INSTITUTO NACIONAL DE ESTATÍSTICA
PORTUGAL

REVSTAT

Statistical Journal

Volume 4, No.2
June 2006

Catalogação Recomendada

REVSTAT. Lisboa, 2003-
Revstat : statistical journal / ed. Instituto Nacional de Estatística. - Vol. 1, 2003- . - Lisboa I.N.E., 2003- . - 30 cm
Semestral. - Continuação de : Revista de Estatística = ISSN 0873-4275. - edição exclusivamente em inglês
ISSN 1645-6726

- **PUBLISHER**
  - *Instituto Nacional de Estatística (INE)*
    Av. António José de Almeida, 2
    1000-043 LISBOA
    PORTUGAL
    Tel.: (0351) 21 842 61 00
    Fax: (0351) 21 844 04 01
    Web site: http://www.ine.pt
    Customer Support Service: 808 201 808
- **COVER DESIGN**
  - *Mário Bouçadas, designed on the stain glass window at INE by the painter Abel Manta*
- **LAYOUT AND GRAPHIC DESIGN**
  - *Carlos Perpétuo*
- **PRINTING**
  - *Instituto Nacional de Estatística*
- **EDITION**
  - *350 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*

PRICE

[VAT 5% included]

- Single issue .... **€ 10**
- Annual subscription ... **€ 16**

INDEX

REVSTAT – Statistical Journal
Volume 4, Number 2, June 2006, 81–109

IMPROVEMENTS IN THE ESTIMATION OF A HEAVY TAIL*

Authors: Orlando Aníbal Oliveira
–

M. Ivette Gomes
– University of Lisbon, FCUL (DEIO) and CEAUL, Portugal
ivette.gomes@fc.ul.pt

M. Isabel Fraga Alves
– University of Lisbon, FCUL (DEIO) and CEAUL, Portugal
isabel.alves@fc.ul.pt

Received: June 2005 Revised: March 2006 Accepted: March 2006

Abstract:

- In this paper, and in a context of regularly varying tails, we suggest new tail index estimators, which provide interesting alternatives to the classical Hill estimator of the tail index γ. They incorporate some extra knowledge on the pattern of scaled top order statistics and seem to work generally pretty well in a semi-parametric context, even for cases where a second order condition does not hold or we are outside Hall's class of models. We shall give particular emphasis to a class of statistics dependent on a *tuning* parameter τ, which is merely a change in the scale of our data, from X to X/τ. Such a statistic is non-invariant both for changes in location and in scale, but compares favourably with the Hill estimator for a class of models where it is not easy to find competitors to this classic tail index estimator. We thus advance with a slight "controversial" argument: it is always possible to take advantage from a non-invariant estimator, playing with particular *tuning* parameters — either a change in the location or in the scale of our data —, improving then the overall performance of the classical estimators of extreme events parameters.

Key-Words:

- *statistics of extremes; semi-parametric estimation; Monte Carlo methods.*

AMS Subject Classification:

- 62G32, 62H12; 65C05.

*Research partially supported by FCT / POCTI and POCI / FEDER. A tribute to Orlando Oliveira, who died prematurely a month after getting his Ph.D. in May 2003.

1. INTRODUCTION AND PRELIMINARIES

Let $X_1, X_2, ..., X_n$ be independent random variables (r.v.'s) with common distribution function (d.f.) F, with a heavy upper tail, i.e., for large x, there exists $\gamma > 0$ such that

$$\overline{F}(x) := 1 - F(x) = x^{-1/\gamma} L_{_F}(x) \ ,$$

where $L_{_F}(x)$ is a slowly varying function, i.e., for every $x > 0$, $L_{_F}(tx)/L_{_F}(t) \to 1$ as $t \to \infty$. F is thus in the max-domain of attraction of an *Extreme Value* (EV) d.f.,

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

with $\gamma > 0$. We shall denote this fact by $F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)$.

Recall that, for $\gamma > 0$,

$$F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma) \quad \text{iff} \quad \overline{F} \in RV_{-1/\gamma} \quad \text{iff} \quad U \in RV_\gamma \ , \tag{1.1}$$

where $U(t) := F^{\leftarrow}(1-1/t),\ t > 1$ (Gnedenko, 1943; de Haan, 1970). RV_α stands for the class of regularly varying functions at infinity with index of regular variation equal to α, i.e., positive functions g with infinite right endpoint, and such that $\lim_{t\to\infty} g(tx)/g(t) = x^\alpha$, for all $x > 0$, and the notation $F^{\leftarrow}$ is used for the generalized inverse function of F, i.e., $F^{\leftarrow}(t) = \inf\{x\colon F(x) \geq t\}$.

The function $A(t)$ measures the rate of convergence of $\{\ln U(tx) - \ln U(t)\}$ towards $\{\gamma \ln x\}$ in (1.1), and it is a function of constant sign, such that

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} \ = \ \frac{x^\rho - 1}{\rho} \ , \tag{1.2}$$

for every $x > 0$, where ρ (≤ 0) is a *second order parameter.* The limit function in (1.2) must be of the stated form, and $|A(t)| \in RV_\rho$ (Geluk and de Haan, 1987).

1.1. The new estimation procedures

Let $X_{i:n}$ denote the i-th ascending order statistic (o.s.), $1 \leq i \leq n$, associated to the sample $\underline{X}_n = (X_1, X_2, ..., X_n)$. Under the validity of the first order framework in (1.1), with $U(t) = t^\gamma\, L_{_U}(t)$, $L_{_U} \in RV_0$, and for *intermediate* k, i.e.,

$$k = k_n \to \infty, \quad k/n \to 0, \qquad \text{as} \quad n \to \infty \ , \tag{1.3}$$

the classic tail index estimator for a positive γ is Hill's estimator (Hill, 1975), with the functional expression

$$\widehat{\gamma}_n^H(k) := \frac{1}{k} \sum_{i=1}^{k} \Big[\ln X_{n-i+1:n} - \ln X_{n-k:n}\Big] . \tag{1.4}$$

For this estimator, and whenever (1.3) holds, we have the validity of the distributional representation,

$$\widehat{\gamma}_n^H(k) \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}} P_k + \frac{1}{1-\rho} A(n/k) \big(1+ o_p(1)\big) ,$$

with P_k asymptotically standard normal (de Haan and Peng, 1998).

Also, under the validity of (1.3), it is possible to scale $X_{n-k:n}$ (or $X_{n-k+1:n}$), with $a_n = U(n)$, so that

$$\ln \frac{X_{n-k+1:n}}{a_n} + \gamma\, \psi(k) \underset{n\to\infty}{\overset{p}{\longrightarrow}} 0 . \tag{1.5}$$

And for every fixed i, $1 \le i < n$, there exists a non-degenerate r.v. ϵ_i, such that $\mathbb{E}[\epsilon_i] = 0$, and

$$\ln \frac{X_{n-i+1:n}}{a_n} + \gamma\, \psi(i) \underset{n\to\infty}{\overset{d}{\longrightarrow}} \epsilon_i . \tag{1.6}$$

As usual, ψ denotes the digamma function, i.e. $\psi(t) = d \ln\Gamma(t)/dt = \Gamma'(t)/\Gamma(t)$, being Γ the complete Gamma function, $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x}\, dx$, $t > 0$. For a justification of these results see Lemma 4.1. For details on the Γ and ψ functions, see Abramowitz and Stegun (1975??).

Let us then think on the least-squares' type estimators of γ and $b := \ln a$, which come from the minimization, jointly in γ and b, of

$$\sum_{i=1}^{k} \Big\{\ln X_{n-i+1:n} - b + \gamma\, \psi(i)\Big\}^2 .$$

Straightforward computations lead us to

$$\widetilde{b}_n(k) = \widetilde{\ln a}(k) = \frac{1}{k} \sum_{i=1}^{k} \ln X_{n-i+1:n} + \widetilde{\gamma}_n(k) \left(\frac{1}{k} \sum_{i=1}^{k} \psi(i)\right), \tag{1.7}$$

with

$$\widetilde{\gamma}_n(k) = \frac{\left(\sum\limits_{i=1}^{k} \psi(i)\right) \left(\sum\limits_{i=1}^{k} \ln X_{n-i+1:n}\right) - k \sum\limits_{i=1}^{k} \psi(i) \ln X_{n-i+1:n}}{k \sum\limits_{i=1}^{k} \psi^2(i) - \left(\sum\limits_{i=1}^{k} \psi(i)\right)^2} . \tag{1.8}$$

Remark 1.1. Notice that the replacement of $\psi(i)$ by $\{\ln i\}$ in the γ-estimator in (1.8) leads us to the estimator, based on a QQ-plot, studied in Kratz and Resnick (1996) and independently in Schultze and Steinbach (1996), and given by

$$\text{(1.9)} \qquad \widetilde{\gamma}_n^{(\mathcal{K})}(k) := \frac{\left(\sum\limits_{i=1}^{k} \ln i\right)\left(\sum\limits_{i=1}^{k} \ln X_{n-i+1:n}\right) - k \sum\limits_{i=1}^{k} (\ln i) \ln X_{n-i+1:n}}{k \sum\limits_{i=1}^{k} \ln^2 i - \left(\sum\limits_{i=1}^{k} \ln i\right)^2} .$$

Since $\psi(x) = \ln x + O(1/x)$, as $x \to \infty$, the difference between the estimators $\widetilde{\gamma}_n$ and $\widetilde{\gamma}_n^{(\mathcal{K})}$ is asymptotically negligible. However, for finite samples, their performance differs significantly, because the approximation in terms of the digamma function $\psi(i)$ is usually better than the use of $\{\ln i\}$ for all i between 1 and k.

We may easily simplify the expressions of $\widetilde{b}_n(k)$ and of $\widetilde{\gamma}_n(k)$ in (1.7) and (1.8), respectively, through the use of the following relations involving the digamma function,

$$\text{(1.10)} \qquad \sum_{j=1}^{k} \psi(j) = k\,\psi(k) - (k-1) = k\big(\psi(k+1) - 1\big) ,$$

$$\sum_{j=1}^{k} \psi^2(j) = k\,\psi^2(k+1) + 2k - (2k+1)\,\psi(k+1) + \psi(1)$$

and

$$k \sum_{j=1}^{j} \psi^2(j) - \left(\sum_{j=1}^{k} \psi(j)\right)^2 = k\Big\{k - \psi(k+1) + \psi(1)\Big\} = k \sum_{j=1}^{k}\left(1 - \frac{1}{j}\right) .$$

We then get the following linear combination of the top log-observations,

$$\text{(1.11)} \qquad \widetilde{\gamma}_n(k) = \frac{\sum\limits_{i=1}^{k}\Big(\psi(k+1) - \psi(i) - 1\Big) \ln X_{n-i+1:n}}{k - \psi(k+1) + \psi(1)} ,$$

and we may also write

$$\text{(1.12)} \qquad \widetilde{a}_n(k) = X_{n-k:n} \exp\Big(\widehat{\gamma}_n^H(k) + \widetilde{\gamma}_n(k)\,\big(\psi(k+1) - 1\big)\Big) ,$$

where $\widehat{\gamma}_n^H(k)$ and $\widetilde{\gamma}_n(k)$ are given in (1.4) and (1.11), respectively.

We shall next assume that we are in Hall's class of models (Hall and Welsh, 1985), where

$$\text{(1.13)} \qquad U(t) = C\, t^{\gamma}\left(1 + \frac{A(t)}{\rho}\,\big(1 + o(1)\big)\right), \qquad A(t) = \gamma\,\beta\, t^{\rho}, \quad \text{as } t \to \infty ,$$

or equivalently that the tail function is of the type

$$1-F(x) \;=\; \left(\frac{x}{C}\right)^{-1/\gamma} \left\{1+\frac{\beta}{\rho}\left(\frac{x}{C}\right)^{\rho/\gamma}+o\big(x^{\rho/\gamma}\big)\right\}, \qquad \text{as} \quad x\to\infty \;,$$

where $\gamma>0$, $C>0$, $\rho<0$ and $\beta\neq 0$.

We may then choose $a=a_n=C\,n^\gamma$, as $n\to\infty$, and, from (1.12), we get a least-squares' estimator of C given by

$$\widetilde{C}_n(k) \;:=\; X_{n-k:n}\,\exp\left\{\widehat{\gamma}_n^H(k)-\widetilde{\gamma}_n(k)\Big(\ln n-\psi(k{+}1)+1\Big)\right\} \tag{1.14}$$

$$\sim\; X_{n-k:n}\left(\frac{k}{n}\right)^{\widetilde{\gamma}_n(k)}\exp\left\{\widehat{\gamma}_n^H(k)-\widetilde{\gamma}_n(k)\right\}, \qquad \text{as} \quad k\to\infty \;, \tag{1.15}$$

again with $\widehat{\gamma}_n^H(k)$ and $\widetilde{\gamma}_n(k)$ given in (1.4) and (1.11), respectively.

Although aware that C is a parameter of the model, which may be estimated for instance through any of the asymptotically equivalent estimators in (1.14) or (1.15), we shall consider $\tau\equiv C$ as a *tuning* parameter. This has been done in a way similar to the one used by Csörgő and Viharos (1998), when they consider a kernel estimator as a function of a tuning parameter $\tau\equiv\rho$, also a model parameter, the second order parameter in (1.2). Notice that if $U_X(t)=C\,t^\gamma(1+o(1))$, then for $Y=X/C$, $U_Y(t)=t^\gamma\,(1+o(1))$. This means that a proper scaling of our data enables us to choose $a=n^\gamma$, i.e., $\gamma=\ln a/\ln n$, a particular situation which will merely help us to build a class of statistics, dependent of the control parameter $\tau=C$, which should be regarded as a possible change in the scale of our data. Such a class is got from the least-squares type estimator of $\{\ln a\}$ in (1.7), and is given by

$$\begin{aligned}\widetilde{\widetilde{\gamma}}_n^{(\tau)}(k) \;&:=\; \frac{1}{k\ln n}\left\{\sum_{i=1}^{k}\ln\frac{X_{n-i+1:n}}{\tau}+\widetilde{\gamma}_n(k)\sum_{i=1}^{k}\psi(i)\right\}\\ &=\; \frac{1}{\ln n}\left\{\ln\frac{X_{n-k:n}}{\tau}+\big(\psi(k{+}1)-1\big)\widetilde{\gamma}_n(k)+\widehat{\gamma}_n^H(k)\right\}.\end{aligned} \tag{1.16}$$

As a particular member of the class in (1.16), we shall consider the estimator

$$\widetilde{\widetilde{\gamma}}_n(k) \;\equiv\; \widetilde{\widetilde{\gamma}}_n^{(1)}(k) \;=\; \frac{\ln X_{n-k:n}+\big(\psi(k{+}1)-1\big)\widetilde{\gamma}_n(k)+\widehat{\gamma}_n^H(k)}{\ln n}\;. \tag{1.17}$$

We shall also consider the estimation of C, and its use in the class of statistics in (1.16), but we are aware that then we are going to get a poorer estimator of the tail index γ, unless the C-estimator is highly efficient. For instance, should we have used $\widetilde{C}_n(k)$, in (1.14), as τ, in (1.16), would we have been led to $\widetilde{\gamma}_n$ in (1.11), i.e., $\widetilde{\widetilde{\gamma}}_n^{(\widetilde{C}_n(k))}(k)\equiv\widetilde{\gamma}_n(k)$. We have here decided to follow Hall and Welsh (1985), and to consider the C-estimator

$$\widehat{C}_n(k) := \left(\frac{k}{n}\right)^{\widehat{\gamma}_n^H(k)} X_{n-k:n}\;. \tag{1.18}$$

Since in Hall's class of models, in (1.13), the mean squared error of both non-degenerate limiting distributions of $\widehat{\gamma}_n^H(k)$ and $\widehat{C}_n(k)$ are minimized by taking

$$k_0 = \left(\frac{(1-\rho)^2}{-2\,\rho\,\beta^2}\; n^{-2\rho}\right)^{1/(1-2\rho)}$$

(Theorem 4.1 in Hall and Welsh, 1985), we shall also consider, in the simulations, and whenever we are in Hall's class of models in (1.13), the estimator of the tail index γ, given by

$$\text{(1.19)} \qquad \widetilde{\widetilde{\gamma}}_n^{(\widehat{C})}(k), \qquad \text{where} \quad \widehat{C} = \widehat{C}_n(\widehat{k}_0), \;\; \widehat{k}_0 = \left(\frac{(1-\widehat{\rho})^2}{-2\,\widehat{\rho}\,\widehat{\beta}^2}\; n^{-2\widehat{\rho}}\right)^{1/(1-2\widehat{\rho})},$$

with $\widehat{C}_n$ given in (1.18) and $\widehat{\rho}$ and $\widehat{\beta}$ adequate estimators of ρ and β, respectively, already considered in Gomes and Martins (2002). In the simulations of models outside Hall's class, due to the difficulties in the estimation of k_0, we shall exhibit the behaviour of

$$\text{(1.20)} \qquad \widetilde{\widetilde{\gamma}}_n^{(\widehat{C}_0)}, \qquad \text{where} \quad \widehat{C}_0 = \widehat{C}_n(k_0), \;\; k_0 = \arg\min_k MSE\Big[\widehat{\gamma}_n^H(k)\Big],$$

again with $\widehat{C}_n$ given in (1.18) and k_0 obtained through simulation.

Remark 1.2. In practice, it is sensible to consider τ in (1.16) as a tuning parameter, choosing τ through a data-driven estimation of the mean squared error of $\widetilde{\widetilde{\gamma}}_n^{(\tau)}(k)$ as a function of k, for adequately chosen fixed values of τ (Oliveira, 2002). The value of τ may be any value τ^* such that

$$\widehat{MSE}\Big[\widetilde{\widetilde{\gamma}}_n^{(\tau^*)}(k)\Big] \le \widehat{MSE}\Big[\widehat{\gamma}_n^H(k)\Big], \quad \text{for every } k\,.$$

When we consider

$$k_{n0}^{(\tau^*)} := \arg\min_k \widehat{MSE}\Big[\widetilde{\widetilde{\gamma}}_n^{(\tau^*)}(k)\Big],$$

it is then sensible to choose the value τ_0^* providing the minimum $\widehat{MSE}\Big[\widetilde{\widetilde{\gamma}}_n^{(\tau^*)}\big(k_{n0}^{(\tau^*)}\big)\Big]$, i.e.,

$$\tau_0^* := \arg\min_{\tau^*} \widehat{MSE}\Big[\widetilde{\widetilde{\gamma}}_n^{(\tau^*)}\Big(k_{n0}^{(\tau^*)}\Big)\Big].$$

We choose then (see also Remark 5.1)

$$\widehat{k}_{n0} = k_{n0}^{\tau_0^*} \qquad \text{and} \qquad \widetilde{\widetilde{\gamma}}_{n0} := \widetilde{\widetilde{\gamma}}_n^{(\tau_0^*)}\big(\widehat{k}_{n0}\big)\,.$$

This is an open problem, beyond the scope of the present paper, where we intend essentially to present the potentialities of the class of statistics in (1.16) to estimate a positive tail index.

In section 2, we shall briefly review the Peaks Over Threshold (POT) methodology, a classical method of estimation of a tail index, to be also compared with the new estimation procedures considered, as well as the estimation of the second order parameters ρ and β in $A(t) = \gamma\,\beta\,t^\rho$. In section 3 we shall compare asymptotically the estimator in (1.11) (or equivalently, the estimator in (1.9)) with the Hill estimator in (1.4). Section 4 is devoted to the asymptotic behaviour of the class of estimators in (1.16). Sections 5 and 6 are devoted to the illustration of the behaviour of these estimators for finite samples, through the use of Monte Carlo simulation techniques.

2. REVIEW OF WELL-ESTABLISHED ESTIMATION PROCEDURES OF FIRST AND SECOND ORDER PARAMETERS

2.1. The link between the Hill estimator and the POT methodology

Let us think on the excesses over a high random threshold $X_{n-k:n}$,

$$V_{ik} := X_{n-i+1:n} - X_{n-k:n}\ , \qquad 1 \le i \le k\ .$$

Since $X \stackrel{d}{=} U(Y)$, Y a standard unit Pareto r.v. with d.f. $1-1/y$, $y \ge 1$, $Y_{n-i+1:n}/Y_{n-k:n} \stackrel{d}{=} Y_{k-i+1:k}$, $1 \le i \le k$, and, for k intermediate, $Y_{n-k:n} = (n/k)(1+o_p(1))$, we may write, under the validity of the first order condition in (1.1),

$$\begin{aligned}
V_{ik} &= X_{n-i+1:n} - X_{n-k:n} \;=\; X_{n-k:n}\big(X_{n-i+1:n}/X_{n-k:n} - 1\big)\\
&\stackrel{d}{=} X_{n-k:n}\Big(U\big(Y_{n-k:n}\,Y_{k-i+1:k}\big)/U\big(Y_{n-k:n}\big) - 1\Big)\\
&\stackrel{d}{=} X_{n-k:n}\Big(Y^\gamma_{k-i+1:k}\big(1+o_p(1)\big) - 1\Big)\\
&= X_{n-k:n}\Big(\big(Y^\gamma_{k-i+1:k} - 1\big)\,\big(1+o_p(1)\big) + o_p(1)\Big)\ .
\end{aligned}$$

Consequently, we may say that there exists δ such that we have approximately $V_{ik}/\delta \approx \big(Y^\gamma_{k-i+1:k}-1\big)/\gamma$, i.e., V_{ik}, $1 \le i \le k$, are approximately the k o.s. of a sample of size k from a Generalized Pareto (GP) model,

$$GP_\gamma(x;\delta) \;=\; 1 - \big(1+\gamma\,x/\delta\big)^{-1/\gamma}\ , \qquad x \ge 0\ \ (\gamma,\delta > 0)\ .$$

The estimation of γ through maximum likelihood (ML) in a GP model has been thoroughly studied in Davison (1984) and Smith (1984a,b). Davison (1984) suggested a re-parameterization of the GP model in $(\gamma,\alpha) = (\gamma,\gamma/\delta)$, which enables us to get only one ML equation to be solved iteratively. Such a re-parameterization has also been used in Gomes and Oliveira (2003a), where a computational study of this methodology has been undertaken. The ML-estimator of γ has, with such a re-parameterization, an explicit expression as a function of the *ML*-estimator $\widehat{\alpha}$ of $\alpha = \gamma/\delta$ and the sample of the excesses. We have

$$\widehat{\gamma}_n^{GP}(k) \;:=\; \frac{1}{k}\sum_{i=1}^{k} \ln\bigl(1 + \widehat{\alpha}\, V_{ik}\bigr) \;, \tag{2.1}$$

and α is such that $\alpha\, V_{ik} \approx Y_{k-i+1:k}^{\gamma} - 1$. Notice that an obvious choice for α is $1/X_{n-k:n}$. Then $1 + \alpha\, V_{ik} = X_{n-i+1:n}/X_{n-k:n}$, and the estimator in (2.1) is the Hill estimator $\widehat{\gamma}_n^H(k)$ in (1.4). Smith (1987) has got the asymptotic behaviour of the estimator in (2.1) for a fixed threshold u. The conclusion of his Theorem 3.2 may be easily rephrased in this set-up (Gomes, 2002; Drees *et al.*, 2004), and, under the second order framework in (1.2), we get the asymptotic distributional representation

$$\widehat{\gamma}_n^{GP}(k) \;\stackrel{d}{=}\; \gamma + \frac{(1+\gamma)}{\sqrt{k}}\, Q_k + \frac{(1+\gamma)\,(\gamma+\rho)\, A(n/k)}{\gamma\,(1-\rho)\,(1-\rho+\gamma)}\, \bigl(1 + o_p(1)\bigr) \;, \tag{2.2}$$

with Q_k asymptotically standard normal.

Remark 2.1. Note that the result in (2.2), although appearing to produce a different bias term, agrees with the one in Drees *et al.* (2004). Indeed, whereas we here assume (1.2), the most common second order condition for heavy-tailed models, Drees *et al.* (2004) consider the general case $\gamma \in \mathbb{R}$, and assume that there exists $a^*(\cdot)$ and $A^*(\cdot)$ such that

$$\frac{\frac{U(tx)-U(t)}{a^*(t)} - \frac{x^\gamma - 1}{\gamma}}{A^*(t)} \;\underset{t\to\infty}{\longrightarrow}\; \frac{1}{\rho^*}\left(\frac{x^{\gamma+\rho^*}-1}{\gamma+\rho^*} - \frac{x^\gamma-1}{\gamma}\right).$$

If we consider $\rho^* < 0$, we may then guarantee that, with $A_0(t) = A^*(t)/\rho^*$ and $a_0(t) = a^*(t)\,(1 - A^*(t)/\rho^*)$, we get,

$$\frac{\frac{U(tx)-U(t)}{a_0(t)} - \frac{x^\gamma - 1}{\gamma}}{A_0(t)} \;\underset{t\to\infty}{\longrightarrow}\; \frac{x^{\gamma+\rho^*}-1}{\gamma+\rho^*} \;. \tag{2.3}$$

For $\gamma > 0$ (and $\rho^* < 0$), condition (2.3) is equivalent to saying that, as $t \to \infty$,

$$U(t) \;=\; C\, t^\gamma \Bigl(1 + A\, t^{\rho^*} + o\bigl(t^{\rho^*}\bigr)\Bigr) \;. \tag{2.4}$$

Then

$$U(tx) - U(t) \;=\; C\,\gamma\, t^\gamma \left(\frac{x^\gamma-1}{\gamma} + \frac{A(\gamma+\rho^*)\, t^{\rho^*}}{\gamma}\left(\frac{x^{\gamma+\rho^*}-1}{\gamma+\rho^*}\right) + o\bigl(t^{\rho^*}\bigr)\right).$$

If $\gamma+\rho^* \neq 0$, we then need to choose $a_0(t) = C\,\gamma\, t^\gamma$, $A_0(t) = A(\gamma+\rho^*)\, t^{\rho^*}/\gamma$. Then

$$\begin{aligned}
\frac{U(tx)}{U(t)} \;&=\; 1 + \frac{U(tx)-U(t)}{C\, t^\gamma}\, \Bigl(1 - A\, t^{\rho^*} + o\bigl(t^{\rho^*}\bigr)\Bigr) \\
&=\; x^\gamma \Bigl(1 + A\, t^{\rho^*}\bigl(x^{\rho^*}-1\bigr) + o\bigl(t^{\rho^*}\bigr)\Bigr) \\
&=\; x^\gamma \left(1 + A\,\rho^*\, t^{\rho^*}\left(\frac{x^{\rho^*}-1}{\rho^*}\right) + o\bigl(t^{\rho^*}\bigr)\right) \;,
\end{aligned}$$

and consequently,

$$\ln U(tx) - \ln U(t) \;=\; \gamma\,\ln x + A\,\rho^*\,t^{\rho^*}\Bigl(\frac{x^{\rho^*}-1}{\rho^*}\Bigr) + o\bigl(t^{\rho^*}\bigr) \;,$$

i.e., provided that $\gamma+\rho^*\neq 0$, and with $A(t)=\frac{\gamma\,\rho^*A_0(t)}{\gamma+\rho^*}$,

$$\frac{\ln U(tx) - \ln U(t) - \gamma\,\ln x}{A(t)} \underset{t\to\infty}{\longrightarrow} \frac{x^{\rho^*}-1}{\rho^*} = \frac{x^{\rho}-1}{\rho} \;, \tag{2.5}$$

i.e., ρ^* in (2.3) is equal to ρ in (1.2). Consequently, if $\sqrt{k}\,A(n/k)\to\lambda$, $\sqrt{k}\,A_0(n/k)\to\lambda(\gamma{+}\rho)/(\gamma\,\rho)$. The bias provided in Drees *et al.* (2004) for the *POT-ML* tail index estimator is then

$$\frac{\lambda(\gamma+\rho)}{\gamma\,\rho}\left(\frac{\rho(1+\gamma)}{(1-\rho)\,(1-\rho+\gamma)}\right) \;=\; \frac{\lambda(1+\gamma)\,(\gamma+\rho)}{\gamma(1-\rho)\,(1-\rho+\gamma)} \;,$$

the values provided in both Smith (1987) and Gomes (2002).

We shall now make explicit the term $o(t^{\rho^*})$ in (2.4), assuming that

$$U(t) \;=\; C\,t^{\gamma}\Bigl(1 + A\,t^{\rho*} + B\,t^{\rho*+\rho'} + o\bigl(t^{\rho*+\rho'}\bigr)\Bigr) \;, \qquad \rho'<0 \;.$$

If $\gamma+\rho^*=0$, i.e., $\rho^*=-\gamma$,

$$U(tx) - U(t) \;\;=\;\; C\,\gamma\,t^{\gamma}\left(\frac{x^{\gamma}-1}{\gamma} + \frac{2B\,\rho^*\,t^{\rho^*+\rho'}}{\gamma}\Bigl(\frac{x^{\rho^*+\rho'}-1}{\rho^*+\rho'}\Bigr) + o\bigl(t^{\rho^*+\rho'}\bigr)\right) \;,$$

and

$$a_0(t) = C\,\gamma\,t^{\gamma} \;, \qquad A_0(t) = \frac{2B\,\rho^*\,t^{\rho^*+\rho'}}{\gamma} \;.$$

But for the model in (2.4), we may choose for any $\rho<0$, $A(t)=\rho\,A\,t^{\rho}$, and we get

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma\,\ln x}{A(t)} \;=\; \frac{x^{\rho}-1}{\rho} \;.$$

If $\sqrt{k}\,A(n/k)\to\lambda$, $\sqrt{k}\,A_0(n/k)\to 0$. So, both from Smith (1987) and from Drees *et al.* (2004), we get a null dominant component for the bias term of the *POT-ML* tail index estimator, whenever $\gamma+\rho^*=0$, as expected.

2.2. Estimators of the second order parameters ρ and β

The estimation of the second order parameter ρ, in $A(t) = \gamma\,\beta\,t^\rho$, is going to be done through particular members of the class of ρ-estimators in Fraga Alves *et al.* (2003). Those estimators are given by

$$\widehat{\rho}_n^{(i)}(k) := \min\left(0\,,\ \frac{3\big(T_n^{(i)}(k)-1\big)}{T_n^{(i)}(k)-3}\right)\,, \qquad i=0,1\,, \tag{2.6}$$

where

$$T_n^{(i)}(k) := \begin{cases} \dfrac{M_n^{(1)}(k) - \big(M_n^{(2)}(k)/2\big)^{1/2}}{\big(M_n^{(2)}(k)/2\big)^{1/2} - \big(M_n^{(3)}(k)/6\big)^{1/3}} & \text{if} \quad i=1 \\[2ex] \dfrac{\ln\big(M_n^{(1)}(k)\big) - \frac{1}{2}\ln\big(M_n^{(2)}(k)/2\big)}{\frac{1}{2}\ln\big(M_n^{(2)}(k)/2\big) - \frac{1}{3}\ln\big(M_n^{(3)}(k)/6\big)} & \text{if} \quad i=0 \end{cases}.$$

The statistics is (2.6) are consistent for the estimation of ρ whenever the second order condition (1.2) holds and k is such that $k\to\infty$, $k=o(n)$ and $\sqrt{k}\,A(n/k)\to\infty$, as $n\to\infty$.

Remark 2.2. The theoretical and simulated results in Fraga Alves *et al.* (2003), together with the use of these estimators in the Generalized Jackknife statistics of Gomes *et al.* (2000), as done in Gomes and Martins (2002), has led these authors to advise the consideration of the level

$$k_1 = \min\Big(n-1,\ \big[2n/\ln\ln n\big]\Big) \tag{2.7}$$

and of the ρ-estimators

$$\widehat{\rho}_0 := \min\Big(0,\ 3\big(T_n^{(0)}(k_1)-1\big)\big/\big(T_n^{(0)}(k_1)-3\big)\Big) \qquad \text{if} \quad \rho\ge -1\,, \tag{2.8}$$

and

$$\widehat{\rho}_1 := \min\Big(0,\ 3\big(T_n^{(1)}(k_1)-1\big)\big/\big(T_n^{(1)}(k_1)-3\big)\Big) \qquad \text{if} \quad \rho< -1\,. \tag{2.9}$$

For the estimation of β we have here considered the estimator of β in Gomes and Martins (2002) and based on the scaled log-spacings $U_i = i\,\{\ln X_{n-i+1:n} - \ln X_{n-i:n}\}$, $1\le i\le k$. Let us denote $\widehat{\rho}$ any of the estimators either in (2.8) or in (2.9) (or even in (2.6)). The β-estimator is given by

$$\widehat{\beta}(k) := \frac{1}{n^{\widehat{\rho}}}\ \frac{\left(\sum\limits_{i=1}^k i^{-\widehat{\rho}}\right)\left(\sum\limits_{i=1}^k U_i\right) - k\left(\sum\limits_{i=1}^k i^{-\widehat{\rho}}\,U_i\right)}{\left(\sum\limits_{i=1}^k i^{-\widehat{\rho}}\right)\left(\sum\limits_{i=1}^k i^{-\widehat{\rho}}\,U_i\right) - k\left(\sum\limits_{i=1}^k i^{-2\widehat{\rho}}\,U_i\right)}\,. \tag{2.10}$$

We have then considered $\widehat{\beta} = \widehat{\beta}(k_1)$, k_1 given in (2.7).

3. ASYMPTOTIC PROPERTIES OF $\widetilde{\gamma}_n$

3.1. The estimator $\widetilde{\gamma}_n$ as a linear combination of Hill's estimators

We first state the following:

Lemma 3.1. *A semi-parametric estimator of the tail index γ which is a linear combination of the k top log-observations, i.e.,*

$$\gamma_n(k) \;=\; \sum_{i=1}^{k} a_i \,\ln X_{n-i+1:n} \tag{3.1}$$

is scale invariant if and only if $\sum\limits_{i=1}^{k} a_i = 0$.

Proof: If we consider a change in scale, moving from X to X/C, $C>0$, $C \neq 1$, the estimator in (3.1) changes to $\sum_{i=1}^{k} a_i \,\ln X_{n-i+1:n} - \ln C \,\sum_{i=1}^{k} a_i$, which equals $\gamma_n(k) = \sum_{i=1}^{k} a_i \,\ln X_{n-i+1:n}$ if and only if $\sum_{i=1}^{k} a_i = 0$. $\square$

Lemma 3.2. *A semi-parametric estimator of the type (3.1) may be expressed as a linear combination of Hill's estimators, i.e.,*

$$\gamma_n(k) \;=\; \sum_{i=1}^{k} a_i \,\ln X_{n-i+1:n} \;=\; \sum_{j=1}^{k-1} b_j \;\widehat{\gamma}_n^H(j) \;, \tag{3.2}$$

where

$$b_j \;=\; -a_{j+1} - \frac{1}{j+1} \sum_{i=j+2}^{k} a_i, \quad j=1,...,k-2 \;, \qquad b_{k-1} = -a_k \;, \tag{3.3}$$

if and only if it is scale invariant, i.e., if and only if $\sum_{i=1}^{k} a_i = 0$.

Proof: We may write

$$\begin{aligned}
\sum_{j=1}^{k-1} b_j \;\widehat{\gamma}_n^H(j) \;&=\; \sum_{j=1}^{k-1} b_j \left\{\frac{1}{j} \sum_{i=1}^{j} \ln X_{n-i+1:n} - \ln X_{n-j:n}\right\} \\
&=\; \sum_{i=1}^{k-1} \left(\sum_{j=i}^{k-1} \frac{b_j}{j}\right) \ln X_{n-i+1:n} \;-\; \sum_{i=1}^{k} b_{i-1} \,\ln X_{n-i+1:n} \quad (b_0 \equiv 0) \\
&=\; \sum_{i=1}^{k-1} \left(\sum_{j=i}^{k-1} \frac{b_j}{j} - b_{i-1}\right) \ln X_{n-i+1:n} -\; b_{k-1} \,\ln X_{n-k+1:n} \;,
\end{aligned}$$

i.e., $a_i = \sum_{j=i}^{k-1} b_j/j - b_{i-1}$, $1 \le i \le k-1$, and $a_k = -b_{k-1}$. This linear system has a unique and possible solution if and only if $\sum_{i=1}^{k} a_i = 0$. Then we just need to solve the linear system of equations:

$$\begin{bmatrix} a_2 \\ a_3 \\ \cdots \\ a_k \end{bmatrix} = \begin{bmatrix} -1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{k-1} \\ 0 & -1 & \frac{1}{3} & \cdots & \frac{1}{k-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_{k-1} \end{bmatrix} =: \mathbf{A}\ \mathbf{b}\ .$$

Since the inverse matrix of $\mathbf{A}$ is

$$\mathbf{A}^{-1} = \begin{bmatrix} -1 & -\frac{1}{2} & -\frac{1}{2} & \cdots & -\frac{1}{2} \\ 0 & -1 & -\frac{1}{3} & \cdots & -\frac{1}{3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix},$$

the result follows. □

Then, from the relation (1.10) and from Lemma 3.2, it follows straightforwardly that:

Proposition 3.1. *The estimator in* (1.11), *which may be written as*

$$\widetilde{\gamma}_n(k) = \sum_{i=1}^{k} a_i \ln X_{n-i+1:n}\ , \qquad a_i = \frac{\psi(k+1) - \psi(i) - 1}{k - \psi(k+1) + \psi(1)},\quad 1 \le i \le k\ , \tag{3.4}$$

is scale invariant, i.e. $\sum\limits_{i=1}^{k} a_i = 0$, *and we may write it as the following linear combination of Hill's estimator,*

$$\widetilde{\gamma}_n(k) = \sum_{j=1}^{k-1} b_j\ \widehat{\gamma}_n^H(j)\ , \qquad b_j = \frac{j}{(j+1)\Big(k - \psi(k+1) + \psi(1)\Big)}\ . \tag{3.5}$$

3.2. The asymptotic behaviour of $\widetilde{\gamma}_n(k)$

Theorem 3.1. *Under the first order framework* (1.1) *and for* k *such that* (1.3) *holds, the estimator in* (1.11) *is a consistent estimator of* γ. *Moreover, under the second order framework in* (1.2), *we have the validity of the following distributional representation,*

$$\widetilde{\gamma}_n(k) \stackrel{d}{=} \gamma + \frac{\gamma\sqrt{2}}{\sqrt{k}}\, P_k + \frac{1}{(1-\rho)^2}\, A(n/k)\big(1 + o_p(1)\big)\ , \tag{3.6}$$

where P_k *is asymptotically standard normal.*

Proof: Since in the linear combination in (3.5), $\sum_{j=1}^{k-1} b_j = 1$, $\widetilde{\gamma}_n(k)$ is, under the conditions of the theorem, a consistent estimator of γ. The linear combination of Hill's estimators, $\sum_{j=1}^{k-1} b_j\, \widehat{\gamma}_n^H(j)$ may be written as

$$\begin{aligned}\sum_{j=1}^{k-1} b_j\, \widehat{\gamma}_n^H(j) \;&=\; \sum_{j=1}^{k-1} \frac{b_j}{j} \sum_{i=1}^{j} i\Big[\ln X_{n-i+1:n} - \ln X_{n-i:n}\Big] \\ &=\; \sum_{i=1}^{k-1} i\left(\sum_{j=i}^{k-1} \frac{b_j}{j}\right)\Big[\ln X_{n-i+1:n} - \ln X_{n-i:n}\Big]\;,\end{aligned}$$

and consequently, with $\{E_i\}_{i\ge 1}$ i.i.d. standard exponential r.v.'s, we may write

$$\begin{aligned}\sum_{j=1}^{k-1} b_j\, \widehat{\gamma}_n^H(j) \;&\stackrel{d}{=}\; \gamma \sum_{i=1}^{k-1}\left(\sum_{j=i}^{k-1} \frac{b_j}{j}\right) E_i \\ &\quad + A(n/k)\, k^{\rho} \sum_{i=1}^{k-1} i^{1-\rho}\left(\sum_{j=i}^{k-1} \frac{b_j}{j}\right) \frac{e^{\rho E_i/i} - 1}{\rho}\big(1 + o_p(1)\big)\;.\end{aligned} \tag{3.7}$$

For the particular linear combination under study we have

$$\sum_{i=1}^{k-1}\left(\sum_{j=i}^{k-1} \frac{b_j}{j}\right) = \sum_{j=1}^{k-1} b_j = 1\;, \qquad \sum_{j=i}^{k-1} \frac{b_j}{j} = \frac{\psi(k+1) - \psi(i+1)}{k - \psi(k+1) + \psi(1)}\;,$$

and

$$\begin{aligned}\sum_{i=1}^{k-1}\left(\sum_{j=i}^{k-1} \frac{b_j}{j}\right)^2 \;&=\; \frac{2k - \psi^2(k+1) + \big(2\psi(1) - 1\big)\,\psi(k+1) + \psi(1) - \psi^2(1)}{\Big(k - \psi(k+1) + \psi(1)\Big)^2} \\ &=\; \frac{2}{k}\,\big(1 + o(1)\big)\;.\end{aligned}$$

Since $\mathbb{E}\big\{\big(e^{\rho E_i/i} - 1\big)/\rho\big\} = 1/(i-\rho)$, and $\sum_{i=1}^{k-1} i^{-\rho}\big\{\psi(k+1) - \psi(i+1)\big\} = O(k^{-\rho+1})/(1-\rho)^2$, we finally get (3.6). □

Remark 3.1. The result in Theorem 3.1 has already been obtained for the estimator in (1.9) by Csörgő and Viharos (1997), who have shown that for intermediate sequences k, and with $\mu_n(k) = -\frac{n}{k}\int_0^{k/n} \big(1 + \ln(ns/k)\big) \ln U(1/s)\, ds$,

$$\sqrt{k}\,\left\{\widetilde{\gamma}_n^{(\mathcal{K})}(k) - \mu_n(k)\right\} \;\xrightarrow[n\to\infty]{d}\; \mathit{Normal}\,(0, 2\gamma^2)\;.$$

But under the second order framework in (1.2), $\mu_n(k)$ may be written as

$$\mu_n(k) \;=\; \gamma + \frac{A(n/k)}{(1-\rho)^2}\,\big(1 + o(1)\big)\;,$$

which agrees with the result in (3.6).

