



A simple estimator for the M-index of functions in M

Meitner Cadena

► To cite this version:

| Meitner Cadena. A simple estimator for the M-index of functions in M. 2015. hal-01142162

HAL Id: hal-01142162

<https://hal.science/hal-01142162>

Preprint submitted on 14 Apr 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A simple estimator for the \mathcal{M} -index of functions in \mathcal{M}

Meitner Cadena*

Abstract

An estimator for the \mathcal{M} -index of functions of \mathcal{M} , a larger class than the class of regularly varying (RV) functions, is proposed. This index is the tail index of RV functions and this estimator is thus a new one on the class of RV functions. This estimator satisfies, assuming suitable conditions, strong consistency. Asymptotic normality of this estimator is proved for a large class of RV functions.

Keywords: estimator; regularly varying function; weak consistency; strong consistency; extreme value theory

AMS classification: 62G05; 62G30

1 Introduction

A measurable function $U : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is a *regularly varying* (RV) function if (see [9] and e.g. [1]), for some $\alpha \in \mathbb{R}$, for any $t > 0$, $U(tx)/U(x) \rightarrow t^\alpha$ ($x \rightarrow \infty$). α is called the tail index of U . If $\alpha = 1$, U is a *slowly varying* (SV) function. RV_α denotes the class of RV functions with tail index α and SV the class of SV functions.

A main concern in extreme value theory (EVT) is the estimation of the tail index of RV functions that are tails of distribution functions. Many estimators of this index have been proposed. The best known estimator among them is one proposed by Hill [8] in 1975, which shows favorable features to be used than other competitors (see e.g. [11], pp. 1181-1182).

Cadena and Kratz [2] introduced in 2014 the class \mathcal{M} which consists of measurable functions $U : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ satisfying

$$\exists \alpha \in \mathbb{R}, \forall \varepsilon > 0, \lim_{x \rightarrow \infty} \frac{U(x)}{x^{\alpha+\varepsilon}} = 0 \quad \text{and} \quad \lim_{x \rightarrow \infty} \frac{U(x)}{x^{\alpha-\varepsilon}} = \infty. \quad (1)$$

One can prove that α in (1) is unique, which we call in this paper the \mathcal{M} -index of U . Furthermore, using Proposition 2.1, one can prove that $RV \subseteq \mathcal{M}$ and that α in (1) is the tail index of U if $U \in RV_\alpha$.

All over this paper X is a positive random variable (r.v.) with distribution $F(x) = P(X \leq x)$ and distribution tail $\bar{F}(x) = 1 - F(x) = P(X > x)$. Suppose that the endpoint of F is

*UPMC Paris 6 & CREAR, ESSEC Business School; E-mail: meitner.cadena@etu.upmc.fr or b00454799@essec.edu

infinite, i.e. $x^* = \sup\{x : F(x) < 1\} = \infty$. Let X_1, \dots, X_n be independent and identically distributed (i.i.d.) r.v.s with distribution F and let $X_{1:n} \leq \dots \leq X_{n:n}$ be the order statistics of this sample. The empirical distribution function of F is defined by, for $x > 0$, $F_n(x) = n^{-1} \sum_{i=1}^n I_{\{X_i \leq x\}}$, where I_A denotes the indicator function of A . We denote by $\bar{F}_n = 1 - F_n$ its empirical tail of distribution.

$\xrightarrow{a.s.}$, \xrightarrow{p} , and \xrightarrow{d} are convergences almost surely (a.s.), in probability, and in distribution respectively, $\log(x)$ is the natural logarithm of x , $\lfloor x \rfloor$ is the highest integer $\leq x$, $o_p(1)$ is a sequence of random values that converges to 0 in probability, and $O_p(1)$ is such sequence bounded in probability (see e.g. [12]).

Assume $\bar{F} \in \mathcal{M}$ with \mathcal{M} -index $-1/\alpha$ (< 0).

