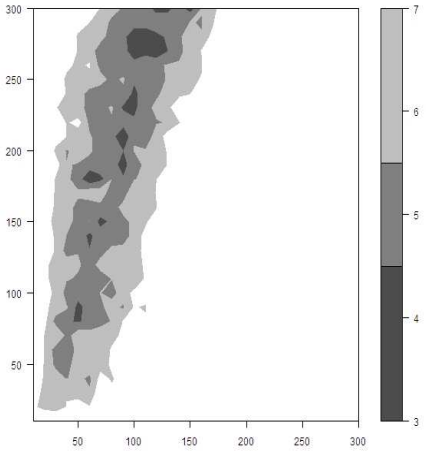
Distr.	$\gamma_1 = \gamma_2$	ρ_1	ρ_2	WW		MT	
				$\hat{\alpha}$ (5%)	$\hat{\alpha}$ (10%)	$\hat{\alpha}$ (5%)	$\hat{\alpha}$ (10%)
tt	0.25	-0.5	-0.5	4.50	8.65	6	17
tt	0.5	-1	-1	5.10	10.85	11	25
tt	1	-2	-2	18.10	25.30	9	20
FF	0.25	-1	-1	6.10	12.25	1	8
FF	0.5	-1	-1	6.50	10.70	2	9
FF	1	-1	-1	6.65	10.10	3	10
BB	0.25	-2	-2	4.25	9.00	3	10
BB	0.25	-1	-2	5.00	10.00	14	28
BB	1	-1	-1	5.40	9.95	9	19
BB	1	-1	-2	5.30	11.35	14	29
tF	0.5	-1	-1	6.15	10.80	33	57
tF	1	-2	-1	7.55	12.60	7	16
tB	0.5	-1	-0.5	15.00	24.80	16	36
tB	1	-2	-1	5.65	11.00	10	22
tB	1	-2	-2	6.60	14.00	8	20
FB	0.25	-1	-1	5.60	10.55	12	28
FB	0.5	-1	-1	5.50	10.85	10	24
FB	1	-1	-1	5.80	10.20	14	26
FB	0.25	-1	-2	5.30	11.90	3	10
FB	0.5	-1	-2	5.80	11.15	2	9
FB	1	-1	-2	6.75	12.35	3	10

We see, through these simulations, that the procedure we propose, though simple to implement, yields generally sharper, more satisfactory results than those from Mougeot and Tribouley (2010) (except for a few cases), even without using any bootstrap calibration for our test. The same conclusion holds when, in their method, the Hill estimator is replaced by the other estimators they considered in their simulations

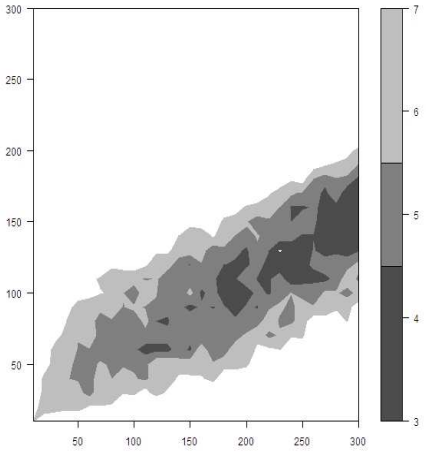
We also present, in Figure 1, some contour graphs of the empirical type I error $\hat{\alpha}$, for a wide range of the sample fractions k_1 and k_2. We can observe that the accuracy of course depends on the choice of these fractions, but there is some kind of permissibility for this choice before obtaining unsatisfactory results (here "satisfactory" was arbitrarily defined as "being included in the interval $[3\%, 7\%]$", for a nominal risk $\alpha = 5\%$) .