Remark 3.2. Notice that, relatively to the Hill estimator, the asymptotic variance of $\widetilde{\gamma}_n(k)$ duplicates, but the bias decreases by a factor $1/(1-\rho)$.

3.3. Asymptotic comparison at optimal levels

Now we proceed to an asymptotic comparison of the estimators $\widetilde{\gamma}_n$, $\widehat{\gamma}_n^H$ and $\widehat{\gamma}_n^{GP}$ at their optimal levels in the lines of de Haan and Peng (1998), Gomes *et al.* (2000, 2002) for sets of Generalized Jackknife statistics, Gomes and Martins (2001) and also Caeiro and Gomes (2002), for specifically built "asymptotically unbiased" estimators of the tail index. Suppose $\gamma_n(k)$ is a general semi-parametric estimator of the tail index, for which the distributional representation

$$\gamma_n(k) \;=\; \gamma + \frac{\sigma}{\sqrt{k}}\, Z_k + b\, A(n/k) + o_p\big(A(n/k)\big) \tag{3.8}$$

holds for any intermediate k, and where Z_k is an asymptotically standard normal r.v.; then we have

$$\sqrt{k}\big[\gamma_n(k) - \gamma\big] \xrightarrow{d} N\big(\lambda b,\, \sigma^2\big), \qquad \text{as} \;\; n \to \infty \;,$$

provided k is such that $\sqrt{k}\, A(n/k) \to \lambda$, finite, as $n \to \infty$. In this situation we write $Bias_\infty[\gamma_n(k)] := b\, A(n/k)$ and $Var_\infty[\gamma_n(k)] := \sigma^2/k$. The so-called Asymptotic Mean Squared Error ($AMSE$) is then given by

$$AMSE\big[\gamma_n(k)\big] \;:=\; \frac{\sigma^2}{k} + b^2 A^2(n/k) \;.$$

Using regular variation theory it may be proved that, whenever $b \neq 0$, there exists a function $\varphi(n)$, dependent only on the underlying model, and not on the estimator, such that

$$\lim_{n\to\infty} \varphi(n)\, AMSE[\gamma_{n0}] \;=\; \frac{2\rho - 1}{2\rho}\, \big(\sigma^2\big)^{-\frac{2\rho}{1-2\rho}}\, \big(b^2\big)^{\frac{1}{1-2\rho}} \;:=\; LMSE[\gamma_{n0}] \;,$$

where $\gamma_{n0} := \gamma_n(k_0(n))$, $k_0(n) := \arg\min_k AMSE[\gamma_n(k)]$.

It is then sensible to consider the following measure of efficiency, defined in a way that the larger such a measure is the better is the estimator.

Definition 3.1. Given two biased estimators $\gamma_n^{(1)}(k)$ and $\gamma_n^{(2)}(k)$, both computed at their optimal levels, and for which distributional representations of the type (3.8) hold, with constants (σ_1, b_1) and (σ_2, b_2), respectively, $b_1, b_2 \neq 0$, the Asymptotic Root Efficiency ($AREFF$) of $\gamma_{n0}^{(2)}$ relatively to $\gamma_{n0}^{(1)}$ is

$$\begin{aligned} AREFF_{2|1} \;\equiv\; AREFF_{\gamma_{n0}^{(2)}|\gamma_{n0}^{(1)}} \;&:=\; \sqrt{LMSE\big[\gamma_{n0}^{(1)}\big] \,/ LMSE\big[\gamma_{n0}^{(2)}\big]} \\ &=\; \left(\left(\frac{\sigma_1}{\sigma_2}\right) \left|\frac{b_1}{b_2}\right|\right)^{\frac{1}{1-2\rho}} \;. \end{aligned}$$

The comparison of the estimator $\widetilde{\gamma}_n$ with the Hill estimator $\widehat{\gamma}_n^H$, both computed at their optimal levels, leads us to the following result:

Proposition 3.2. *The asymptotic root efficiency of $\widetilde{\gamma}_n$ relatively to the Hill estimator $\widehat{\gamma}_n^H$, both computed at their optimal levels, is given by*

$$AREFF_{\widetilde{\gamma}_n|\widehat{\gamma}_n^H} \;=\; \big(2^\rho(1-\rho)\big)^{\frac{1}{1-2\rho}} \;, \tag{3.9}$$

being thus greater than 1 iff $\rho > -1$, and equal to 1 at $\rho = 0$ and $\rho = -1$.

The comparison of the three estimators $\widetilde{\gamma}_n$, $\widehat{\gamma}_n^H$ and $\widehat{\gamma}_n^{GP}$ is done graphically in Figure 1, where the "best" estimator, in terms of minimum *LMSE* at the optimal level, is exhibited. As expected, all depends on the region (γ,ρ), but for values of ρ close to 0, say $\rho > -1$, a region where Hill's estimator exhibits "disturbing" sample paths, the new estimator $\widetilde{\gamma}_n$, at its optimal level, not only overpasses the Hill estimator for all γ, as stated in Proposition 3.2, but also overpasses the *GP*-estimator, at their respective optimal levels, for a wide region of (γ,ρ)-values.

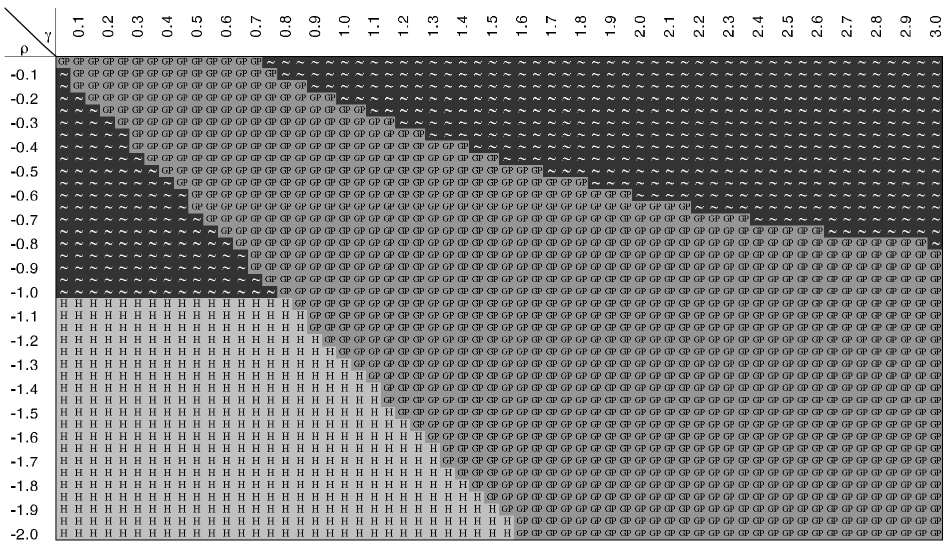

Figure 1: Mimum *LMSE* among the estimators $\widehat{\gamma}^H$, $\widetilde{\gamma}$ and $\widehat{\gamma}^{GP}$ in (1.4), (1.11) and (2.1), respectively.

4. THE ASYMPTOTIC BEHAVIOUR OF $\widetilde{\widetilde{\gamma}}_n^{(\tau)}(k)$

Notice first of all that we no longer have linear combinations of the top log-observations, unless $\tau = 1$, and then:

Proposition 4.1. *If we consider $\tau = 1$ in* (1.16), *the statistic $\widetilde{\widetilde{\gamma}}(k) \equiv \widetilde{\widetilde{\gamma}}_n^{(1)}(k)$, in* (1.17), *may be written as*

$$\widetilde{\widetilde{\gamma}}_n(k) \;=\; \sum_{i=1}^{k} a_i^* \,\ln X_{n-i+1:n} \;, \tag{4.1}$$

where

$$a_i^* \;=\; \frac{1}{k\,\ln n}\left\{1+\frac{k\big(\psi(k+1)-1\big)\,\big(\psi(k+1)-\psi(i)-1\big)}{k-\psi(k+1)+\psi(1)}\right\}. \tag{4.2}$$

The statistic $\widetilde{\widetilde{\gamma}}_n(k)$ is only asymptotically scale invariant, and consequently cannot be expressed as a linear combination of Hill's estimators.

Proof: To get the coefficients of the linear combination in (4.1) we just need to use again Lemma 3.2. Since $\sum_{i=1}^k a_i^* = \frac{1}{\ln n} \neq 0$, but converging towards 0, as $n\to\infty$, $\widetilde{\widetilde{\gamma}}_n(k)$ is not scale invariant, but it is asymptotically scale invariant. □

The asymptotic behaviour of $\widetilde{\widetilde{\gamma}}_n^{(\tau)}(k)$ in (1.16) is not directly related to that of the Hill estimator. Indeed the dominant term of $\widetilde{\widetilde{\gamma}}_n^{(\tau)}(k)$ is $\{\ln X_{n-k:n}\}$, and we shall base the proof of the asymptotic behaviour of this estimator on the following:

Lemma 4.1. *If $i \geq 1$ is fixed, and under the first order condition* (1.1),

$$\ln\frac{X_{n-i+1:n}}{U(n)} \;\underset{n\to\infty}{\overset{\mathrm{d}}{\longrightarrow}}\; \gamma\,W_i\;, \tag{4.3}$$

where W_i is a non-degenerate r.v. with a probability density function (p.d.f.) $g_i(w) = \Lambda(w)\,(-\ln\Lambda(w))^i/\Gamma(i)$, $\Lambda(w) = e^{-e^{-w}}$, $w \in \mathbb{R}$. For k intermediate, and under the validity of the second order condition (1.2), *the distributional representation*

$$\ln\frac{X_{n-k:n}}{U(n/k)} \;=\; \frac{\gamma}{\sqrt{k}}\,B_k + o_p\big(A(n/k)\big) \tag{4.4}$$

holds, with B_k an asymptotically standard normal r.v.

Proof: The result in (4.3) is well-known from the field of Extreme Value Theory (see, for instance, Galambos, 1987). Indeed, since $Y_{n-i+1:n}/n$ converges towards a non-degenerate r.v. $Z_i = \exp(W_i)$, and

$$\ln\frac{X_{n-i+1:n}}{U(n)} \;=\; \ln\frac{U\big(n(Y_{n-i+1:n}/n)\big)}{U(n)} \;=\; \gamma\,\ln Z_i + o_p(1)\;,$$

(4.3) follows.

For k intermediate (Ferreira *et al.*, 2003),

$$\frac{X_{n-k:n}}{U(n/k)} \;=\; \frac{U(Y_{n-k:n})}{U(n/k)} \;=\; 1+\frac{\gamma}{\sqrt{k}}\,B_k + o_p\big(A(n/k)\big)\;,$$

with B_k asymptotically standard normal r.v., and consequently (4.4) holds true. □

Remark 4.1. Notice that $W_i \stackrel{d}{=} -\ln Gama(i)$, where $Gama(i)$ denotes a gamma r.v., with p.d.f. $f(w) = w^{i-1}\exp(-w)/\Gamma(i)$, $w \geq 0$. Consequently $\mathbb{E}(W_i) = -\psi(i)$, and hence (1.6). The relation (1.5) is also a direct consequence of (4.4), together with the fact that $\psi(k) = \ln k + O(1/k)$, as $k \to \infty$.

We thus have, for every $\tau > 0$, consistency of $\widetilde{\widetilde{\gamma}}{}_n^{(\tau)}(k)$ for the estimation of the tail index γ, but we cannot guarantee asymptotic normality. We may however state the following:

Theorem 4.1. *In Hall's class of models, where* (1.13) *holds, and both for fixed and intermediate k, $\widetilde{\widetilde{\gamma}}{}_n^{(\tau)}(k)$ is consistent for the estimation of γ, for every $\tau > 0$. For intermediate k we have*

$$\ln n\left\{\widetilde{\widetilde{\gamma}}{}_n^{(\tau)}(k) - \gamma\right\} \xrightarrow[n\to\infty]{p} \ln\{C/\tau\}\ , \tag{4.5}$$

i.e., $\widetilde{\widetilde{\gamma}}{}_n^{(\tau)}(k)$ exhibits a degenerate behaviour. For models where $C=1$ (*or if we scale our data, dividing them by the appropriate scale $C\neq 1$, so that we have a unit scale*), *we get*

$$\frac{\sqrt{k}\,\ln n}{\ln k}\left(\widetilde{\widetilde{\gamma}}{}_n^{(1)}(k) - \gamma\right) \stackrel{d}{=} \gamma\sqrt{2}\,P_k + \frac{\sqrt{k}\,A(n/k)}{(1-\rho)^2}\big(1 + o_p(1)\big)\ , \tag{4.6}$$

i.e., $\widetilde{\widetilde{\gamma}}{}_n^{(1)}(k)$ is asymptotically normal, at a rate of convergence of the order of $\ln k/(\sqrt{k}\,\ln n)$, with an asymptotic variance equal to $2\gamma^2$ and an asymptotic bias equal to $\lambda/(1-\rho)^2$, whenever $\sqrt{k}\,A(n/k) \underset{n\to\infty}{\longrightarrow} \lambda$, finite.

Proof: The expression of $\widetilde{\widetilde{\gamma}}{}_n^{(\tau)}(k)$ in (1.16) enables us to get, for fixed k,

$$\widetilde{\widetilde{\gamma}}{}_n^{(\tau)}(k) \stackrel{d}{=} \gamma + \frac{\ln C - \ln\tau + \gamma\big(W_k + \widetilde{H}_k + H_k\big) + o_p(1)}{\ln n}\ ,$$

with W_k, $\widetilde{H}_k$ and H_k non-degenerate r.v.'s. Hence, consistency follows.
For intermediate k,

$$\widetilde{\widetilde{\gamma}}{}_n^{(\tau)}(k) \stackrel{d}{=} \gamma + \frac{\ln C - \ln\tau}{\ln n} + \gamma\sqrt{2}\,\frac{\ln k}{\sqrt{k}\,\ln n}\,P_k + \frac{\ln k\ A(n/k)}{(1-\rho)^2\,\ln n}\big(1 + o_p(1)\big)\ .$$

Consequently (4.5) follows and, for $C = 1$, (4.6) follows, as well as the remaining of the theorem. □

Remark 4.2. Note again that the value $C = 1$ may be achieved through a change in the scale of our data. Indeed, as said from the beginning, if for the original r.v. X we have a quantile function $U_X(t) = C\,t^\gamma(1 + o(1))$, for $Y = X/C$, $U_Y(t) = U_X(t)/C = t^\gamma(1 + o(1))$, and (4.6) holds.

Remark 4.3. Note also that the rate of convergence in (4.6) is of the order of $\ln k/(\sqrt{k}\,\ln n)$, which is a $o(1/\sqrt{k})$, for k intermediate and such that $\ln k = o(\ln n)$. The rate of convergence $1/\sqrt{k}$ is the usual rate of convergence for the most common tail index estimators. The rate of convergence here is also the usual one, whenever $k = O(n^\epsilon)$.

5. PATTERNS OF MEAN VALUES AND MEAN SQUARE ERRORS OF THE ESTIMATORS

From Figure 2 till Figure 7, and with the obvious notation H, GP, $\widetilde{H}$ and $\widetilde{\widetilde{H}}$ instead of $\widehat{\gamma}_n^H$, $\widehat{\gamma}_n^{GP}$, $\widetilde{\gamma}_n$ and $\widetilde{\widetilde{\gamma}}_n$, respectively, we present, in the top, the simulated mean values and MSE's of $\widetilde{\gamma}_n^{(\mathcal{K})}$, $\widetilde{\gamma}_n$ and $\widehat{\gamma}_n^{GP}$ in (1.9), (1.11) and (2.1), respectively. In the bottom part of each figure we picture the same characteristics of $\widetilde{\widetilde{\gamma}}_n^{(1)}|C=0.5$, $\widetilde{\widetilde{\gamma}}_n^{(C)} \equiv \widetilde{\widetilde{\gamma}}_n^{(1)}|C=1$ and $\widetilde{\widetilde{\gamma}}_n^{(1)}|C=2$, as well as of $\widetilde{\widetilde{\gamma}}_n^{(\widehat{C})}$, in (1.19), for models in Hall's class and $\widetilde{\widetilde{\gamma}}_n^{(\widehat{C}_0)}$, in (1.20), for models outside Hall's class. We place in all Figures the same characteristics of the Hill estimator $\widehat{\gamma}_n^H$, for an easier comparison. Simulations related to these Figures are based on 2000 runs, due to the computational time associated to the Peaks Over Threshold methodology. For some of the models, and due the erratic behaviour of the GP estimator for small values of k, we picture its mean value only for $k \geq 100$. The sample size is $n = 1000$ and we have considered the following set of models:

1. *the Fréchet model*, $F(x;C) = \exp\big(-(x/C)^{-1/\gamma}\big)$, $x \geq 0$, with $\gamma = 1$ and $C = 0.5$, 1 and 2, for which $\rho = -1$ (Figure 2);

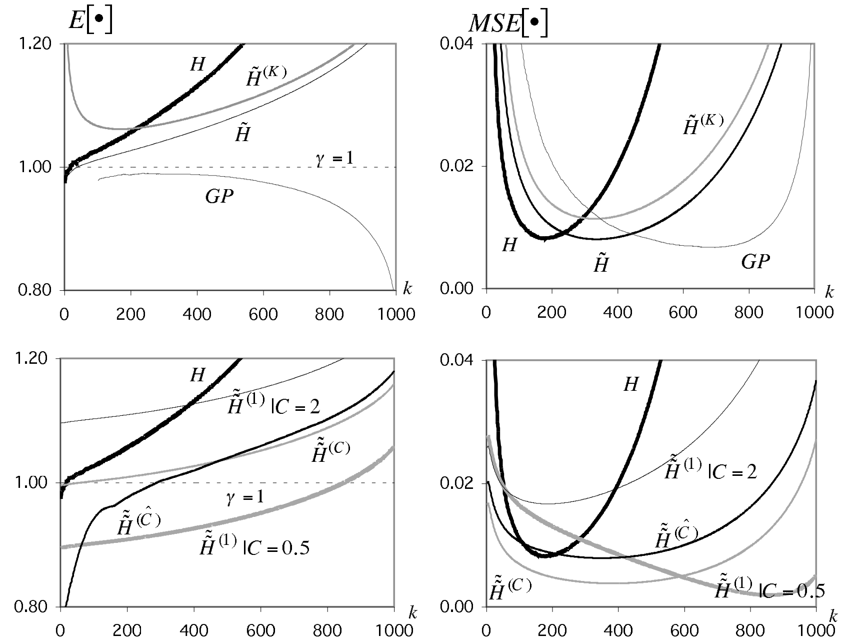

Figure 2: *Fréchet* parent with $\gamma = 1$.

When we look at Figure 2, we immediately notice that the expected changes have obviously occurred, despite the asymptotic scale invariance of $\widetilde{\widetilde{\gamma}}_n(k)$. Indeed, the changes in C induce a "shift" in the sample paths of our estimator. Note that for a scale C, we should get a dominant term of bias given by $\ln C/\ln n = -0.10, 0$ and $+0.10$ for $C = 0.5$, 1 and 2, respectively, which really agrees with the simulated mean values' patterns presented in Figure 2 (*bottom, left*).

Remark 5.1. Looking at the mean squared error patterns, presented also at the Figure 2 (*bottom, right*), we think that we may play with the *tuning* parameter τ in our benefit, in the lines of the work developed in Gomes and Oliveira (2003*b*), where, the use of a control parameter $\{a\}$, which is merely a shift, artificially imposed to the data, and the choice of the adequate value of $\{a\}$ improves greatly the performance of our original estimator. The criterion used there for the choice of $\{a\}$ is a stability criterion of sample paths. Here the methodology must be different, and further research is under development, but we have already the adequate methodologies to deal with this problem. As said before, in Remark 1.2, we think that the best way to proceed (Oliveira, 2002) is to estimate the mean squared errors of our estimators as functions of k, merely on the basis of the available sample, proceeding next to the adaptive choice of the k and τ-values providing the minimum mean squared error: we already have access to suitable procedures of estimation of *MSE*(k), either through the regression diagnostic methodology of Beirlant *et al.* (1996*a*, 1996*b*) or through the use of the bootstrap methodology in Draisma *et al.* (1999), Danielsson *et al.* (2001) and Gomes and Oliveira (2001). Such a computer intensive study is however beyond the scope of this paper.

2. *the Burr model*, $F(x;C) = 1 - \left(1+(x/C)^{-\rho/\gamma}\right)^{1/\rho}$, $x\geq 0$, $\gamma>0$, $\rho<0$, with $\gamma=1$, $C=0.5$, 1 and 2 and for $\rho=-0.5$, -1 and -2 (Figures 3, 4 and 5);

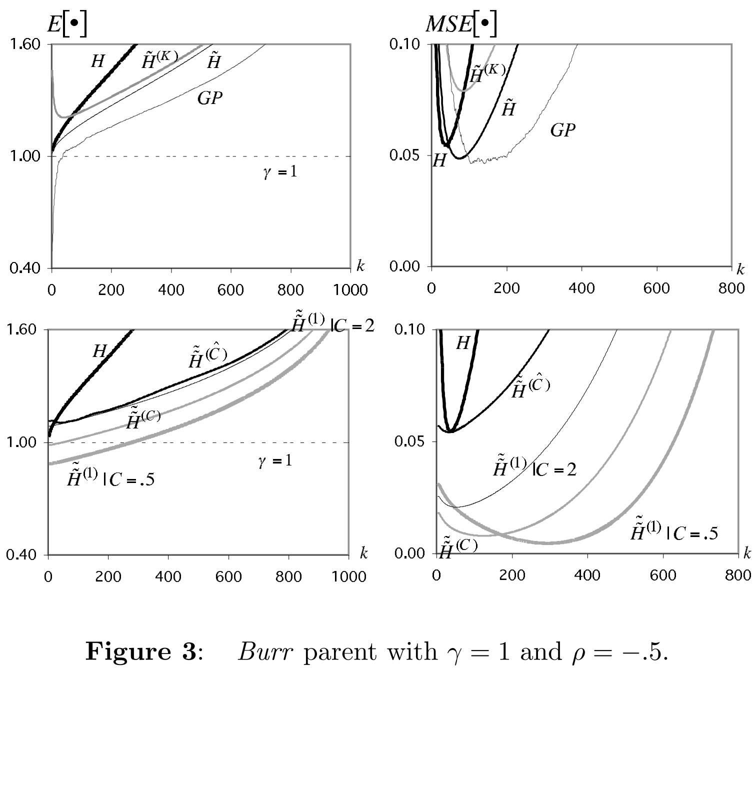

Figure 3: *Burr* parent with $\gamma = 1$ and $\rho = -.5$.

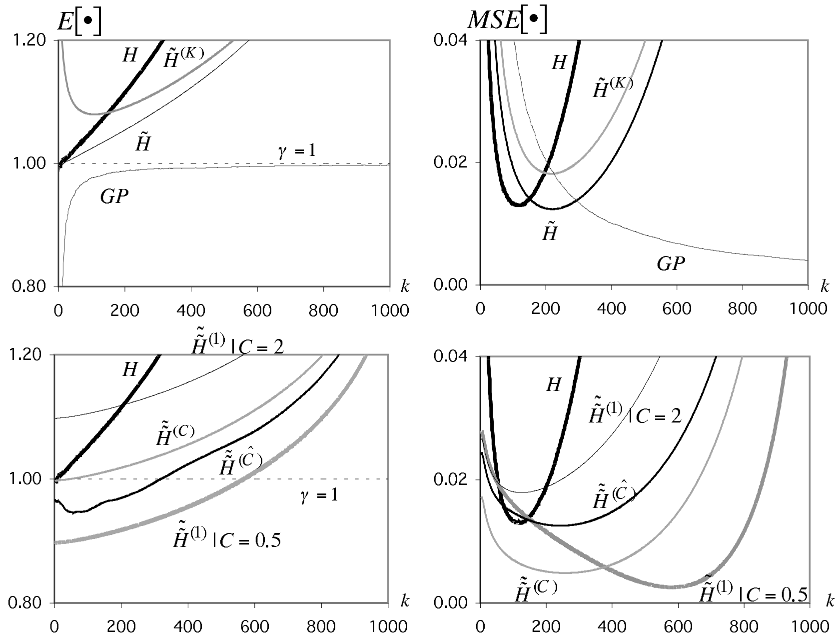

Figure 4: *Burr* parent with $\gamma = 1$ and $\rho = -1$.

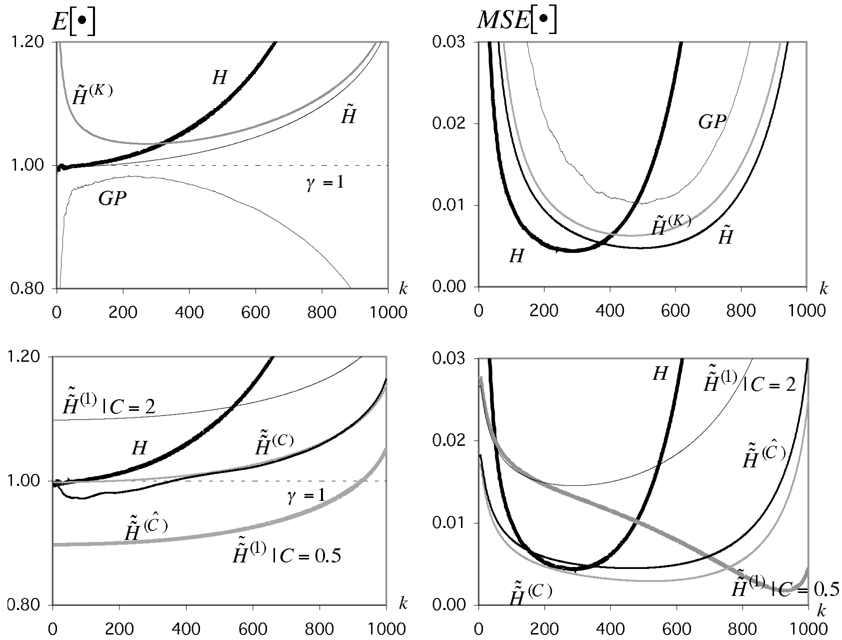

Figure 5: *Burr* parent with $\gamma = 1$ and $\rho = -2$.

a model outside Hall's class,

3. *the Out-Hall model*, with a quantile function $F^{\leftarrow}(1-t) = C\, t^{-1}\, e^{-2t(\ln t - 1)}$, for all $0 < t \leq 1$, $C = 0.5$, 1 and 2 (Figure 6);

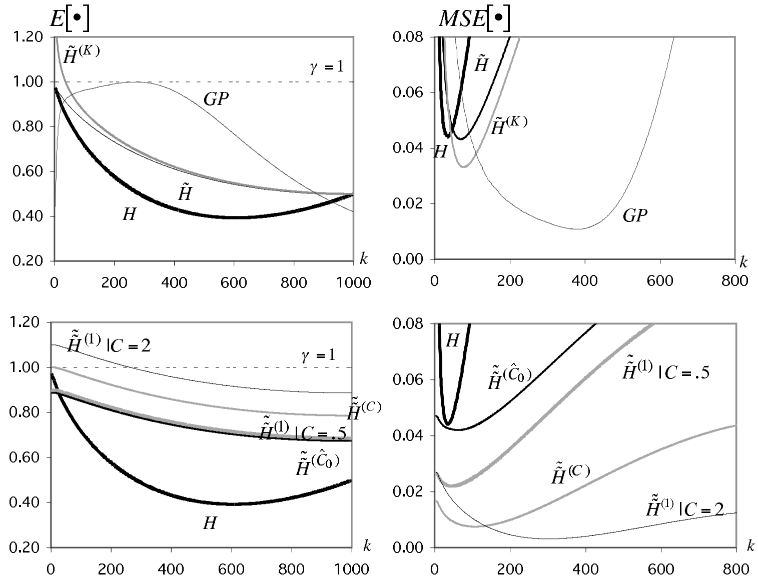

Figure 6: *Out-Hall* parent with $\gamma = 1$.

and the following model for which the second order condition in (1.2) does not hold:

4. *the sin-Burr model*, with a quantile function given by $F^{\leftarrow}(1-t) = C\big(t^\rho - \sin(t^\rho)\big)^{-\gamma/\rho}$, $0 < t \leq 1$, with $\gamma = 1$, $C = 0.5$, 1 and 2 and for $\rho = -0.5$ (Figure 7).

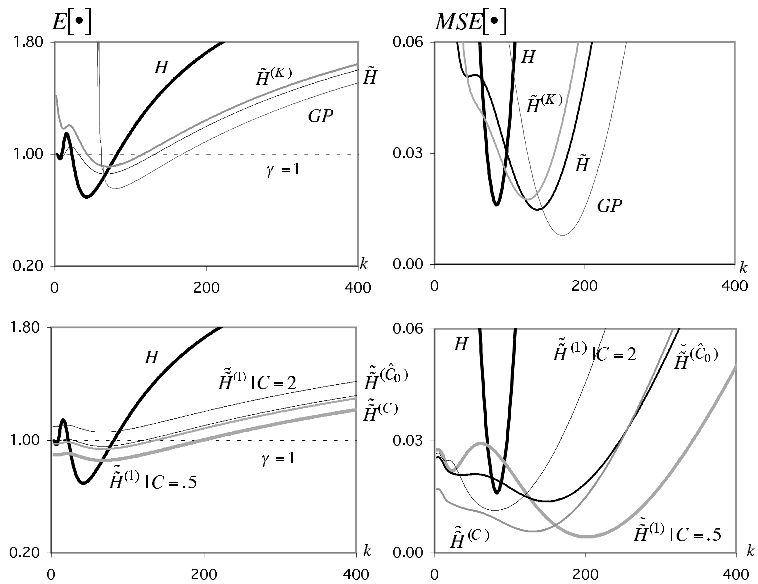

Figure 7: *Sin-Burr* parent with $\gamma = 1$ and $\rho = -.5$.

Figures 8 and 9 are equivalent to the previous figures, but for standard models with $C \neq 1$, in (1.13) — the Student models with $\nu = 4$, and 2 degrees of freedom. Notice that for the Student model with ν degrees of freedom,

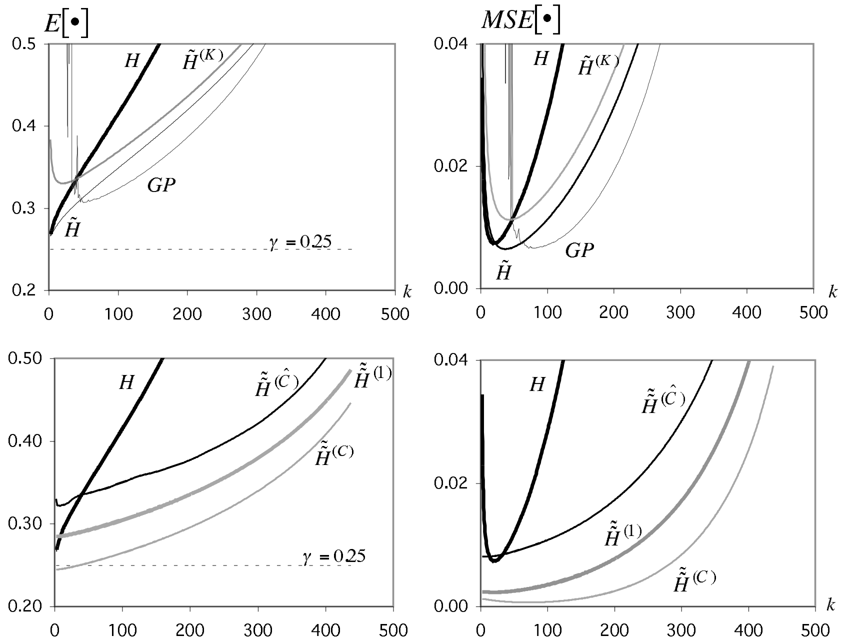

Figure 8: *Student*(4) parent with $\gamma = .25$ and $\rho = -.5$ ($C = 1.32$).

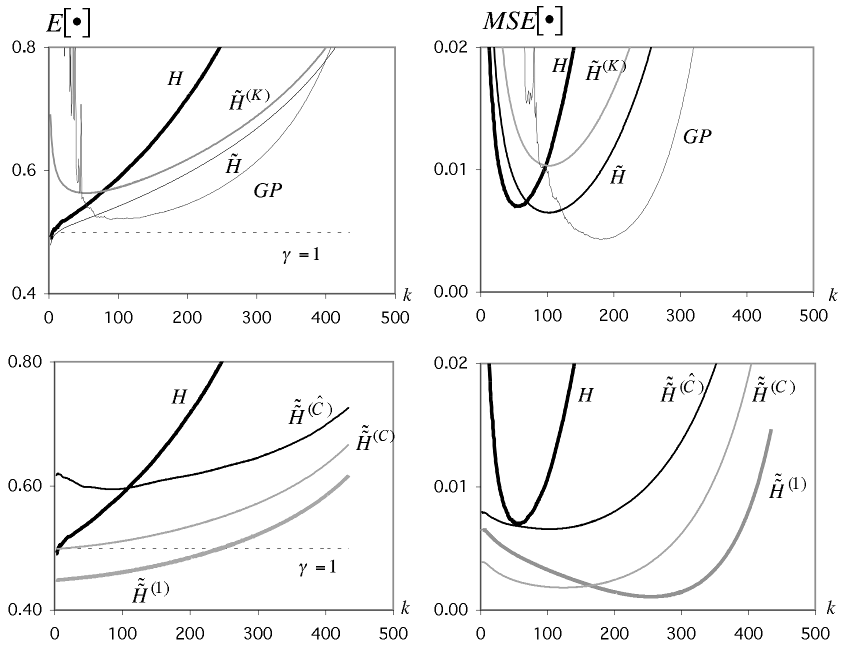

Figure 9: *Student*(2) parent with $\gamma = .5$ and $\rho = -1$ ($C = 0.71$).

$C = (-c_\nu\, \nu^{\nu/2})^{1/\nu}$, where c_ν is given for instance in Martins (2000). For $\nu=4$, and 2 we have $c_4=-3/16$, and $-1/4$, respectively. Consequently, for these models $C=$ 1.32 and 0.71, respectively. Such as in Figure 2, it is clear the existence of a bias close now to $\ln C/\ln n = 0.04$, and -0.05 for $\nu=4$ (Figure 8) and $\nu=2$ (Figure 9), respectively. For a more exhaustive simulation study see Oliveira (2002).