Assume $k = k(n)$ is a sequence of positive integers satisfying

$$1 \leq k \leq n-1, \quad k \rightarrow \infty, \quad \text{and} \quad k/n \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty. \quad (\text{k})$$

We aim to formulate an estimator for α . To this aim, we notice that, applying Proposition 2.1, $\log(\bar{F}_n(x))/\log(x) \rightarrow -1/\alpha$ ($x \rightarrow \infty$). Then, a natural estimator of α is, for $0 \leq k < n$,

$$\hat{\alpha} = -\frac{\log(X_{n-k:n})}{\log(\bar{F}_n(X_{n-k:n}))}.$$

This may be rewritten as

$$\hat{\alpha} = \frac{\log(X_{n-k:n})}{\log(n/k)}. \quad (2)$$

Then, in this note, the following results are proved.

Theorem 1.1 (Weak consistency of $\hat{\alpha}$). *Let X_1, \dots, X_n be i.i.d. r.v.s with continuous distribution F such that its tail \bar{F} belongs to \mathcal{M} with \mathcal{M} -index $-1/\alpha$ ($\alpha > 0$). Let $X_{1:n} \leq \dots \leq X_{n:n}$ be its order statistics. Then, if (k) is satisfied,*

$$\hat{\alpha} \xrightarrow{p} \alpha \quad (n \rightarrow \infty). \quad (3)$$

Theorem 1.2 (Strong consistency of $\hat{\alpha}$). *Let X_1, \dots, X_n be i.i.d. r.v.s with continuous distribution F such that its tail \bar{F} belongs to \mathcal{M} with \mathcal{M} -index $-1/\alpha$ ($\alpha > 0$). Let $X_{1:n} \leq \dots \leq X_{n:n}$ be its order statistics. Then, if $k = k(n) = \lfloor n^\delta \rfloor$, $0 < \delta < 1$,*

$$\hat{\alpha} \xrightarrow{a.s.} \alpha \quad (n \rightarrow \infty).$$

Theorem 1.3 (Asymptotic normality of $\hat{\alpha}$). *Assume that F is a continuous distribution with tail \bar{F} satisfying $\bar{F}(x) = x^{-\alpha}(1 + o(x^{-\beta}))$ as $x \rightarrow \infty$, for positive constants α and β . Let X_1, \dots, X_n be i.i.d. r.v.s with distribution F , and let $X_{1:n} \leq \dots \leq X_{n:n}$ be its order statistics. Then, if (k) is satisfied,*

$$\sqrt{k} \log\left(\frac{n}{k}\right) \left(\frac{1}{\hat{\alpha}} - \frac{1}{\alpha}\right) \xrightarrow{d} \mathcal{N}(0, \alpha^{-2}) \quad (n \rightarrow \infty). \quad (4)$$

2 Proofs

Proposition 2.1. *Let $U : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be a measurable function. Then, $U \in \mathcal{M}$ with \mathcal{M} -index $\alpha = \tau$ iff $\log(U(x))/\log(x) \rightarrow \eta$ ($x \rightarrow \infty$).*

This result was given by [2] and for the sake of completeness of this note, we give its proof, part of it being copied from [2].

Proof. Let $U : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be a measurable function and let $\epsilon > 0$.

Assume $U \in \mathcal{M}$ with \mathcal{M} -index $\alpha = \tau$. Then, by definition,

$$\lim_{x \rightarrow \infty} \frac{U(x)}{x^{\tau+\epsilon}} = 0 \quad \text{and} \quad \lim_{x \rightarrow \infty} \frac{U(x)}{x^{\tau-\epsilon}} = \infty.$$

Hence, there exists $x_0 \geq 1$ such that, for $x \geq x_0$,

$$U(x) \leq \epsilon x^{\tau+\epsilon} \quad \text{and} \quad U(x) \geq \frac{1}{\epsilon} x^{\tau-\epsilon}.$$