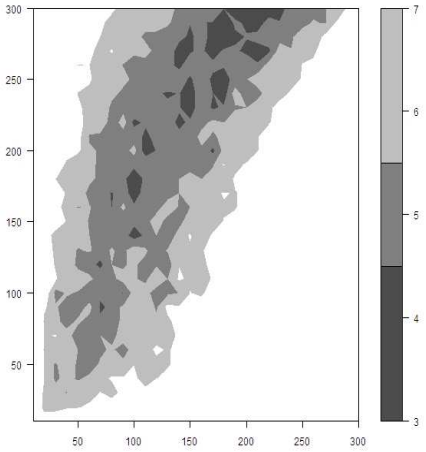
(a) $X_1 \sim \text{Burr}(1/4, -2)$, $X_2 \sim \text{Burr}(1/4, -2)$



(b) $X_1 \sim \text{Burr}(1/4, -1)$, $X_2 \sim \text{Burr}(1/4, -2)$



(c) $X_1 \sim \text{Frechet}(1/2)$, $X_2 \sim \text{Burr}(1/2, -1)$



(d) $X_1 \sim \text{Frechet}(1/2)$, $X_2 \sim \text{Burr}(1/2, -2)$

Figure 1: Estimated type I error $\hat{\alpha}$ as a function of the sample fractions k_1 and k_2, for $\alpha = 5\%$. White areas correspond to cases where $\hat{\alpha}$ turned out to be greater than 7%

3.2. Power function estimation

In this part, we get the extreme value index γ_1 of X_1 fixed and calculate the power associated to our test procedure for different values of γ_2, the extreme value index of X_2. For example, if $\gamma_1 = 1/2$, we take γ_2 between 0.1 and 1. Moreover, ρ_1 and ρ_2 can be taken equal or different. We use again the method of Hall and Welsh (1985) to estimate the optimal k_1 and k_2 values. Figure 2 gives two examples of estimated power functions for Burr and Fréchet distributions.

In both examples, the graphic on the left is such that ρ_1 is always equal to ρ_2, whereas the graphic on the right is such that ρ_1 is always different from ρ_2. We remarked through these two examples and others not presented here, that when the two distributions of X_1 and X_2 have the same second order parameters ρ_1 and ρ_2, the estimated power function is, as expected, minimal when the tail indices γ_1 and γ_2 are equal. This is not the case when the second order parameters ρ_1 and ρ_2 are different.

The empirical results look rather satisfactory, though no comparison with other methods can be made due to the fact that the extreme value indices comparison topic has hardly been addressed in the literature before.

4. Conclusion

In this work, we presented a first attempt to address the problem of comparing extreme value indices of different heavy-tailed samples, through some statistical test with prescribed theoretical accuracy. Exception made of Mougeot and Tribouley (2010), we know of no work addressing explicitly and methodologically this topic in the extreme value literature. The method we propose (i) appears to perform relatively well in terms of empirical type I error accuracy and power, (ii) is valid for comparison of more than 2 samples, and (iii) allows for testing linear hypotheses more general than just the equality of all the extreme value indices. Its inspiration stems from the empirical likelihood work on linear models in Owen (1991) and from J.C. Lu and L. Peng (2002), which yields confidence intervals for the tail index based on the Hill estimator.

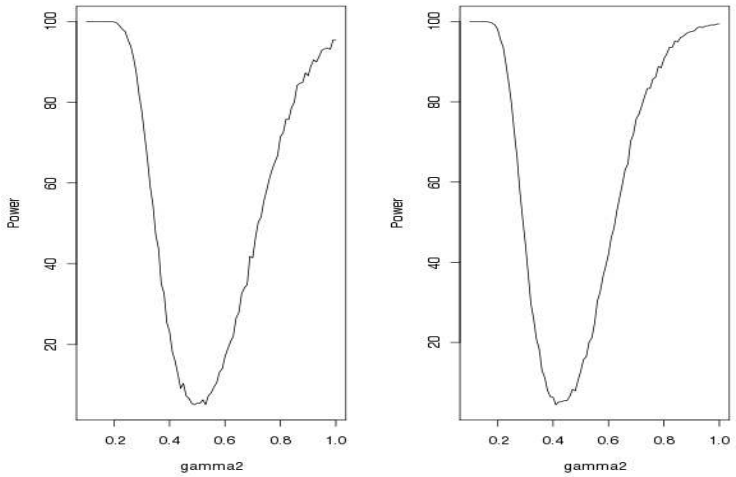
One possibility to generalize this work would be to define a similar test statistic as ours, but based on more efficient tail index estimators than the Hill estimator ; it should though be stressed that the particular structure of the Hill estimator has been used here to prove the asymptotic distribution of the test statistic (the Hill estimator is the mean of log-spacings and its asymptotic variance is a simple function of the tail index alone). Another path to pursue the study of this topic could be to use known results on regression models for the tail index.