6. FINITE SAMPLE BAHAVIOUR AND ROBUSTNESS OF THE ESTIMATORS

The estimators $\widehat{\gamma}_n^H$, $\widetilde{\gamma}_n$, $\widetilde{\gamma}_n^{(\mathcal{K})}$ and $\widetilde{\widetilde{\gamma}}_n^{(\widehat{C})}$ (or $\widetilde{\widetilde{\gamma}}_n^{(\widehat{C}_0)}$, whenever we are outside Hall's class) will be also denoted $\gamma_n^{(1)}$, $\gamma_n^{(2)}$, $\gamma_n^{(3)}$ and $\gamma_n^{(4)}$, respectively. The r.v. $\widetilde{\widetilde{\gamma}}^{(C)}$ will be denoted $\gamma_n^{(5)}$. For the comparison of $\gamma_n^{(j)}$, $j=1,2,3,4$ and 5, at their optimal levels, we have implemented a multi-sample simulation of size 5000×10 in order to guarantee small standard errors (not presented in the tables, but available from the authors) for the simulated characteristics, the Mean Value ($E_\bullet$), the Mean Squared Error ($MSE_\bullet$), the Optimal Sample Fraction, $k_0^\bullet/n$, with $k_0^\bullet := \arg\min_k MSE_\bullet(k)$, and the Relative Efficiency ($REFF_\bullet$), defined as

$$REFF_\bullet = REFF\big[\gamma_{n0}^\bullet\big] = \sqrt{MSE_s\big[\gamma_n^{(1)}\big(k_{0s}^{(1)}(n)\big)\big]\big/MSE_s\big[\gamma_n^\bullet\big(k_{0s}^\bullet(n)\big)\big]}\ , \tag{6.1}$$

with $\gamma_{n0}^\bullet = \gamma_n^\bullet(k_{0s}^\bullet(n))$, and where MSE_s denotes the simulated MSE of the estimator at its simulated optimal level. The simulator of for instance $k_0^\bullet(n)$, denoted by $k_{0s}^\bullet(n)$, is $\widehat{E}_{10}\big[\,\overline{k}_0^\bullet(n)\big]$, the average of the 10 independent replicates of $\overline{k}_0^\bullet(n) = \arg\min_k \sum_{j=1}^{5000}\big(\gamma_{nj}^\bullet(k)-\gamma\big)^2$. The simulated mean values of these five estimators, at their optimal levels, are presented in Table 1 (Fréchet and Burr parents), Table 3 (Student parents) and Table 5 (Out-Hall, Sin-Fréchet and Sin-Burr parents). Tables 2, 4 and 6 are equivalent to tables 1, 3 and 5, respectively, with simulated mean values replaced by simulated mean squared errors.

Finally in Table 4 we present the $REFF$'s of the estimator $\widetilde{\widetilde{\gamma}}^{(1)}$ at its optimal level, for models with a scale $C\neq 1$. Note that the lost in efficiency is very high for the Fréchet, Sin-Fréchet and Burr models with a large scale, as it is the scale $C=2$, used here for illustration.

Table 1: Simulated mean values and mean squared errors of $\widehat{\gamma}_n^H$, $\widetilde{\gamma}_n$, $\widetilde{\gamma}_n^{(K)}$, $\widetilde{\widetilde{\gamma}}_n^{(\widehat{C})}$, and $\widetilde{\widetilde{\gamma}}_n^{(C)}$ at the simulated optimal levels, for Fréchet and Burr parents.

n	$E_{(1)}$	$E_{(2)}$	$E_{(3)}$	$E_{(4)}$	$E_{(5)}$	$MSE_{(1)}$	$MSE_{(2)}$	$MSE_{(3)}$	$MSE_{(4)}$	$MSE_{(5)}$
Fréchet parent: $\rho = -1$, $\gamma = 1$										
100	1.0987	1.0863	1.1682	1.0877	1.0347	0.0423	0.0365	0.0624	0.0393	0.0188
500	1.0628	1.0597	1.0958	1.1062	1.0297	0.0135	0.0130	0.0192	0.0131	0.0062
1000	1.0490	1.0487	1.0739	1.0501	1.0252	0.0083	0.0082	0.0115	0.0081	0.0039
2000	1.0380	1.0388	1.0565	1.0453	1.0207	0.0051	0.0050	0.0068	0.0050	0.0024
5000	1.0294	1.0286	1.0397	1.0375	1.0159	0.0027	0.0027	0.0035	0.0027	0.0013
Burr parent: $\rho = -0.5$, $\gamma = 1$										
100	1.2920	1.2825	1.4944	1.5005	1.0308	0.2286	0.1966	0.3745	0.2342	0.0329
500	1.1851	1.1785	1.2876	1.2718	1.0342	0.0834	0.0736	0.1245	0.0826	0.0121
1000	1.1545	1.1488	1.2321	1.2147	1.0330	0.0557	0.0500	0.0808	0.0555	0.0083
2000	1.1329	1.1236	1.1872	1.1861	1.0303	0.0374	0.0339	0.0527	0.0379	0.0057
5000	1.1021	1.0980	1.1424	1.0858	1.0266	0.0228	0.0207	0.0306	0.0228	0.0035
Burr parent: $\rho = -1$, $\gamma = 1$										
100	1.1361	1.1331	1.2424	1.2034	1.0452	0.0705	0.0648	0.1138	0.0664	0.0246
500	1.0782	1.0776	1.1266	1.1207	1.0336	0.0216	0.0206	0.0316	0.0207	0.0080
1000	1.0640	1.0625	1.0968	1.0699	1.0286	0.0132	0.0128	0.0188	0.0129	0.0050
2000	1.0498	1.0494	1.0741	1.0664	1.0238	0.0082	0.0080	0.0112	0.0079	0.0032
5000	1.0373	1.0368	1.0523	1.0330	1.0185	0.0043	0.0043	0.0057	0.0042	0.0017
Burr parent: $\rho = -2$, $\gamma = 1$										
100	1.0660	1.0648	1.1301	1.0179	1.0333	0.0294	0.0295	0.0469	0.0319	0.0181
500	1.0376	1.0385	1.0639	1.0655	1.0232	0.0077	0.0085	0.0118	0.0098	0.0052
1000	1.0290	1.0299	1.0465	1.0534	1.0186	0.0044	0.0049	0.0065	0.0059	0.0030
2000	1.0218	1.0228	1.0338	1.0290	1.0146	0.0025	0.0028	0.0036	0.0030	0.0017
5000	1.0148	1.0160	1.0222	0.9970	1.0103	0.0012	0.0013	0.0016	0.0013	0.0008

Table 2: Simulated mean values and mean squared errors of $\widehat{\gamma}_n^H$, $\widetilde{\gamma}_n$, $\widetilde{\gamma}_n^{(K)}$, $\widetilde{\widetilde{\gamma}}_n^{(\widehat{C})}$, and $\widetilde{\widetilde{\gamma}}_n^{(C)}$ at the simulated optimal levels, for Student parents.

n	$E_{(1)}$	$E_{(2)}$	$E_{(3)}$	$E_{(4)}$	$E_{(5)}$	$MSE_{(1)}$	$MSE_{(2)}$	$MSE_{(3)}$	$MSE_{(4)}$	$MSE_{(5)}$
Student(4) parent: $\rho = -0.5$, $\gamma = 0.25$										
100	0.3559	0.3548	0.4400	0.4435	0.2539	0.0316	0.0269	0.0557	0.0395	0.0029
500	0.3171	0.3137	0.3580	0.3287	0.2575	0.0109	0.0095	0.0174	0.0123	0.0010
1000	0.3037	0.3027	0.3365	0.3355	0.2576	0.0072	0.0063	0.0111	0.0081	0.0007
2000	0.2956	0.2939	0.3197	0.3228	0.2574	0.0048	0.0043	0.0072	0.0054	0.0005
5000	0.2860	0.2844	0.3025	0.2947	0.2569	0.0029	0.0026	0.0041	0.0031	0.0003
Student(2) parent: $\rho = -1$, $\gamma = 0.5$										
100	0.5988	0.5937	0.6832	0.6047	0.5235	0.0374	0.0329	0.0640	0.0381	0.0081
500	0.5556	0.5545	0.5959	0.5295	0.5187	0.0114	0.0104	0.0174	0.0107	0.0029
1000	0.5448	0.5425	0.5721	0.5426	0.5155	0.0069	0.0063	0.0101	0.0064	0.0018
2000	0.5356	0.5335	0.5546	0.5238	0.5127	0.0042	0.0039	0.0059	0.0040	0.0011
5000	0.5253	0.5235	0.5370	0.5240	0.5096	0.0022	0.0020	0.0029	0.0021	0.0006
Student(1) parent: $\rho = -2$, $\gamma = 1$										
100	1.0929	1.0642	1.1866	0.7481	1.0334	0.0608	0.0601	0.1019		0.0261
500	1.0557	1.0551	1.0985	1.0848	1.0305	0.0163	0.0174	0.0258	0.0187	0.0081
1000	1.0410	1.0431	1.0722	1.0340	1.0239	0.0093	0.0101	0.0143	0.0112	0.0048
2000	1.0314	1.0332	1.0526	1.0656	1.0184	0.0053	0.0059	0.0079	0.0066	0.0028
5000	1.0217	1.0231	1.0342	1.0151	1.0132	0.0025	0.0028	0.0036	0.0031	0.0014

Table 3: Simulated mean values and mean squared errors of $\widehat{\gamma}_n^H$, $\widetilde{\gamma}_n$, $\widetilde{\gamma}_n^{(K)}$, $\widetilde{\widetilde{\gamma}}_n^{(\hat{C}_0)}$, and $\widetilde{\widetilde{\gamma}}_n^{(C)}$ at the simulated optimal levels, for Out-Hall, Sin-Fréchet and Sin-Burr parents.

n	$E_{(1)}$	$E_{(2)}$	$E_{(3)}$	$E_{(4)}$	$E_{(5)}$	$MSE_{(1)}$	$MSE_{(2)}$	$MSE_{(3)}$	$MSE_{(4)}$	$MSE_{(5)}$
Out-Hall parent: $\rho=-1$, $\gamma=1$										
100	0.7178	0.7228	0.7881	0.7320	0.9248	0.1574	0.1568	0.1289	0.1517	0.0146
500	0.8255	0.8325	0.8808	0.8368	0.9569	0.0653	0.0644	0.0502	0.0625	0.0103
1000	0.8613	0.8657	0.9062	0.8704	0.9618	0.0437	0.0431	0.0333	0.0418	0.0076
2000	0.8892	0.8922	0.9252	0.8961	0.9669	0.0291	0.0286	0.0220	0.0278	0.0055
5000	0.9166	0.9189	0.9436	0.9220	0.9728	0.0169	0.0166	0.0128	0.0161	0.0034
Sin-Fréchet parent: $\gamma=1$										
100	1.0284	1.0680	1.1263	1.0629	1.0391	0.0359	0.0380	0.0593	0.0347	0.0209
500	1.0067	1.0223	1.0372	1.0184	1.0155	0.0073	0.0091	0.0114	0.0078	0.0056
1000	1.0025	1.0133	1.0204	1.0107	1.0098	0.0036	0.0047	0.0055	0.0040	0.0031
2000	1.0011	1.0075	1.0104	1.0060	1.0058	0.0018	0.0024	0.0027	0.0021	0.0016
5000	1.0007	1.0033	1.0041	1.0027	1.0027	0.0007	0.0010	0.0010	0.0009	0.0007
Sin-Burr parent: $\rho=-0.5$, $\gamma=1$										
100	1.0459	1.1088	1.2267	1.0873	1.0382	0.1601	0.1314	0.2380	0.1498	0.0296
500	1.0116	1.0301	1.0480	1.0286	1.0184	0.0323	0.0290	0.0378	0.0277	0.0100
1000	1.0042	1.0168	1.0235	1.0150	1.0115	0.0163	0.0151	0.0179	0.0140	0.0059
2000	1.0030	1.0091	1.0110	1.0078	1.0066	0.0081	0.0077	0.0085	0.0070	0.0033
5000	1.0011	1.0037	1.0041	1.0033	1.0028	0.0032	0.0031	0.0033	0.0028	0.0015

Table 4: Relative efficiencies of $\widetilde{\widetilde{\gamma}}_n^{(1)}$ relatively to $\widehat{\gamma}_n^H$, at their optimal levels, for models with $C \neq 1$.

n	100	500	1000	2000	5000
Fréchet ($\gamma=1$, $\rho=-1$, $C=.5$)	1.5713	1.9344	2.0609	2.1933	2.4099
Fréchet ($\gamma=1$, $\rho=-1$, $C=2$)	0.9726	0.7781	0.7039	0.6304	0.5422
Burr ($\gamma=1$, $\rho=-.5$, $C=.5$)	2.8678	3.2438	3.4208	3.5979	3.9156
Burr ($\gamma=1$, $\rho=-.5$, $C=2$)	2.0830	1.7470	1.6127	1.4894	1.3433
Burr ($\gamma=1$, $\rho=-1$, $C=.5$)	1.9624	2.1431	2.2474	2.3686	2.5641
Burr ($\gamma=1$, $\rho=-1$, $C=2$)	1.1404	0.9347	0.8472	0.7660	0.6569
Burr ($\gamma=1$, $\rho=-2$, $C=.5$)	1.3344	1.4941	1.5549	1.6196	1.7241
Burr ($\gamma=1$, $\rho=-2$, $C=2$)	0.8069	0.6189	0.5436	0.4719	0.3808
Student(4) ($\gamma=0.25$, $\rho=-0.5$, $C=1.32$)	2.5488	1.9430	1.7554	1.5875	1.3802
Student(2) ($\gamma=0.5$, $\rho=-1$, $C=.71$)	2.0645	2.4684	2.5236	2.6085	2.7371
Student(1) ($\gamma=1$, $\rho=-2$, $C=.32$)	0.9243	1.0394	1.3273	1.8245	1.9260
Out-Hall ($\gamma=1$, $C=.5$)	1.9073	1.5357	1.4122	1.2991	1.1562
Out-Hall ($\gamma=1$, $C=2$)	3.2856	3.8498	3.6741	3.6192	3.6691
Sin-Fréchet ($\gamma=1$, $C=.5$)	1.4649	1.4346	1.3653	1.3087	1.2586
Sin-Fréchet ($\gamma=1$, $C=2$)	0.8766	0.6537	0.5592	0.4720	0.3658
Sin-Burr ($\gamma=1$, $\rho=-.5$, $C=.5$)	2.5707	2.0930	1.9252	1.7945	1.6664
Sin-Burr ($\gamma=1$, $\rho=-.5$, $C=2$)	1.6669	1.2783	1.1763	1.1088	1.0713

A few final remarks:

1. The class of statistics $\widetilde{\widetilde{\gamma}}_n^{(\tau)}(k)$ revealed a surprisingly good behaviour among the estimators considered both for small and for large sample sizes, and for all the simulated models (most of them in the class of models where (1.13) holds, with $C = 1$). Indeed, for every k, we have got $MSE\big[\widetilde{\widetilde{\gamma}}_n(k)\big]$ smaller than $MSE\big[\gamma_n^H(k)\big]$ and also smaller than $MSE\big[\widetilde{\gamma}_n(k)\big]$ for all models simulated. This class of statistics also enables us to find an estimator of the tail index γ, which behaves better than the maximum lilelihood estimator based on the Generalized Pareto excesses, for most of the models simulated.

2. Particularly astonishing is the behaviour of $\widetilde{\widetilde{\gamma}}_n^{(\tau)}(k)$ for small values of ρ, like the value $\rho = -2$ used herein for illustration, a region where has been claimed to be difficult to find good competitors for the Hill estimator. Also, the results obtained for models for which the second order condition does not hold deserve further investigation, and are interesting from a point of view of a more general application.

3. It may be claimed that such a good behaviour is due to the fact that $\widetilde{\widetilde{\gamma}}_n(k)$ is not only non-invariant for location, like the Hill statistic, but also non-invariant for scale. The adequate estimation of the parameter C is a possible way out, but that induces an increase in the variance of our final tail index estimator, and the nice features of this estimator will disappear. Alternatively, the best way to proceed is to estimate the mean squared errors of our estimators as functions of k, merely on the basis of the available sample, proceeding next to the adaptive choice of the k and a τ-value providing a mean squared error smaller than that of the Hill estimator for every k. It is perhaps also sensible to use an extra *tuning* parameter a, a shift in the location of our data, like in Gomes and Oliveira (2003b). All this work is essentially computational, and as said before, overpasses the scope of this paper.

REFERENCES

[1] Beirlant, J.; Vynckier, P. and Teugels, J.L. (1996a). Excess function and estimation of the extreme-value index, *Bernoulli*, **2**(2), 293–318.

[2] Beirlant, J.; Vynckier, P. and Teugels, J.L. (1996b). Tail index estimation, Pareto quantile plots, and regression diagnostics, *J. Amer. Statist. Assoc.*, **91**, 1659–1667.

[3] Caeiro, F. and Gomes, M.I. (2002). A new class of asymptotically unbiased estimators of the index of regular variation, *Test*, **11**(2), 345–364.

[4] Csörgő, S. and Viharos, L. (1997). Asymptotic normality of least squares estimators of tail indices, *Bernoulli*, **3**, 351–370.

[5] Csörgő, S. and Viharos, L. (1998). *Estimating the tail index.* In "Asymptotic Methods in Probability and Statistics" (Szyszkowicz, Ed.), North-Holland, Amsterdam, 833–881.

[6] Danielsson, J.; de Haan, L.; Peng, L. and de Vries, C.G. (2001). Using a bootstrap method to choose the sample fraction in the tail index estimation, *J. Multivariate Analysis*, **76**, 226–248.

[7] Davison, A. (1984). *Modelling excesses over high threshold with an application.* In "Statistical Extremes and Applications" (J. Tiago de Oliveira, Ed.), D. Reidel, 461–482.

[8] Draisma, G.; de Haan, L.; Peng, L. and Pereira, T.T. (1999). A bootstrap-based method to achieve optimality in estimating the extreme-value index, *Extremes*, **2**(4), 367–404.

[9] Drees, H.; Ferreira, A. and de Haan, L. (2004). On maximum likelihood estimation of the extreme value index, *Ann. Appl. Prob.*, **14**, 1179–1201.

[10] Ferreira, A.; de Haan, L. and Peng, L. (2003). On optimising the estimation of high quantiles of a probability distribution, *Statistics*, **37**(5), 401–434.

[11] Fraga Alves, M.I.; Gomes, M.I. and de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter, *Portugaliae Mathematica*, **60**(1), 193–213.

[12] Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics* (2nd edition), Krieger.

[13] Geluk, J. and de Haan, L. (1987). *Regular Variation, Extensions and Tauberian Theorems*, CWI Tract 40, Center for Mathematics and Computer Science, Amsterdam, Netherlands.

[14] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire, *Ann. Math.*, **44**, 423–453.

[15] Gomes, M.I. (2002). *A note on the excesses over a high threshold*, Notas e Comunicações, CEAUL 10/2002.

[16] Gomes, M.I. and Martins, M.J. (2001). Alternatives to Hill's estimator — asymptotic versus finite sample behaviour, *J. Statist. Planning and Inference*, **93**, 161–180.

[17] Gomes, M.I. and Martins, M.J. (2002). "Asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter, *Extremes*, **5**(1), 5–31.

[18] Gomes, M.I. and Martins, M.J. (2004). Bias reduction and explicit estimation of the tail index, *J. Statist. Planning and Inference*, **124**, 361–378.

[19] Gomes, M.I.; Martins, M.J. and Neves, M. (2000). Alternatives to a semi-parametric estimator of parameters of rare events — the Jackknife methodology, *Extremes*, **3**(3), 207–229.

[20] Gomes, M.I.; Martins, M.J. and Neves, M. (2002). Generalized Jackknife semi-parametric estimators of the tail index, *Portugaliae Mathematica*, **59**(4), 393–408.

[21] Gomes, M.I. and Oliveira, O. (2001). The bootstrap methodology in Statistics of Extremes — choice of the optimal sample fraction, *Extremes*, **4**(4), 331–358.

[22] Gomes, M.I. and Oliveira, O. (2003a). Maximum likelihood revisited under a semi-parametric context — estimation of the tail index, *J. Statist. Comp. and Simul.*, **73**(4), 285–301.

[23] Gomes, M.I. and Oliveira, O. (2003b). How can non-invariant statistics work in our benefit in the semi-parametric estimation of parameters of rare events, *Commun. in Statist. — Simulation and Computation*, **32**(4), 1005–1028.

[24] de Haan, L. (1970). *On Regular Variation and its Application to the Weak Convergence of Sample Extremes*, Mathematical Centre Tract 32, Amsterdam.

[25] de Haan, L. and Peng, L. (1998). Comparison of tail index estimators, *Statistica Neerlandica*, **52**, 60–70.

[26] Hall, P. and Welsh, A.H. (1985). Adaptive estimates of parameters of regular variation, *Ann. Statist.*, **13**, 331–341.

[27] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Ann. Statist.*, **3**, 1163–1174.

[28] Kratz, M. and Resnick, S. (1996). The *QQ*-estimator of the index of regular variation, *Commun. in Statist. — Stochastic Models*, **12**, 699–724.

[29] Martins, M.J. (2000). *Estimação de Caudas Pesadas — Variantes ao Estimador de Hill*, Ph. D. Thesis, DEIO, FCUL.

[30] Oliveira, O. (2002). *In Extremis*, Ph. D. Thesis, DEIO, FCUL.

[31] Schultze, J. and Steinbach, J. (1996). On least squares estimates of an exponential tail coefficient, *Statistics and Decisions*, **14**, 353–372.

[32] Smith, R.L. (1984a). *Threshold methods for sample extremes.* In "Statistical Extremes and Applications" (J. Tiago de Oliveira, Ed.), D. Reidel, 621–638.

[33] Smith, R.L. (1984b). Maximum likelihood estimation in a class of nonregular cases, *Biometrika*, **72**, 67–90.

[34] Smith, R.L. (1987). Estimating tails of probability distributions, *The Annals of Statistics*, **15**(3), 1174–1207.

REVSTAT – Statistical Journal
Volume 4, Number 2, June 2006, 111–130

PERFORMANCE OF THE *EM* ALGORITHM ON THE IDENTIFICATION OF A MIXTURE OF WATSON DISTRIBUTIONS DEFINED ON THE HYPERSPHERE

Authors: Adelaide Figueiredo
– Faculdade de Economia da Universidade do Porto e LIACC,
Rua Dr. Roberto Frias, 4200-464 Porto, Portugal
Adelaide@fep.up.pt

Paulo Gomes
– Instituto Nacional de Estatística,
Av. António José Almeida, 1000-043 Lisboa, Portugal
Paulo.Gomes@ine.pt

Received: April 2004 Revised: August 2005 Accepted: March 2006

Abstract:

- We consider a set of n individuals described by p standardised variables, and we suppose that the individuals are previously selected from a population and the variables are a sample of variables assumed to come from a mixture of k bipolar Watson distributions defined on the hypersphere. In this context we provide the identification of the mixture through the *EM* algorithm and we also carry out a simulation study to compare the maximum likelihood estimates obtained from samples of moderate size with the respective asymptotic estimates. Our simulation results revealed good performance of the *EM* algorithm for moderate sample sizes.

Key-Words:

- *EM algorithm; mixture; principal components; Watson distribution.*

AMS Subject Classification:

- 62H11, 62H12, 62H25.

1. INTRODUCTION

We consider multivariate data with n individuals described by p variables. In the classical approach it is usual to assume that the p variables are fixed and the n individuals are randomly selected from a population of individuals. Now, we consider that the n individuals are fixed and the p variables are randomly selected from a population of variables. We standardise the variables to be points on the unit sphere in $\mathbb{R}^n$, denoted by $S_{n-1} = \{\mathbf{x} \in \mathbb{R}^n\colon\ \mathbf{x}'\mathbf{x} = 1\}$.

We suppose that the group of available variables on S_{n-1} is composed of k subgroups of variables and each subgroup comes from a bipolar Watson distribution. So we associate the sample of variables to a mixture of k bipolar Watson distributions defined on the hypersphere, as in Gomes [9]. This author considers an approach, based on the sampling of variables, and introduces some new results concerning the estimation of the parameters of the bipolar Watson distribution, taking into account not a sample of individuals but, a sample of variables. This type of ideas was referred to by Hotelling [10] who, in the context of Principal Components, studied the convergence of the eigenvalues and eigenvectors of the covariance matrix of groups of variables randomly chosen from a population of variables, when the dimension of the groups increases. Escoufier [5] also proposed a new coefficient for evaluating the proximity of two groups of variables, but supposing that the variables are observed.

For the identification of the mixture, we use the well-known *EM* algorithm proposed in Dempster, Laird and Rubin [3] (see Redner and Homer [14]).[1] This algorithm was developed to solve the likelihood equations in problems of incomplete data and we apply it to estimate the parameters of a mixture of k bipolar Watson distributions (see Figueiredo [7]).

The bipolar Watson distribution has been much used for axial data on the sphere (see Watson [16], Fisher, Lewis and Embleton [8] and Mardia and Jupp [13]). This distribution is denoted by $W_n(\mathbf{u}, \xi)$ and it has density probability function given by

$$f(\mathbf{x}) = \left\{ {}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \xi\right) \right\}^{-1} \exp\left\{ \xi\left(\mathbf{u}'\mathbf{x}\right)^2 \right\}, \qquad \mathbf{x} \in S_{n-1},\ \ \mathbf{u} \in S_{n-1},\ \ \xi > 0\,, \tag{1.1}$$

where the normalising constant is the reciprocal of a confluent hipergeometric function defined by

$$ {}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \xi\right) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)} \int_0^1 e^{\xi t}\, t^{-0.5} (1-t)^{(n-3)/2}\, dt\ . \tag{1.2}$$

[1] Another possible method for the identification of the mixture is the k-means method proposed in Diday and Schroeder [15] (see Gomes [9]).

This distribution has two parameters: a directional parameter $\mathbf{u}$ and a concentration parameter ξ, which measures the concentration about $\pm\mathbf{u}$. As ξ increases, the distribution becomes more concentrated about $\pm\mathbf{u}$. This is a rotationally symmetric distribution about the principal axis $\pm\mathbf{u}$ and it is bimodal, with modes $\mathbf{u}$ and $-\mathbf{u}$.

Let $X = \left[\mathbf{x}^1|\mathbf{x}^2|...|\mathbf{x}^p\right]$ be a random sample of variables from the bipolar Watson distribution $W_n(\mathbf{u},\xi)$. The maximum likelihood estimator of $\mathbf{u}$ is the eigenvector associated with the largest eigenvalue $\widehat{w}$ of $XX' = \sum_{i=1}^{p} \mathbf{x}^i\,\mathbf{x}^{i\,\prime}$, that is, $\widehat{\mathbf{u}}$ is defined by $(XX')\,\widehat{\mathbf{u}} = \widehat{w}\,\widehat{\mathbf{u}}$. So, it follows that the maximum likelihood estimator of the directional parameter $\mathbf{u}$ based on the sample of variables is the first principal component of the sample. The maximum likelihood estimator of ξ is the solution of the equation $Y(\widehat{\xi}) = \widehat{w}/p$, where the function $Y(\xi)$ is defined by $Y(\xi) = \frac{d}{d\xi} \ln {}_1F_1(1/2, n/2, \xi)$.

The estimators $\widehat{\xi}$ and $\widehat{w}$ have asymptotic Gaussian distribution (see Gomes [9] and Bingham [1]):

$$\widehat{\xi} \;\dot\sim\; N\left(\xi, \frac{1}{p\,Y_{11}^2(\xi)}\right) \qquad \text{and} \qquad \frac{\widehat{w}}{p} \;\dot\sim\; N\left(Y(\xi), \frac{Y_{11}^2(\xi)}{p}\right). \tag{1.3}$$

where the function $Y_{11}^2(\xi)$ is defined by $Y_{11}^2(\xi) = \frac{d^2}{d\xi^2} \ln {}_1F_1(\frac{1}{2}, \frac{n}{2}, \xi)$.

In this study we consider the particular case of a bipolar Watson distribution. If we had assumed $\xi < 0$ in (1.1), we would obtain a girdle Watson distribution and the study of this distribution would be similar to the one that is done in this paper.

In Section 2 we present the identification of the mixture of k bipolar Watson distributions through the *EM* algorithm. In Section 3 we carry out a simulation study to compare the behaviour of the estimators obtained through the *EM* algorithm for moderate samples with the respective asymptotic estimators. In Section 4 we give some concluding remarks.

2. IDENTIFICATION OF A MIXTURE OF k BIPOLAR WATSON DISTRIBUTIONS DEFINED ON THE HYPERSPHERE

The density function of a mixture of k bipolar Watson components $C_1, ..., C_k$ defined on the hypersphere, whose identifiability was proved by Kent [12], is given by

$$\begin{aligned} g(\mathbf{x}|\phi) = \sum_{j=1}^{k} \pi_j f(\mathbf{x}|\theta_j)\,, \qquad & \mathbf{x}\in S_{n-1}, \quad 0<\pi_j<1, \quad j=1,...,k, \quad \sum_{j=1}^{k}\pi_j = 1\,, \\ & \phi = \big(\mathbf{u}_1,...,\mathbf{u}_k,\xi_1,...,\xi_k,\pi_1,...,\pi_k\big), \quad \theta_j = (\mathbf{u}_j,\xi_j)\;, \end{aligned} \tag{2.1}$$

where $(\pi_1, ..., \pi_k)$ are the proportions of the mixture and $f(\mathbf{x}|\theta_j)$ is the density function corresponding to the C_j component.

As a mixture of distributions may be seen as a problem of incomplete data (see Everitt and Hand [6]), the *EM* algorithm may be applied to solve the likelihood equations in the estimation of the parameters of a mixture of k bipolar Watson distributions.

Let $\left[\mathbf{x}^1|\mathbf{x}^2|...|\mathbf{x}^p\right]$ be a random sample from the mixture and let $Z = \left[\mathbf{z}_1|...|\mathbf{z}_p\right]$ be the missing data, where the indicator vector $\mathbf{z}_i = (Z_{i1}, Z_{i2}, ..., Z_{ik})$ with $Z_{ij} = \begin{cases} 1 & \text{if } \mathbf{x}^i \in C_j \\ 0 & \text{if } \mathbf{x}^i \notin C_j \end{cases}$, $\sum\limits_{j=1}^{k} Z_{ij} = 1$ indicates the component of the variable $\mathbf{x}^i$ of the mixture.

The log likelihood associated with the complete sample $\left[\mathbf{x}^1|...|\mathbf{x}^p|Z\right]$ is given by

$$L\big(\phi|\mathbf{x}^1, ..., \mathbf{x}^p, Z\big) = \sum_{i=1}^{p}\sum_{j=1}^{k} t_j(\mathbf{x}^i)\,\ln\Big\{\pi_j f\big(\mathbf{x}^i|\theta_j\big)\Big\}, \tag{2.2}$$

where $t_j(\mathbf{x}^i)$ is the *posterior* probability of $\mathbf{x}^i$ belonging to C_j defined by

$$t_j(\mathbf{x}^i) = \frac{\pi_j\, f(\mathbf{x}^i|\theta_j)}{\sum\limits_{h=1}^{k} \pi_h\, f(\mathbf{x}^i|\theta_h)} . \tag{2.3}$$

The log likelihood associated with the complete sample given by (2.2) may be written as

$$L\big(\phi|\mathbf{x}^1, ..., \mathbf{x}^p, Z\big) = L\big(\phi_1|\mathbf{x}^1, ..., \mathbf{x}^p, Z\big) + L\big(\phi_2|\mathbf{x}^1, ..., \mathbf{x}^p, Z\big) ,$$

where

$$L\big(\phi_1|\mathbf{x}^1, ..., \mathbf{x}^p, Z\big) = \sum_{i=1}^{p}\sum_{j=1}^{k} t_j(\mathbf{x}^i)\,\ln f(\mathbf{x}^i|\theta_j)\,, \qquad \phi_1 = (\theta_1, ..., \theta_k)$$

and

$$L\big(\phi_2|\mathbf{x}^1, ..., \mathbf{x}^p, Z\big) = \sum_{i=1}^{p}\sum_{j=1}^{k} t_j(\mathbf{x}^i)\,\ln \pi_j\,, \qquad \phi_2 = (\pi_1, ..., \pi_k) .$$

To estimate the vector of unknown parameters ϕ of the mixture, the *EM* algorithm proceeds iteratively in two steps:

$$E - \text{Estimation} \quad \text{and} \quad M - \text{Maximisation}\,.$$

The algorithm starts with the initial solution:

$$\phi^0 = \big(\mathbf{u}_1^0, ..., \mathbf{u}_k^0, \xi_1^0, ..., \xi_k^0, \pi_1^0, ..., \pi_k^0\big) .$$

In the m-th iteration, the two steps are:

E-Step

Use estimates $\phi^{(m)}$ of the parameters of the mixture in the m-th iteration for $j=1,...,k$ and $i=1,...,p$ to estimate the *posterior* probability of $\mathbf{x}^i$ belonging to the j-th component of the mixture

$$t_j^{(m)}(\mathbf{x}^i) \;=\; \frac{\pi_j^{(m)} f\big(\mathbf{x}^i|\theta_j^{(m)}\big)}{\sum\limits_{h=1}^{k} \pi_h^{(m)} f\big(\mathbf{x}^i|\theta_h^{(m)}\big)} \;. \tag{2.4}$$

M-Step

Use estimates $t_j^{(m)}(\mathbf{x}^i)$ to maximise the logarithm of the likelihood function $L(\phi_1|\mathbf{x}^1,...,\mathbf{x}^p,Z)$.

First, we consider the function $L(\phi_1)$, subject to the constraint $\mathbf{u}_j'\mathbf{u}_j = 1$:

$$L(\phi_1) \;=\; \sum_{i=1}^{p}\sum_{j=1}^{k} t_j^{(m)}(\mathbf{x}^i)\Big[-\ln\big\{{}_1F_1(1/2,n/2,\xi_j)\big\} + \xi_j(\mathbf{u}_j'\mathbf{x}^i)^2\Big] - \lambda_1(\mathbf{u}_j'\mathbf{u}_j - 1) \;,$$

where λ_1 is a Lagrange multiplier and $t_j^{(m)}(\mathbf{x}^i)$ is defined in (2.4).

The maximum likelihood estimate of $\mathbf{u}_j$ is the solution of the following equation:

$$\frac{\partial L(\phi_1)}{\partial \mathbf{u}_j} \;=\; \sum_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)\, 2\,\xi_j\, \mathbf{x}^i\mathbf{x}^{i\,\prime}\mathbf{u}_j - 2\,\lambda_1\mathbf{u}_j \;=\; 0 \;. \tag{2.5}$$

We premultiply the last expression by $\mathbf{u}_j\prime$ to obtain

$$\lambda_1 \;=\; \xi_j \sum_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)\, \mathbf{u}_j\prime\, \mathbf{x}^i\mathbf{x}^{i\,\prime}\mathbf{u}_j \;.$$

Then, the maximum likelihood estimator of $\mathbf{u}_j\prime$ in the $(m{+}1)$-th iteration, $\widehat{\mathbf{u}}_j^{(m+1)}$ is the eigenvector associated with the eigenvalue $\widehat{w}_j$, that is

$$\left(\sum_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)\, \mathbf{x}^i\mathbf{x}^{i\,\prime}\right)\widehat{\mathbf{u}}_j^{(m+1)} \;=\; \widehat{w}_j\, \widehat{\mathbf{u}}_j^{(m+1)} \;, \qquad j=1,..,k \;, \tag{2.6}$$

where $\widehat{w}_j$ is a eigenvalue of $\sum\limits_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)\, \mathbf{x}^i\mathbf{x}^{i\,\prime}$ and it is given by

$$\widehat{w}_j \;=\; \sum_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)\, \widehat{\mathbf{u}}_j^{(m+1)}\prime\, \mathbf{x}^i\mathbf{x}^{i\,\prime}\, \widehat{\mathbf{u}}_j^{(m+1)} \;.$$

Next, we show that we maximise $L(\phi_1)$ if we consider the largest eigenvalue of the matrix. In fact, the function $L(\phi_1)$ can be written in the form

$$L(\phi_1) \;=\; -\sum_{i=1}^{p}\sum_{j=1}^{k} t_j^{(m)}(\mathbf{x}^i)\;\ln\bigl\{{}_1F_1(1/2,n/2,\xi_j)\bigr\} + \sum_{j=1}^{k}\xi_j\,\widehat{w}_j\;.$$

As $\ln{}_1F_1(1/2,n/2,\xi_j)>0$, we have $\sum\limits_{i=1}^{p}\sum\limits_{j=1}^{k} t_j^{(m)}(\mathbf{x}^i)\;\ln{}_1F_1(1/2,n/2,\xi_j)>0$. We also have $\widehat{w}_j\ge 0$ because $\sum\limits_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)\,\mathbf{x}^i\mathbf{x}^{i\,\prime}$ is a positive definite matrix. Consequently, the function $L(\phi_1)$ is maximised if $\widehat{w}_j$ is maximum.

Second, the maximum likelihood estimator of ξ_j is the solution of the following equation

$$\frac{\partial L(\phi_1)}{\partial\xi_j} \;=\; \sum_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)\,\Bigl\{-Y(\xi_j)+(\mathbf{u}_j{\prime}\mathbf{x}^i)^2\Bigr\} \;=\; 0\;,$$

where the function $Y(.)$ is defined in Section 1. The solution of this equation leads to the maximum of $L(\phi_1)$ as we show that $\partial^2L(\phi_1)/\partial\xi_j^2<0$, $\forall_{\xi_j}$. In fact, $\partial^2L(\phi_1)/\partial\xi_j^2=-\sum\limits_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)\,dY(\xi_j)/\xi_j$ and $Y(\xi)$ is an increasing function (see Gomes [9]).

Then, the maximum likelihood estimator of ξ_j in the $(m{+}1)$-th iteration, $\widehat{\xi}_j^{(m+1)}$, is the solution of the equation

$$Y\bigl(\widehat{\xi}_j^{(m+1)}\bigr) \;=\; \frac{\widehat{w}_j}{\sum\limits_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)}\;, \qquad j=1,...,k\;. \tag{2.7}$$

Third, we consider the function $L(\phi_2)$, subject to the constraint $\sum\limits_{j=1}^{k}\pi_j=1$:

$$L(\phi_2) \;=\; \sum_{i=1}^{p}\sum_{j=1}^{k} t_j^{(m)}(\mathbf{x}^i)\;\ln\pi_j \;-\; \lambda_2\biggl(\sum_{j=1}^{k}\pi_j-1\biggr)\,,$$

where λ_2 is a Lagrange multiplier. The maximum likelihood estimator of π_j is the solution of the following equation

$$\frac{\partial L(\phi_2)}{\partial\pi_j} \;=\; \frac{\sum\limits_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)}{\pi_j} - \lambda_2 \;=\; 0\;.$$

We sum the last equation for j from 1 to k to obtain $\lambda_2=p$. Then, the maximum likelihood estimator of π_j in the $(m{+}1)$-th iteration, $\widehat{\pi}_j^{(m+1)}$ is given by

$$\widehat{\pi}_j^{(m+1)} \;=\; \frac{\sum\limits_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)}{p}\;, \qquad j=1,...,k\;. \tag{2.8}$$

The estimation of the parameters $\mathbf{u}_j$ and ξ_j associated with the j-th component gives us a privileged direction as well as a measure of dispersion of the j-th cluster around this direction.

A partition $(P_1, ..., P_k)$ of the sample of variables is obtained assigning the variable $\mathbf{x}^j$ to the component for which the *posterior* probability is the largest, that is,

$$P_j = \left\{\mathbf{x}^i\colon\ t_j(\mathbf{x}^i) = \max_h t_h(\mathbf{x}^i),\ h=1, ..., k\right\} \tag{2.9}$$

and when $t_j(\mathbf{x}^i) = t_h(\mathbf{x}^i)$ consider $\mathbf{x}^i \in P_j$ if $j<h$.