Applying the logarithm function to these inequalities and dividing them by $\log(x)$ (with $x > 1$) provide

$$\frac{\log(U(x))}{\log(x)} \leq \frac{\log(\epsilon)}{\log(x)} + \tau + \epsilon \quad \text{and} \quad \frac{\log(U(x))}{\log(x)} \geq -\frac{\log(\epsilon)}{\log(x)} + \tau - \epsilon,$$

and, one then has

$$\overline{\lim}_{x \rightarrow \infty} \frac{\log(U(x))}{\log(x)} \leq \tau + \epsilon \quad \text{and} \quad \underline{\lim}_{x \rightarrow \infty} \frac{\log(U(x))}{\log(x)} \geq \tau - \epsilon,$$

from which, taking ϵ arbitrary, the assertion follows.

Conversely, assume $\log(U(x))/\log(x) \rightarrow \eta$ ($x \rightarrow \infty$). Then, there exists $x_0 > 1$ such that, for $x \geq x_0$, $|\log(U(x))/\log(x) - \tau| \leq \epsilon/2$.

Writing, for $w \in \{\epsilon, -\epsilon\}$,

$$\frac{U(x)}{x^{\tau+w}} = \exp \left\{ \log(x) \times \left(\frac{\log(U(x))}{\log(x)} - \tau - w \right) \right\}$$

gives

$$\exp \left\{ \log(x) \times \left(-\frac{\epsilon}{2} - w \right) \right\} \leq \frac{U(x)}{x^{\tau+w}} \leq \exp \left\{ \log(x) \times \left(\frac{\epsilon}{2} - w \right) \right\},$$

and then,

$$\lim_{x \rightarrow \infty} \frac{U(x)}{x^{\tau+\epsilon}} \leq \lim_{x \rightarrow \infty} \exp \left\{ \log(x) \times \left(\frac{\epsilon}{2} - \epsilon \right) \right\} = 0$$

and

$$\lim_{x \rightarrow \infty} \frac{U(x)}{x^{\tau-\epsilon}} \geq \lim_{x \rightarrow \infty} \exp \left\{ \log(x) \times \left(\frac{\epsilon}{2} + \epsilon \right) \right\} = \infty.$$

These two limits provide $U \in \mathcal{M}$ with \mathcal{M} -index $\alpha = \tau$. This completes the proof.

Lemma 2.1. Let X_1, X_2, \dots be i.i.d. r.v.s with distribution $F: \mathbb{R}^+ \rightarrow \mathbb{R}^+$. Let $X_{1:n} \leq \dots \leq X_{n:n}$ be its order statistics. Then, if (k) is satisfied,

$$X_{n-k:n} \xrightarrow{a.s.} \infty \quad (n \rightarrow \infty).$$

Proof. Applying the proof of Lemma 3.2.1 given in [5], which is independent of the distribution F .

Lemma 2.2. Let E_1, \dots, E_n be i.i.d. r.v.s with standard exponential distribution F . Let $E_{1:n} \leq \dots \leq E_{n:n}$ be its order statistics. Then, if (k) is satisfied,

$$\sqrt{k} \left(E_{n-k:n} - \log \left(\frac{n}{k} \right) \right)$$

is asymptotically standard normal as $n \rightarrow \infty$.

Proof. Let $U = (1 - F)^-$, $U(x) = \inf \{y : F(y) \geq 1 - 1/y\}$. Since $1 - F(x) = e^{-x}$ we have $U(x) = \log(x)$. Then, by Theorem 2.1.1 of [5],

$$\sqrt{k} \frac{E_{n-k:n} - U\left(\frac{n}{k}\right)}{\frac{n}{k} U'\left(\frac{n}{k}\right)} = \sqrt{k} \frac{E_{n-k:n} - \log\left(\frac{n}{k}\right)}{\frac{n}{k} \times \frac{k}{n}} = \sqrt{k} \left(E_{n-k:n} - \log\left(\frac{n}{k}\right) \right)$$

is asymptotically standard normal as $n \rightarrow \infty$, supposing (k).