5. Appendix

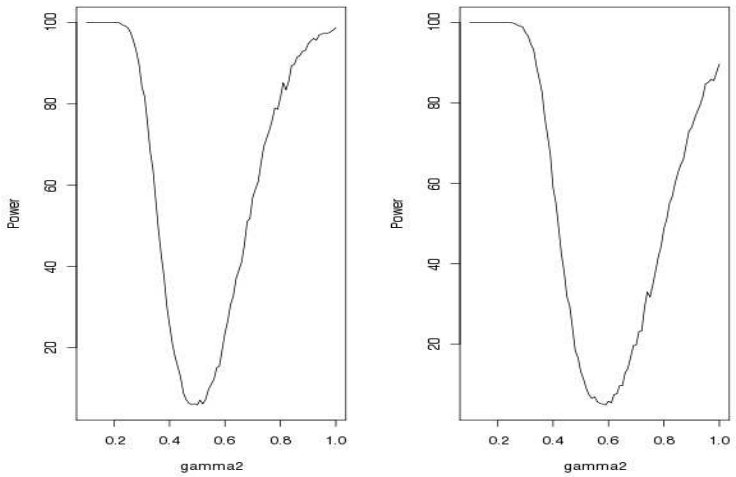
5.1. Proof of (5)

Let $m_0 = 0$ and $m_j = k_1 + \ldots + k_j$ (for $j \leq K$). We denote by $(Z_{l,n}(\gamma))_{\iota \leq K}$ the K-dimensional vector for which, if $l \in \{m_{j-1} + 1, \ldots, m_j\}$,

$$(Z_{l,n}(\gamma))_\iota := \begin{cases} 0 & \text{if} \quad \iota \neq j \\ Y_i^{(j)} - \gamma_j & \text{if} \quad \iota = j, \text{ with } i = l - m_{j-1} \end{cases}$$



(a) $X_1 \sim \mathrm{B}(1/2, -1)$, and $X_2 \sim \mathrm{B}(1/2, \rho_2)$, $\rho_2 = -1$ or $-1/2$



(b) $X_1 \sim \mathrm{B}(1/2, \rho_1)$, $\rho_1 = -1$ or $-1/2$, and $X_2 \sim \mathrm{F}(1/2)$

Figure 2: Estimated power function as γ_2 varies, when γ_1 is kept fixed

$ELR(\gamma)$ can then be rewritten as (remind that $k = \sum_{j=1}^{K} k_j$)

$$ELR(\gamma) := \sup_{(p_l)} \left\{ \prod_{l=1}^{k} (k p_l) \ ; \ p_l \geq 0, \ \sum_{l=1}^{k} p_l = 1, \ \sum_{l=1}^{k} p_l Z_{l,n}(\gamma) = 0 \right\},$$

The classical Lagrange multipliers method yields the following optimal weights p_l in the above description of $ELR(\gamma)$

$$p_{l,n}(\gamma) := \frac{1}{k(1+ < \lambda_n(\gamma), Z_{l,n}(\gamma) >)}$$

where $\lambda_n(\gamma)$ is determined as the solution of

$$\sum_{l=1}^{k} (1+ < \lambda_n(\gamma), Z_{l,n}(\gamma) >)^{-1} Z_{l,n}(\gamma) = 0, \tag{7}$$