3. SIMULATION STUDY

We considered a mixture with equal proportions ($\pi_1 = \pi_2 = 0.5$) of two bipolar Watson distributions: $W_n(\mathbf{u}_1, \xi_1)$ and $W_n(\mathbf{u}_2, \xi_2)$, with $\xi_1 = \xi_2 = \xi$, $\mathbf{u}_1 = (0, ..., 0, 1)$ and $\mathbf{u}_2 = (0, ...0, (1-\cos^2\theta)^{1/2}, \cos\theta)$, where θ is the angle between $\mathbf{u}_1$ and $\mathbf{u}_2$. The bipolar Watson distribution is rotationally symmetric about the directional parameter, so if we had used, for each θ, other directional parameters $\mathbf{u}_1$ and $\mathbf{u}_2$, we would have obtained the same results in our study. For the simulation of the bipolar Watson distribution we used a rejection-type method (see Huo [11] and Bingham [2]). We considered two dimensions of the sphere $n = \mathit{10{,}30}$. For each n, we assumed equal samples size $p_1 = p_2 = p = \mathit{30(10)100}$, several values of the concentration parameter $\xi = \mathit{10(10)50{,}100}$ and several values of the angle $\theta = 18°, 54°, 90°$. For each case, we considered *2500* replicates of the *EM* algorithm. In each replicate, we used a randomly chosen initial solution and a sufficiently large number of iterations (*100*) to obtain the final solution. We supposed that the algorithm converged, in a certain replicate, if the condition:

$$\left|\Big(L\big(\phi^{(m+1)}\big) - L\big(\phi^{(m)}\big)\Big)\Big/L\big(\phi^{(m+1)}\big)\right| \le 10^{-5}$$

holds in the last five iterations, where $L(\phi^{(m)})$ denotes the likelihood of the sample in the m-th iteration. For each n and p, the *EM* algorithm converged in most part of the replicates, it did not converge only in very few replicates when ξ is very small or θ is small.

In each replicate we determined the following estimates $\widehat{\xi}_j$, $\widehat{w}_j/p_j$, $j=1,2$, $\widehat{\theta}$, $\widehat{\pi}_j$, $j=1,2$ of the parameters ξ_j, $Y(\xi_j)$, $j=1,2$, θ, π_j, $j=1,2$, respectively, where p_j is the dimension of the j-th group, which is equal to $\sum\limits_{i=1}^{p} t_j(\mathbf{x}^i)$. Then, we calculated the average and the standard deviation of the estimates obtained in all replicates, denoted by $\overline{\widehat{\xi}}_j$, $\overline{\widehat{w}_j/p_j}$, $j=1,2$, $\overline{\widehat{\theta}}$, $\overline{\widehat{\pi}}_j$, $j=1,2$ and $s(\widehat{\xi}_j)$, $s(\widehat{w}_j/p_j)$, $j=1,2$, $s(\widehat{\theta})$, $s(\widehat{\pi}_j)$, $j=1,2$, respectively. If in a replicate the *EM* algorithm did not converge we excluded that replicate for calculating the average and the standard deviation of the estimates.

By (1.3) the asymptotic expected value of $\widehat{\xi}_j$ and $\widehat{w}_j/p_j$ are ξ_j and $Y(\xi_j)$ respectively, $j=1,2$. In Table 1 and Figure 1, we indicate the values of $Y(\xi)$ [2] for each n and ξ.

Table 1: Values of $Y(\xi)$ for each n and ξ.

$n\diagdown\xi$	10	20	30	40	50	60	70	80	90	100
10	0.500	0.766	0.847	0.886	0.909	0.924	0.935	0.943	0.950	0.955
30	0.074	0.241	0.496	0.630	0.706	0.756	0.791	0.817	0.838	0.854

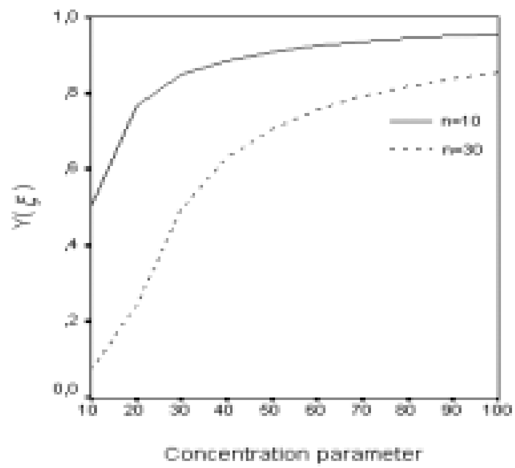

Figure 1: Values of $Y(\xi)$ for $n=10$ and $n=30$.

As expected for each n, $Y(\xi)$ is an increasing function with ξ, which tends to *1*, when ξ increases (see Gomes [9], p. 43–45). For each ξ, the function $Y(\xi)$ increases when n decreases.

We determined the estimated relative bias of the estimators given by the expressions: $(\overline{\widehat{\xi}}_j-\xi_j)/\xi_j, (\overline{\widehat{w}}_j/p_j-Y(\xi_j))/Y(\xi_j),\ j=1,2,\ (\overline{\widehat{\theta}}-\theta)/\theta,\ (\overline{\widehat{\pi}}_j-\pi_j)/\pi_j$, $j=1{,}2$ and the estimated mean squared error (MSE) given by: $s^2(\widehat{\xi}_j)+(\overline{\widehat{\xi}}_j-\xi_j)^2$, $s^2(\widehat{w}_j/p_j)+(\overline{\widehat{w}}_j/p_j-Y(\xi_j))^2,\ j=1,2,\ \ s^2(\widehat{\theta})+(\overline{\widehat{\theta}}-\theta)^2,\ s^2(\widehat{\pi}_j)+(\overline{\widehat{\pi}}_j-\pi_j)^2,\ j=1,2.$

[2] We obtained the function $Y(\xi)$ using the Kummer function, which is defined by $M(a,b,z) = 1+\sum_{i=1}^{\infty}\left\{a(a+1)...(a+i-1)z^i\right\}/\left\{b(b+1)...(b+i-1)\,i!\right\}$ or by the integral $M(a,b,z) = \Gamma(b)/\left\{\Gamma(b-a)\Gamma(a)\right\}\int_0^1 e^{zt}\,t^{a-1}(1-t)^{b-a-1}dt$, where ${}_1F_1(1/2,n/2,\xi)=M(1/2,n/2,\xi)$.

We indicate the results of our simulation study in the Tables A1–A4 of the Appendix and in the Figures 2–8. In the tables of the Appendix, the algorithm converged in all replicates for each case. We have also produced another 4 tables, which were not included: two tables for the relative bias (for $n=10$ and $n=30$) and two tables for the *MSE* (for $n=10$ and $n=30$) of the estimators when the concentration parameter ξ varies.

In Figure 2 we observe that

- As expected, the estimators $\widehat{\xi}_1$ and $\widehat{\xi}_2$ are asymptotically unbiased, that is the estimated relative bias of these estimators tends to 0 as the sample size p increases. For fixed ξ and p, the relative bias of $\widehat{\xi}_1$ and $\widehat{\xi}_2$ tends to decrease when θ increases. For an angle $\theta = 90°$ or $\theta = 54°$, the bias of the estimators $\widehat{\xi}_1$ and $\widehat{\xi}_2$ is relatively small and when $\theta = 90°$ the bias is not greater than 10% of the true value of the concentration parameter (for $n=10,30$, $\xi=30,100$ and $p=30(10)100$).

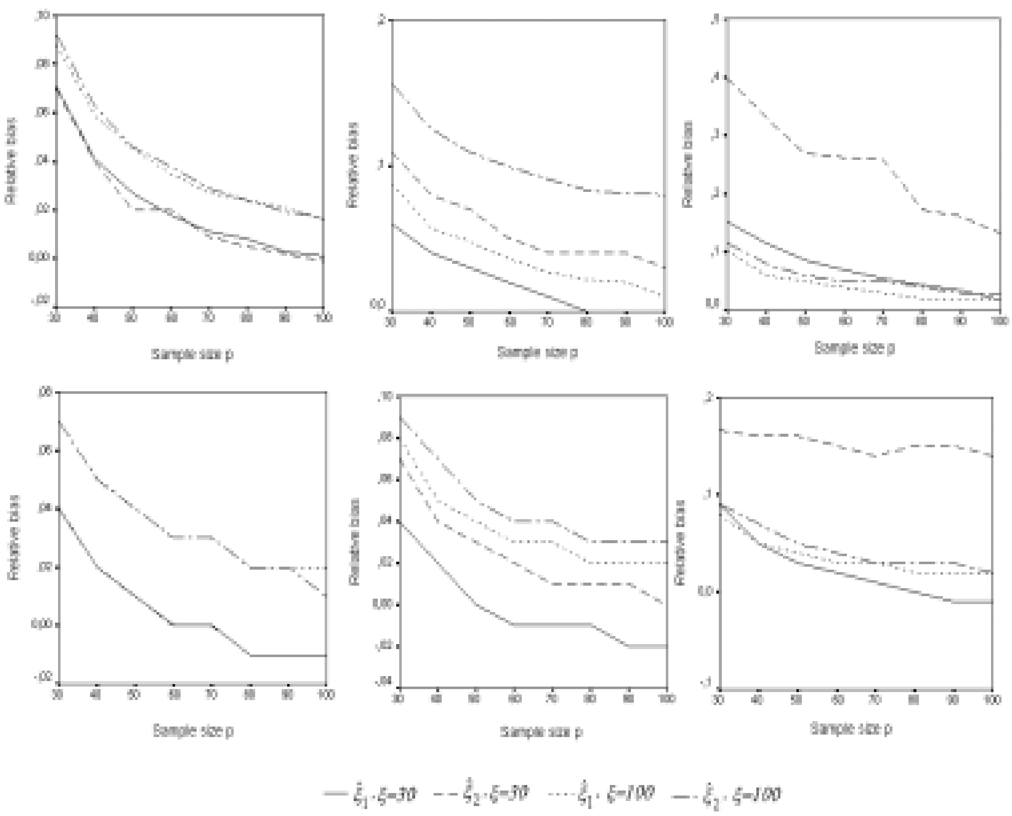

Figure 2: Relative bias of the estimators $\widehat{\xi}_1$ and $\widehat{\xi}_2$ when p varies (in top: $n = 10$, in bottom: $n = 30$ and from left to right: angle $90°$, $54°$, $18°$).

In Figure 3 we observe that

- As expected, in general the estimators $\widehat{\xi}_1$ and $\widehat{\xi}_2$ become more efficient as p increases. When the angle is large or moderate ($\theta = 90°$ or $\theta = 54°$) and $\xi = 30$, these estimators have relatively small *MSE* and become less efficient when ξ increases.

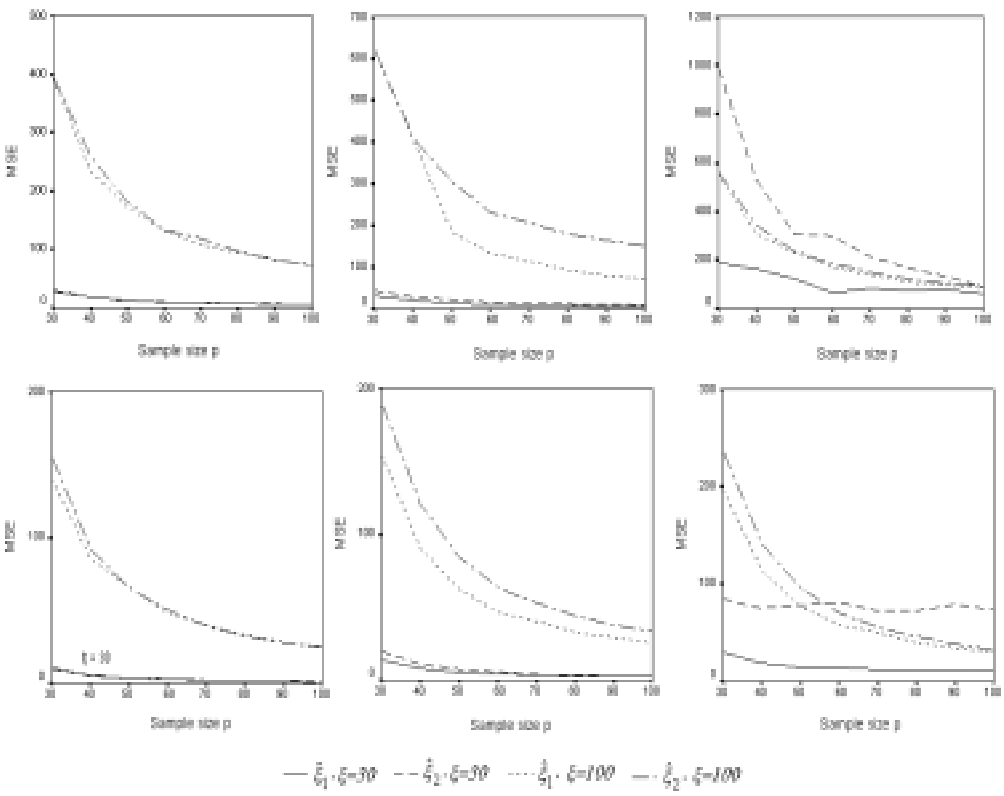

Figure 3: Mean squared error of the estimators $\widehat{\xi}_1$ and $\widehat{\xi}_2$ when p varies (in top: $n = 10$, in bottom: $n = 30$ and from left to right: angle $90°$, $54°$, $18°$).

In Figure 4 we observe that

- When the angle is moderate or large ($\theta = 54°$ or $\theta = 90°$), the bias of $\widehat{\xi}_1$ and $\widehat{\xi}_2$ is very small and maintains approximately constant or increases slightly as ξ increases for $\xi \geq 20$ when $n = 10$ and for $\xi \geq 30$ when $n = 30$. When $n = 10$ and $\theta = 18°$, the bias of the estimators is relatively large, but it decreases when ξ increases.

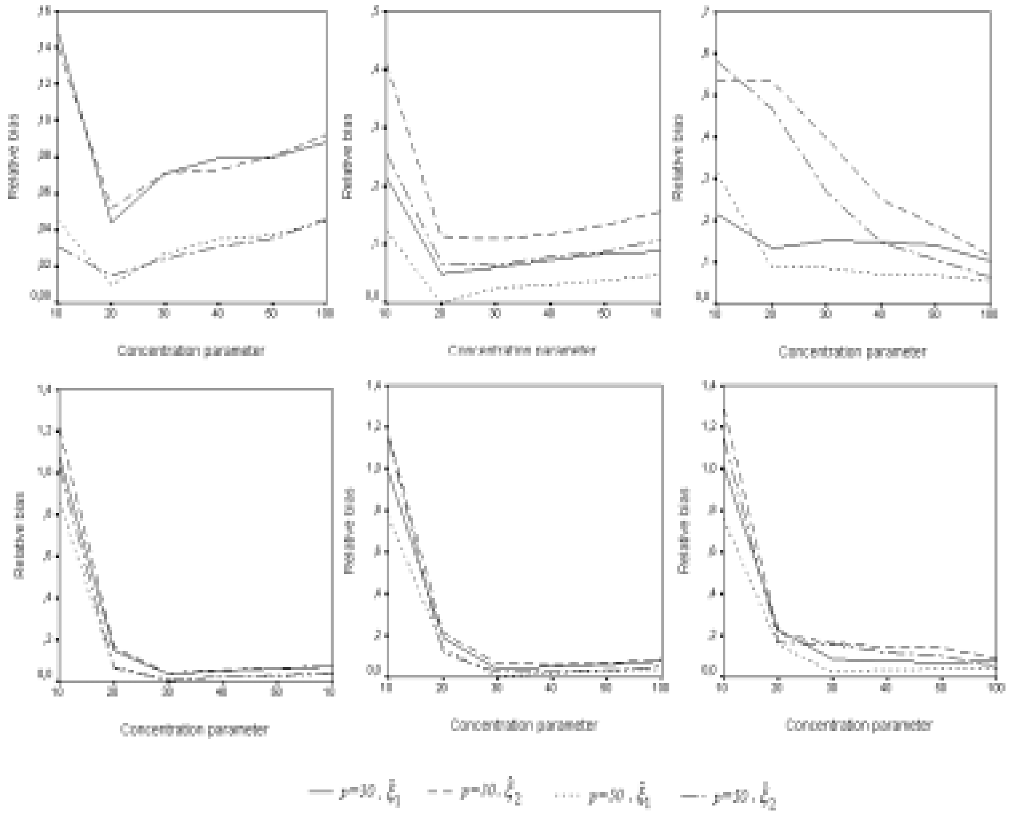

Figure 4: Relative bias of the estimators $\widehat{\xi}_1$ and $\widehat{\xi}_2$ when ξ varies (in top: $n = 10$, in bottom: $n = 30$ and from left to right: angle $90°$, $54°$, $18°$).

In Figure 5 we observe that

- When the angle is moderate or large ($\theta = 54°$ or $\theta = 90°$), the *MSE* of the estimators $\widehat{\xi}_1$ and $\widehat{\xi}_2$ increases when ξ increases for $\xi \geq 30$ and so these estimators become less efficient.
- The estimators $\widehat{w}_1/p_1$ and $\widehat{w}_2/p_2$ are unbiased or have very small bias for every p and ξ. When $\theta = 90°$ the bias of these estimators is not greater than approximately 3% of the respective parameter. The estimators $\widehat{w}_1/p_1$ and $\widehat{w}_2/p_2$ are asymptotically unbiased, that is, the estimated relative bias of the estimators tends to *0* as the sample size p increases. See Tables A1–A2 of the Appendix.

In Figure 6 we observe that

- The estimators $\widehat{w}_1/p_1$ and $\widehat{w}_2/p_2$ have bias approximately equal to *0* for $\xi \geq 20$ when $n = 10$ and for $\xi \geq 30$ when $n = 30$.
- As the *MSE* of the estimators $\widehat{w}_1/p_1$ and $\widehat{w}_2/p_2$ are *0* or approximately *0*, these estimators are very efficient. See Tables A3–A4 of the Appendix.

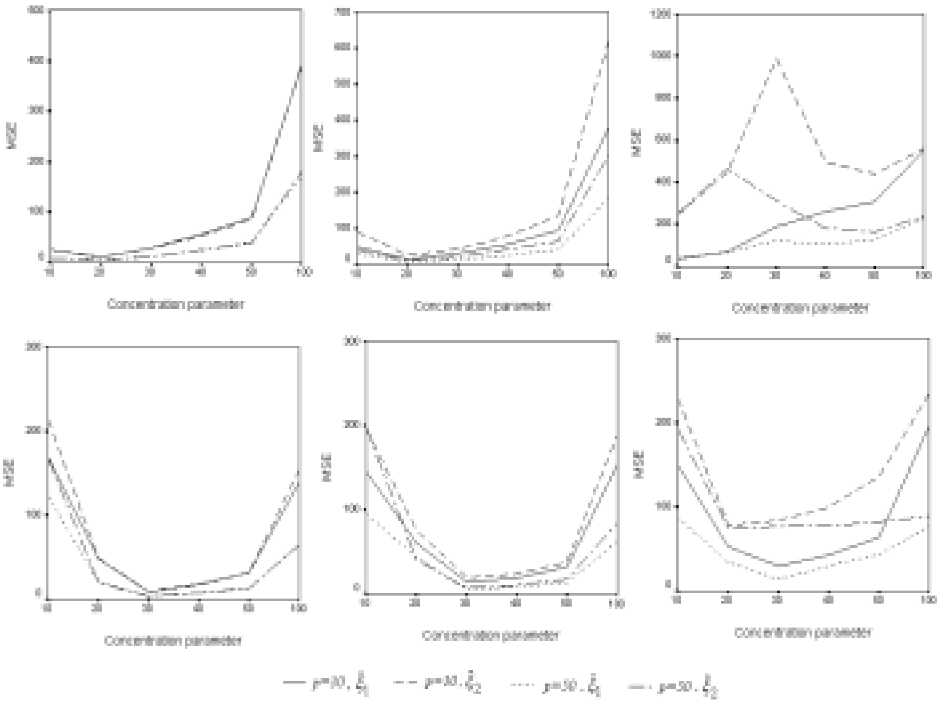

Figure 5: Mean squared error of the estimators $\widehat{\xi_1}$ and $\widehat{\xi_2}$ when ξ varies (in top: $n=10$, in bottom: $n = 30$ and from left to right: angle 90°, 54°, 18°).

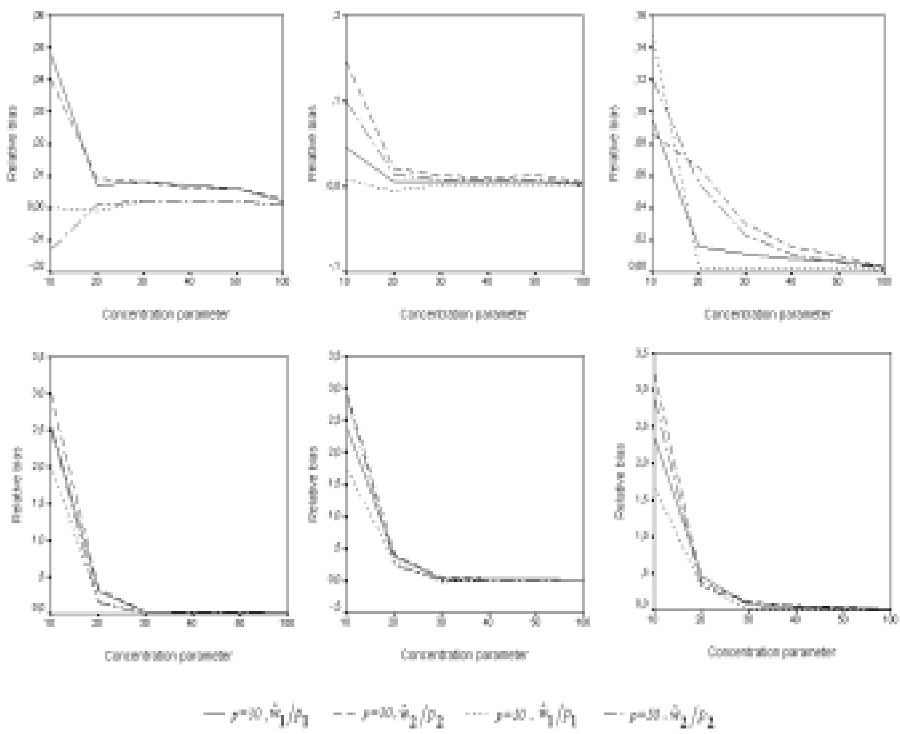

Figure 6: Relative bias of the estimators $\widehat{w}_1/p_1$ and $\widehat{w}_2/p_2$ when ξ varies (in top: $n=10$, in bottom: $n = 30$ and from left to right: angle 90°, 54°, 18°).

In Figure 7 we observe that

- The estimator $\widehat{\theta}$ has relatively small *MSE*, except for $n = 30$ and $\xi = 30$ when the relative bias and the standard deviation of $\widehat{\theta}$ are relatively large. The *MSE* of the estimator $\widehat{\theta}$ decreases when p increases.

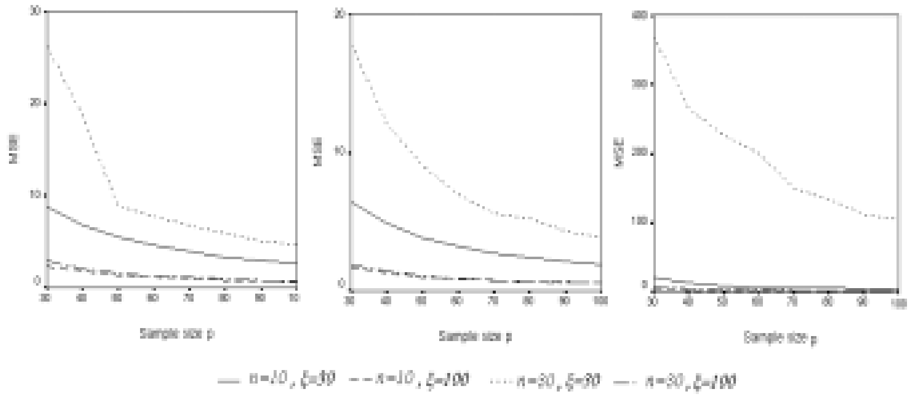

Figure 7: Mean squared error of the estimator $\widehat{\theta}$ when p varies (from left to right: angle 90°, 54°, 18°).

In Figure 8 we observe that

- For every θ and $\xi \geq 20$, the *MSE* of the estimator $\widehat{\theta}$ decreases when ξ increases.
- The estimators $\widehat{\pi}_1$ and $\widehat{\pi}_2$ are unbiased or present very small bias for the analysed cases, except in some cases when $\theta = 18°$. See Tables A1–A2 of the Appendix.
- The estimators $\widehat{\pi}_1$ and $\widehat{\pi}_2$ have *MSE* equal to *0* or approximately *0*, and so these estimators are very efficient. See Tables A3–A4 of the Appendix.

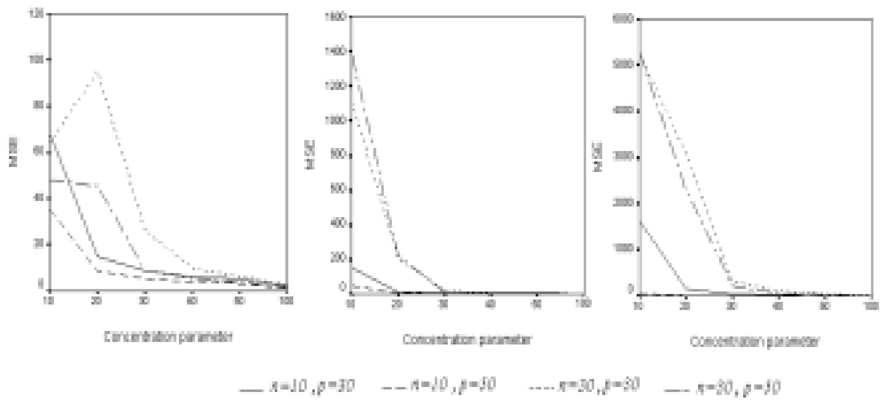

Figure 8: Mean squared error of the estimator $\widehat{\theta}$ when ξ varies (from left to right: angle 90°, 54°, 18°).

4. CONCLUSION

The simulation study has revealed a good identification of a mixture of bipolar Watson distributions defined on the hypersphere through the *EM* algorithm.

The performance of this algorithm is good for moderate sample sizes, essentially on the estimation of the prior probabilities and on the estimation of the directional parameters of the mixture. For a large or moderate angle θ between the directional parameters of the mixture, the efficiency of the estimators of the concentration parameters of the mixture is better for moderate values (neither very small nor very large) of the true concentration parameters. The estimation of the angle θ is very efficient in general and the efficiency of $\widehat{\theta}$ improves as the concentration parameter increases.

REFERENCES

[1] Bingham, C. (1964). *Distribution on the sphere and on the projective plane*, Ph. D. Thesis, Yale University.

[2] Bingham, C. (1983). *Simulation on the n-sphere: The Bingham distribution*, Rapport Technique, University of Minnesota.

[3] Dempster, A.P.; Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion), *Journal of the Royal Statistical Society*, **B**, 3, 1–38.

[4] Diday, E. (1972). *Nouveaux concepts et nouvelles méthodes en classification automatique*, Thèse de Doctorat d'État, Université Paris VI.

[5] Escoufier, Y. (1968). Le traitement des variables vectorielles, *Biometrics*, **29**, 31–47.

[6] Everitt, B.S. and Hand, D.J. (1981). *Finite Mixture Distributions*, Chapman and Hall, Ltd, London.

[7] Figueiredo, A. (2000). *Classificação de variáveis no contexto de um modelo probabilístico definido na n-esfera*, Tese de Doutoramento, Faculdade de Ciências da Universidade de Lisboa.

[8] Fisher, N.I.; Lewis, T. and Embleton, B.J.J. (1987). *Statistical Analysis of Spherical Data*, Cambridge University Press.

[9] Gomes, P. (1987). *Distribution de Bingham sur la n-sphere: une nouvelle approche de l'Analyse Factorielle*, Thèse d'État, Université des Sciences et Techniques du Languedoc-Montpellier.

[10] Hotteling, H. (1933). Analysis of a complex of statistical variables into principal components, *J. Educational Psychology*, **24**, 417–441.

[11] Kent, J.T. (1983). Identifiability of finite mixtures for directional data, *Annals of Statistics*, **11**, 984–988.

[12] Huo, V. (1984). *Small samples from Bingham distributions*, Ph. D. thesis, University of Minnesota.

[13] Mardia, K.V. and Jupp, P.E. (2000). *Directional Statistics*, 2nd Ed., John Wiley and Sons, Chichester.

[14] Redner, R. and Homer, W. (1984). Mixture Densities, Maximum Likelihood, EM algorithm, *Journal of the American Statistical Association*, **B**, 39, 1–38.

[15] Schroeder, A. (1981). Analyse d'un mélange de distributions de probabilité de même type, *Revue de Statistique Appliquée*, **XXIV**, 1, 39–62.

[16] Watson, G.S. (1983). *Statistics on Spheres*, John Wiley and Sons, New York.

APPENDIX

Table 2: Relative bias of the estimators for $n=10$ with the sample size p

ξ	θ	p	$\widehat{\xi}_1$	$\widehat{\xi}_2$	$\widehat{w}_1/p_1$	$\widehat{w}_2/p_2$	$\widehat{\theta}$	$\widehat{\pi}_1$	$\widehat{\pi}_2$
30	90°	30	0.071	0.071	0.008	0.008	0.000	0.005	−0.005
		40	0.041	0.041	0.004	0.004	−0.001	−0.001	0.001
		50	0.027	0.024	0.002	0.002	−0.001	0.004	−0.004
		60	0.018	0.017	0.001	0.001	0.000	−0.005	0.005
		70	0.011	0.009	0.000	0.000	0.000	0.001	−0.001
		80	0.008	0.005	0.000	−0.001	0.001	0.001	−0.001
		90	0.003	0.002	−0.001	−0.001	0.000	0.000	0.000
		100	0.001	−0.001	−0.001	−0.002	0.002	0.001	−0.001
	54°	30	0.061	0.109	0.006	0.013	−0.002	0.002	−0.002
		40	0.042	0.080	0.004	0.010	−0.002	0.001	−0.001
		50	0.025	0.065	0.002	0.008	−0.004	0.000	0.000
		60	0.020	0.051	0.001	0.006	−0.005	−0.002	0.002
		70	0.011	0.044	0.000	0.005	−0.006	0.000	0.000
		80	0.002	0.040	−0.001	0.005	−0.007	0.000	0.000
		90	0.002	0.037	−0.001	0.005	−0.007	−0.001	0.001
		100	−0.004	0.033	−0.002	0.005	−0.007	−0.001	0.001
	18°	30	0.152	0.398	0.011	0.030	0.256	0.116	−0.116
		40	0.116	0.331	0.006	0.027	0.192	0.086	−0.086
		50	0.086	0.269	0.003	0.023	0.157	0.070	−0.070
		60	0.070	0.262	0.004	0.022	0.125	0.061	−0.061
		70	0.055	0.259	0.000	0.018	0.107	0.049	−0.050
		80	0.044	0.173	−0.001	0.016	0.086	0.030	−0.030
		90	0.034	0.161	−0.002	0.016	0.077	0.031	−0.031
		100	0.018	0.132	−0.004	0.014	0.068	0.020	−0.020
100	90°	30	0.088	0.092	0.002	0.003	0.000	0.000	0.000
		40	0.059	0.063	0.002	0.002	0.000	−0.003	0.003
		50	0.045	0.046	0.001	0.001	0.000	0.003	−0.003
		60	0.035	0.038	0.001	0.001	0.000	0.001	−0.001
		70	0.027	0.029	0.001	0.001	0.000	0.000	0.000
		80	0.024	0.024	0.000	0.000	0.000	0.000	0.000
		90	0.021	0.019	0.000	0.000	0.000	0.002	−0.002
		100	0.016	0.017	0.000	0.000	0.000	−0.001	0.001
	54°	30	0.088	0.156	0.002	0.005	−0.003	0.000	0.000
		40	0.057	0.125	0.001	0.004	−0.003	0.002	−0.002
		50	0.048	0.109	0.001	0.004	−0.004	−0.001	0.001
		60	0.036	0.099	0.001	0.003	−0.004	0.000	0.000
		70	0.027	0.091	0.001	0.003	−0.004	0.001	−0.001
		80	0.022	0.083	0.000	0.003	−0.005	0.000	0.000
		90	0.018	0.081	0.000	0.003	−0.005	−0.001	0.001
		100	0.014	0.080	0.000	0.003	−0.005	−0.001	0.001
	18°	30	0.103	0.116	0.003	0.003	0.041	−0.001	0.001
		40	0.064	0.081	0.002	0.002	0.032	0.000	0.000
		50	0.054	0.064	0.001	0.002	0.027	−0.002	0.002
		60	0.040	0.052	0.001	0.001	0.023	−0.001	0.001
		70	0.029	0.046	0.001	0.001	0.021	0.001	−0.001
		80	0.024	0.037	0.000	0.001	0.019	0.000	0.000
		90	0.020	0.034	0.000	0.001	0.018	−0.002	0.002
		100	0.015	0.032	0.000	0.001	0.017	−0.002	0.002

Table 3: Relative bias of the estimators for $n=30$ with the sample size p

ξ	θ	p	$\widehat{\xi}_1$	$\widehat{\xi}_2$	$\widehat{w}_1/p_1$	$\widehat{w}_2/p_2$	$\widehat{\theta}$	$\widehat{\pi}_1$	$\widehat{\pi}_2$
30	90°	30	0.036	0.038	0.029	0.031	−0.001	0.001	−0.001
		40	0.017	0.018	0.012	0.012	0.000	0.000	0.000
		50	0.009	0.007	0.005	0.003	0.001	0.003	−0.003
		60	0.001	0.000	−0.004	−0.005	0.000	0.004	−0.004
		70	−0.004	−0.002	−0.008	−0.006	0.000	−0.001	0.001
		80	−0.006	−0.008	−0.011	−0.020	0.001	−0.005	0.005
		90	−0.011	−0.010	−0.015	−0.014	0.000	0.001	−0.001
		100	−0.011	−0.013	−0.016	−0.017	0.000	0.001	−0.001
	54°	30	0.040	0.067	0.028	0.053	0.048	0.009	−0.009
		40	0.017	0.043	0.008	0.035	0.038	0.008	−0.008
		50	0.003	0.026	−0.012	0.019	0.031	0.006	−0.006
		60	−0.005	0.019	−0.012	0.013	0.025	0.007	−0.007
		70	−0.010	0.011	−0.017	0.007	0.022	0.006	−0.006
		80	−0.014	0.007	−0.021	0.002	0.021	0.006	−0.006
		90	−0.018	0.005	−0.025	0.001	0.017	0.006	−0.006
		100	−0.020	0.001	−0.027	−0.003	0.016	0.004	−0.004
	18°	30	0.085	0.166	0.065	0.113	0.994	0.218	−0.218
		40	0.049	0.156	0.033	0.106	0.857	0.243	−0.243
		50	0.025	0.157	0.011	0.106	0.773	0.274	−0.274
		60	0.020	0.153	0.006	0.100	0.700	0.274	−0.274
		70	0.009	0.144	−0.006	0.096	0.629	0.286	−0.286
		80	−0.003	0.148	−0.018	0.100	0.581	0.292	−0.292
		90	−0.005	0.151	−0.021	0.100	0.531	0.290	−0.290
		100	−0.006	0.140	−0.023	0.092	0.493	0.275	−0.275
100	90°	30	0.070	0.074	0.010	0.011	0.000	0.001	−0.001
		40	0.049	0.053	0.007	0.008	−0.001	0.005	−0.005
		50	0.038	0.039	0.006	0.006	0.000	0.003	−0.003
		60	0.030	0.031	0.004	0.005	0.000	−0.001	0.001
		70	0.026	0.025	0.004	0.004	0.000	0.001	−0.001
		80	0.020	0.021	0.003	0.003	0.000	0.003	−0.003
		90	0.018	0.017	0.003	0.003	0.000	−0.002	0.002
		100	0.015	0.014	0.002	0.002	0.000	0.000	0.000
	54°	30	0.075	0.087	0.011	0.012	0.003	0.004	−0.004
		40	0.051	0.065	0.008	0.009	0.002	−0.004	0.004
		50	0.039	0.052	0.006	0.008	0.001	0.002	−0.002
		60	0.029	0.042	0.004	0.006	−0.001	0.002	−0.002
		70	0.025	0.036	0.004	0.006	−0.001	0.002	−0.002
		80	0.021	0.033	0.003	0.005	−0.002	0.000	0.000
		90	0.018	0.028	0.003	0.004	−0.002	0.000	0.000
		100	0.016	0.027	0.002	0.004	−0.002	−0.002	0.002
	18°	30	0.078	0.091	0.011	0.013	0.119	0.006	−0.007
		40	0.054	0.067	0.008	0.010	0.095	−0.005	0.005
		50	0.040	0.052	0.006	0.008	0.080	0.001	−0.001
		60	0.030	0.041	0.004	0.006	0.066	0.001	−0.001
		70	0.026	0.034	0.004	0.005	0.059	0.001	−0.001
		80	0.021	0.030	0.003	0.005	0.052	−0.002	0.002
		90	0.016	0.025	0.002	0.004	0.047	0.001	−0.001
		100	0.015	0.022	0.002	0.003	0.044	−0.004	0.004

Table 4: Mean squared error of the estimators for $n = 10$ with the sample size p