Lemma 2.3. Let E_1, \dots, E_n be i.i.d. r.v.s with standard exponential distribution F . Let $E_{1:n} \leq \dots \leq E_{n:n}$ be its order statistics. Then, if (k) is satisfied,

$$\frac{E_{n-k:n}}{\log\left(\frac{n}{k}\right)} \xrightarrow{p} 1 \quad (n \rightarrow \infty).$$

Proof. Application of Lemma 2.2 and e.g. [12], exercice 18, page 24.

Proof of Theorem 1.1. By hypothesis we have $\log(\bar{F}(x))/\log(x) \rightarrow -1/\alpha$ ($x \rightarrow \infty$). Then, noting that $X_{n-k:n} \rightarrow \infty$ ($n \rightarrow \infty$) by Lemma 2.1, we have

$$\lim_{n \rightarrow \infty} \frac{\log(X_{n-k:n})}{\log\left(\frac{n}{k}\right)} = \lim_{n \rightarrow \infty} \frac{\log(X_{n-k:n})}{\log(\bar{F}(X_{n-k:n}))} \times \frac{\log(\bar{F}(X_{n-k:n}))}{\log\left(\frac{n}{k}\right)} \stackrel{a.s.}{=} -\alpha \lim_{n \rightarrow \infty} \frac{\log(\bar{F}(X_{n-k:n}))}{\log\left(\frac{n}{k}\right)},$$

from which, noting that $-\log(\bar{F}(X))$ is a r.v. following a standard exponential distribution G (see e.g. [6], page 18), $-\log(\bar{F}(X_1)), \dots, -\log(\bar{F}(X_n))$ is a sample of i.i.d. r.v.s following G , and $-\log(\bar{F}(X_{1:n})), \dots, -\log(\bar{F}(X_{n:n}))$ is its order statistics (see e.g. [6], page 20), then applying Lemma 2.3, follows

$$\lim_{n \rightarrow \infty} \hat{\alpha} = \lim_{n \rightarrow \infty} \frac{\log(X_{n-k:n})}{\log\left(\frac{n}{k}\right)} \stackrel{p}{=} \alpha.$$

This concludes the proof.

Lemma 2.4 (Lemma 5.1 given by Davis and Resnick (1984) [4]). Let E_1, \dots, E_n be i.i.d. r.v.s with standard exponential distribution F . Let $E_{1:n} \leq \dots \leq E_{n:n}$ be its order statistics. Then, if $k = k(n) = \lfloor n^\delta \rfloor$, $0 < \delta < 1$,

$$E_{n-k:n} - \log\left(\frac{n}{k}\right) \xrightarrow{a.s.} 0 \quad (n \rightarrow \infty).$$

Proof of Theorem 1.2. Using Lemma 2.4 instead of Lemma 2.3 in the proof of Theorem 1.2.

Lemma 2.5 (Lemma given in [7], page 40). *Suppose $\lambda > 0$. If (k), then*

$$-\frac{1}{\lambda} \sum_{j=1}^{n-k} \frac{E_{n-j+1}}{n-j+1} = -\frac{1}{\lambda} \log\left(\frac{n}{k}\right) + \log(1 + O_p(k^{-1/2})).$$

Proof of Theorem 1.3. Applying the transformation $x \mapsto x^{-1}$ used by Hall in [7] to have simpler notations, $\bar{F}(x) = x^{-\alpha}(1 + o(x^{-\beta}))$ as $x \rightarrow \infty$ becomes $F(x) = x^{\alpha}(1 + o(x^{\beta}))$ as $x \rightarrow 0^+$. This expression we may write, as $t \rightarrow 0^+$,

$$F^{-1}(t) = t^{1/\alpha} e^{o(t^{\beta})}. \quad (5)$$

Let E_1, \dots, E_n be independent standard exponential r.v.s. Define

$$S_k = \sum_{j=1}^{n-k+1} \frac{E_{n-j+1}}{n-j+1} = \sum_{j=1}^{n-k+1} \frac{E_{n-j+1} - 1}{n-j+1} + \log\left(\frac{k+1}{n}\right) + \log(1 + O(k^{-1})).$$