so that

$$l(\gamma) = 2 \sum_{l=1}^{k} log(1+ < \lambda_n(\gamma), Z_{l,n}(\gamma) >).$$

Now, if we note $\Delta_n := diag(k_1/k, \ldots, k_K/k)$ and

$$\bar{Z}_n(\gamma) := \frac{1}{k} \sum_{l=1}^{k} Z_{l,n}, \quad \hat{V}_n(\gamma) := \frac{1}{k} \sum_{l=1}^{k} Z_{l,n} Z_{l,n}^t, \quad Z_n^*(\gamma) := \max_{1 \leq l \leq k} ||Z_{l,n}(\gamma)||,$$

then we have (remind the definitions of $G_n, \tilde{G}_n, B_n$ and M_n before the statement of Lemma 1), for any given $C > 0$ and $\gamma \in B_{n,C}$,

$$\begin{aligned}
\bar{Z}_n(\gamma) &= \Delta_n G_n(\gamma) &&= \Delta_n G_n(\gamma_0) + \Delta_n(\gamma_0 - \gamma) && (8)\\
\sqrt{k}\bar{Z}_n(\gamma) &= \Delta_n^{1/2} \tilde{G}_n(\gamma) &&= \Delta_n^{1/2} \tilde{G}_n(\gamma_0) + \Delta_n \sqrt{k}(\gamma_0 - \gamma) && (9)\\
\hat{V}_n(\gamma) &= \Delta_n B_n(\gamma) &&= \Delta_n B_n(\gamma_0) + k^{-1/2} \Delta_n^{1/2} O_n(\gamma_0 - \gamma) && (10)\\
Z_n^*(\gamma) &= M_n(\gamma) &&\leq M_n(\gamma_0) + \kappa \, ||\gamma_0 - \gamma|| && (11)
\end{aligned}$$

where O_n is a $K \times K$ matrix such that $\|O_n\| = O_{\mathbb{P}}(1)$, and κ is an absolute constant.

Please note that, in the sequel, all the asymptotic results (in $o_{\mathbb{P}}$ or $O_{\mathbb{P}}$) are established *uniformly* for $\gamma \in B_{n,C}$.

Now, as usual in empirical likelihood methodology, starting from the definition of $\lambda_n(\gamma)$ we obtain

$$||\lambda_n(\gamma)|| \, (u_n^t(\gamma) \hat{V}_n(\gamma) u_n(\gamma) - Z_n^*(\gamma) ||\bar{Z}_n(\gamma)||) \ \leq \ ||\bar{Z}_n(\gamma)||$$

where $u_n(\gamma) := \lambda_n(\gamma)/||\lambda_n(\gamma)||$. It is clear from Lemma 1, relations (9) and (11), and the definition of the ball $B_{n,C}$ that

$$\sqrt{k}||\bar{Z}_n(\gamma)|| = O_{\mathbb{P}}(1) \quad \text{and} \quad Z_n^*(\gamma) = o_{\mathbb{P}}(\sqrt{k}).$$

Consequently,

$$||\lambda_n(\gamma)|| = O_{\mathbb{P}}(k^{-1/2})$$

since $u_n^t(\gamma) \hat{V}_n(\gamma) u_n(\gamma)$ is bounded away from 0, which is due to the assumption that k_j/k is bounded away from 0 (which implies that $\|\Delta_n\| = O_{\mathbb{P}}(1)$) and that $\hat{V}_n(\gamma) = \Delta_n^{1/2} B \Delta_n^{1/2} + o_{\mathbb{P}}(1)$, with B invertible (since the indices γ_{0j} are positive).

Therefore, setting $g_{l,n}(\gamma) := < \lambda_n(\gamma), Z_{l,n}(\gamma) >$ we have

$$g_n^* \ := \ \sup\{||g_{l,n}(\gamma)||, \ \gamma \in B_{n,C}\} \ = \ o_{\mathbb{P}}(1). \tag{12}$$

It follows classically (from the definition of $\lambda_n(\gamma)$) that

$$\hat{V}_n(\gamma)\lambda_n(\gamma) = \bar{Z}_n(\gamma) + R_n(\gamma),$$

where $||R_n(\gamma)|| \leq 2(g_n^*)^2||\bar{Z}_n(\gamma)|| = o_{\mathbb{P}}(k^{-1/2})$. By Lemma 1, it is simple to see that $B_n(\gamma) = B_n(\hat{\gamma}) + o_{\mathbb{P}}(1)$ and thus (10) leads to

$$\sqrt{k}\lambda_n(\gamma) \ = \ (B_n(\hat{\gamma}))^{-1}\Delta_n^{-1/2}\widetilde{G}_n(\gamma) + o_{\mathbb{P}}(1). \tag{13}$$