ξ	θ	p	$\widehat{\xi}_1$	$\widehat{\xi}_2$	$\widehat{w}_1/p_1$	$\widehat{w}_2/p_2$	$\widehat{\theta}$	$\widehat{\pi}_1$	$\widehat{\pi}_2$
		30	28.39	29.34	0.001	0.001	8.8	0.01	0.01
		40	17.49	18.51	0	0	6.81	0.01	0.01
		50	12.99	12.57	0	0	5.47	0.01	0.01
	90°	60	10.00	10.14	0	0	4.67	0	0
		70	8.41	8.52	0	0	3.92	0	0
		80	7.14	6.90	0	0	3.25	0	0
		90	6.40	6.53	0	0	2.92	0	0
		100	5.77	5.47	0	0	2.67	0	0
		30	29.61	42.68	0.001	0.001	6.43	0.01	0.01
		40	20.20	29.07	0	0.001	4.89	0.01	0.01
		50	13.51	21.59	0	0.001	3.80	0	0
30	54°	60	11.28	16.18	0	0.001	3.22	0	0
		70	8.94	14.29	0	0.001	2.70	0	0
		80	7.32	11.97	0	0.001	2.40	0	0
		90	6.38	9.97	0	0.001	2.09	0	0
		100	5.87	8.59	0	0.001	1.94	0	0
		30	188.76	987.30	0.001	0.002	34.48	0.04	0.04
		40	163.08	526.42	0.001	0.002	22.30	0.03	0.03
		50	120.85	309.95	0.001	0.002	16.37	0.03	0.03
	18°	60	66.77	299.22	0.001	0.002	10.97	0.03	0.03
		70	83.52	214.79	0.001	0.001	9.52	0.03	0.03
		80	75.39	167.57	0.001	0.001	7.45	0.02	0.02
		90	77.56	133.97	0.001	0.001	6.14	0.02	0.02
		100	47.93	86.04	0.001	0.001	5.42	0.02	0.02
		30	391.95	393.74	0	0	2.34	0.01	0.01
		40	233.19	258.01	0	0	1.82	0.01	0.01
		50	172.30	181.35	0	0	1.37	0	0
	90°	60	132.61	133.23	0	0	1.17	0	0
		70	108.59	118.57	0	0	0.98	0	0
		80	94.51	96.45	0	0	0.81	0	0
		90	81.51	82.20	0	0	0.81	0	0
		100	71.97	72.44	0	0	0.66	0	0
		30	380.48	620.53	0	0	1.70	0.01	0.01
		40	239.14	407.87	0	0	1.30	0.01	0.01
		50	183.84	304.88	0	0	0.98	0.01	0.01
100	54°	60	140.48	263.35	0	0	0.88	0	0
		70	112.76	209.09	0	0	0.79	0	0
		80	93.25	179.25	0	0	0.64	0	0
		90	79.89	164.52	0	0	0.64	0	0
		100	71.10	150.27	0	0	0.56	0	0
		30	554.58	560.82	0	0	2.79	0.01	0.01
		40	309.45	342.75	0	0	2.07	0.01	0.01
		50	231.62	238.22	0	0	1.51	0.01	0.01
	18°	60	169.17	186.41	0	0	1.27	0	0
		70	138.87	151.15	0	0	1.13	0	0
		80	110.62	121.23	0	0	0.92	0	0
		90	97.90	106.03	0	0	0.82	0	0
		100	83.85	92.72	0	0	0.77	0	0

Table 5: Mean squared error of the estimators for $n = 30$ with the sample size p

ξ	θ	p	$\widehat{\xi}_1$	$\widehat{\xi}_2$	$\widehat{w}_1/p_1$	$\widehat{w}_2/p_2$	$\widehat{\theta}$	$\widehat{\pi}_1$	$\widehat{\pi}_2$
30	90°	30	8.99	9.33	0.002	0.002	26.27	0.01	0.01
		40	5.36	5.82	0.002	0.002	18.66	0.01	0.01
		50	3.83	3.78	0.001	0.001	8.85	0	0
		60	3.03	3.13	0.001	0.001	7.78	0	0
		70	2.52	2.52	0.001	0.001	6.81	0	0
		80	2.17	2.10	0.001	0.001	5.92	0	0
		90	1.97	1.96	0.001	0.001	5.06	0	0
		100	1.74	1.85	0.001	0.001	4.67	0	0
	54°	30	13.87	19.54	0.004	0.004	17.86	0.01	0.01
		40	8.14	11.07	0.002	0.003	12.06	0.01	0.01
		50	5.24	7.29	0.002	0.002	9.08	0.01	0.01
		60	4.14	5.72	0.001	0.002	6.97	0	0
		70	3.48	4.12	0.001	0.001	5.58	0	0
		80	3.02	3.55	0.001	0.001	5.23	0	0
		90	2.92	3.19	0.001	0.001	4.33	0	0
		100	2.61	2.73	0.001	0.001	3.83	0	0
	18°	30	29.40	84.02	0.005	0.011	370.63	0.03	0.03
		40	19.16	74.94	0.004	0.011	265.15	0.03	0.03
		50	14.18	77.88	0.003	0.011	226.23	0.04	0.04
		60	14.27	80.57	0.003	0.011	200.45	0.04	0.04
		70	12.62	72.60	0.003	0.011	148.84	0.04	0.04
		80	11.84	72.02	0.003	0.010	134.46	0.05	0.05
		90	11.80	78.76	0.003	0.011	112.28	0.23	0.23
		100	12.62	73.23	0.003	0.010	104.27	0.05	0.05
100	90°	30	139.69	154.52	0	0	2.92	0.01	0.01
		40	85.91	92.13	0	0	2.08	0.01	0.01
		50	65.74	65.13	0	0	1.59	0.01	0.01
		60	48.48	50.47	0	0	1.17	0	0
		70	39.77	39.20	0	0	1.17	0	0
		80	33.22	32.60	0	0	0.98	0	0
		90	28.12	27.72	0	0	0.81	0	0
		100	25.01	24.67	0	0	0.81	0	0
	54°	30	154.06	189.17	0	0	1.86	0	0
		40	89.81	120.39	0	0	1.43	0.01	0.01
		50	62.23	84.41	0	0	1.06	0.01	0.01
		60	46.53	63.51	0	0	0.84	0	0
		70	40.38	52.90	0	0	0.75	0	0
		80	32.60	44.35	0	0	0.67	0	0
		90	29.15	37.53	0	0	0.58	0	0
		100	25.36	33.66	0	0	0.58	0	0
	18°	30	196.39	234.80	0	0	7.18	0.01	0.01
		40	113.25	141.25	0	0	4.67	0.01	0.01
		50	76.84	96.11	0	0	3.39	0.01	0.01
		60	58.06	69.951	0	0	2.57	0	0
		70	49.88	55.35	0	0	2.15	0	0
		80	39.48	45.93	0	0	1.74	0	0
		90	33.79	38.03	0	0	1.50	0	0
		100	30.11	32.01	0	0	1.37	0	0

REVSTAT – Statistical Journal
Volume 4, Number 2, June 2006, 131–142

A LOGNORMAL MODEL FOR INSURANCE CLAIMS DATA

Authors: Daiane Aparecida Zuanetti
– Departamento de Estatística, Universidade Federal de São Carlos, São Paulo, Brazil

Carlos A. R. Diniz
– Departamento de Estatística, Universidade Federal de São Carlos, São Paulo, Brazil
dcad@power.ufscar.br

José Galvão Leite
– Departamento de Estatística, Universidade Federal de São Carlos, São Paulo, Brazil

Received: June 2005 Revised: February 2006 Accepted: February 2006

Abstract:

- In the insurance area, especially based on observations of the number of claims, $N(w)$, corresponding to an exposure w, and on observations of the total amount of claims incurred, $Y(w)$, the risk theory arises to quantify risks and to fit models of pricing and insurance company ruin. However, the main problem is the complexity to obtain the distribution function of $Y(w)$ and, consequently, the likelihood function used to calculate the estimation of the parameters.

  This work considers the Poisson$(w\lambda)$, $\lambda>0$, for $N(w)$ and lognormal(μ,σ^2), $-\infty<\mu<\infty$, and $\sigma^2>0$, for Z_i, the individual claims, and presents maximum-likelihood estimates for λ, μ and σ^2.

Key-Words:

- *lognormal distribution; maximum-likelihood estimation; number of claims; total amount of claims.*

AMS Subject Classification:

- 62J02, 62F03.

1. INTRODUCTION

In the insurance area, the main goals of the risk theory are to study, analyze, specify dimensions and quantify risks. The risk theory is also responsible for fitting models of pricing and insurance company ruin, especially based on observations of the random variables for the number of claims, $N(w)$, and the total amount of claims incurred, $Y(w)$, defined as

$$Y(w) = \sum_{i=1}^{N(w)} Z_i \, I_{(N(w)>0)} \tag{1.1}$$

where the Z_i's are random variables representing the individual claims, $w = vt$ corresponds to the exposure, v denotes the value insured and t is the period during which the value v is exposed to the risk of claims.

Assuming that $N(w)$, $Z_1, Z_2, ...$ are independent and the individual claims are identically distributed, Jorgensen and Souza ([4]) discussed the estimation and inference problem concerning the parameters considering the situation in which the number of claims follows a Poisson process and the individual claims follow a gamma distribution.

Using the properties of the Tweedie family for exponential dispersion models ([8]; [3]), Jorgensen and Souza ([4]) determined, using the convolution formula, that $Y(w) \,|\, N(w)$ follows an exponential dispersion model and the joint distribution of $N(w)$ and $Y(w)/w$ follows a Tweedie compound Poisson distribution. For more details about exponential dispersion models read [2] and [3].

In spite of the distribution of the individual claim values being very well represented in some situations by the gamma distribution, in other cases it could be more suitable to attribute a lognormal distribution for $Z_1, Z_2, ...$. For instance, in collision situations in car insurances and in common fires, where the individual claim values can increase almost without limits but cannot fall below zero, with most of the values near the lower limit and where the natural logarithm of the individual claim variable yields a normal distribution.

The aim of this paper is to estimate the parameters of $Y(w) = \sum_{i=1}^{N(w)} Z_i I_{(N(w)>0)}$ and $N(w)$ distributions, where $N(w)$, $Z_1, Z_2, ...$ are independent, $Z_1, Z_2, ...$ is a sequence of random variables with lognormal(μ, σ^2) distribution and $N(w)$ follows a Poisson distribution with rate λ.

Simulated examples are given to illustrate the methodology. The use of a real dataset is not possible due to the high confidentiality with which the companies deal with their database.

2. LOGNORMAL MODEL

A positive random variable Z is lognormally distributed if the logarithm of the random variable is normally distributed. Hence Z follows a lognormal(μ, σ^2) distribution if its density function is given by

$$f_Z(z; \mu, \sigma^2) \;=\; \frac{(2\pi\,\sigma^2)^{-\frac{1}{2}}}{z}\,\exp\left\{-\frac{1}{2\,\sigma^2}\big(\log(z)-\mu\big)^2\right\}, \tag{2.1}$$

for $z > 0$, $-\infty < \mu < \infty$ and $\sigma > 0$.

The moments of the lognormal distribution can be calculated from the moment generating function of the normal distribution and are defined as

$$E[Z^k] \;=\; \exp\left(k\mu + \frac{1}{2}\,k^2\sigma^2\right). \tag{2.2}$$

Thus, the mean of the lognormal distribution is given by

$$E[Z] \;=\; \exp\left(\mu + \frac{1}{2}\,\sigma^2\right) \tag{2.3}$$

and the variance is given by

$$\mathrm{Var}[Z] \;=\; \exp\big(2\mu + 2\sigma^2\big) - \exp\big(2\mu + \sigma^2\big)\;. \tag{2.4}$$

Products and quotients of lognormally distributed variables are themselves lognormally distributed, as well as Z^b and bZ, for $b \neq 0$ and Z following a lognormal(μ, σ^2) distribution ([1]). However, the distribution of the sum of independent lognormally distributed variables, that appears in many practical problems and describes the distribution of $Y(w)|N(w)$, is not lognormally distributed and does not present a recognizable probability density function ([7]).

Approximations for the distribution of the sum of lognormally distributed random variables are suggested by Levy ([5]) and Milevsky and Posner ([6]).

3. PARAMETER ESTIMATION

As described in the previous section, the distribution function for $Y(w)$, where the claims Z_i are independently and identically lognormal(μ,σ^2) distributed, is not known. Consequently, the joint distribution for $(N(w), Y(w))$ and the corresponding likelihood function for the parameters μ, σ^2 and λ cannot be exactly defined.

However, since the lognormal distribution was defined with reference to the normal distribution, estimate μ, σ^2 and λ from the likelihood function for these parameters considering the variables $N(w)$ and $Y(w)$ is equivalent to estimate μ, σ^2 and λ from the likelihood function based on the variables $N(w)$ and

$$X_+(w) \;=\; \sum_{i=1}^{N(w)} X_i \, I_{(N(w)>0)} \;,$$

where $N(w)$ follows a Poisson$(w\lambda)$, $X_i = \log(Z_i)$ follows a Normal(μ,σ^2) and the Z_i's are independent identically lognormal(μ,σ^2).

Then, we have

$$X_+(w)|N(w) \;=\; n \;\sim\; \text{Normal}(n\mu, n\sigma^2), \qquad \text{for} \quad n \geq 1 \;. \tag{3.1}$$

The joint density of $X_+(w)$ and $N(w)$, for $n \geq 1$, is defined as

$$\begin{aligned} &f_{(X_+(w),N(w))}\big(x_+, n; \mu, \sigma^2, \lambda\big) \;= \\ &\quad=\; \frac{(w\lambda)^n}{n!\,\sqrt{2\pi n\,\sigma^2}}\, \exp\!\left\{-\frac{1}{2n\,\sigma^2}(x_+ - n\mu)^2 - w\lambda\right\} I_{(0,\infty)}(x_+) \end{aligned} \tag{3.2}$$

and

$$f_{(X_+(w),N(w))}\big(x_+, 0; \mu, \sigma^2, \lambda\big) \;=\; \exp\{-w\lambda\}\, I_{(0,\infty)}(x_+) \;. \tag{3.3}$$

In this work, without loss of generality, w is assumed to be equal to 1. Considering $(x_{+1}, n_1), (x_{+2}, n_2), ..., (x_{+m}, n_m)$ observations from the independent random vectors $(X_{+1}, N_1), (X_{+2}, N_2) ..., (X_{+m}, N_m)$, where $N_i \sim$ Poisson(λ), $X_{+i} \,|\, (N_i = n_i) \sim$ Normal$(n_i\mu, n_i\sigma^2)$, $i = 1, 2, ..., m$, and m is the number of groups present in the portfolio and considering $\delta_i = 0$ for $N_i = 0$ and $\delta_i = 1$ for $N_i > 0$, the log likelihood function for the parameters μ, σ^2 and λ is given by

$$\begin{aligned} &l(\mu, \sigma^2, \lambda) \;= \\ &= \sum_{i=1}^{m} \left\{ \delta_i \left(-\frac{1}{2} \log\big(2\pi n_i\,\sigma^2\big) + n_i \log(\lambda) - \frac{1}{2n_i\,\sigma^2}\big(x_{+i} - n_i\mu\big)^2 - \lambda \right) + (1-\delta_i)\,(-\lambda) \right\} \end{aligned} \tag{3.4}$$

If $\sigma^2 = \sigma_0^2$ is known the maximum likelihood estimates of μ and λ are given by

$$\hat{\mu} \;=\; \frac{\sum\limits_{i=1}^{m} \delta_i X_{+i}}{\sum\limits_{j=1}^{m} \delta_j N_j} \;=\; \frac{\sum\limits_{i=1}^{m} X_{+i}}{\sum\limits_{j=1}^{m} N_j} \qquad \text{if} \quad \sum_{j=1}^{m} N_j > 0 \;, \tag{3.5}$$

and

$$\hat{\lambda} \;=\; \frac{\sum\limits_{i=1}^{m} \delta_i N_i}{m} \;=\; \frac{\sum\limits_{i=1}^{m} N_i}{m} \;. \tag{3.6}$$

Let $S = \sum_{i=1}^{m} N_i$ be the total number of claims and $U = \sum_{i=1}^{m} X_{+i}$. Hence, S follows a Poisson$(m\lambda)$ and $U\,|\,(\mathbf{N}=\mathbf{n})$ follows a Normal$\Big(\mu \sum_{i=1}^{m} n_i,\ \sigma_0^2 \sum_{i=1}^{m} n_i\Big)$, where $\mathbf{N} = (N_1, N_2, ..., N_m)$ and $\mathbf{n} = (n_1, n_2, ..., n_m)$ is the observed vector of number of claims for m groups. Thus $U\,|\,(S=s)$ follows a Normal$(\mu s, \sigma_0^2 s)$ and the exact distribution of $\overset{\wedge}{\lambda}$ is given by

$$P\left(\overset{\wedge}{\lambda} = \frac{c}{m}\right) = P\big(S=c\big) = \frac{\exp(-m\lambda)\,(m\lambda)^c}{c!} \qquad \text{for} \quad c = 0, 1, 2, ... \,. \tag{3.7}$$

The cumulative distribution function of $\overset{\wedge}{\mu}$ given $S>0$, $F_{\overset{\wedge}{\mu}|S>0}(v)$, for $v \in R$ is

$$\begin{aligned}
P\left(\overset{\wedge}{\mu} \le v|S>0\right) &= P\left[\big(\overset{\wedge}{\mu} \le v\big) \cap \bigcup_{j=1}^{\infty} (S=j) | S>0\right] \\
&= \frac{P\left(\overset{\wedge}{\mu} \le v,\ \bigcup_{j=1}^{\infty} (S=j),\ S>0\right)}{P\big(S>0\big)} \\
&= \frac{P\left(\overset{\wedge}{\mu} \le v,\ \bigcup_{j=1}^{\infty} (S=j)\right)}{P\big(S>0\big)} \\
&= \frac{\sum_{j=1}^{\infty} P\Big(\overset{\wedge}{\mu} \le v,\ S=j\Big)}{P\big(S>0\big)} \\
&= \frac{\sum_{j=1}^{\infty} P\Big(\overset{\wedge}{\mu} \le v|S=j\Big)\, P\big(S=j\big)}{P\big(S>0\big)} \\
&= \frac{\sum_{j=1}^{\infty} P\Big(\frac{U}{S} \le v|S=j\Big)\, P\big(S=j\big)}{P\big(S>0\big)} \\
&= \frac{\sum_{j=1}^{\infty} P\Big(\frac{U}{j} \le v|S=j\Big)\, P\big(S=j\big)}{P\big(S>0\big)} \\
&= \frac{\sum_{j=1}^{\infty} P\Big(U \le jv|S=j\Big)\, P\big(S=j\big)}{P\big(S>0\big)} \\
&= \sum_{j=1}^{\infty} F_U(jv)\, \frac{\exp(-m\lambda)\,(m\lambda)^j}{j!\,\big(1-\exp(-m\lambda)\big)} \,,
\end{aligned} \tag{3.8}$$

where F_U is the cumulative distribution function of the Normal$(\mu j, \sigma_0^2 j)$ distribution, for $j = 1, 2, ...$.

The corresponding probability density function is defined as

$$
\begin{aligned}
f_{\overset{\wedge}{\mu}|S>0}(v) &= \frac{dF_{\overset{\wedge}{\mu}|S>0}(v)}{dv} \\
&= \sum_{j=1}^{\infty} f_U(jv)\, \frac{\exp(-m\lambda)\,(m\lambda)^j}{j!\,\bigl(1-\exp(-m\lambda)\bigr)}\; j \\
&= \frac{\exp(-m\lambda)}{1-\exp(-m\lambda)} \sum_{j=1}^{\infty} f_U(jv)\, \frac{(m\lambda)^j}{(j-1)!} \qquad (3.9) \\
&= \frac{\exp(-m\lambda)}{1-\exp(-m\lambda)} \sum_{r=0}^{\infty} f_U\bigl((r+1)v\bigr)\, \frac{(m\lambda)^{r+1}}{(r)!} \\
&= \frac{(m\lambda)\exp(-m\lambda)}{1-\exp(-m\lambda)} \sum_{r=0}^{\infty} f_U\bigl((r+1)v\bigr)\, \frac{(m\lambda)^r}{(r)!}\;,
\end{aligned}
$$

where f_U is the probability density function of the Normal$\bigl(\mu(r+1), \sigma_0^2(r+1)\bigr)$ distribution.

Let k be the number of groups with number of claims greater that zero. If σ^2 is unknown, the maximum likelihood estimate of σ^2 is

$$
\begin{aligned}
\overset{\wedge}{\sigma}{}^2 &= \frac{\sum_{i=1}^{m} \delta_i \left(\frac{\bigl(X_{+i} - N_i\,\overset{\wedge}{\mu}\bigr)^2}{N_i}\right)}{\sum_{i=1}^{m} \delta_i} \qquad (3.10) \\
&= \frac{\sum_{j=1}^{k} \left(\frac{\bigl(X_{+j} - N_j\,\overset{\wedge}{\mu}\bigr)^2}{N_j}\right)}{k} \qquad \text{if} \quad N_j > 0, \quad \text{for all } j = 1, 2, ..., k\;.
\end{aligned}
$$

Using the invariant principle of maximum-likelihood estimation, the estimates of $E[Z]$, $\mathrm{Var}[Z]$, $E[N]$ and $\mathrm{Var}[N]$, where Z represents the individual claims and N the number of claims, are, respectively

$$
\begin{aligned}
\widehat{E}[Z] &= \exp\left(\overset{\wedge}{\mu} + \frac{1}{2}\,\overset{\wedge}{\sigma}{}^2\right), \\
\widehat{\mathrm{Var}}[Z] &= \exp\left(2\,\overset{\wedge}{\mu} + 2\,\overset{\wedge}{\sigma}{}^2\right) - \exp\left(2\,\overset{\wedge}{\mu} + \overset{\wedge}{\sigma}{}^2\right), \\
\widehat{E}[N] &= \overset{\wedge}{\lambda}
\end{aligned}
$$

and

$$
\widehat{\mathrm{Var}}[N] = \overset{\wedge}{\lambda}\;.
$$

4. THE LOCATION PARAMETER μ AS A FUNCTION OF A COVARIATE

Suppose that $(x_{+1}, n_1), (x_{+2}, n_2), ..., (x_{+m}, n_m)$ are observations of the independent random vectors $(X_{+1}, N_1), (X_{+2}, N_2) ..., (X_{+m}, N_m)$, m is the number of groups present in the insurance portfolio, $N_i \sim \text{Poisson}(\lambda)$, and $X_{+i} \,|\, (N_i = n_i)$ $\sim \text{Normal}(\mu_i, n_i\sigma^2)$, $i = 1, 2, ..., m$, with the following regression structure for the location parameter

$$\mu_i \;=\; \alpha\, n_i + \beta \sum_{j=1}^{n_i} v_{ij} \;,$$

where v_{ij} represents the covariate of the j-th individual claims of the i-th group, for $i = 1, 2, ..., m$ and $j = 1, 2, ..., n_i$.

Defining $r_i = \sum\limits_{j=1}^{n_i} v_{ij}$, the log likelihood function for the parameters α, β, σ^2 and λ is given by

$$l\big(\alpha, \beta, \sigma^2, \lambda\big) \;=$$
$$= \sum_{i=1}^{m} \left\{ \delta_i \left(-\frac{1}{2} \log\big(2\pi n_i\, \sigma^2\big) + n_i \log(\lambda) - \frac{1}{2n_i\, \sigma^2} \big(x_{+i} - \alpha\, n_i - \beta\, r_i\big)^2 \right) + (-\lambda) \right\}.$$

Let k, $k \leq m$, be the number of groups with the number of claims greater that zero, so that $\sum\limits_{j=1}^{k} N_j > 0$. The maximum likelihood estimates of α, β, σ^2 are obtained through the data of only these k groups and are given by

$$\hat{\alpha} \;=\; \frac{\sum\limits_{j=1}^{k} X_{+j} \;-\; \hat{\beta} \sum\limits_{j=1}^{k} r_j}{\sum\limits_{j=1}^{k} N_j} \;, \tag{4.1}$$

$$\hat{\beta} \;=\; \frac{\sum\limits_{j=1}^{k} \frac{X_{+j}\, r_j}{N_j} \;-\; \frac{\sum\limits_{j=1}^{k} X_{+j} \sum\limits_{j=1}^{k} r_j}{\sum\limits_{j=1}^{k} N_j}}{\sum\limits_{j=1}^{k} \frac{r_j^2}{N_j} \;-\; \frac{\left(\sum\limits_{j=1}^{k} r_j\right)^2}{\sum\limits_{j=1}^{k} N_j}} \;, \tag{4.2}$$

$$\hat{\sigma}^2 \;=\; \frac{\sum\limits_{j=1}^{k} \left(\frac{\big(X_{+j} - \hat{\mu}_j\big)^2}{N_j} \right)}{k} \;, \qquad N_j > 0, \quad \text{for all } j \;, \tag{4.3}$$

where $\hat{\mu}_j = \hat{\alpha}\, n_j + \hat{\beta}\, r_j$.

The maximum likelihood estimates of λ is defined as (3.6).

5. APPLICATIONS

In order to illustrate the methods outlined in this article, two simulated data set, with 20 insurance groups each, are presented. For the i-th group we generated one observation of N following Poisson(λ) and n_i observations of Z following lognormal(μ, σ^2) and we considered $X_j = \log(Z_j)$ for $j = 1, 2, ..., N_i$ and $X_{+i} = \sum_{j=1}^{N_i} X_j$ for each group. These observations, together with the values of N, are used in the estimation of the parameters. The first data set was simulated considering a small rate of occurrence of claims in each insurance group and, consequently, a large probability of groups with zero claims. The second data set was simulated considering a large rate of occurrence of claims and, consequently, a small number of groups with zero claims. In both cases the values of μ and σ^2 considered in the simulation of the data was 7.1 and 0.1, respectively. Thus

$$E[Z] \;=\; 1274.11 \qquad \text{and} \qquad \mathrm{Var}[Z] = 170728.8 \;,$$

that is, the expected individual claim value is 1274.11 MU with a variance of 170728.8 MU.

5.1. Portfolio with small rate of occurrence of claim

Considering $m = 20$, $N \sim$ Poisson(2), and the Z_i's iid lognormal$(7.1, 0.1)$, we have

$$P\big[N=0\big] \;=\; \exp(-2) \;=\; 0.135 \;,$$

that is, the probability of occurrence of no claims in each group is equal to 0.135.

The simulated individual claim values vary between 634.48 MU and 2819.6 MU and the observed values of N, X_+ and δ are presented in Table 1. Note that four of the twenty groups have no occurrence of claims.

Table 1: Observed values of N, X_+ and δ for a simulated insurance portfolio

	N	X_+	δ		N	X_+	δ		N	X_+	δ
1	1	6.79	1	8	2	14.63	1	15	3	21.58	1
2	3	21.12	1	9	1	6.89	1	16	0	0.00	0
3	0	0.00	0	10	1	7.29	1	17	2	13.54	1
4	3	20.98	1	11	1	6.54	1	18	2	14.48	1
5	2	13.56	1	12	3	21.97	1	19	0	0.00	0
6	0	0.00	0	13	1	7.03	1	20	2	14.24	1
7	3	21.82	1	14	4	27.69	1	Total	34	240.17	16

The estimates of λ, μ and σ^2, calculated by (3.6), (3.5) and (3.10), respectively, as well as a comparison between the true values of the parameters and their estimates are presented in Table 2.

Table 2: The parameters true values and their estimates

	True value	Estimate	Difference
λ	2	1.7	0.3
μ	7.1	7.06	0.04
σ^2	0.1	0.09	0.01

From the distribution function of $\hat{\mu}$ given $S>0$, defined in (3.8), we can calculate $P\big(\hat{\mu} \leq v|S>0\big)$ for different values of $v \in R$. Table 3 shows the probability of $\hat{\mu} \leq v$ given $S>0$, considering $\lambda=2$, $\mu=7.1$, $\sigma^2=0.1$ (used for the data simulation) and $s=34$ (observed in this dataset).

Table 3: $P\big(\hat{\mu} \leq v|S>0\big)$ for $\lambda=2$, $\mu=7.1$, $\sigma^2=0.1$ and $s=34$

v	$P\big(\hat{\mu} \leq v\|S>0\big)$	v	$P\big(\hat{\mu} \leq v\|S>0\big)$	v	$P\big(\hat{\mu} \leq v\|S>0\big)$
4	0.0013	7	0.8063	10	0.9960
4.25	0.0060	7.25	0.8529	10.25	0.9974
4.5	0.0193	7.5	0.8913	10.5	0.9980
4.75	0.0484	7.75	0.9208	10.75	0.9986
5	0.0980	8	0.9424	11	0.9990
5.25	0.1721	8.25	0.9585	11.25	0.9993
5.5	0.2656	8.5	0.9709	11.5	0.9995
5.75	0.3687	8.75	0.9804	11.75	0.9996
6	0.4718	9	0.9863	12	0.9997
6.25	0.5787	9.25	0.9892	12.25	0.9998
6.5	0.6635	9.5	0.9925	2.5	0.9998
6.75	0.7473	9.75	0.9952	12.75	0.9999

Note that, from the results of Table 3,

$$P\big[4.5 \leq \hat{\mu} \leq 9.25\big] \;=\; 0.9699 \;.$$

5.2. Portfolio with large rate of occurrence of claim

In the second dataset, twenty observations of N were generated from the Poisson(100) distribution and the Z_i's were generated from the lognormal$(7.1, 0.1)$ distribution. Consequently,

$$P\big[N\!=0\big] \;=\; \exp\big(-100\big) \;\simeq\; 0 \;,$$

that is, the probability of occurrence of no claims in each group is practically null. The observed values of N, X_+ and δ are presented in Table 4 and the simulated individual claim values vary between 440.91 MU and 3212.9 MU.

Table 4: Observed values of N, X_+ and δ for a simulated insurance portfolio

	N	X_+	δ		N	X_+	δ		N	X_+	δ
1	95	676.0	1	8	86	606.6	1	15	79	553.4	1
2	104	739.3	1	9	108	762.0	1	16	95	674.5	1
3	85	601.6	1	10	85	601.1	1	17	87	619.8	1
4	92	652.6	1	11	98	695.8	1	18	105	747.6	1
5	106	749.6	1	12	86	612.0	1	19	101	717.3	1
6	111	791.9	1	13	83	586.1	1	20	100	714.6	1
7	85	600.8	1	14	100	709.8	1	Total	1891	13412.5	20

The estimates of λ, μ and σ^2, calculated by (3.6), (3.5) and (3.10), respectively, as well as a comparison between the true values of the parameters and its estimates are displayed in in Table 5.

Table 5: The parameters true values and their estimates

	True value	Estimate	Difference
λ	100	94.55	5.45
μ	7.1	7.093	0.007
σ^2	0.1	0.096	0.004

Table 6 shows the probability of $\overset{\wedge}{\mu} \le v$ given $S > 0$, considering $\lambda = 100$, $\mu = 7.1$, $\sigma^2 = 0.1$ (used for the data simulation) and $s = 1891$ (observed in this dataset).

Table 6: $P\big(\overset{\wedge}{\mu} \le v | S > 0\big)$ for $\lambda = 100$, $\mu = 7.1$, $\sigma^2 = 0.1$ and $s = 1891$

v	$P\big(\overset{\wedge}{\mu} \le v\|S>0\big)$	v	$P\big(\overset{\wedge}{\mu} \le v\|S>0\big)$	v	$P\big(\overset{\wedge}{\mu} \le v\|S>0\big)$
6.1	0.0000	6.6	0.2216	7.1	0.9929
6.2	0.0001	6.7	0.4639	7.2	0.9989
6.3	0.0019	6.8	0.7152	7.3	0.9999
6.4	0.0150	6.9	0.8873	7.4	1.0000
6.5	0.0722	7.0	0.9671	7.5	1.0000

From the results of Table 6 we have $P[6.4 \le \overset{\wedge}{\mu} \le 7.1] = 0.978$.

6. CONCLUDING REMARKS

The theory for exponential dispersion models cannot be applied to estimate the parameters μ, σ^2, that specify the lognormal distribution of the individual claim value (Z), and λ, the occurrence rate of claims, because the lognormal distribution and, consequently, the joint distribution of $Y(w) = \sum_{i=1}^{N(w)} Z_i \, I_{(N(w)>0)}$ and $N(w)$ does not belong to the class of the exponential dispersion model. However, from the joint distribution of $X_+(w) = \sum\limits_{i=1}^{N(w)} \log(Z_i) \, I_{(N(w)>0)}$ and $N(w)$, maximum likelihood estimates of μ, σ^2 and λ can be defined and applied to an insurance portfolio dataset, in which $N(w)$ follows a Poisson($w\lambda$) distribution and Z is lognormally distributed.

ACKNOWLEDGMENTS

We would like to thank the associate editor and referees for their valuable suggestions and comments on an earlier version of this paper.

REFERENCES

[1] Crow, E.L. and Shimizu, K. (Eds.) (1988). *Lognormal Distributions: Theory and Application*, Dekker, New York.

[2] Jorgensen, Bent (1986). Some properties of exponential dispersion models, *Scand. J. Statist.*, **13**, 187–198.

[3] Jorgensen, Bent (1987). Exponential dispersion models (with discussion), *J. Royal Statistical Society Ser. B*, **49**, 127–162.

[4] Jorgensen, B. and Souza, M.C.P. de (1994). Fitting Tweedie's compound poisson model to insurance claims data, *Scandinavian Actuarial Journal*, 69–93.

[5] Levy, E. (1992). Pricing European average rate currency options, *Journal of International Money and Finance*, **14**, 474–491.

[6] Milevsky, M.A. and Posner, S.E. (1998). Asian Options, the sum of lognormal, and the Reciprocal Gamma distribution, *Journal of Financial and Quantitative Analysis*, **33**, 3.

[7] Slimanc, S.B. (2001). Bounds to the distribution of a sum of independent lognormal random variables, *IEEE Transations on Communications*, **20**, 6.

[8] Tweedie, M.C.K. (1947). Functions of a statistical variate with given means, with special reference to Laplacian distributions, *Proc. Camb. Phil. Soc.*, **49**, 41–49.

REVSTAT – Statistical Journal
Volume 4, Number 2, June 2006, 143–151

A NEW DEPENDENCE CONDITION FOR TIME SERIES AND THE EXTREMAL INDEX OF HIGHER-ORDER MARKOV CHAINS

Authors: Helena Ferreira
– Department of Mathematics, University of Beira Interior,
6200 Covilhã, Portugal
ferreira@fenix2.ubi.pt

Received: June 2005 Revised: August 2005 Accepted: August 2005

Abstract:

- We present a new dependence condition for time series and extend the extremal types theorem.

  The dependence structure of a stationary sequence is described by a sequence of extremal functions. Under a stability condition for the sequence of extremal functions, we obtain the asymptotic distribution of the sample maximum.

  As a corollary, we derive a surprisingly simple method for computing the extremal index through a limit of a sequence of extremal coefficients.

  The results may be used to determine the asymptotic distribution of extreme values from stationary time series based on copulas. We illustrate it with the study of the extremal behaviour of d^{th}-order stationary Markov chains in discrete time with continuous state space. For such sequences we present a way to compute the extremal index from the upper extreme value limit for its joint distribution of $d+1$ consecutive variables.

Key-Words:

- *extremal coefficient; dependence; extremes; extremal index; higher-order stationary Markov sequences.*

1. INTRODUCTION

Let $\mathbf{X} = \{X_n\}_{n\geq 1}$ be a stationary sequence with common distribution function F in the domain of attraction of an extreme value distribution G. Therefore there exist real sequences $\mathbf{a} = \{a_n > 0\}_{n\geq 1}$ and $\mathbf{b} = \{b_n\}_{n\geq 1}$ such that

$$F^n\big(u_n(x)\big) \xrightarrow[n\to\infty]{} G(x) , \qquad x \in \mathbb{R} ,$$

where $u_n(x) = a_n x + b_n$.

Let $\{\varepsilon_n^{\mathbf{X}}(.)\}_{n\geq 1}$ be the sequence of functions satisfying

$$P\big(X_1 \leq y, ..., X_n \leq y\big) = F^{\varepsilon_n^{\mathbf{X}}(y)}(y) , \qquad y \in (\alpha_F, \omega_F), \quad n \geq 1 ,$$

where α_F and ω_F denote the left and right end points of F.

This sequence of extremal functions $\{\varepsilon_n^{\mathbf{X}}(.)\}_{n\geq 1}$ associated to $\mathbf{X}$ is inspired by the extremal coefficients considered in Buishand (1984), Tiago de Oliveira (1989) and Smith (1990), among others, to model the dependence of marginals of a multivariate extreme value distribution.

Here we will consider a stability condition for this sequence of extremal functions in order to obtain limiting results for the distribution of maxima $M_n = \max\{X_1, ..., X_n\}$ and the existence of the extremal index of $\mathbf{X}$.

We first point out some properties of $\varepsilon_n^{\mathbf{X}}(.)$ coming directly from the definition.

The Fréchet bounds for $F_n(y) = P(X_1 \leq y, ..., X_n \leq y)$, given by the inequalities $\max\big\{0,\, nF(y) - (n-1)\big\} \leq F_n(y) \leq F(y)$, enables the conclusion that $\varepsilon_n^{\mathbf{X}}(y) \geq 1,\ y \in \mathbb{R}$.

In particular, if $\mathbf{X}$ has a positive dependence structure (Joe (1997)) then

$$F^{\varepsilon_n^{\mathbf{X}}(y)}(y) \geq F^n(y) ,$$

and it would follow that $\varepsilon_n^{\mathbf{X}}(y) \leq n,\ y \in \mathbb{R}$.

Finally, if $(X_1, ..., X_n)$ has a multivariate extreme value distribution then the stability equation for its dependence function D_{F_n} (Deheuvels (1978), Hsing (1989)),

$$D^t_{F_n}(y_1, ..., y_n) = D_{F_n}(y_1^t, ..., y_n^t) ,$$

$t > 0,\ y_1, ..., y_n \in [0, 1]$, leads to $\varepsilon_n^{\mathbf{X}}(y) = \varepsilon_n^{\mathbf{X}},\ y \in \mathbb{R}$. Moreover, this constant $\varepsilon_n^{\mathbf{X}} \in [1, n]$ takes the extreme values 1 or n if and only if F_n has perfect positive dependence or independent marginals, respectively.

In this paper we will only assume that the sequences $\{\varepsilon_n^{\mathbf{X}}(u_n(x))\}_{n\geq 1}$, $x\in\mathbb{R}$, satisfy a stability condition introduced in section 2. Such condition is sufficient to conclude that if $F_n(u_n(x))$ converges to a non degenerate distribution G_* then G_* is in the class of max-stable distributions.

Moreover, we recall the definition of extremal index θ and prove that it can be computed from the limit of $\varepsilon_n^{\mathbf{X}}(u_n^{\tau_0})/n$, for some $\tau_0 > 0$, where $\{u_n^\tau\}_{n\geq 1}$ denotes a real sequence such that $n(1-F(u_n^\tau)) \xrightarrow[n\to\infty]{} \tau > 0$.

In section 3 we apply the results to Markov chains in discrete time with continuous state space. After the calculation of the extremal index of a Markov chain of order 1 based on a given bivariate dependence (copula) function, we demonstrate a sufficient condition for the existence of extremal index of a d^{th}-order Markov chain and compute its value. For such sequences, when the distribution of $d+1$ consecutive variables is in the domain of attraction of a $(d+1)$-multivariate extreme distribution H_{d+1}, it holds

$$\theta \;=\; -\ln D_{H_{d+1}}(e^{-1},...,e^{-1}) + \ln D_{H_d}(e^{-1},...,e^{-1})\ ,$$

where $D_{H_{d+1}}$, D_{H_d} denote the dependence functions of the multivariate distribution functions H_{d+1}, H_d, respectively, and

$$H_d(y_1,...,y_d) = H_{d+1}(y_1,...,y_d,+\infty)\ .$$

The notation introduced in this paragraph will be used throughout the paper.