Then, by Rényi's representation of order statistics (see [10] and e.g. [3], page 21), we may write $1/X_{k:n} = F^{-1}(e^{-S_k})$ for $k = 1, \dots, n$. It now follows from (5) that

$$\log(X_{k:n}) = \frac{1}{\alpha} S_k + o_p(e^{-\beta S_k}), \quad (6)$$

Applying Lemma 2.5 to (6) gives

$$\begin{aligned} \log(X_{k:n}) &= \frac{1}{\alpha} \sum_{j=1}^{n-k+1} \frac{E_{n-j+1} - 1}{n-j+1} + \frac{1}{\alpha} \log\left(\frac{n}{k+1}\right) \\ &\quad + \frac{1}{\alpha} \log(1 + O_p((k+1)^{-1/2})) + o_p(((k+1)/n)^{\beta}(1 + O_p((k+1)^{-1/2})^{\beta})) \\ &= \frac{1}{\alpha} \sum_{j=1}^{n-k+1} \frac{E_{n-j+1} - 1}{n-j+1} + \frac{1}{\alpha} \log\left(\frac{n}{k+1}\right) + \left(1 + \frac{1}{\alpha}\right) o_p(1). \end{aligned} \quad (7)$$

The mean of $\alpha^{-1} \sum_{j=1}^{n-k+1} (E_{n-j+1} - 1)/(n-j+1)$ is 0 and its variance

$$\frac{1}{\alpha^2} \sum_{j=k}^n \frac{1}{j^2} \sim \frac{1}{k\alpha^2} \quad (n \rightarrow \infty),$$

where $f(x) \sim g(x)$ ($x \rightarrow \infty$) means $f(x)/g(x) \rightarrow 1$ ($x \rightarrow \infty$). Hence, $\log(X_{k:n}) - \alpha^{-1} \log(n/k)$ is asymptotically $\mathcal{N}(0, k^{-1}\alpha^{-2})$ as $n \rightarrow \infty$, and (4) then follows. This completes the proof.

Acknowledgments

I am indebted to Prof. Einmahl for having pointed out an error in the formulation of (4). I acknowledges the support of SWISS LIFE through its ESSEC research program on 'Consequences of the population ageing on the insurances loss'.

References

- [1] N. Bingham, C. Goldie, J. Teugels, Regular Variation. *Cambridge University Press* (1989).
- [2] M. Cadena and M. Kratz, An extension of the class of regularly varying functions. *arXiv:1411.5276 [math.PR]*, (2014).
- [3] H. David, Order Statistics. *John Wiley & Sons, Inc.* (1981).
- [4] R. Davis, S. Resnick, Tail Estimates Motivated by Extreme Value Theory. *Ann. Stat.* **12**, (1984) 1467-1487.
- [5] L. de Haan, A. Ferreira, Extreme Value Theory. An Introduction. *Springer*, (2006).
- [6] J. Galambos, S. Kotz, Characterizations of Probability Distributions. (Lecture Notes in Mathematics). *Springer*, (1978).
- [7] P. Hall, On Some Simple Estimates of an Exponent of Regular Variation. *Journal of the Royal Statistical Society. Series B (Methodological)* **44**, (1982) 37-42.
- [8] B. Hill, A Simple General Approach to Inference About the Tail of a Distribution. *Ann. Stat.* **3**, (1975) 1163-1174.
- [9] J. Karamata, Sur un mode de croissance régulière des fonctions. *Mathematica (Cluj)* **4**, (1930) 38-53.
- [10] A. Rényi, On the theory of order statistics. *Acta Mathematica Academiae Scientiarum Hungaricae* **3-4**, (1953) 191-231.
- [11] R. Smith, Estimating Tails of Probability Distributions. *Ann. Stat.* **15**, (1987) 1174-1207.
- [12] A. van der Vaart, Asymptotic Statistics. *Cambridge University Press*, (1998).