Using (12), a Taylor expansion yields

$$l(\gamma) \ = \ 2\sum_{l=1}^{k}\log(1+g_{l,n}(\gamma)) \ = \ 2\sum_{l=1}^{k}g_{l,n}(\gamma) - \sum_{l=1}^{k}(g_{l,n}(\gamma))^2 + R_n'(\gamma),$$

where $R_n'(\gamma) = \frac{2}{3}\sum_{l=1}^{k}(g_{l,n}(\gamma))^3(1+\xi_{l,n}(\gamma))^{-3}$ for some $\xi_{l,n}(\gamma) \in]0, g_{l,n}(\gamma)[$. Moreover, by (9), (10) and (13),

$$\sum_{l=1}^{k}g_{l,n}(\gamma) \ = \ (\sqrt{k}\lambda_n(\gamma))^t(\sqrt{k}\bar{Z}_n(\gamma)) \ = \ (\widetilde{G}_n(\gamma))^t(B_n(\hat{\gamma}))^{-1}\widetilde{G}_n(\gamma) + o_{\mathbb{P}}(1)$$

and

$$\begin{aligned}
\textstyle\sum_{l=1}^{k}(g_{l,n}(\gamma))^2 &= (\sqrt{k}\lambda_n(\gamma))^t\hat{V}_n(\gamma)(\sqrt{k}\lambda_n(\gamma)) \\
&= (\widetilde{G}_n(\gamma) + o_{\mathbb{P}}(1))^t\Delta_n^{-1/2}(B_n(\hat{\gamma}))^{-1}(\Delta_n B_n(\hat{\gamma}) + o_{\mathbb{P}}(1))(B_n(\hat{\gamma}))^{-1}\Delta_n^{-1/2}(\widetilde{G}_n(\gamma) + o_{\mathbb{P}}(1)) \\
&= (\widetilde{G}_n(\gamma))^t(B_n(\hat{\gamma}))^{-1}\widetilde{G}_n(\gamma) + o_{\mathbb{P}}(1).
\end{aligned}$$

Finally, by (12) it comes $|R_n'(\gamma)| \leq o_{\mathbb{P}}(1)\sum_{l=1}^{k}(g_{l,n}(\gamma))^2 = o_{\mathbb{P}}(1)$ and

$$l(\gamma) \ = \ (\widetilde{G}_n(\gamma))^t(B_n(\hat{\gamma}))^{-1}\widetilde{G}_n(\gamma) + o_{\mathbb{P}}(1) \ = \ Q(\gamma) + o_{\mathbb{P}}(1),$$

which concludes the first step of the proof of Theorem 1. Note that in the last equation, $B_n(\hat{\gamma})$ can be replaced indifferently by $B_n(\gamma)$ or $B_n(\gamma_0)$ without invalidating the approximation.

5.2. Proof of (6)

Recalling that $\mathcal{A} = \{\gamma \in]0, +\infty[^K, \ \gamma_1 = \ldots = \gamma_K\}$ and that $\delta_j = \sqrt{k_j/S_j^2(\hat{\gamma})}$, we readily have

$$\inf_{\gamma\in\mathcal{A}} Q(\gamma) = Q(\tilde{\gamma}, \ldots, \tilde{\gamma}), \quad \text{where} \ \ \tilde{\gamma} = \textstyle\sum_{j=1}^{K}\delta_j^2\hat{\gamma}_j / \sum_{j=1}^{K}\delta_j^2. \tag{14}$$

Moreover,

$$\inf_{\gamma\in\mathcal{A}} Q(\gamma) = ||W - W'||^2,$$

where $W = (\sqrt{k_j}(\hat{\gamma}_j - \gamma_{0j})/S_j(\hat{\gamma}))_{j=1\ldots K}$ and $W' = (\delta_j(\tilde{\gamma} - \gamma_{0j}))_{j=1\ldots K}$. By Lemma 1, positiveness of the γ_{0j}, and the fact that $S_j^2(\hat{\gamma}) - S_j^2(\gamma_0) \xrightarrow{\mathbb{P}} 0$, we see that W is asymptotically distributed as $N_K(0, I)$.