2. STABLE EXTREMAL FUNCTIONS

We now introduce a stability condition for the sequence $\{\varepsilon_n^{\mathbf{X}}(.)\}_{n\geq 1}$ under which we can, asymptotically, relate the dependence measure $\varepsilon_n^{\mathbf{X}}(.)$ for $(X_1,...,X_n)$ to the analogous measure $\varepsilon_{[n/k]}^{\mathbf{X}}(.)$ for $\big(X_{(i-1)[n/k]+1},...,X_{i[n/k]}\big)$, $1\leq i\leq k$.

Definition. The sequence $\{\varepsilon_n^{\mathbf{X}}(.)\}_{n\geq 1}$ is stable over the real sequence $\{u_n\}_{n\geq 1}$ if, for each $k\geq 1$, it holds

$$\left|\varepsilon_n^{\mathbf{X}}(u_n) - k\,\varepsilon_{[n/k]}^{\mathbf{X}}(u_n)\right| \xrightarrow[n\to\infty]{} \varepsilon_k \geq 0\ . \tag{2.1}$$

We shall pursue the direction of this dependence condition and extend the extremal types theorem (Leadbetter *et al.* (1983)). Although the dependence between X_i and X_j does not necessarily fall off when $|i-j|$ increases, as occurs in the condition $D(u_n)$ of Leadbetter (1974), the condition (2.1) is still appropriate for the argument of extremes.

Proposition 2.1. *Let* $\mathbf{X} = \{X_n\}_{n\geq 1}$ *be a stationary sequence with common distribution function* F *and* $\mathbf{a} = \{a_n > 0\}_{n\geq 1}$, $\mathbf{b} = \{b_n\}_{n\geq 1}$ *real sequences such that* $F^n(u_n(x)) \xrightarrow[n\to\infty]{} G(x)$, $x \in \mathbb{R}$, *where* $u_n(x) = a_n x + b_n$ *and* G *is a non degenerate distribution function.*

If $F_n(u_n(x)) \xrightarrow[n\to\infty]{} G_*(x)$, $x \in \mathbb{R}$, *for some non degenerate distribution* G_* *and* $\{\varepsilon_n^{\mathbf{X}}(.)\}_{n\geq 1}$ *is stable over the real sequence* $\{u_n(x)\}_{n\geq 1}$, *for all* $x \in \mathbb{R}$, *then* G_* *is of extreme value type.*

Proof: Since every max-stable distribution is of extreme value type, it is sufficient to prove that there are real sequences $\{\alpha_n > 0\}_{n\geq 1}$ and $\{\beta_n\}_{n\geq 1}$ such that

$$G_*^n(\alpha_n x + \beta_n) = G_*(x) , \qquad n \geq 1 . \tag{2.2}$$

We follow essentially the proof of Theorem 1.3.1 of Leadbetter *et al.* (1983): if $F_n(u_{nk}(x)) \xrightarrow[n\to\infty]{} G_*^{1/k}(x)$, $x \in \mathbb{R}$, $k \geq 1$, then (2.2) holds. To obtain this last convergence we note that $F_{nk}(u_{nk}(x)) \xrightarrow[n\to\infty]{} G_*(x)$ and

$$\begin{aligned}
&\Big|F_{nk}\big(u_{nk}(x)\big) - F_n^k\big(u_{nk}(x)\big)\Big| \;=\\
&\qquad = \Big|F^{\varepsilon_{nk}^{\mathbf{X}}(u_{nk}(x))}\big(u_{nk}(x)\big) - F^{k\varepsilon_n^{\mathbf{X}}(u_{nk}(x))}\big(u_{nk}(x)\big)\Big|\\
&\qquad = F^{\varepsilon_{nk}^{\mathbf{X}}(u_{nk}(x))}\big(u_{nk}(x)\big)\Big|1 - F^{k\varepsilon_n^{\mathbf{X}}(u_{nk}(x)) - \varepsilon_{nk}^{\mathbf{X}}(u_{nk}(x))}\big(u_{nk}(x)\big)\Big| \;=\; o(1) ,
\end{aligned}$$

by applying (2.1). □

The proof points out that the convergence in (2.1) can be weakned. The result holds for bounded sequences $\big|\varepsilon_n^{\mathbf{X}}(u_n(x)) - k\,\varepsilon_{[n/k]}^{\mathbf{X}}(u_n(x))\big|$, $x \in \mathbb{R}$, $k \geq 1$.

As a corollary we provide a relation between the sequence of extremal coefficients $\{\varepsilon_n^{\mathbf{X}}(u_n^\tau)\}_{n\geq 1}$ and the extremal index θ of $\mathbf{X}$.

Specifically, $\mathbf{X}$ has extremal index θ (Leadbetter *et al.* (1983)) if, for each $\tau > 0$, there exists $\{u_n^\tau\}_{n\geq 1}$ such that $\lim\limits_{n\to+\infty} n(1 - F(u_n^\tau)) = \tau$ and $\lim\limits_{n\to+\infty} F_n(u_n^\tau) = e^{-\theta\tau}$. If θ exists then is given by

$$\theta \;=\; \frac{\ln \lim\limits_{n\to+\infty} F_n(u_n^\tau)}{\ln \lim\limits_{n\to+\infty} F^n(u_n^\tau)} \;.$$

Proposition 2.2. *Let* $\mathbf{X}$ *be a stationary sequence with common distribution function* F *such that, for each* $\tau > 0$, *there exists* $\{u_n^\tau\}_{n\geq 1}$ *satisfying* $n(1-F(u_n^\tau)) \xrightarrow[n\to\infty]{} \tau > 0$. *If, for each* $\tau > 0$, $\{\varepsilon_n^{\mathbf{X}}(.)\}_{n\geq 1}$ *is stable over* $\{u_n^\tau\}_{n\geq 1}$ *then:*

(**i**) *there are constants* θ' *and* θ'' *satisfying* $\liminf_{n\to+\infty} F_n(u_n^\tau) = e^{-\theta'\tau}$ *and* $\limsup_{n\to+\infty} F_n(u_n^\tau) = e^{-\theta''\tau}$, *for all* $\tau > 0$;

(ii) *the convergence of* $\{F_n(u_n^{\tau_0})\}_{n\geq 1}$*, for some* $\tau_0 > 0$*, implies* $\theta' = \theta''$ *and* $\lim_{n\to+\infty} F_n(u_n^\tau) = e^{-\theta\tau}$*, for all* $\tau > 0$. □

We omit the proof since it follows the same discussion used in Theorem 3.7.1 of Leadbetter *et al.* (1983) from the result

$$\left|F_n(u_n^\tau) - F^k_{[n/k]}(u_n^\tau)\right| = o(1) .$$

Since $\lim_{n\to+\infty} F_n(u_n^\tau) = \lim_{n\to+\infty} F^{\varepsilon_n^{\mathbf{X}}(u_n^\tau)}(u_n^\tau)$ and $\lim_{n\to+\infty} F^n(u_n^\tau) = e^{-\tau}$ the second statement of the above result can be rewritten as follows.

Corollary 2.1. *Let* $\mathbf{X}$ *be a stationary sequence with common distribution function* F *such that, for each* $\tau > 0$*, there exists* $\{u_n^\tau\}_{n\geq 1}$ *satisfying* $n(1-F(u_n^\tau)) \underset{n\to\infty}{\longrightarrow} \tau > 0$.

If, for each $\tau > 0$*,* $\{\varepsilon_n^{\mathbf{X}}(.)\}_{n\geq 1}$ *is stable over* $\{u_n^\tau\}_{n\geq 1}$ *then* $\mathbf{X}$ *has extremal index* θ *if and only if* $\theta = \lim_{n\to+\infty} \frac{\varepsilon_n^{\mathbf{X}}(u_n^{\tau_0})}{n}$*, for some* $\tau_0 > 0$. □

This surprisingly simple result presents a new method for computing the extremal index, through a limit of a sequence of extremal coefficients, and relates the extremal index with the dependence structure of $\mathbf{X}$.

3. CALCULATING THE EXTREMAL INDEX OF MARKOV CHAINS

The stationary Markov chains are important both from the applied and theoretical points of view and a sizeable literature on its extremal behaviour is available. There are stationary Markov sequences for which the condition $D(u_n^\tau)$ fails and, in general, it is not easy to show directly from the functional form of its distributions that $D(u_n^\tau)$ holds for each $\tau > 0$. O'Brien (1987) and Rootzen (1988) propose instead a general method by considering $\mathbf{X}$ as a measurable function of a Harris chain.

Since

$$\frac{\varepsilon_n^{\mathbf{X}}(u_n^\tau)}{n} = \frac{\ln D_{F_n}\big(F(u_n^\tau), ..., F(u_n^\tau)\big)}{\ln F^n(u_n^\tau)} ,$$

the above corollary seems to be suitable for the computation of θ in stationary sequences constructed from a given dependence function and a univariate margin.

We will apply the previous results to Markov models which can be defined from families of dependence functions. We start by illustrating the results with a Markov chain of order 1.

Example 3.1. Let $\mathbf{X} = \{X_n\}_{n\geq 1}$ be a stationary Markov chain of order 1 with common distribution function F such that, for each $\tau > 0$, there exists $\{u_n^\tau\}_{n\geq 1}$ satisfying $n(1-F(u_n^\tau)) \xrightarrow[n\to\infty]{} \tau > 0$.

Suppose that the dependence function D_{F_2} of (X_1, X_2) is defined (Kimeldorf and Sampson (1975)) by

$$D_{F_2}(u,v) \;=\; u+v-1+\Big((1-u)^{-1}+(1-v)^{-1}-1\Big)^{-1}\,, \qquad u,\, v \in [0,1]\ .$$

We get, for each $\tau > 0$,

$$\begin{aligned}
\varepsilon_n^{\mathbf{X}}(u_n^\tau) \;&=\; \frac{\ln D_{F_2}^{n-1}\big(F(u_n^\tau), F(u_n^\tau)\big) - \ln F^{n-2}(u_n^\tau)}{\ln F(u_n^\tau)} \\
&=\; \frac{\ln F^n(u_n^\tau) - \ln\Big(\frac{1+F(u_n^\tau)}{2}\Big)^{n-1}}{\ln F(u_n^\tau)}
\end{aligned}$$

and, for each $k \geq 1$,

$$\lim_{n\to+\infty}\Big|\varepsilon_{nk}^{\mathbf{X}}(u_{nk}^\tau) - k\,\varepsilon_n^{\mathbf{X}}(u_{nk}^\tau)\Big| \;=\; \lim_{n\to+\infty}\frac{(k-1)\,\ln\Big(\frac{1+F(u_{nk}^\tau)}{2}\Big)}{\ln F(u_{nk}^\tau)} \;=\; \frac{k-1}{2}\ .$$

Therefore, for each $\tau > 0$, $\{\varepsilon_n^{\mathbf{X}}(.)\}_{n\geq 1}$ is stable over $\{u_n^\tau\}_{n\geq 1}$ and

$$\theta \;=\; \lim_{n\to+\infty}\frac{\ln F^n(u_n^\tau) - \ln\Big(\frac{1+F(u_n^\tau)}{2}\Big)^{n-1}}{\ln F^n(u_n^\tau)} \;=\; \frac{1}{2}\ .$$

The following result is a contribution to compute θ for the special cases where the dependence struture of $\mathbf{X}$ is given. Smith (1992) and Perfekt (1994), among others, present a technique for calculating the extremal index of Markov chains under the assumption that a multivariate extreme limit distribution exists for the joint distribution of sucessive variables and suitable conditions on the transition probabilities.

We also assume here that the joint distribution of $d{+}1$ consecutive variables is in the domain of attraction of some multivariate extreme value distribution H_{d+1} and prove that this is sufficient for the stability condition to hold and compute θ from H_{d+1}.

Proposition 3.1. *Let* $\mathbf{X}$ *be a* d^{th} *order stationary Markov chain with the joint distribution* F_{d+1} *of* $d{+}1$ *sucessive variables in the domain of attraction of a* $(d{+}1)$*-multivariate extreme value distribution* H_{d+1}*. Then:*

(i) $\{\varepsilon_n^{\mathbf{X}}(.)\}_{n\geq 1}$ *is stable over* $\{u_n^\tau\}_{n\geq 1}$*, for each* $\tau > 0$*;*

(ii) $\mathbf{X}$ *has extremal index* $\theta = -\ln D_{H_{d+1}}(e^{-1}, ..., e^{-1}) + \ln D_{H_d}(e^{-1}, ..., e^{-1})$.

Proof: We first note that if F_{d+1} is in the domain of attraction of an extreme value distribution then the same holds for the common distribution F of variables in $\mathbf{X}$ and for each $\tau>0$ there exists $\{u_n^\tau\}_{n\geq 1}$ satisfying $n(1-F(u_n^\tau)) \xrightarrow[n\to\infty]{} \tau > 0$.

It follows from the Markov property (Joe, 1997) that

$$\lim_{n\to+\infty}\left|\varepsilon_{nk}^{\mathbf{X}}(u_{nk}^\tau) - k\,\varepsilon_n^{\mathbf{X}}(u_{nk}^\tau)\right| =$$

$$= \lim_{n\to+\infty}\left|\frac{\ln D_{F_{d+1}}^{(k-1)d}\big(F(u_{nk}^\tau),...,F(u_{nk}^\tau)\big) - \ln D_{F_d}^{(k-1)(d+1)}\big(F(u_{nk}^\tau),...,F(u_{nk}^\tau)\big)}{\ln F(u_{nk}^\tau)}\right|$$

$$= \lim_{n\to+\infty}(k-1)\left|\frac{d\ln D_{F_{d+1}}^{nk}\big(F(u_{nk}^\tau),...,F(u_{nk}^\tau)\big) - (d+1)\ln D_{F_d}^{nk}\big(F(u_{nk}^\tau),...,F(u_{nk}^\tau)\big)}{-\tau}\right| .$$

Since

$$D_{F_{d+1}}^{nk}\big(F(u_{nk}^\tau),...,F(u_{nk}^\tau)\big) = D_{F_{d+1}^{nk}}\big(F^{nk}(u_{nk}^\tau),...,F^{nk}(u_{nk}^\tau)\big)$$

converges to $D_{H_{d+1}}(e^{-\tau},...,e^{-\tau})$, we find

$$\begin{aligned}\varepsilon_k &= (k-1)\left|\frac{d\,\ln D_{H_{d+1}}(e^{-\tau},...,e^{-\tau}) - (d+1)\,\ln D_{H_d}(e^{-\tau},...,e^{-\tau})}{-\tau}\right| \\ &= (k-1)\big(-d\,\ln D_{H_{d+1}}(e^{-1},...,e^{-1})\big) + (d+1)\,\ln D_{H_d}(e^{-1},...,e^{-1}) .\end{aligned}$$

Then, by applying the corollary 2.1, we get

$$\begin{aligned}\theta &= \lim_{n\to+\infty}\frac{\ln D_{F_{d+1}}^{n-d}\big(F(u_n^\tau),...,F(u_n^\tau)\big) - \ln D_{F_d}^{n-d-1}\big(F(u_n^\tau),...,F(u_n^\tau)\big)}{\ln F^n(u_n^\tau)} \\ &= \lim_{n\to+\infty}\frac{\ln D_{F_{d+1}^n}\big(F^n(u_n^\tau),...,F^n(u_n^\tau)\big) - \ln D_{F_d^n}\big(F^n(u_n^\tau),...,F^n(u_n^\tau)\big)}{-\tau} \\ &= -\ln D_{H_{d+1}}(e^{-1},...,e^{-1}) + \ln D_{H_d}(e^{-1},...,e^{-1}) .\end{aligned}$$

□

One can easily construct examples to illustrate the result. We note instead that F_2 in the previous example defined by

$$\begin{aligned}F_2(x,y) &= D_{F_2}\big(F(x),F(y)\big) \\ &= F(x)+F(y)-1+\Big(\big(1-F(x)\big)^{-1}+\big(1-F(y)\big)^{-1}-1\Big)^{-1},\end{aligned}$$

$x,y\in\mathbb{R}$, is in the domain of attraction of

$$\begin{aligned}H_2(x,y) &= D_{H_2}\big(G(x),G(y)\big) \\ &= G(x)\,G(y)\,\exp\Big(\big(-\ln G(x)\big)^{-1}+\big(-\ln G(y)\big)^{-1}\Big)^{-1} ,\end{aligned}$$

where $G(x)=H_2(x,+\infty)=H_2(+\infty,x)$ (Joe (1997)).

Therefore, we can apply directly (ii) above and find $\theta=-\ln D_{H_2}(e^{-1},e^{-1})-1 = \frac{3}{2}-1=\frac{1}{2}$.

ACKNOWLEDGMENTS

I am grateful to the referee for his corrections and rigorous report.

REFERENCES

[1] Buishand, T.A. (1984). Bivariate extreme-value data and the station-year method, *J. Hydrol.*, **69**, 77–95.

[2] Deheuvels, P. (1978). Caractérisation complète des lois extrèmes multivariées et de la convergence aux types extrèmes, *Publ. Inst. Statist. Univ. Paris*, **23**.

[3] Hsing, T. (1989). Extreme value theory for multivariate stationary sequences, *J. Multivariate Anal.*, **29**, 274–291.

[4] Joe, H. (1997). *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.

[5] Kimeldorf, G. and Sampson, A.R. (1975). Uniform representations of bivariate distributions, *Comm. Statist.*, **4**, 617–627.

[6] Leadbetter, M.R. (1974). On extreme values in stationary sequences, *Z. Wahrsch. Verw. Gebiete*, **28**, 289–303.

[7] Leadbetter, M.R.; Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, Berlin.

[8] O'Brien, G.L. (1987). Extreme values for stationary and Markov sequences, *Ann. Probab.*, **15**, 281–291.

[9] Perfekt, R. (1994). Extremal behaviour of stationary Markov chains with applications, *Ann. Appl. Probab.*, **4**, 529–548.

[10] Rootzén, H. (1988). Maxima and exceedances of stationary Markov chains, *Adv. in Appl. Probab.*, **20**, 371–390.

[11] Smith, R.L. (1992). The extremal index for a Markov chain, *J. Appl. Probab.*, **29**, 37–45.

[12] Tiago de Oliveira, J. (1962/63). Structure theory of bivariate extremes, extensions, *Est. Mat., Estat. e Econ.*, **7**, 165–195.

REVSTAT – Statistical Journal
Volume 4, Number 2, June 2006, 153–161

DECOMPOSITIONS OF SYMMETRY MODEL INTO MARGINAL HOMOGENEITY AND DISTANCE SUBSYMMETRY IN SQUARE CONTINGENCY TABLES WITH ORDERED CATEGORIES

Authors: Sadao Tomizawa
– Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science, Noda City, Chiba, 278-8510, Japan
tomizawa@is.noda.tus.ac.jp

Nobuko Miyamoto
– Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science, Noda City, Chiba, 278-8510, Japan
miyamoto@is.noda.tus.ac.jp

Masami Ouchi
– Graduate School of Sciences and Technology, Tokyo University of Science, Noda City, Chiba, 278-8510, Japan

Received: July 2005 Revised: October 2005 Accepted: October 2005

Abstract:

- For square contingency tables with ordered categories, this paper proposes some distance subsymmetry models. The one model indicates that the cumulative probability that an observation will fall in row category i or below and column category $i+k$ ($k \geq 2$) or above, is equal to the probability that it falls in column category i or below and row category $i+k$ or above. This paper also gives the decomposition of the symmetry model into the marginal homogeneity model and some distance subsymmetry models. The father-son occupational mobility data in Britain and the women's unaided vision data in Britain are analyzed.

Key-Words:

- *decomposition; distance subsymmetry model; marginal homogeneity model; ordered category; square contingency table; symmetry model.*

AMS Subject Classification:

- 62H17.

1. INTRODUCTION

For an $r \times r$ square contingency table with ordered categories, let p_{ij} denote the probability that an observation will fall in the i^{th} row and j^{th} column of the table ($i = 1, 2, ..., r$; $j = 1, 2, ..., r$). The symmetry (S) model is defined by

$$p_{ij} = p_{ji} \qquad \text{for} \quad i = 1, 2, ..., r; \quad j = 1, 2, ..., r \ .$$

See Bishop, Fienberg, and Holland ([2], p. 282). This model indicates that the probability that an observation will fall in cell (i, j) of the table is equal to the probability that it falls in cell (j, i). Namely, this describes a structure of symmetry of the cell probabilities $\{p_{ij}\}$ with respect to the main diagonal of the table.

Let X_1 and X_2 denote the row and column variables, respectively. The marginal homogeneity (MH) model is defined by

$$\Pr(X_1 = i) \ = \ \Pr(X_2 = i) \qquad \text{for} \quad i = 1, 2, ..., r \ ,$$

namely

$$p_{i\cdot} = p_{\cdot i} \qquad \text{for} \quad i = 1, 2, ..., r \ ,$$

where $p_{i\cdot} = \sum_{t=1}^{r} p_{it}$ and $p_{\cdot i} = \sum_{s=1}^{r} p_{si}$ (Stuart, [8]). This indicates that the row marginal distribution is identical with the column marginal distribution.

Let

$$G_{ij} \ = \ \Pr\bigl(X_1 \leq i,\, X_2 \geq j\bigr) = \sum_{s=1}^{i} \sum_{t=j}^{r} p_{st} \qquad \text{for} \quad i < j \ ,$$

and

$$G^*_{ij} \ = \ \Pr\bigl(X_1 \geq i,\, X_2 \leq j\bigr) = \sum_{s=i}^{r} \sum_{t=1}^{j} p_{st} \qquad \text{for} \quad i > j \ .$$

Then the S model may be expressed as

$$G_{ij} = G^*_{ji} \qquad \text{for} \quad i < j \ . \tag{1.1}$$

The MH model may be expressed as

$$G_{i,i+1} = G^*_{i+1,i} \qquad \text{for} \quad i = 1, 2, ..., r-1 \ . \tag{1.2}$$

The S model implies the MH model. So, from (1.1) and (1.2), we are interested in decomposing (1.1) into (1.2) and the structure of

$$G_{ij} = G^*_{ji} \qquad \text{for} \quad j - i = 2, 3, ..., r-1; \quad i < j \ .$$

The purpose of this paper is to give the decompositions of the S model into some new models. The decompositions may be useful for seeing the reason for the poor fit when the S model fits the data poorly.

2. DECOMPOSITIONS OF SYMMETRY MODEL

This section proposes some new models based on $\{G_{ij}\}$ and based on $\{p_{ij}\}$, and gives the decompositions of the S model.

2.1. Distance Cumulative Subsymmetry Model

Consider a model defined by

$$G_{ij} = G^*_{ji} \qquad \text{for} \quad j-i = 2,3,...,r-1; \quad i<j \ , \tag{2.1}$$

which is equivalent to

$$p_{ij} = p_{ji} \qquad \text{for} \quad j-i = 2,3,...,r-1; \quad i<j \ .$$

This model indicates that the probability that an observation will fall in cell (i,j), which is one of cells such that the distance from the main diagonal is greater than or equal to 2, is equal to the probability that the observation falls in cell (j,i). We shall refer to (2.1) as the subsymmetry (SS) model.

Next, for fixed k ($k = 2,3,...,r-1$), consider a model defined by

$$G_{i,i+k} = G^*_{i+k,i} \qquad \text{for} \quad i = 1,2,...,r-k \ . \tag{2.2}$$

This model indicates that the cumulative probability that an observation will fall in row category i or below and column category $i+k$ or above, is equal to the cumulative probability that the observation falls in column category i or below and row category $i+k$ or above. We shall refer to (2.2) as the model of the distance cumulative subsymmetry with the difference k between the diagonal containing the cutpoint [i and $i+k$] and the main diagonal (denoted by the DCS-k model).

2.2. Distance Subsymmetry Model

For fixed k ($k = 1,2,...,r-1$), consider a model defined by

$$p_{ij} = p_{ji} \qquad \text{for} \quad j-i = k; \quad i<j \ . \tag{2.3}$$

This model indicates that the probability that an observation will fall in cell (i,j) with the distance k from the main diagonal, is equal to the probability that the observation falls in cell (j,i) with the same distance k. We shall refer to (2.3) as the distance subsymmetry with distance k (DS-k) model. We obtain the following theorem.

Theorem 2.1. *The following four statements are equivalent:*

(1) *the S model holds,*

(2) *the MH and SS models hold,*

(3) *the MH and {DCS-k} ($k = 2, 3, ..., r-1$) models hold,*

(4) *all the {DS-k} ($k = 1, 2, ..., r-1$) models hold.*

2.3. Goodness-of-Fit Test

Assume that a multinomial distribution is applied to the $r \times r$ table. The maximum likelihood estimates (MLEs) of expected frequencies under the S, SS and DS-k models are obtained in the closed-forms. The MLEs of them under the MH and DCS-k models could not be obtained in the closed-forms, however, they could be obtained using the Newton–Raphson methods in the log-likelihood equations.

The likelihood ratio statistic for testing the goodness-of-fit of the model is

$$G^2 = 2 \sum_{i=1}^{r} \sum_{j=1}^{r} n_{ij} \log\left(\frac{n_{ij}}{\hat{m}_{ij}}\right),$$

with the corresponding degrees of freedom (df), where n_{ij} is the observed frequency in cell (i, j), and $\hat{m}_{ij}$ is the MLE of expected frequency m_{ij} under the model. The numbers of df for the MH and SS models are $r-1$ and $(r-1)(r-2)/2$. Also the numbers of df for the DCS-k model ($k = 2, 3, ..., r-1$) are $r-k$, and those for the DS-k model ($k = 1, 2, ..., r-1$) are $r-k$.

3. EXAMPLES

We shall analyze the data in Tables 1 and 2.

3.1. Analysis of Table 1

Consider the data in Table 1, taken directly from Goodman [5]. These data relate the father's and his son's occupational status category in Britain. These data have been analyzed by some statisticians, including Agresti ([1], p. 206) and Tomizawa [9].

Table 1: The father's and son's occupational mobility data in Britain; from Goodman [5].

Father's status	**Son's status** (1)	(2)	(3)	(4)	(5)	**Total**
(1)	50	45	8	18	8	129
(2)	28	174	84	154	55	495
(3)	11	78	110	223	96	518
(4)	14	150	185	714	447	1510
(5)	3	42	72	320	411	848
Total	106	489	459	1429	1017	3500

Table 2: Unaided distance vision of 7477 women aged 30-39 employed in Royal Ordnance factories in Britain from 1943 to 1946; from Stuart [8].

Right eye grade	**Left eye grade** Best (1)	Second (2)	Third (3)	Worst (4)	**Total**
Best (1)	1520	266	124	66	1976
Second (2)	234	1512	432	78	2256
Third (3)	117	362	1772	205	2456
Worst (4)	36	82	179	492	789
Total	1907	2222	2507	841	7477

Table 3 presents the likelihood ratio chi-square values G^2 for the models applied to these data. The S model fits these data poorly, yielding $G^2 = 37.46$ with 10 df (Table 3). By using the decompositions of the S model, we shall consider the reason why the S model fits these data poorly.

The MH model fits these data poorly, however, the SS model fits these data well (Table 3). Therefore we can see from Theorem 2.1 that the poor fit of the S model is caused by the poor fit of the MH model (rather than the SS model).

Moreover, all the DCS-k ($k = 2, 3, 4$) models fit the data in Table 1 well. Therefore we can also see from Theorem 2.1 that the poor fit of the S model is caused by the poor fit of the MH model (rather than the DCS-k ($k = 2, 3, 4$) models). The DCS-k ($k = 2, 3, 4$) models provide that the probability that the occupational status category of the father in a pair is k or above higher than that of his son, is estimated to equal the probability that the status category of the son is k or above higher than that of his father.

Table 3: Likelihood ratio chi-square values for models applied to the data in Table 1.

Applied models	Degrees of freedom	Likelihood ratio chi-square
S	10	37.46*
MH	4	32.80*
SS	6	8.58
DCS-2	3	6.89
DCS-3	2	4.29
DCS-4	1	2.36
DS-1	4	28.89*
DS-2	3	3.97
DS-3	2	2.25
DS-4	1	2.36

* means significant at the 0.05 level.

In addition, the DS-k ($k=2,3,4$) models fit these data well, but the DS-1 model fits these data poorly. Therefore we can see from Theorem 2.1 that the poor fit of the S model is caused by the poor fit of the DS-1 model (rather than the DS-k ($k = 2, 3, 4$) models). The DS-k ($k = 2, 3, 4$) models provide that the probability that the occupational status of the father in a pair is k categories higher than that of his son, is estimated to equal the probability that the status of the son is k categories higher than that of his father.

3.2. Analysis of Table 2

Consider the data in Table 2, taken directly from Stuart [8]. These data are constructed from the unaided distance vision of 7477 women aged 30–39 employed in Royal Ordnance factories in Britain from 1943 to 1946. These data have been analyzed by many statisticians, including Caussinus [3], Bishop *et al.* ([2], p. 284), McCullagh [6], Goodman [4], Tomizawa [10], and Miyamoto, Ohtsuka and Tomizawa [7].

From Table 4, we see that the S model fits these data poorly, yielding $G^2 = 19.25$ with 6 df. By using the decompositions of the S model, we shall consider the reason why the S model fits these data poorly.

Both the MH and SS models, being the decomposed models of the S model, fit these data poorly. So, in order to analyze these data in more details, we shall apply Theorem 2.1. The DCS-2 model fits these data well, however, the DCS-3 model fits them poorly (Table 4). Therefore we can see from Theorem 2.1 that the poor fit of the S model is caused by the poor fits of the MH and DCS-3 models (rather than the DCS-2 model). The DCS-2 model provides that the probability that a woman's right eye is 2 or 3 grades better than her left eye is estimated to equal the probability that the woman's left eye is 2 or 3 grades better than her right eye.

Table 4: Likelihood ratio chi-square values for models applied to the data in Table 2.

Applied models	**Degrees of freedom**	**Likelihood ratio chi-square**
S	6	19.25*
MH	3	11.99*
SS	3	9.26*
DCS-2	2	5.00
DCS-3	1	8.96*
DS-1	3	9.99*
DS-2	2	0.30
DS-3	1	8.96*

* means significant at the 0.05 level.

The DS-2 model fits these data very well, however, the DS-1 and DS-3 models fit them poorly (Table 4). Therefore we can see from Theorem 2.1 that the poor fit of the S model is caused by the poor fits of the DS-1 and DS-3 models (rather than the DS-2 model). The DS-2 model provides that the probability that a woman's right eye is 2 grades better than her left eye is estimated to equal the probability that the woman's left eye is 2 grades better than her right eye.

Therefore, these indicate that there are the structures of subsymmetry (not the complete symmetry) in these data.

4. CONCLUDING REMARKS

Theorem 2.1 gives the decompositions of the S model into some distance subsymmetry models including the MH model. These decompositions would be useful for seeing which structures of distance subsymmetry are lacking when the S model does not hold for analyzing the data.

Finally we note that Caussinus [3] gave the decomposition of the S model into the quasi-symmetry model, which indicates the symmetry of odds-ratios, and the MH model. Caussinus's decomposition would be useful for seeing which of the structure of symmetry of odds-ratios and the structure of marginal homogeneity is lacking when the S model does not hold for analyzing the data (although Caussinus's decomposition could not see which structures of some distance subsymmetry are lacking).

ACKNOWLEDGMENTS

The authors thank the referee and the associate editor for their useful comments.

REFERENCES

[1] Agresti, A. (1984). *Analysis of Ordinal Categorical Data*, John Wiley & Sons, New York.

[2] Bishop, Y.M.M.; Fienberg, S.E. and Holland, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice*, The MIT Press, Cambridge, Massachusetts.

[3] Caussinus, H. (1965). Contribution à l'analyse statistique des tableaux de corrélation. *Annales de la Faculté des Sciences de l'Université de Toulouse*, **29**, 77–182.

[4] Goodman, L.A. (1979). Multiplicative models for square contingency tables with ordered categories, *Biometrika*, **66**, 413–418.

[5] Goodman, L.A. (1981). Association models and the bivariate normal for contingency tables with ordered categories, *Biometrika*, **68**, 347–355.

[6] McCullagh, P. (1978). A class of parametric models for the analysis of square contingency tables with ordered categories, *Biometrika*, **65**, 413–418.

[7] Miyamoto, N.; Ohtsuka, W. and Tomizawa, S. (2004). Linear diagonals-parameter symmetry and quasi-symmetry models for cumulative probabilities in square contingency tables with ordered categories, *Biometrical Journal*, **46**, 664–674.

[8] Stuart, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification, *Biometrika*, **42**, 412–416.

[9] Tomizawa, S. (1990). Another linear diagonals-parameter symmetry model for square contingency tables with ordered categories, *South African Statistical Journal*, **24**, 117–125.

[10] Tomizawa, S. (1993). Diagonals-parameter symmetry model for cumulative probabilities in square contingency tables with ordered categories, *Biometrics*, **49**, 883–887.

REVSTAT – Statistical Journal
Volume 4, Number 2, June 2006, 163–188

COMPARISON OF WEIBULL TAIL-COEFFICIENT ESTIMATORS

Authors: Laurent Gardes
– LabSAD, Université Grenoble 2, France
Laurent.Gardes@upmf-grenoble.fr

Stéphane Girard
– LMC-IMAG, Université Grenoble 1, France
Stephane.Girard@imag.fr

Received: September 2005 Revised: December 2005 Accepted: January 2006

Abstract:

- We address the problem of estimating the Weibull tail-coefficient which is the regular variation exponent of the inverse failure rate function. We propose a family of estimators of this coefficient and an associate extreme quantile estimator. Their asymptotic normality are established and their asymptotic mean-square errors are compared. The results are illustrated on some finite sample situations.

Key-Words:

- *Weibull tail-coefficient; extreme quantile; extreme value theory; asymptotic normality.*

AMS Subject Classification:

- 62G32, 62F12, 62G30.

1. INTRODUCTION

Let $X_1, X_2, ..., X_n$ be a sequence of independent and identically distributed random variables with cumulative distribution function F. We denote by $X_{1,n} \leq ... \leq X_{n,n}$ their associated order statistics. We address the problem of estimating the Weibull tail-coefficient $\theta > 0$ defined when the distribution tail satisfies

$$\textbf{(A.1)} \qquad 1 - F(x) = \exp\big(-H(x)\big) , \qquad H^{\leftarrow}(t) = \inf\big\{x,\ H(x) \geq t\big\} = t^{\theta}\ell(t) ,$$

where ℓ is a slowly varying function, *i.e.*,

$$\ell(\lambda x)/\ell(x) \to 1 \qquad \text{as} \quad x \to \infty \quad \text{for all} \ \ \lambda > 0 .$$

The inverse cumulative hazard function $H^{\leftarrow}$ is said to be regularly varying at infinity with index θ and this property is denoted by $H^{\leftarrow} \in \mathcal{R}_\theta$, see [7] for more details on this topic. As a comparison, Pareto type distributions satisfy $(1/(1-F))^{\leftarrow} \in \mathcal{R}_\gamma$, and $\gamma > 0$ is the so-called extreme value index. Weibull tail-distributions include for instance Gamma, Gaussian and, of course, Weibull distributions.

Let (k_n) be a sequence of integers such that $1 \leq k_n < n$ and (T_n) be a positive sequence. We examine the asymptotic behavior of the following family of estimators of θ:

$$\hat{\theta}_n = \frac{1}{T_n}\frac{1}{k_n}\sum_{i=1}^{k_n}\Big(\log(X_{n-i+1,n}) - \log(X_{n-k_n+1,n})\Big) . \tag{1.1}$$

Following the ideas of [10], an estimator of the extreme quantile x_{p_n} can be deduced from (1.1) by:

$$\hat{x}_{p_n} = X_{n-k_n+1,n}\left(\frac{\log(1/p_n)}{\log(n/k_n)}\right)^{\hat{\theta}_n} =: X_{n-k_n+1,n}\,\tau_n^{\hat{\theta}_n} . \tag{1.2}$$

Recall that an extreme quantile x_{p_n} of order p_n is defined by the equation

$$1 - F(x_{p_n}) = p_n , \qquad \text{with} \quad 0 < p_n < 1/n .$$

The condition $p_n < 1/n$ is very important in this context. It usually implies that x_{p_n} is larger than the maximum observation of the sample. This necessity to extrapolate sample results to areas where no data are observed occurs in reliability [8], hydrology [21], finance [9], ... We establish in Section 2 the asymptotic normality of $\hat{\theta}_n$ and $\hat{x}_{p_n}$. The asymptotic mean-square error of some particular members of (1.1) are compared in Section 3. In particular, it is shown that family (1.1) encompasses the estimator introduced in [12] and denoted by $\hat{\theta}_n^{(2)}$ in the sequel. In this paper, the asymptotic normality of $\hat{\theta}_n^{(2)}$ is obtained under weaker conditions. Furthermore, we show that other members of family (1.1) should be preferred in some typical situations. We also quote some other estimators of θ which do not belong to family (1.1): [4, 3, 6, 19]. We refer to [12] for a comparison with $\hat{\theta}_n^{(2)}$. The asymptotic results are illustrated in Section 4 on finite sample situations. Proofs are postponed to Section 5.

2. ASYMPTOTIC NORMALITY

To establish the asymptotic normality of $\hat{\theta}_n$, we need a second-order condition on ℓ:

(A.2) There exist $\rho \le 0$ and $b(x) \to 0$ such that uniformly locally on $\lambda \ge 1$

$$\log\left(\frac{\ell(\lambda x)}{\ell(x)}\right) \sim b(x)\, K_\rho(\lambda)\,, \qquad \text{when} \quad x \to \infty\,,$$

with $K_\rho(\lambda) = \int_1^\lambda u^{\rho-1} du$.

It can be shown [11] that necessarily $|b| \in \mathcal{R}_\rho$. The second order parameter $\rho \le 0$ tunes the rate of convergence of $\ell(\lambda x)/\ell(x)$ to 1. The closer ρ is to 0, the slower is the convergence. Condition (A.2) is the cornerstone in all proofs of asymptotic normality for extreme value estimators. It is used in [18, 17, 5] to prove the asymptotic normality of estimators of the extreme value index γ. In regular case, as noted in [13], one can choose $b(x) = x\,\ell'(x)/\ell(x)$ leading to

$$b(x) \;=\; \frac{x\, e^{-x}}{F^{-1}(1-e^{-x})\, f\big(F^{-1}(1-e^{-x})\big)} - \theta\,, \tag{2.1}$$

where f is the density function associated to F. Let us introduce the following functions: for $t > 0$ and $\rho \le 0$,

$$\mu_\rho(t) \;=\; \int_0^\infty K_\rho\Big(1+\frac{x}{t}\Big)\, \mathrm{e}^{-x}\, dx\,,$$

$$\sigma_\rho^2(t) \;=\; \int_0^\infty K_\rho^2\Big(1+\frac{x}{t}\Big)\, \mathrm{e}^{-x}\, dx \;-\; \mu_\rho^2(t)\,,$$

and let $a_n = \mu_0(\log(n/k_n))/T_n - 1$. As a preliminary result, we propose an asymptotic expansion of $(\hat{\theta}_n - \theta)$:

Proposition 2.1. *Suppose (A.1) and (A.2) hold. If $k_n \to \infty$, $k_n/n \to 0$, $T_n \log(n/k_n) \to 1$ and $k_n^{1/2} b(\log(n/k_n)) \to \lambda \in \mathbb{R}$ then,*

$$\begin{aligned} k_n^{1/2}(\hat{\theta}_n - \theta) \;&=\\ &= \theta\xi_{n,1} + \theta\mu_0\big(\log(n/k_n)\big)\xi_{n,2} + k_n^{1/2}\,\theta a_n + k_n^{1/2}\, b\big(\log(n/k_n)\big)\big(1 + o_{\mathrm{P}}(1)\big)\,, \end{aligned}$$

where $\xi_{n,1}$ and $\xi_{n,2}$ converge in distribution to a standard normal distribution.

Similar distributional representations exist for various estimators of the extreme value index γ. They are used in [16] to compare the asymptotic properties of several tail index estimators. In [15], a bootstrap selection of k_n is derived from such a representation. It is also possible to derive bias reduction method as in [14]. The asymptotic normality of $\hat{\theta}_n$ is a straightforward consequence of Proposition 2.1.

Theorem 2.1. *Suppose (A.1) and (A.2) hold. If* $k_n \to \infty$, $k_n/n \to 0$, $T_n \log(n/k_n) \to 1$ *and* $k_n^{1/2}\, b(\log(n/k_n)) \to \lambda \in \mathbb{R}$ *then,*

$$k_n^{1/2}\Big(\hat{\theta}_n - \theta - b\big(\log(n/k_n)\big) - \theta a_n\Big) \xrightarrow{d} \mathcal{N}(0, \theta^2) \ . \qquad \square$$

Theorem 2.1 implies that the Asymptotic Mean Square Error (AMSE) of $\hat{\theta}_n$ is given by:

$$AMSE(\hat{\theta}_n) \;=\; \Big(\theta a_n + b\big(\log(n/k_n)\big)\Big)^2 + \frac{\theta^2}{k_n} \ . \tag{2.2}$$

It appears that all estimators of family (1.1) share the same variance. The bias depends on two terms $b(\log(n/k_n))$ and θa_n. A good choice of T_n (depending on the function b) could lead to a sequence a_n cancelling the bias. Of course, in the general case, the function b is unknown making difficult the choice of a "universal" sequence T_n. This is discussed in the next section.

Clearly, the best rate of convergence in Theorem 2.1 is obtained by choosing $\lambda \neq 0$. In this case, the expression of the intermediate sequence (k_n) is known.

Proposition 2.2. *If* $k_n \to \infty$, $k_n/n \to 0$ *and* $k_n^{1/2}\, b(\log(n/k_n)) \to \lambda \neq 0$,

$$k_n \sim \left(\frac{\lambda}{b\big(\log(n)\big)}\right)^2 \;=\; \lambda^2 \big(\log(n)\big)^{-2\rho}\, L\big(\log(n)\big) \ ,$$

where L is a slowly varying function.

The "optimal" rate of convergence is thus of order $(\log(n))^{-\rho}$, which is entirely determined by the second order parameter ρ: small values of $|\rho|$ yield slow convergence. The asymptotic normality of the extreme quantile estimator (1.2) can be deduced from Theorem 2.1:

Theorem 2.2. *Suppose (A.1) and (A.2) hold. If moreover,* $k_n \to \infty$, $k_n/n \to 0$, $T_n \log(n/k_n) \to 1$, $k_n^{1/2} b(\log(n/k_n)) \to 0$ *and*

$$1 \;\leq\; \liminf \tau_n \;\leq\; \limsup \tau_n \;<\; \infty \tag{2.3}$$

then,

$$\frac{k_n^{1/2}}{\log \tau_n}\left(\frac{\hat{x}_{p_n}}{x_{p_n}} - \tau_n^{\theta a_n}\right) \xrightarrow{d} \mathcal{N}(0, \theta^2) \ .$$

3. COMPARISON OF SOME ESTIMATORS

First, we propose some choices of the sequence (T_n) leading to different estimators of the Weibull tail-coefficient. Their asymptotic distributions are provided, and their AMSE are compared.

3.1. Some examples of estimators

– The natural choice is clearly to take

$$T_n = T_n^{(1)} =: \mu_0\big(\log(n/k_n)\big) ,$$

in order to cancel the bias term a_n. This choice leads to a new estimator of θ defined by:

$$\hat{\theta}_n^{(1)} = \frac{1}{\mu_0\big(\log(n/k_n)\big)} \frac{1}{k_n} \sum_{i=1}^{k_n} \Big(\log(X_{n-i+1,n}) - \log(X_{n-k_n+1,n})\Big) .$$

Remarking that

$$\mu_\rho(t) = \mathrm{e}^t \int_1^\infty \mathrm{e}^{-tu}\, u^{\rho-1}\, du$$

provides a simple computation method for $\mu_0(\log(n/k_n))$ using the Exponential Integral (EI), see for instance [1], Chapter 5, pages 225–233.

– Girard [12] proposes the following estimator of the Weibull tail-coefficient:

$$\hat{\theta}_n^{(2)} = \sum_{i=1}^{k_n} \Big(\log(X_{n-i+1,n}) - \log(X_{n-k_n+1,n})\Big) \Big/ \sum_{i=1}^{k_n} \Big(\log_2(n/i) - \log_2(n/k_n)\Big) ,$$

where $\log_2(x) = \log(\log(x))$, $x > 1$. Here, we have

$$T_n = T_n^{(2)} =: \frac{1}{k_n} \sum_{i=1}^{k_n} \log\left(1 - \frac{\log(i/k_n)}{\log(n/k_n)}\right) .$$

It is interesting to remark that $T_n^{(2)}$ is a Riemann's sum approximation of $\mu_0(\log(n/k_n))$ since an integration by parts yields:

$$\mu_0(t) = \int_0^1 \log\left(1 - \frac{\log(x)}{t}\right) dx .$$

– Finally, choosing T_n as the asymptotic equivalent of $\mu_0(\log(n/k_n))$,

$$T_n = T_n^{(3)} =: 1/\log(n/k_n)$$

leads to the estimator:

$$\hat{\theta}_n^{(3)} = \frac{\log(n/k_n)}{k_n} \sum_{i=1}^{k_n} \Big(\log(X_{n-i+1,n}) - \log(X_{n-k_n+1,n})\Big) .$$

For $i = 1, 2, 3$, let us denote by $\hat{x}_{p_n}^{(i)}$ the extreme quantile estimator built on $\hat{\theta}_n^{(i)}$ by (1.2). Asymptotic normality of these estimators is derived from Theorem 2.1 and Theorem 2.2. To this end, we introduce the following conditions:

(C.1) $k_n/n \to 0$,

(C.2) $\log(k_n)/\log(n) \to 0$,

(C.3) $k_n/n \to 0$ and $k_n^{1/2}/\log(n/k_n) \to 0$.

Our result is the following:

Corollary 3.1. *Suppose (A.1) and (A.2) hold together with* $k_n \to \infty$ *and* $k_n^{1/2}\, b(\log(n/k_n)) \to 0$*. For* $i = 1, 2, 3$*:*

i) *If (C.i) hold then*

$$k_n^{1/2}\big(\hat{\theta}_n^{(i)} - \theta\big) \xrightarrow{d} \mathcal{N}(0, \theta^2) \;.$$

ii) *If (C.i) and (2.3) hold, then*

$$\frac{k_n^{1/2}}{\log \tau_n}\left(\frac{\hat{x}_{p_n}^{(i)}}{x_{p_n}} - 1\right) \xrightarrow{d} \mathcal{N}(0, \theta^2) \;.$$

In view of this corollary, the asymptotic normality of $\hat{\theta}_n^{(1)}$ is obtained under weaker conditions than $\hat{\theta}_n^{(2)}$ and $\hat{\theta}_n^{(3)}$, since (C.2) implies (C.1). Let us also highlight that the asymptotic distribution of $\hat{\theta}_n^{(2)}$ is obtained under less assumptions than in [12], Theorem 2, the condition $k_n^{1/2}/\log(n/k_n) \to 0$ being not necessary here. Finally, note that, if b is not ultimately zero, condition $k_n^{1/2} b(\log(n/k_n)) \to 0$ implies (C.2) (see Lemma 5.1).

3.2. Comparison of the AMSE of the estimators

We use the expression of the AMSE given in (2.2) to compare the estimators proposed previously.

Theorem 3.1. *Suppose (A.1) and (A.2) hold together with* $k_n \to \infty$, $\log(k_n)/\log(n) \to 0$ *and* $k_n^{1/2}\, b(\log(n/k_n)) \to \lambda \in \mathbb{R}$*. Several situations are possible:*

i) *b is ultimately non-positive. Introduce* $\alpha = -4 \lim\limits_{n\to\infty} b(\log n) \dfrac{k_n}{\log k_n} \in [0,+\infty]$*.*
If $\alpha > \theta$*, then, for n large enough,*

$$AMSE\big(\hat{\theta}_n^{(2)}\big) \;<\; AMSE\big(\hat{\theta}_n^{(1)}\big) \;<\; AMSE\big(\hat{\theta}_n^{(3)}\big) \;.$$

If $\alpha < \theta$*, then, for n large enough,*

$$AMSE\big(\hat{\theta}_n^{(1)}\big) \;<\; \min\Big(AMSE\big(\hat{\theta}_n^{(2)}\big), AMSE\big(\hat{\theta}_n^{(3)}\big)\Big) \;.$$

ii) *b is ultimately non-negative. Let us introduce $\beta = 2 \lim_{x\to\infty} xb(x) \in [0, +\infty]$. If $\beta > \theta$ then, for n large enough,*

$$AMSE\big(\hat{\theta}_n^{(3)}\big) \;<\; AMSE\big(\hat{\theta}_n^{(1)}\big) \;<\; AMSE\big(\hat{\theta}_n^{(2)}\big) \;.$$

If $\beta < \theta$ then, for n large enough,

$$AMSE\big(\hat{\theta}_n^{(1)}\big) \;<\; \min\Big(AMSE\big(\hat{\theta}_n^{(2)}\big), AMSE\big(\hat{\theta}_n^{(3)}\big)\Big) \;.$$

It appears that, when b is ultimately non-negative (case ii)), the conclusion does not depend on the sequence (k_n). The relative performances of the estimators is entirely determined by the nature of the distribution: $\hat{\theta}_n^{(1)}$ has the best behavior, in terms of AMSE, for distributions close to the Weibull distribution (small b and thus, small β). At the opposite, $\hat{\theta}_n^{(3)}$ should be preferred for distributions far from the Weibull distribution.

The case when b is ultimately non-positive (case i)) is different. The value of α depends on k_n, and thus, for any distribution, one can obtain $\alpha = 0$ by choosing small values of k_n(for instance $k_n = -1/b(\log n)$) as well as $\alpha = +\infty$ by choosing large values of k_n (for instance $k_n = (1/b(\log n))^2$ as in Proposition 2.2).

4. NUMERICAL EXPERIMENTS

4.1. Examples of Weibull tail-distributions

Let us give some examples of distributions satisfying (A.1) and (A.2).

Absolute Gaussian distribution: $|\mathcal{N}(\mu, \sigma^2)|$, $\sigma > 0$.

From [9], Table 3.4.4, we have $H^{\leftarrow}(x) = x^{\theta}\ell(x)$, where $\theta = 1/2$ and an asymptotic expansion of the slowly varying function is given by:

$$\ell(x) \;=\; 2^{1/2}\sigma - \frac{\sigma}{2^{3/2}}\,\frac{\log x}{x} + O(1/x) \;.$$

Therefore $\rho = -1$ and $b(x) = \log(x)/(4x) + O(1/x)$. b is ultimately positive, which corresponds to case ii) of Theorem 3.1 with $\beta = +\infty$. Therefore, one always has, for n large enough:

$$AMSE\big(\hat{\theta}_n^{(3)}\big) \;<\; AMSE\big(\hat{\theta}_n^{(1)}\big) \;<\; AMSE\big(\hat{\theta}_n^{(2)}\big) \;. \tag{4.1}$$

Gamma distribution: $\Gamma(a,\lambda),\ a,\lambda>0$.

We use the following parameterization of the density

$$f(x) \;=\; \frac{\lambda^a}{\Gamma(a)}\, x^{a-1}\, \exp\left(-\lambda x\right) \,.$$

From [9], Table 3.4.4, we obtain $H^{\leftarrow}(x)=x^\theta \ell(x)$ with $\theta=1$ and

$$\ell(x) \;=\; \frac{1}{\lambda} + \frac{a-1}{\lambda}\,\frac{\log x}{x} + O(1/x) \,.$$

We thus have $\rho=-1$ and $b(x)=(1-a)\log(x)/x+O(1/x)$. If $a>1$, b is ultimately negative, corresponding to case i) of Theorem 3.1. The conclusion depends on the value of k_n as explained in the preceding section. If $a<1$, b is ultimately positive, corresponding to case ii) of Theorem 3.1 with $\beta=+\infty$. Therefore, we are in situation (4.1).

Weibull distribution: $\mathcal{W}(a,\lambda),\ a,\lambda>0$.

The inverse failure rate function is $H^{\leftarrow}(x)=\lambda x^{1/a}$, and then $\theta=1/a$, $\ell(x)=\lambda$ for all $x>0$. Therefore $b(x)=0$ and we use the usual convention $\rho=-\infty$. One may apply either i) or ii) of Theorem 3.1 with $\alpha=\beta=0$ to get for n large enough,

$$AMSE\big(\hat\theta_n^{(1)}\big) \;<\; \min\Big(AMSE\big(\hat\theta_n^{(2)}\big), AMSE\big(\hat\theta_n^{(3)}\big)\Big) \,. \tag{4.2}$$

4.2. Numerical results

The finite sample performance of the estimators $\hat\theta_n^{(1)}$, $\hat\theta_n^{(2)}$ and $\hat\theta_n^{(3)}$ are investigated on 5 different distributions: $\Gamma(0.5,1)$, $\Gamma(1.5,1)$, $|\mathcal{N}(0,1)|$, $\mathcal{W}(2.5,2.5)$ and $\mathcal{W}(0.4,0.4)$. In each case, $N=200$ samples $(\mathcal{X}_{n,i})_{i=1,\ldots,N}$ of size $n=500$ were simulated. On each sample $(\mathcal{X}_{n,i})$, the estimates $\hat\theta_{n,i}^{(1)}(k)$, $\hat\theta_{n,i}^{(2)}(k)$ and $\hat\theta_{n,i}^{(3)}(k)$ are computed for $k=2,\ldots,150$. Finally, the associated Mean Square Error (MSE) plots are built by plotting the points

$$\left(k,\ \frac{1}{N}\sum_{i=1}^{N}\Big(\hat\theta_{n,i}^{(j)}(k)-\theta\Big)^2\right), \qquad j=1,2,3 \,.$$

They are compared to the AMSE plots (see (2.2) for the definition of the AMSE):

$$\left(k,\ \Big(\theta a_n^{(j)} + b\big(\log(n/k)\big)\Big)^2 + \frac{\theta^2}{k}\right), \qquad j=1,2,3 \,,$$

and where b is given by (2.1). It appears on Figure 1 – Figure 5 that, for all the above mentioned distributions, the MSE and AMSE have a similar qualitative behavior. Figure 1 and Figure 2 illustrate situation (4.1) corresponding to ultimately positive bias functions. The case of an ultimately negative bias function is presented on Figure 3 with the $\Gamma(1.5,1)$ distribution. It clearly appears that the MSE associated to $\hat{\theta}_n^{(3)}$ is the largest. For small values of k, one has $MSE(\hat{\theta}_n^{(1)}) < MSE(\hat{\theta}_n^{(2)})$ and $MSE(\hat{\theta}_n^{(1)}) > MSE(\hat{\theta}_n^{(2)})$ for large value of k. This phenomenon is the illustration of the asymptotic result presented in Theorem 3.1i). Finally, Figure 4 and Figure 5 illustrate situation (4.2) of asymptotically null bias functions. Note that, the MSE of $\hat{\theta}_n^{(1)}$ and $\hat{\theta}_n^{(2)}$ are very similar. As a conclusion, it appears that, in all situations, $\hat{\theta}_n^{(1)}$ and $\hat{\theta}_n^{(2)}$ share a similar behavior, with a small advantage to $\hat{\theta}_n^{(1)}$. They provide good results for null and negative bias functions. At the opposite, $\hat{\theta}_n^{(3)}$ should be preferred for positive bias functions.

5. PROOFS

For the sake of simplicity, in the following, we note k for k_n. We first give some preliminary lemmas. Their proofs are postponed to the appendix.

5.1. Preliminary lemmas

We first quote a technical lemma.

Lemma 5.1. *Suppose that b is ultimately non-zero. If $k \to \infty$, $k/n \to 0$ and $k^{1/2}\, b(\log(n/k)) \to \lambda \in \mathbb{R}$, then $\log(k)/\log(n) \to 0$.*

The following two lemmas are of analytical nature. They provide first-order expansions which will reveal useful in the sequel.

Lemma 5.2. *For all $\rho \le 0$ and $q \in \mathbb{N}^*$, we have*

$$\int_0^\infty K_\rho^q\left(1+\frac{x}{t}\right)\mathrm{e}^{-x}\,dx \sim \frac{q!}{t^q} \qquad \text{as} \quad t \to \infty\ .$$

Let $a_n^{(i)} = \mu_0(\log(n/k_n))/T_n^{(i)} - 1$, for $i = 1,2,3$.

Lemma 5.3. *Suppose $k \to \infty$ and $k/n \to 0$.*

i) $T_n^{(1)} \log(n/k) \to 1$ *and* $a_n^{(1)} = 0$.

ii) $T_n^{(2)} \log(n/k) \to 1$. *If moreover* $\log(k)/\log(n) \to 0$ *then* $a_n^{(2)} \sim \log(k)/(2k)$.

iii) $T_n^{(3)} \log(n/k) = 1$ *and* $a_n^{(3)} \sim -1/\log(n/k)$.

The next lemma presents an expansion of $\hat{\theta}_n$.

Lemma 5.4. *Suppose $k \to \infty$ and $k/n \to 0$. Under (A.1) and (A.2), the following expansions hold:*

$$\hat{\theta}_n = \frac{1}{T_n}\Big(\theta U_n^{(0)} + b\big(\log(n/k)\big)\, U_n^{(\rho)}\big(1+o_{\mathrm{P}}(1)\big)\Big) ,$$

where

$$U_n^{(\rho)} = \frac{1}{k}\sum_{i=1}^{k-1} K_\rho\bigg(1+\frac{F_i}{E_{n-k+1,n}}\bigg) , \qquad \rho \le 0$$

and where $E_{n-k+1,n}$ is the $(n-k+1)$-th order statistics associated to n independent standard exponential variables and $\{F_1, ..., F_{k-1}\}$ are independent standard exponential variables and independent from $E_{n-k+1,n}$.

The next two lemmas provide the key results for establishing the asymptotic distribution of $\hat{\theta}_n$. Their describe they asymptotic behavior of the random terms appearing in Lemma 5.4.

Lemma 5.5. *Suppose $k \to \infty$ and $k/n \to 0$. Then, for all $\rho \le 0$,*

$$\mu_\rho(E_{n-k+1,n}) \overset{P}{\sim} \sigma_\rho(E_{n-k+1,n}) \overset{P}{\sim} \frac{1}{\log(n/k)} .$$

Lemma 5.6. *Suppose $k \to \infty$ and $k/n \to 0$. Then, for all $\rho \le 0$,*

$$\frac{k^{1/2}}{\sigma_\rho(E_{n-k+1,n})}\Big(U_n^{(\rho)} - \mu_\rho(E_{n-k+1,n})\Big) \xrightarrow{d} \mathcal{N}(0,1) .$$

5.2. Proofs of the main results

Proof of Proposition 2.1: Lemma 5.6 states that for $\rho \le 0$,

$$\frac{k^{1/2}}{\sigma_\rho(E_{n-k+1,n})}\Big(U_n^{(\rho)} - \mu_\rho(E_{n-k+1,n})\Big) = \xi_n(\rho) ,$$

where $\xi_n(\rho) \xrightarrow{d} \mathcal{N}(0,1)$ for $\rho \le 0$. Then, by Lemma 5.4

$$\begin{aligned}
k^{1/2}(\hat{\theta}_n - \theta) &= \\
&= \theta\,\frac{\sigma_0(E_{n-k+1,n})}{T_n}\,\xi_n(0) + k^{1/2}\theta\bigg(\frac{\mu_0(E_{n-k+1,n})}{T_n} - 1\bigg) \\
&\quad + k^{1/2}\, b\big(\log(n/k)\big)\bigg(\frac{\sigma_\rho(E_{n-k+1,n})}{T_n}\,\frac{\xi_n(\rho)}{k^{1/2}} + \frac{\mu_\rho(E_{n-k+1,n})}{T_n}\bigg)\big(1+o_{\mathrm{P}}(1)\big) .
\end{aligned}$$

Since $T_n \sim 1/\log(n/k)$ and from Lemma 5.5, we have

$$
\begin{aligned}
&k^{1/2}(\hat{\theta}_n - \theta) = \\
&\quad = \theta\xi_{n,1} + k^{1/2}\,\theta\left(\frac{\mu_0(E_{n-k+1,n})}{T_n} - 1\right) + k^{1/2}\,b\big(\log(n/k)\big)\big(1 + o_{\mathrm{P}}(1)\big)\ ,
\end{aligned}
\tag{5.1}
$$

where $\xi_{n,1} \xrightarrow{d} \mathcal{N}(0,1)$. Moreover, a first-order expansion of μ_0 yields

$$
\frac{\mu_0(E_{n-k+1,n})}{\mu_0\big(\log(n/k)\big)} = 1 + \Big(E_{n-k+1,n} - \log(n/k)\Big)\,\frac{\mu_0^{(1)}(\eta_n)}{\mu_0\big(\log(n/k)\big)}\ ,
$$

where $\eta_n \in \big]\min(E_{n-k+1,n}, \log(n/k)),\, \max(E_{n-k+1,n}, \log(n/k))\big[$ and

$$
\mu_0^{(1)}(t) = \frac{d}{dt}\int_0^\infty \log\Big(1 + \frac{x}{t}\Big)\mathrm{e}^{-x}\,dx =: \frac{d}{dt}\int_0^\infty f(x,t)\,dx\ .
$$

Since for $t \ge T > 0$, $f(.,t)$ is integrable, continuous and

$$
\left|\frac{\partial f(x,t)}{\partial t}\right| = \frac{x}{t^2}\Big(1 + \frac{x}{t}\Big)^{-1}\mathrm{e}^{-x} \le x\,\frac{\mathrm{e}^{-x}}{T^2}\ ,
$$

we have that

$$
\mu_0^{(1)}(t) = -\int_0^\infty \frac{x}{t^2}\Big(1 + \frac{x}{t}\Big)^{-1}\mathrm{e}^{-x}\,dx\ .
$$

Then, Lebesgue Theorem implies that $\mu_0^{(1)}(t) \sim -1/t^2$ as $t \to \infty$. Therefore, $\mu_0^{(1)}$ is regularly varying at infinity and thus

$$
\frac{\mu_0^{(1)}(\eta_n)}{\mu_0\big(\log(n/k)\big)} \overset{P}{\sim} \frac{\mu_0^{(1)}\big(\log(n/k)\big)}{\mu_0\big(\log(n/k)\big)} \sim -\frac{1}{\log(n/k)}\ .
$$

Since $k^{1/2}(E_{n-k+1,n} - \log(n/k)) \xrightarrow{d} \mathcal{N}(0,1)$ (see [12], Lemma 1), we have

$$
\frac{\mu_0(E_{n-k+1,n})}{\mu_0\big(\log(n/k)\big)} = 1 - \frac{k^{-1/2}}{\log(n/k)}\,\xi_{n,2}\ ,
\tag{5.2}
$$

where $\xi_{n,2} \xrightarrow{d} \mathcal{N}(0,1)$. Collecting (5.1), (5.2) and taking into account that $T_n \log(n/k) \to 1$ concludes the proof. □

Proof of Proposition 2.2: Lemma 5.1 entails $\log(n/k) \sim \log(n)$. Since $|b|$ is a regularly varying function, $b(\log(n/k)) \sim b(\log(n))$ and thus, $k^{1/2} \sim \lambda/b(\log(n))$. □

Proof of Theorem 2.2: The asymptotic normality of $\hat{x}_{p_n}$ can be deduced from the asymptotic normality of $\hat{\theta}_n$ using Theorem 2.3 of [10]. We are in the situation, denoted by (S.2) in the above mentioned paper, where the limit distribution of $\hat{x}_{p_n}/x_{p_n}$ is driven by $\hat{\theta}_n$. Following, the notations of [10], we denote by $\alpha_n = k_n^{1/2}$ the asymptotic rate of convergence of $\hat\theta_n$, by $\beta_n = \theta a_n$ its asymptotic bias, and by $\mathcal{L} = \mathcal{N}(0, \theta^2)$ its asymptotic distribution. It suffices to verify that

$$\log(\tau_n) \log(n/k) \to \infty \; . \tag{5.3}$$

To this end, note that conditions (2.3) and $p_n < 1/n$ imply that there exists $0 < c < 1$ such that

$$\log(\tau_n) \;>\; c(\tau_n - 1) \;>\; c\left(\frac{\log(n)}{\log(n/k)} - 1\right) \;=\; c\,\frac{\log(k)}{\log(n/k)} \;,$$

which proves (5.3). We thus have

$$\frac{k^{1/2}}{\log \tau_n}\, \tau_n^{-\theta a_n} \left(\frac{\hat x_{p_n}}{x_{p_n}} - \tau_n^{\theta a_n}\right) \xrightarrow{d} \mathcal{N}(0, \theta^2) \; .$$

Now, remarking that, from Lemma 5.2, $\mu_0(\log(n/k)) \sim 1/\log(n/k) \sim T_n$, and thus $a_n \to 0$ gives the result. $\square$

Proof of Corollary 3.1: Lemma 5.3 shows that the assumptions of Theorem 2.1 and Theorem 2.2 are verified and that, for $i = 1, 2, 3$, $k^{1/2} a_n^{(i)} \to 0$. $\square$

Proof of Theorem 3.1:

i) First, from (2.2) and Lemma 5.3 iii), since b is ultimately non-positive,

$$AMSE\big(\hat\theta_n^{(1)}\big) - AMSE\big(\hat\theta_n^{(3)}\big) \;=\; -\theta\big(a_n^{(3)}\big)^2 \left(\theta + 2\,\frac{b\big(\log(n/k)\big)}{a_n^{(3)}}\right) \;<\; 0 \; . \tag{5.4}$$

Second, from (2.2),

$$AMSE\big(\hat\theta_n^{(2)}\big) - AMSE\big(\hat\theta_n^{(1)}\big) \;=\; \theta\big(a_n^{(2)}\big)^2 \left(\theta + 2\,\frac{b\big(\log(n/k)\big)}{a_n^{(2)}}\right) \; . \tag{5.5}$$

If b is ultimately non-zero, Lemma 5.1 entails that $\log(n/k) \sim \log(n)$ and consequently, since $|b|$ is regularly varying, $b(\log(n/k)) \sim b(\log(n))$. Thus, from Lemma 5.3 ii),

$$2\,\frac{b\big(\log(n/k)\big)}{a_n^{(2)}} \;\sim\; 4\, b(\log n)\, \frac{k}{\log(k)} \;\to\; -\alpha \; . \tag{5.6}$$

Collecting (5.4)–(5.6) concludes the proof of i).

ii) First, (5.5) and Lemma 5.3 ii) yields

$$AMSE\big(\hat\theta_n^{(2)}\big) - AMSE\big(\hat\theta_n^{(1)}\big) \;>\; 0 \;, \tag{5.7}$$

since b is ultimately non-negative. Second, if b is ultimately non-zero, Lemma 5.1 entails that $\log(n/k) \sim \log(n)$ and consequently, since $|b|$ is regularly varying, $b(\log(n/k)) \sim b(\log(n))$. Thus, observe that in (5.4),

$$2\,\frac{b\big(\log(n/k)\big)}{a_n^{(3)}} \;\sim\; -2\,b(\log n)(\log n) \;\to\; -\beta\;. \tag{5.8}$$

Collecting (5.4), (5.7) and (5.8) concludes the proof of ii). The case when b is ultimately zero is obtained either by considering $\alpha = 0$ in (5.6), or $\beta = 0$ in (5.8). □

REFERENCES

[1] Abramowitz, M. and Stegan, J.A. (1972). *Exponential integral and related functions.* In "Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables", 9^{th} printing, New-York, Bover.

[2] Arnold, B.C.; Balakrishnan, N. and Nagaraja, H.N. (1992). *A First Course in Order Statistics*, Wiley and Sons.

[3] Beirlant, J.; Bouquiaux, C. and Werker, B.. (2006). Semiparametric lower bounds for tail index estimation, *Journal of Statistical Planning and Inference*, **136**, 705–729.

[4] 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.

[5] Beirlant, J.; Dierckx, G.; Goegebeur, Y. and Matthys, G. (1999). Tail index estimation and an exponential regression model, *Extremes*, **2**, 177–200.

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

[7] Bingham, N.H.; Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*, Cambridge University Press.

[8] Ditlevsen, O. (1994). *Distribution arbitrariness in structural reliability.* In "Structural Safety and Reliability", Balkema, Rotterdam, pp. 1241–1247.

[9] Embrechts, P.; Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events*, Springer.

[10] Gardes, L. and Girard, S. (2005). Estimating extreme quantiles of Weibull tail-distributions, *Communication in Statistics – Theory and Methods*, **34**, 1065–1080.

[11] Geluk, J.L. and de Haan, L. (1987). *Regular variation, extensions and Tauberian theorems.* In "Math Centre Tracts", **40**, Centre for Mathematics and Computer Science, Amsterdam.

[12] Girard, S. (2004). A Hill type estimate of the Weibull tail-coefficient, *Communication in Statistics – Theory and Methods*, **33**(2), 205–234.

[13] Gomes, I.; de Haan, L. and Peng, L. (2003). Semi-parametric estimation of the second order parameter in statistics of extremes, *Extremes*, **5**(4), 387–414.

[14] Gomes, I. and Caeiro, F. (2002). A class of asymptotically unbiased semi-parametric estimators of the tail index, *Test*, **11**(2), 345–364.

[15] Gomes, I. and Oliveira, O. (2001). The bootstrap methodology in statistics of extremes. Choice of the optimal sample fraction, *Extremes*, **4**(4), 331–358.

[16] de Haan, L. and Peng, L. (1998). Comparison of tail index estimators, *Statistica Neerlandica*, **52**(1), 60–70.

[17] Häusler, E. and Teugels, J.L. (1985). On asymptotic normality of Hill's estimator for the exponent of regular variation, *The Annals of Statistics*, **13**, 743–756.

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

[19] Klüppelberg, C. and Villaseñor, J.A. (1993). Estimation of distribution tails – a semiparametric approach, *Deutschen Gesellschaft für Versicherungsmathematik*, **XXI**(2), 213–235.

[20] Petrov, V.V. (1975). *Sums of Independent Random Variables*, Springer-Verlag, Berlin, Eidelberg, New York.

[21] Smith, J. (1991). Estimating the upper tail of flood frequency distributions, *Water Resour. Res.*, **23**(8), 1657–1666.

APPENDIX: PROOF OF LEMMAS

Proof of Lemma 5.1: Remark that, for n large enough,

$$\left|k^{1/2}\, b\big(\log(n/k)\big)\right| \;\le\; \left|k^{1/2}\, b\big(\log(n/k)\big)-\lambda\right| + |\lambda| \;\le\; 1+|\lambda| \;,$$

and thus, if b is ultimately non-zero,

$$0 \;\le\; \frac{1}{2}\,\frac{\log(k)}{\log(n/k)} \;\le\; \frac{\log\,(1+|\lambda|)}{\log(n/k)} - \frac{\log\big|b\big(\log(n/k)\big)\big|}{\log(n/k)} \;. \tag{5.9}$$

Since $|b|$ is a regularly varying function, we have that (see [7], Proposition 1.3.6.)

$$\frac{\log\big|b\big(\log(x)\big)\big|}{\log(x)} \to 0 \qquad \text{as} \quad x\to\infty \;.$$

Then, (5.9) implies $\log(k)/\log(n/k)\to 0$ which entails $\log(k)/\log(n)\to 0$. □

Proof of Lemma 5.2: Since for all $x,t>0$, $tK_\rho(1{+}x/t)<x$, Lebesgue Theorem implies that

$$\begin{aligned}\lim_{t\to\infty}\int_0^\infty \Big(tK_\rho\Big(1+\frac{x}{t}\Big)\Big)^q\, \mathrm{e}^{-x}\,dx \;&=\; \int_0^\infty \lim_{t\to\infty}\Big(tK_\rho\Big(1+\frac{x}{t}\Big)\Big)^q\,\mathrm{e}^{-x}\,dx\\ &=\; \int_0^\infty x^q\,\mathrm{e}^{-x}\,dx \;=\; q!\;,\end{aligned}$$

which concludes the proof. □

Proof of Lemma 5.3:

i) Lemma 5.2 shows that $\mu_0(t) \sim 1/t$ and thus $T_n^{(1)} \log(n/k) \to 1$. By definition, $a_n^{(1)} = 0$.

ii) The well-known inequality $-x^2/2 \leq \log(1+x) - x \leq 0$, $x > 0$ yields

$$-\frac{1}{2}\,\frac{1}{\log(n/k)}\,\frac{1}{k}\sum_{i=1}^{k}\log^2(k/i) \;\leq\; \log(n/k)\,T_n^{(2)} - \frac{1}{k}\sum_{i=1}^{k}\log(k/i) \;\leq\; 0\ . \tag{5.10}$$

Now, since when $k \to \infty$,

$$\frac{1}{k}\sum_{i=1}^{k}\log^2(k/i) \to \int_0^1 \log^2(x)\,dx = 2 \quad \text{and} \quad \frac{1}{k}\sum_{i=1}^{k}\log(k/i) \to -\int_0^1 \log(x)\,dx = 1\ ,$$

it follows that $T_n^{(2)} \log(n/k) \to 1$. Let us now introduce the function defined on $(0,1]$ by:

$$f_n(x) \;=\; \log\left(1 - \frac{\log(x)}{\log(n/k)}\right)\ .$$

We have:

$$\begin{aligned}
a_n^{(2)} \;&=\; -\frac{1}{T_n^{(2)}}\Big(T_n^{(2)} - \mu_0\big(\log(n/k)\big)\Big)\\
&=\; -\frac{1}{T_n^{(2)}}\left(\frac{1}{k}\sum_{i=1}^{k-1} f_n(i/k) \;-\; \int_0^1 f_n(t)\,dt\right)\\
&=\; -\frac{1}{T_n^{(2)}}\sum_{i=1}^{k-1}\int_{i/k}^{(i+1)/k}\big(f_n(i/k) - f_n(t)\big)\,dt \;+\; \frac{1}{T_n^{(2)}}\int_0^{1/k} f_n(t)\,dt\ .
\end{aligned}$$

Since

$$f_n(t) \;=\; f_n(i/k) + (t - i/k)\,f_n^{(1)}(i/k) + \int_{i/k}^{t}(t-x)\,f_n^{(2)}(x)\,dx\ ,$$

where $f_n^{(p)}$ is the pth derivative of f_n, we have:

$$\begin{aligned}
a_n^{(2)} \;&=\; \frac{1}{T_n^{(2)}}\sum_{i=1}^{k-1}\int_{i/k}^{(i+1)/k}(t - i/k)\,f_n^{(1)}(i/k)\,dt\\
&\quad+\; \frac{1}{T_n^{(2)}}\sum_{i=1}^{k-1}\int_{i/k}^{(i+1)/k}\int_{i/k}^{t}(t-x)\,f_n^{(2)}(x)\,dx\,dt \;+\; \frac{1}{T_n^{(2)}}\int_0^{1/k} f_n(t)\,dt\\
&=:\; \Psi_1 + \Psi_2 + \Psi_3\ .
\end{aligned}$$

Let us focus first on the term Ψ_1:

$$\begin{aligned}
\Psi_1 &= \frac{1}{T_n^{(2)}}\,\frac{1}{2k^2}\sum_{i=1}^{k-1} f_n^{(1)}(i/k) \\
&= \frac{1}{2k\,T_n^{(2)}}\int_{1/k}^{1} f_n^{(1)}(x)\,dx \;+\; \frac{1}{2k\,T_n^{(2)}}\left(\frac{1}{k}\sum_{i=1}^{k-1} f_n^{(1)}(i/k) - \int_{1/k}^{1} f_n^{(1)}(x)\,dx\right) \\
&= \frac{1}{2k\,T_n^{(2)}}\Big(f_n(1)-f_n(1/k)\Big) - \frac{1}{2k\,T_n^{(2)}}\sum_{i=1}^{k-1}\int_{i/k}^{(i+1)/k}\Big(f_n^{(1)}(x)-f_n^{(1)}(i/k)\Big)\,dx \\
&=: \Psi_{1,1}-\Psi_{1,2}\ .
\end{aligned}$$

Since $T_n^{(2)}\sim 1/\log(n/k)$ and $\log(k)/\log(n)\to 0$, we have:

$$\Psi_{1,1} \;=\; -\frac{1}{2k\,T_n^{(2)}}\,\log\left(1+\frac{\log(k)}{\log(n/k)}\right) \;=\; -\frac{\log(k)}{2k}\Big(1+o(1)\Big)\ .$$

Furthermore, since, for n large enough, $f_n^{(2)}(x)>0$ for $x\in[0,1]$,

$$\begin{aligned}
O \;\le\; \Psi_{1,2} &\le \frac{1}{2k\,T_n^{(2)}}\sum_{i=1}^{k-1}\int_{i/k}^{(i+1)/k}\Big(f_n^{(1)}\big((i+1)/k\big)-f_n^{(1)}(i/k)\Big)\,dx \\
&= \frac{1}{2k^2\,T_n^{(2)}}\Big(f_n^{(1)}(1)-f_n^{(1)}(1/k)\Big) \\
&= \frac{1}{2k^2\,T_n^{(2)}}\left(-\frac{1}{\log(n/k)}+\frac{k}{\log(n/k)}\left(1+\frac{\log(k)}{\log(n/k)}\right)^{-1}\right) \\
&\sim \frac{1}{2k} \;=\; o\!\left(\frac{\log(k)}{k}\right)\ .
\end{aligned}$$