Moreover, setting (under H_0) $\gamma^* = \gamma_{01} = \ldots = \gamma_{0K}$, if we introduce $p = \sum_{j=1}^{K} \delta_j^2$, $\delta = (\delta_j)_{j=1,\ldots K}$ and $v_n = \frac{\delta}{\sqrt{p}}$, then, under H_0,

$$< W, v_n > v_n = \sum_{j=1}^{K} (\frac{\delta_j}{\sqrt{p}} \delta_j (\hat{\gamma}_j - \gamma_{0j})) \frac{v_n}{\sqrt{p}} = \tilde{\gamma}\delta - \gamma^*\delta = W',$$

so that W' is the orthogonal projection of W on a subspace of dimension 1. Equation (6) follows.

Finally, if the hypothesis tested is H_0' : "$C\gamma_0 = 0$" as in the second part of Theorem 1, then let us set $\Delta = diag(\delta_1, \ldots, \delta_K)$, $\tilde{C} = C\Delta^{-1}$, $w_0 = \Delta\gamma_0$. Since $W = \Delta(\hat{\gamma} - \gamma_0)$ is asymptotically distributed as $N_K(0, I)$ and, under H_0', we have $\tilde{C}w_0 = 0$, it comes

$$\inf_{\gamma \in \mathcal{A}'} Q(\gamma) = \inf_{\gamma} \{ \, \|\Delta(\hat{\gamma} - \gamma)\| \; ; \; C\gamma = 0 \, \} = \inf_{w} \{ \, \|\Delta(\hat{\gamma} - \gamma_0) - w\| \; ; \; \tilde{C}w = 0 \, \}$$

Because the subspace $\{ \, w \; ; \; \tilde{C}w = 0 \, \}$ is $(K - d)$-dimensional, the result thus comes by Cochran's theorem.

References

L. Bottolo, G. Consonni, P. Dellaportas and A. Lijoi (2003). Bayesian Analysis of Extreme Values by Mixture Modelling In *Extremes* 6, pages 25-47.

G. Ciuperca and C. Mercadier (2010). Semi-parametric estimation for heavy tailed distributions. In *Extremes* 13(1), pages 55-87.

I. Fraga Alves, M.I. Gomes and L. de Haan (2003). A new class of semi-parametric estimators of the second order parameter. In *Portugaliae Mathematica* **60**, pages 193213.

Y. Goegebeur, J. Beirlant and T. de Wet (2010). Kernel estimators for the second order parameter in extreme value statistics. In *Journal of statistical Planning and Inference* **140**, pages 26322652.

L. de Haan and L. Peng (1998). Comparison of tail index estimators. In *Statistica Neerlandica* 52(1), pages 60-70.

P. Hall and A. Welsh (1985). Adaptive estimates of parameters of regular variation. In *Annals of Statistics* 13, pages 331341.

J.C. Lu and L. Peng (2002). Likelihood based confidence intervals for the tail index. In *Extremes* 5, pages 337-352.

M. Mougeot and C. Tribouley (2010). Procedure of test to compare tail indices. In *Ann Inst Stat Math* 62, pages 383-412.

A. Owen (1991). Empirical Likelihood for Linear Models In *Annals of Statistics* 19 (4), pages 1725-1747.

T. de Wet, Y. Goegebeur and M. Reimat Munch (2012). Asymptotically unbiased estimation of the second order tail parameter. In *Statistics and Probability Letters* 82, pages 565-573.

J. Worms and R. Worms (2012). Estimation of second order parameters using Probability Weighted Moments. To appear in *ESAIM P&S* DOI : 10.1051/ps/2010017