Thus,

$$\Psi_1 \;=\; -\frac{\log(k)}{2k}\Big(1+o(1)\Big)\ . \tag{5.11}$$

Second, let us focus on the term Ψ_2. Since, for n large enough, $f_n^{(2)}(x)>0$ for $x\in[0,1]$,

$$\begin{aligned}
0 \;\le\; \Psi_2 &\le \frac{1}{T_n^{(2)}}\sum_{i=1}^{k-1}\int_{i/k}^{(i+1)/k}\int_{i/k}^{(i+1)/k}(t-i/k)\,f_n^{(2)}(x)\,dx\,dt \\
&= \frac{1}{2k^2\,T_n^{(2)}}\Big(f_n^{(1)}(1)-f_n^{(1)}(1/k)\Big) \;=\; o\!\left(\frac{\log(k)}{k}\right)\ .
\end{aligned} \tag{5.12}$$

Finally,

$$\Psi_3 \;=\; \frac{1}{T_n^{(2)}}\int_0^{1/k} -\frac{\log(t)}{\log(n/k)}\,dt + \frac{1}{T_n^{(2)}}\int_0^{1/k}\left(f_n(t)+\frac{\log(t)}{\log(n/k)}\right)dt \;=:\; \Psi_{3,1}+\Psi_{3,2}\ ,$$

and we have:

$$\Psi_{3,1} \;=\; \frac{1}{\log(n/k)\,T_n^{(2)}}\,\frac{1}{k}\big(\log(k)+1\big) \;=\; \frac{\log(k)}{k}\big(1+o(1)\big)\;.$$

Furthermore, using the well known inequality: $|\log(1+x)-x| \le x^2/2,\ x>0$, we have:

$$\begin{aligned}
|\Psi_{3,2}| \;&\le\; \frac{1}{2\,T_n^{(2)}}\int_0^{1/k}\bigg(\frac{\log(t)}{\log(n/k)}\bigg)^2 dt\\
&=\; \frac{1}{2\,T_n^{(2)}}\,\frac{1}{k\big(\log(n/k)\big)^2}\Big(\big(\log(k)\big)^2+2\log(k)+2\Big)\\
&\sim\; \frac{\big(\log(k)\big)^2}{2k\log(n/k)} \;=\; o\bigg(\frac{\log(k)}{k}\bigg)\;,
\end{aligned}$$

since $\log(k)/\log(n)\to 0$. Thus,

$$\Psi_3 \;=\; \frac{\log(k)}{k}\,\big(1+o(1)\big)\;. \tag{5.13}$$

We conclude the proof of i) by collecting (5.11)–(5.13).

iii) First, $T_n^{(3)}\log(n/k)=1$ by definition. Besides, we have

$$\begin{aligned}
a_n^{(3)} \;&=\; \frac{\mu_0\big(\log(n/k)\big)}{T_n^{(3)}}-1\\
&=\; \log(n/k)\,\mu_0\big(\log(n/k)\big)-1\\
&=\; \int_0^\infty \log(n/k)\log\bigg(1+\frac{x}{\log(n/k)}\bigg)\mathrm{e}^{-x}\,dx \;-\; 1\\
&=\; \int_0^\infty x\,\mathrm{e}^{-x}\,dx \;-\; \frac{1}{2}\int_0^\infty \frac{x^2}{\log(n/k)}\,\mathrm{e}^{-x}\,dx \;-\; 1+R_n \;=\; -\frac{1}{\log(n/k)}+R_n\;,
\end{aligned}$$

where

$$R_n \;=\; \int_0^\infty \log(n/k)\Bigg(\log\bigg(1+\frac{x}{\log(n/k)}\bigg)-\frac{x}{\log(n/k)}+\frac{x^2}{2\big(\log(n/k)\big)^2}\Bigg)\mathrm{e}^{-x}\,dx\;.$$

Using the well known inequality: $|\log(1{+}x)-x+x^2/2| \le x^3/3,\ x>0,$ we have,

$$|R_n| \;\le\; \frac{1}{3}\int_0^\infty \frac{x^3}{\big(\log(n/k)\big)^2}\,\mathrm{e}^{-x}\,dx \;=\; o\bigg(\frac{1}{\log(n/k)}\bigg)\;,$$

which finally yields $a_n^{(3)}\sim -1/\log(n/k)$. □

Proof of Lemma 5.4: Recall that

$$\hat\theta_n \;=:\; \frac{1}{T_n}\,\frac{1}{k}\sum_{i=1}^{k-1}\Big(\log(X_{n-i+1,n})-\log(X_{n-k+1,n})\Big)\;,$$

and let $E_{1,n}, ..., E_{n,n}$ be ordered statistics generated by n independent standard exponential random variables. Under (A.1), we have

$$\begin{aligned}
\hat{\theta}_n &\stackrel{d}{=} \frac{1}{T_n}\frac{1}{k}\sum_{i=1}^{k-1}\Big(\log H^{\leftarrow}(E_{n-i+1,n}) - \log H^{\leftarrow}(E_{n-k+1,n})\Big)\\
&\stackrel{d}{=} \frac{1}{T_n}\left(\theta\,\frac{1}{k}\sum_{i=1}^{k-1}\log\left(\frac{E_{n-i+1,n}}{E_{n-k+1,n}}\right) + \frac{1}{k}\sum_{i=1}^{k-1}\log\left(\frac{\ell(E_{n-i+1,n})}{\ell(E_{n-k+1,n})}\right)\right) .
\end{aligned}$$

Define $x_n = E_{n-k+1,n}$ and $\lambda_{i,n} = E_{n-i+1,n}/E_{n-k+1,n}$. It is clear, in view of [12], Lemma 1 that $x_n \stackrel{P}{\to} \infty$ and $\lambda_{i,n} \stackrel{P}{\to} 1$. Thus, (A.2) yields that uniformly in $i = 1, ..., k-1$:

$$\hat{\theta}_n \stackrel{d}{=} \frac{1}{T_n}\left(\theta\frac{1}{k}\sum_{i=1}^{k-1}\log\left(\frac{E_{n-i+1,n}}{E_{n-k+1,n}}\right) + \big(1+o_p(1)\big)\, b(E_{n-k+1,n})\,\frac{1}{k}\sum_{i=1}^{k-1}K_\rho\left(\frac{E_{n-i+1,n}}{E_{n-k+1,n}}\right)\right) .$$

The Rényi representation of the Exp(1) ordered statistics (see [2], p. 72) yields

$$\left\{\frac{E_{n-i+1,n}}{E_{n-k+1,n}}\right\}_{i=1,...,k-1} \stackrel{d}{=} \left\{1+\frac{F_{k-i,k-1}}{E_{n-k+1,n}}\right\}_{i=1,...,k-1} , \tag{5.14}$$

where $\{F_{1,k-1}, ..., F_{k-1,k-1}\}$ are ordered statistics independent from $E_{n-k+1,n}$ and generated by $k-1$ independent standard exponential variables $\{F_1, ..., F_{k-1}\}$. Therefore,

$$\begin{aligned}
\hat{\theta}_n \stackrel{d}{=} \frac{1}{T_n}&\left(\theta\,\frac{1}{k}\sum_{i=1}^{k-1}\log\left(1+\frac{F_i}{E_{n-k+1,n}}\right)\right.\\
&\left. + \big(1+o_p(1)\big)\, b(E_{n-k+1,n})\,\frac{1}{k}\sum_{i=1}^{k-1}K_\rho\left(1+\frac{F_i}{E_{n-k+1,n}}\right)\right) .
\end{aligned}$$

Remarking that $K_0(x) = \log(x)$ concludes the proof. □

Proof of Lemma 5.5: Lemma 5.2 implies that,

$$\mu_\rho(E_{n-k+1,n}) \stackrel{P}{\sim} \frac{1}{E_{n-k+1,n}} \stackrel{P}{\sim} \frac{1}{\log(n/k)} ,$$

since $E_{n-k+1,n}/\log(n/k) \stackrel{P}{\to} 1$ (see [12], Lemma 1). Next, from Lemma 5.2,

$$\begin{aligned}
\sigma_\rho^2(E_{n-k+1,n}) &= \frac{2}{E^2_{n-k+1,n}}\big(1+o_{\mathrm{P}}(1)\big) - \frac{1}{E^2_{n-k+1,n}}\big(1+o_{\mathrm{P}}(1)\big)\\
&= \frac{1}{E^2_{n-k+1,n}}\big(1+o_{\mathrm{P}}(1)\big) = \frac{1}{\big(\log(n/k)\big)^2}\big(1+o_{\mathrm{P}}(1)\big) ,
\end{aligned}$$

which concludes the proof. □

Proof of Lemma 5.6: Remark that

$$\frac{k^{1/2}}{\sigma_\rho(E_{n-k+1,n})}\Big(U_n^{(\rho)} - \mu_\rho(E_{n-k+1,n})\Big) =$$
$$= \frac{k^{-1/2}}{\sigma_\rho(E_{n-k+1,n})} \sum_{i=1}^{k-1}\left(K_\rho\left(1+\frac{F_i}{E_{n-k+1,n}}\right) - \mu_\rho(E_{n-k+1,n})\right) - k^{-1/2}\,\frac{\mu_\rho(E_{n-k+1,n})}{\sigma_\rho(E_{n-k+1,n})} .$$

Let us introduce the following notation:

$$S_n(t) = \frac{(k-1)^{-1/2}}{\sigma_\rho(t)} \sum_{i=1}^{k-1}\left(K_\rho\left(1+\frac{F_i}{t}\right) - \mu_\rho(t)\right) .$$

Thus,

$$\frac{k^{1/2}}{\sigma_\rho(E_{n-k+1,n})}\Big(U_n^{(\rho)} - \mu_\rho(E_{n-k+1,n})\Big) = S_n(E_{n-k+1,n})\big(1+o(1)\big) + o_{\mathrm{P}}(1) ,$$

from Lemma 5.5. It remains to prove that for $x \in \mathbb{R}$,

$$\mathrm{P}\Big(S_n(E_{n-k+1,n}) \le x\Big) - \Phi(x) \to 0 \qquad \text{as} \quad n\to\infty ,$$

where Φ is the cumulative distribution function of the standard Gaussian distribution. Lemma 5.2 implies that for all $\varepsilon \in\,]0,1[$, there exists T_ε such that for all $t \ge T_\varepsilon$,

$$\text{(5.15)} \qquad \frac{q!}{t^q}\,(1-\varepsilon) \le \mathbb{E}\left(\left(K_\rho\left(1+\frac{F_1}{t}\right)\right)^q\right) \le \frac{q!}{t^q}(1+\varepsilon) .$$

Furthermore, for $x\in\mathbb{R}$,

$$\mathrm{P}\Big(S_n(E_{n-k+1,n}) \le x\Big) - \Phi(x) =$$
$$= \int_0^{T_\varepsilon}\Big(\mathrm{P}\big(S_n(t)\le x\big) - \Phi(x)\Big)\,h_n(t)\,dt + \int_{T_\varepsilon}^\infty \Big(\mathrm{P}\big(S_n(t)\le x\big) - \Phi(x)\Big)\,h_n(t)\,dt$$
$$=: A_n + B_n ,$$

where h_n is the density of the random variable $E_{n-k+1,n}$. First, let us focus on the term A_n. We have,

$$|A_n| \le 2\,\mathrm{P}\big(E_{n-k+1,n} \le T_\varepsilon\big) .$$

Since $E_{n-k+1,n}/\log(n/k) \xrightarrow{P} 1$ (see [12], Lemma 1), it is easy to show that $A_n \to 0$. Now, let us consider the term B_n. For the sake of simplicity, let us denote:

$$\left\{Y_i = K_\rho\left(1+\frac{F_i}{t}\right) - \mu_\rho(t), \quad i=1,...,k-1\right\} .$$

Clearly, $Y_1, ..., Y_{k-1}$ are independent, identically distributed and centered random variables. Furthermore, for $t \geq T_\varepsilon$,

$$\begin{aligned}
\mathbb{E}\big(|Y_1|^3\big) \;&\leq\; \mathbb{E}\left(\left(K_\rho\left(1+\frac{F_1}{t}\right)+\mu_\rho(t)\right)^3\right)\\
&=\; \mathbb{E}\left(\left(K_\rho\left(1+\frac{F_1}{t}\right)\right)^3\right)+\big(\mu_\rho(t)\big)^3+3\,\mathbb{E}\left(\left(K_\rho\left(1+\frac{F_1}{t}\right)\right)^2\right)\mu_\rho(t)\\
&\quad+3\,\mathbb{E}\left(K_\rho\left(1+\frac{F_1}{t}\right)\right)\big(\mu_\rho(t)\big)^2\\
&\leq\; \frac{1}{t^3}\,C_1(q,\varepsilon) \;<\; \infty\;,
\end{aligned}$$

from (5.15) where $C_1(q,\varepsilon)$ is a constant independent of t. Thus, from Esseen's inequality (see [20], Theorem 3), we have:

$$\sup_x\Big|\mathrm{P}\big(S_n(t)\leq x\big)-\Phi(x)\Big| \;\leq\; C_2\,L_n\;,$$

where C_2 is a positive constant and

$$L_n \;=\; \frac{(k-1)^{-1/2}}{\big(\sigma_\rho(t)\big)^3}\;\mathbb{E}\big(|Y_1|^3\big)\;.$$

From (5.15), since $t \geq T_\varepsilon$,

$$\big(\sigma_\rho(t)\big)^2 \;=\; \mathbb{E}\left(\left(K_\rho\left(1+\frac{F_1}{t}\right)\right)^2\right)-\left(\mathbb{E}\left(K_\rho\left(1+\frac{F_1}{t}\right)\right)\right)^2 \;\geq\; \frac{1}{t^2}\,C_3(\varepsilon)\;,$$

where $C_3(\varepsilon)$ is a constant independent of t. Thus, $L_n \leq (k-1)^{-1/2}\,C_4(q,\varepsilon)$ where $C_4(q,\varepsilon)$ is a constant independent of t, and therefore

$$|B_n| \;\leq\; C_4(q,\varepsilon)\,(k-1)^{-1/2}\,\mathrm{P}\big(E_{n-k+1,n}\geq T_\varepsilon\big) \;\leq\; C_4(q,\varepsilon)\,(k-1)^{-1/2} \;\to\; 0\;,$$

which concludes the proof. $\square$

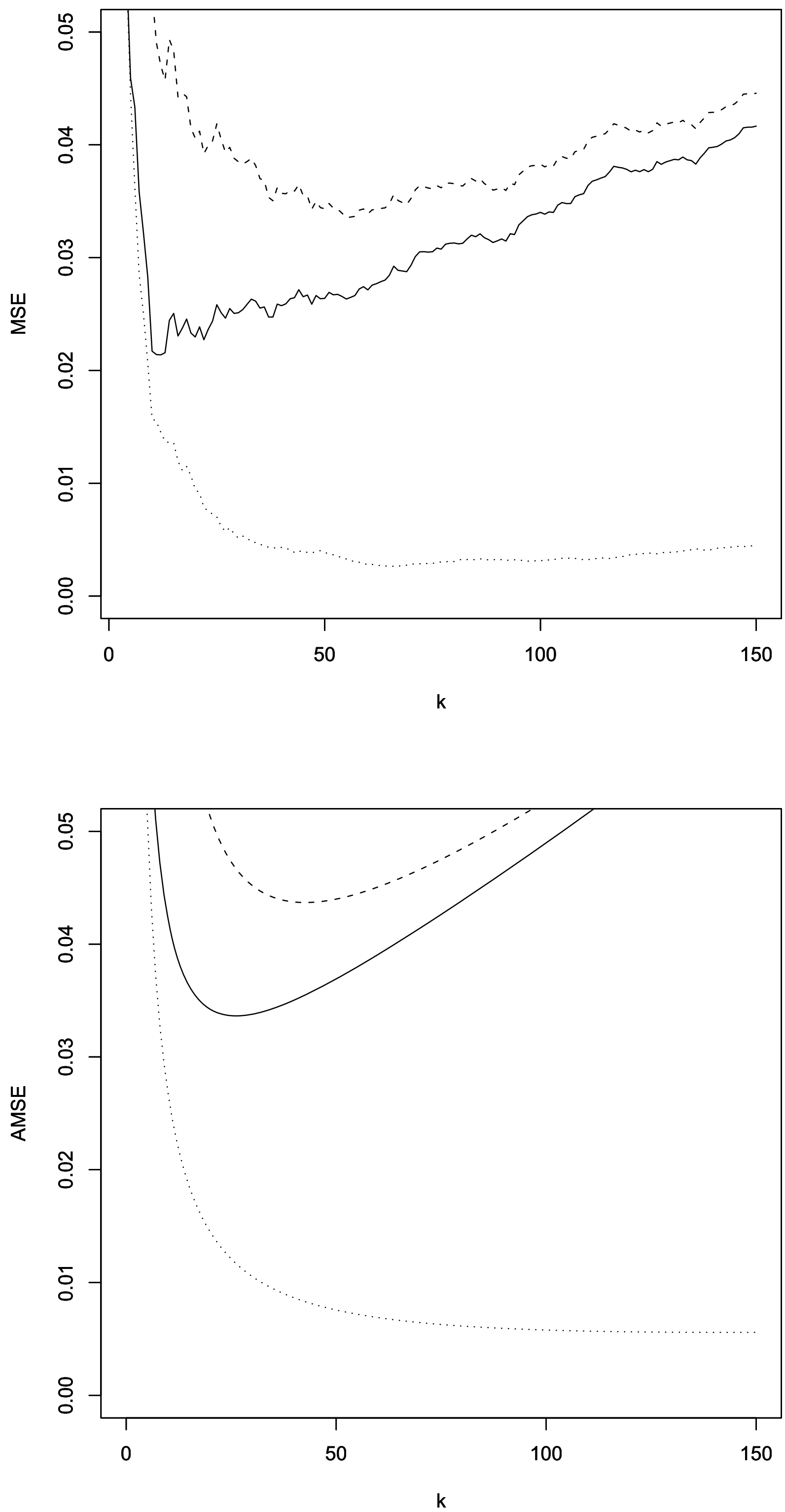

Figure 1: Comparison of estimates $\hat{\theta}_n^{(1)}$ (solid line), $\hat{\theta}_n^{(2)}$ (dashed line) and $\hat{\theta}_n^{(3)}$ (dotted line) for the $|\mathcal{N}(0,1)|$ distribution. Up: MSE, down: AMSE.

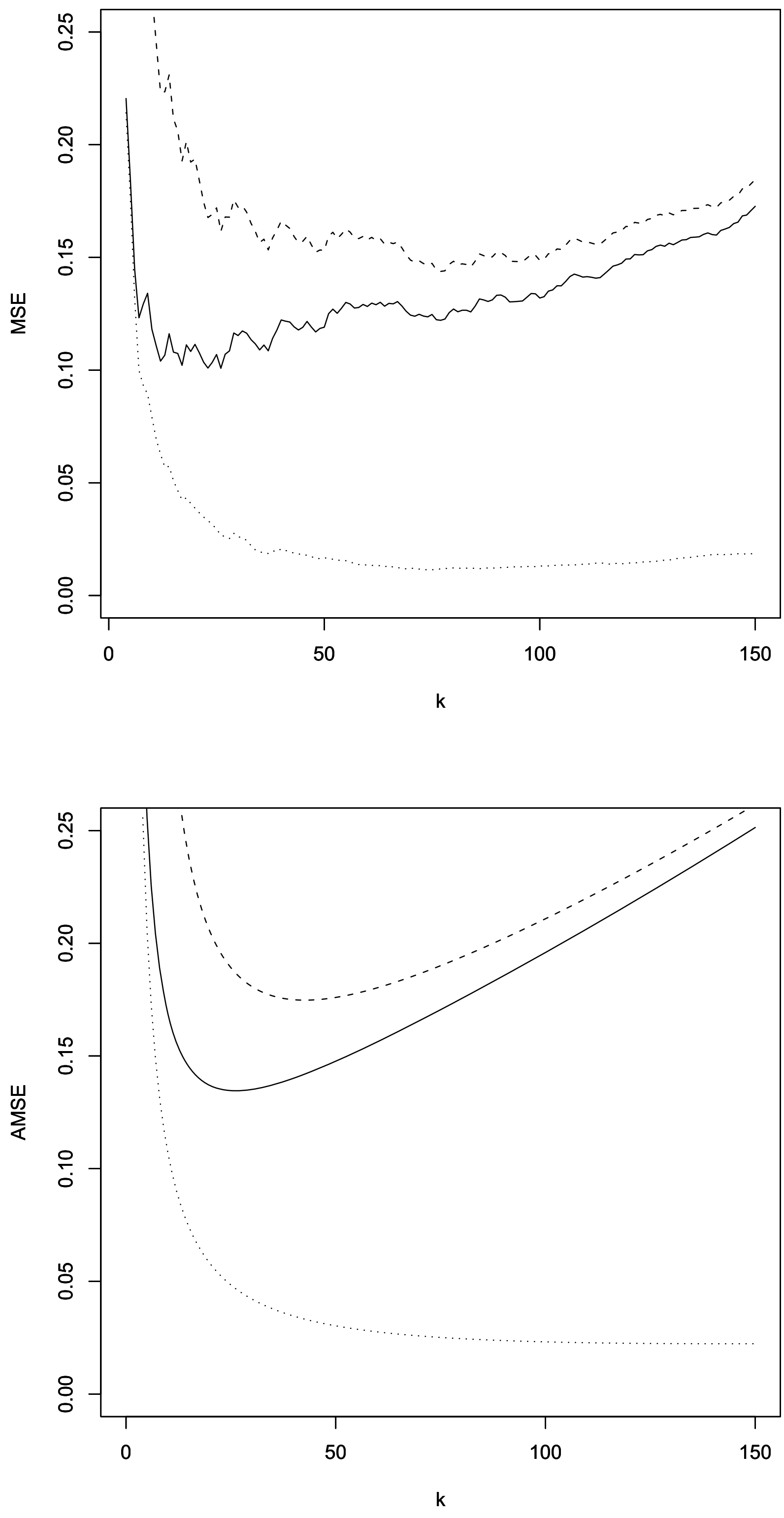

Figure 2: Comparison of estimates $\hat{\theta}_n^{(1)}$ (solid line), $\hat{\theta}_n^{(2)}$ (dashed line) and $\hat{\theta}_n^{(3)}$ (dotted line) for the $\Gamma(0.5, 1)$ distribution. Up: MSE, down: AMSE.

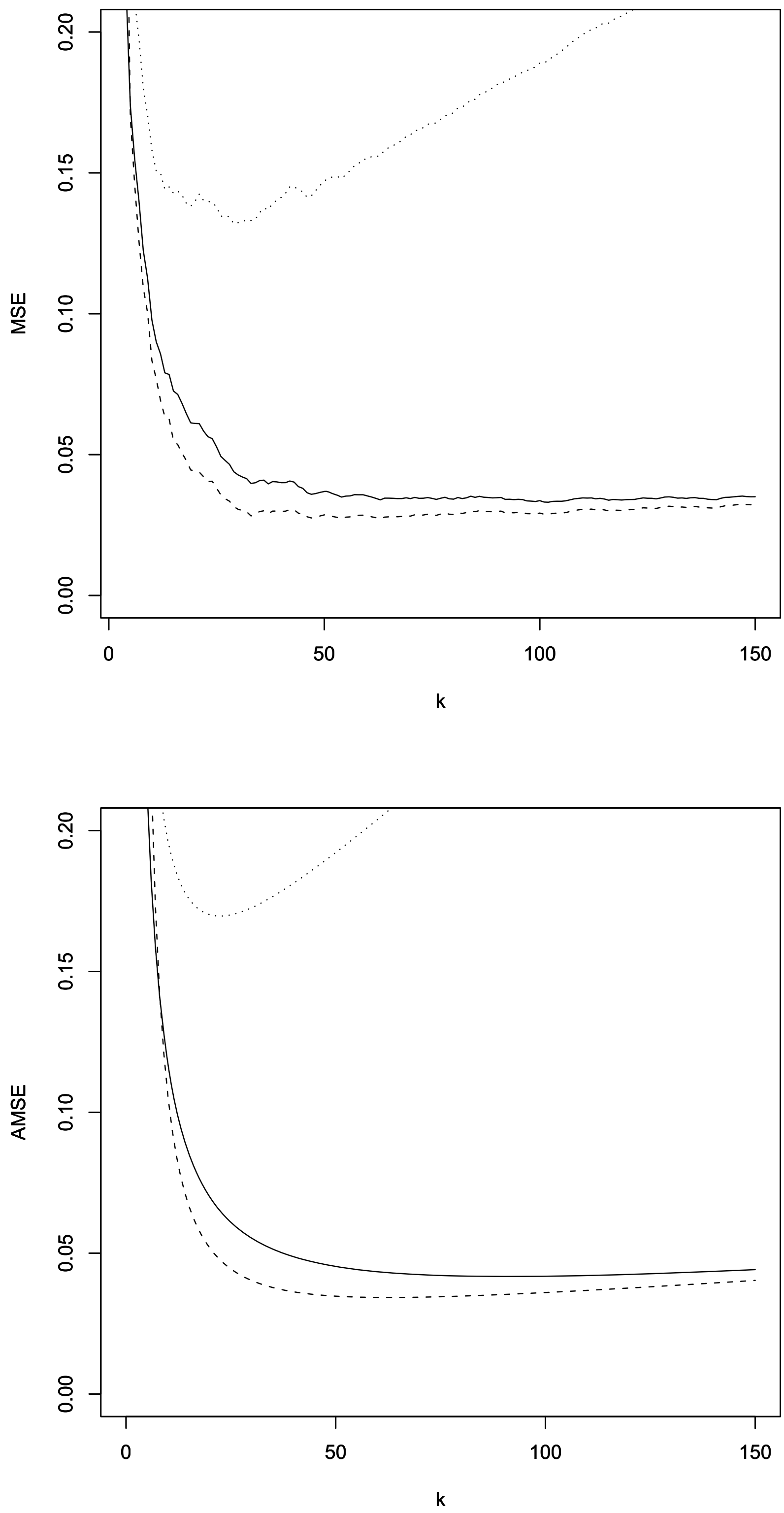

Figure 3: Comparison of estimates $\hat{\theta}_n^{(1)}$ (solid line), $\hat{\theta}_n^{(2)}$ (dashed line) and $\hat{\theta}_n^{(3)}$ (dotted line) for the $\Gamma(1.5, 1)$ distribution. Up: MSE, down: AMSE.

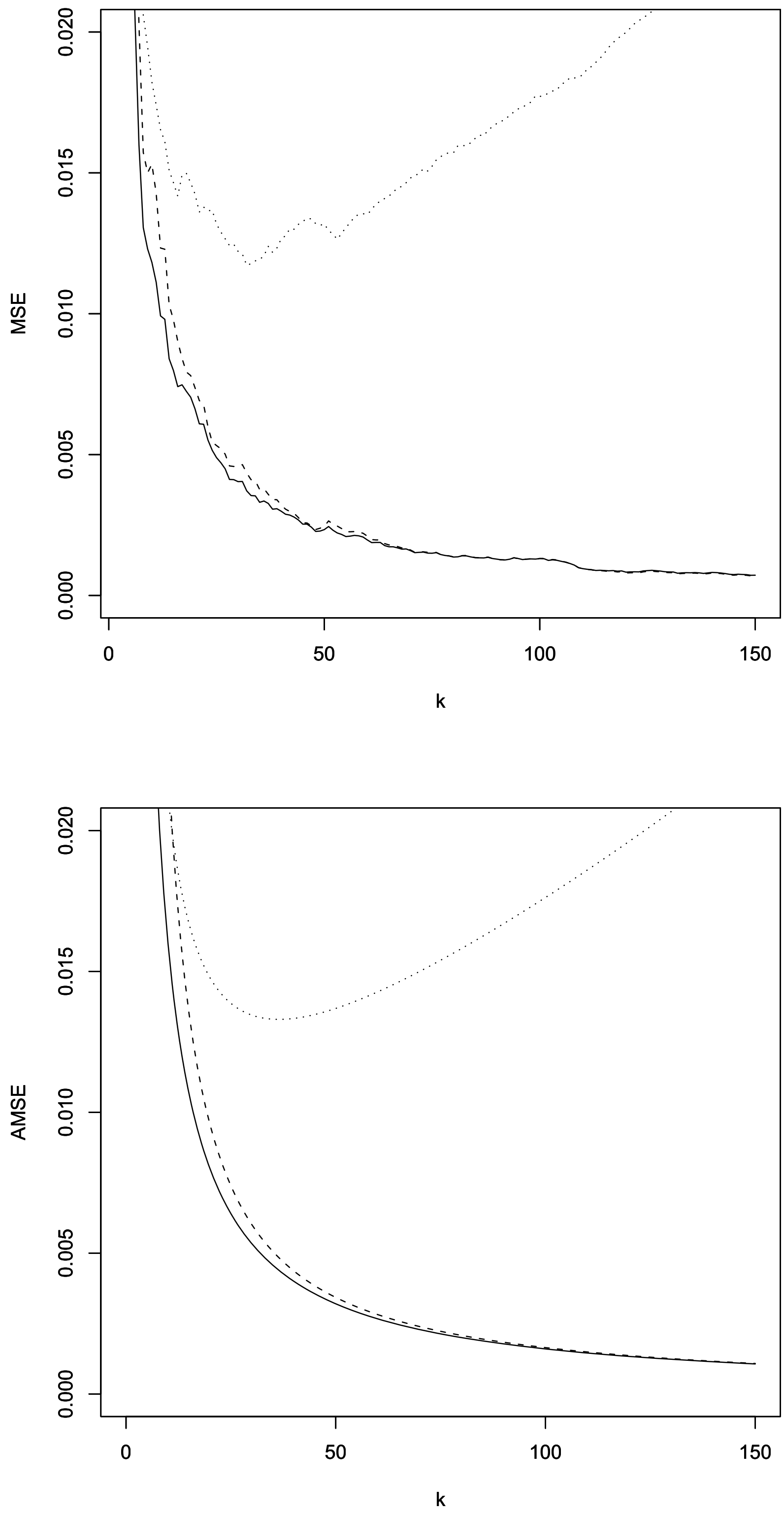

Figure 4: Comparison of estimates $\hat{\theta}_n^{(1)}$ (solid line), $\hat{\theta}_n^{(2)}$ (dashed line) and $\hat{\theta}_n^{(3)}$ (dotted line) for the $\mathcal{W}(2.5, 2.5)$ distribution. Up: MSE, down: AMSE.

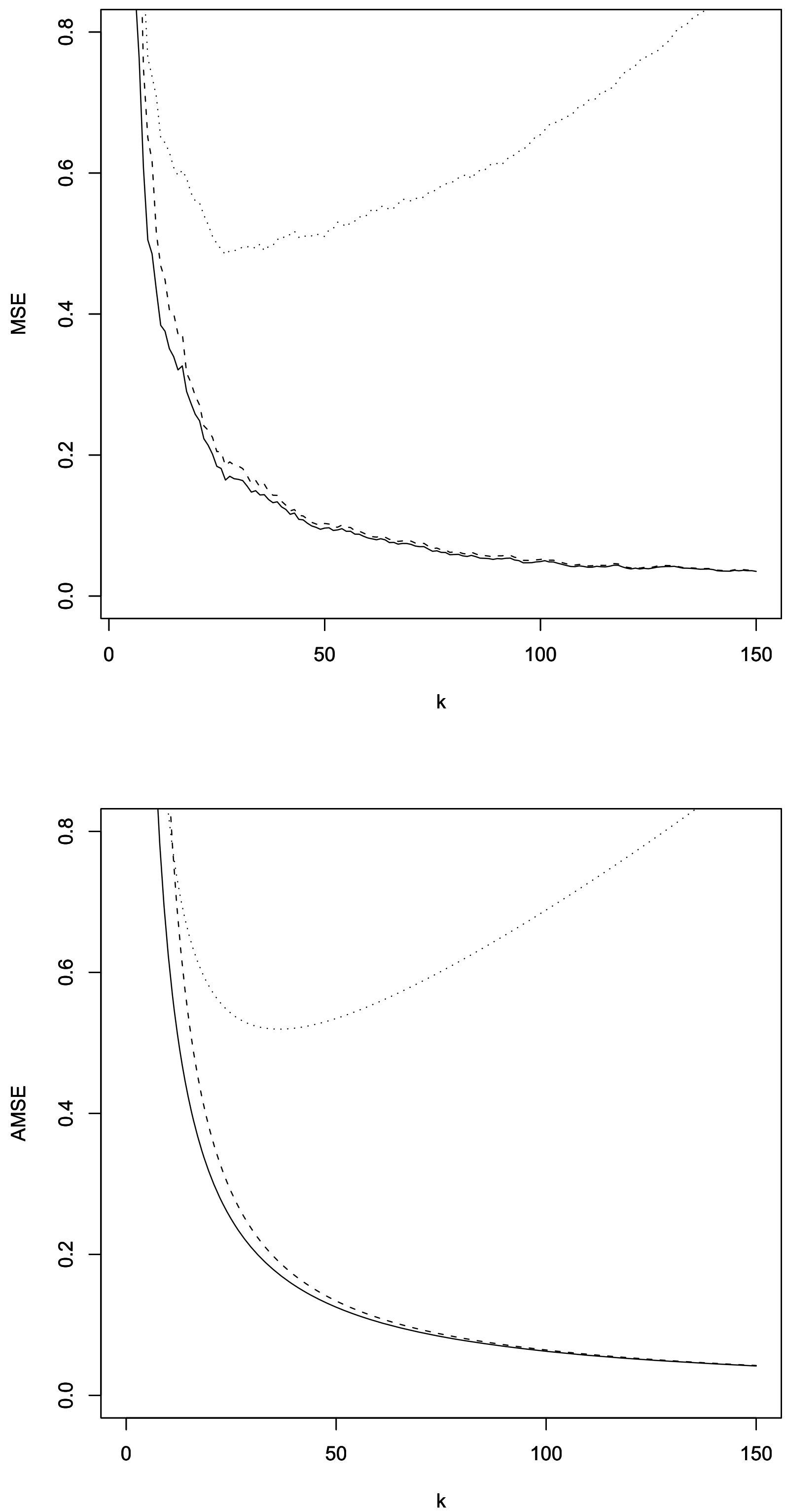

Figure 5: Comparison of estimates $\hat{\theta}_n^{(1)}$ (solid line), $\hat{\theta}_n^{(2)}$ (dashed line) and $\hat{\theta}_n^{(3)}$ (dotted line) for the $\mathcal{W}(0.4, 0.4)$ distribution. Up: MSE, down: AMSE.

REVSTAT – STATISTICAL JOURNAL

Background

Statistical Institute of Portugal (INE), well aware of how vital a statistical culture is in understanding most phenomena in the present-day world, and of its responsibility in disseminating statistical knowledge, started the publication of the scientific statistical journal *Revista de Estatística*, in Portuguese, publishing three times a year papers containing original research results, and application studies, namely in the economic, social and demographic fields.

In 1998 it was decided to publish papers also in English. This step has been taken to achieve a larger diffusion, and to encourage foreign contributors to submit their work.

At the time, the Editorial Board was mainly composed by Portuguese university professors, being now composed by national and international university professors, and this has been the first step aimed at changing the character of *Revista de Estatística* from a national to an international scientific journal.

In 2001, the *Revista de Estatística* published three volumes special issue containing extended abstracts of the invited contributed papers presented at the 23rd European Meeting of Statisticians.

The name of the Journal has been changed to REVSTAT – STATISTICAL JOURNAL, published in English, with a prestigious international editorial board, hoping to become one more place where scientists may feel proud of publishing their research results.

— The editorial policy will focus on publishing research articles at the highest level in the domains of Probability and Statistics with emphasis on the originality and importance of the research.

— All research articles will be refereed by at least two persons, one from the Editorial Board and another, external.

— The only working language allowed will be English.

— Two volumes are scheduled for publication, one in June and the other in November.

— On average, four articles will be published per issue.

Aims and Scope

The aim of REVSTAT is to publish articles of high scientific content, in English, developing innovative statistical scientific methods and introducing original research, grounded in substantive problems.

REVSTAT covers all branches of Probability and Statistics. Surveys of important areas of research in the field are also welcome.

Abstract/indexed in

REVSTAT is expected to be abstracted/indexed at least in *Current Index to Statistics*, *Mathematical Reviews*, *Statistical Theory and Method Abstracts*, and *Zentralblatt für Mathematic.*

Instructions to Authors, special-issue editors and publishers

Papers may be submitted in two different ways:

— By sending a paper copy to the Executive Editor and one copy to one of the two Editors or Associate Editors whose opinion the author(s) would like to be taken into account, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

— By sending a paper copy to the Executive Editor, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

Submission of a paper means that it contains original work that has not been nor is about to be published elsewhere in any form.

Submitted manuscripts (text, tables and figures) should be typed only in black, on one side, in double spacing, with a left margin of at least 3 cm and not have more than 30 pages.

The first page should include the name, affiliation and address of the author(s) and a short abstract with the maximum of 100 words, followed by the key words up to the limit of 6, and the AMS 2000 subject classification.

Authors are obliged to write the final version of accepted papers using LaTeX, in the REVSTAT style.

This style (REVSTAT.sty), and examples file (REVSTAT.tex), which may be download to *PC Windows System* (Zip format), *Mackintosh, Linux* and *Solaris Systems* (StuffIt format), and *Mackintosh System* (BinHex Format), are available in the REVSTAT link of the National Statistical Institute's Website: http://www.ine.pt/revstat.html

Additional information for the authors may be obtained in the above link.

Accepted papers

Authors of accepted papers are requested to provide the LaTeX files and also a postscript (PS) or an acrobat (PDF) file of the paper to the Secretary of REVSTAT: liliana.martins@ine.pt.

Such e-mail message should include the author(s)'s name, mentioning that it has been accepted by REVSTAT.

The authors should also mention if encapsulated postscript figure files were included, and submit electronics figures separately in .tiff, .gif, .eps or .ps format. Figures must be a minimum of 300 dpi.

Also send always the final paper version to:

Maria José Carrilho
Executive Editor, REVSTAT – STATISTICAL JOURNAL
Instituto Nacional de Estatística
Av. António José de Almeida
1000-043 LISBOA
PORTUGAL

Copyright and Reprints

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, the INE, in order to ensure the widest possible dissemination of information, namely through the National Statistical Institute's Website (http://www.ine.pt).

After assigning the transfer copyright form, authors may use their own material in other publications provided that the REVSTAT is acknowledged as the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.

Authors of articles published in the REVSTAT will be entitled to one free copy of the respective issue of the Journal and twenty-five reprints of the paper are provided free. Additional reprints may be ordered at expenses of the author(s), and prior to publication.