ISSN 1645-6726

Instituto Nacional de Estatística
Statistics Portugal

REVSTAT
Statistical Journal

Volume 11, No.3
November 2013

REVSTAT

Statistical Journal

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

CREDITS

- **PUBLISHER**
  - *Instituto Nacional de Estatística, I.P. (INE, I.P.)*
    Av. António José de Almeida, 2
    1000-043 LISBOA
    PORTUGAL
    Tel.: +351 218 426 100
    Fax: +351 218 454 084
    Web site: http://www.ine.pt
    Customer Support Service
    (National network): 808 201 808
    (Other networks): +351 218 440 695
- **COVER DESIGN**
  - *Mário Bouçadas, designed on the stain glass window at INE, I.P., by the painter Abel Manta*
- **LAYOUT AND GRAPHIC DESIGN**
  - *Carlos Perpétuo*
- **PRINTING**
  - *Instituto Nacional de Estatística, I.P.*
- **EDITION**
  - *190 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*

PRICE

[VAT included]

- Single issue ………………………………………………………….. **€ 11**
- Annual subscription (No. 1 Special Issue, No. 2 and No.3)………. **€ 26**
- Annual subscription (No. 2, No. 3) ……………………………….. **€ 18**

INDEX

Abstracted/indexed in: *Current Index to Statistics, DOAJ, Google Scholar, Journal Citation Reports/Science Edition, Mathematical Reviews, Science Citation Index Expanded®, SCOPUS* and *Zentralblatt für Mathematic.*

REVSTAT – Statistical Journal
Volume 11, Number 3, November 2013, 231–254

AN ACCURATE APPROXIMATION TO THE DISTRIBUTION OF A LINEAR COMBINATION OF NON-CENTRAL CHI-SQUARE RANDOM VARIABLES

Authors: Hyung-Tae Ha
– Applied Statistics, Gachon University,
Sung-Nam, South Korea, 461-701
htha@gachon.ac.kr

Serge B. Provost
– Statistical & Actuarial Sciences, The University of Western Ontario,
London, Canada, N6A 5B7
provost@stats.uwo.ca

Received: May 2011 Revised: June 2012 Accepted: January 2013

Abstract:

- This paper provides an accessible methodology for approximating the distribution of a general linear combination of non-central chi-square random variables. Attention is focused on the main application of the results, namely the distribution of positive definite and indefinite quadratic forms in normal random variables. After explaining that the moments of a quadratic form can be determined from its cumulants by means of a recursive formula, we propose a moment-based approximation of the density function of a positive definite quadratic form, which consists of a gamma density function that is adjusted by a linear combination of Laguerre polynomials or, equivalently, by a single polynomial. On expressing an indefinite quadratic form as the difference of two positive definite quadratic forms, explicit representations of approximations to its density and distribution functions are obtained in terms of confluent hypergeometric functions. The proposed closed form expressions converge rapidly and provide accurate approximations over the entire support of the distribution. Additionally, bounds are derived for the integrated squared and absolute truncation errors. An easily implementable algorithm is provided and several illustrative numerical examples are presented. In particular, the methodology is applied to the Durbin–Watson statistic. Finally, relevant computational considerations are discussed. Linear combinations of chi-square random variables and quadratic forms in normal variables being ubiquitous in statistics, the distribution approximation technique introduced herewith should prove widely applicable.

Key-Words:

- *chi-square random variables; linear combinations; quadratic forms; cumulants; moments; density approximation; Durbin–Watson statistic.*

AMS Subject Classification:

- 62E17, 62Q05.

1. INTRODUCTION

The distribution of linear combinations of chi-square random variables and that of quadratic forms in normal vectors have already received a lot of attention in the statistical literature. Box (1954) considered a linear combination of chi-square variables having even degrees of freedom. Some representations of the density function of linear combinations of chi-square variables were derived by Mathai and Saxena (1978). Various representations of the distribution function of a quadratic form are available, and several procedures have been proposed for computing percentage points and preparing tables. Gurland (1948, 1953, 1956), Pachares (1955), Ruben (1960, 1962), Shah and Khatri (1961), and Kotz *et al.* (1967a,b) among others, have given representations of the distribution function of quadratic forms in terms of MacLaurin series and the distribution function of chi-square variables. Gurland (1956) and Shah (1963) considered respectively central and non-central indefinite quadratic forms, but as pointed by Shah (1963), the expansions obtained are not practical. Press (1967) provided infinite series representations of the density and distribution functions of an indefinite quadratic form in normal variables. Other representations of the exact density and distribution functions of indefinite quadratic forms have been given by Imhof (1961), Davis (1973) and Rice (1980). As pointed out in Mathai and Provost (1992), a wide array of statistics can be expressed in terms of quadratic forms in normal random vectors. For example, one may consider the lagged regression residuals developed by De Gooijer and MacNeill (1999) and discussed in Provost *et al.* (2005), or certain change point test statistics derived by MacNeill (1978). Hillier (2001) expressed the density function of a ratio of quadratic forms in normal random variables in terms of top-order zonal polynomials involving difference quotients of the characteristic roots of the matrix in the numerator quadratic form. The sample serial correlation coefficient as defined in Anderson (1990) and discussed in Provost and Rudiuk (1995) as well as the sample innovation cross-correlation function for an ARMA time series whose asymptotic distribution was derived by McLeod (1979) have such a structure.

Monte Carlo simulations, whereby artificial data are generated and sampling distributions and moments then are estimated, can be implemented more easily on an extensive array of models. These simulations may, however, result in some limitations such as sampling variations and simulation inadequacies, and their results may be specific to the set of parameter values assumed in the simulations. Hendry and Harrison (1974), Dempster *et al.* (1977), Hendry (1979), and Hendry and Mizon (1980) among others, have attempted to cope with these issues. On the other hand, the analytical approach derives results which hold over the entire parameter space but may find some limitations in terms of simplifications on the model, which are imposed to render the problem tractable. The analytical approach has been applied to various statistics involving quadratic forms. Examples in this area include certain heteroscedastic models studied by Taylor (1977, 1978), the first-order autoregressive process considered by Sawa (1978) and Phillips (1977, 1978), the regression models analyzed by Dwivedi and Srivastava (1979), a linear model with unknown covariance structure studied by Yamamoto (1979), as well as the Bayesian analysis of simultaneous equations models carried out by Zellner (1971) and Dreze (1976).

A novel and accessible moment-based approach is proposed in this paper for approximating the density function of positive definite quadratic forms in normal random variables in terms of a gamma density function and a linear combination of Laguerre polynomials, which is re-expressed as a single polynomial so that analytic expressions could also be worked out for the case of indefinite quadratic forms. The resulting closed form density and distribution functions converge rapidly and provide accurate approximations over the entire support of the distribution.

Existing expansions that are expressed in terms of rescaled chi-square density functions and Laguerre polynomials such as those discussed in Kotz *et al.* (1967a,b) for the case of positive definite quadratic forms, were derived by making use of a different technique. As in the case of Edgeworth-type expansions whose leading terms are Gaussian density or distribution functions, such representations cannot converge as quickly as the proposed expansion, which is more appropriately based on a gamma density function whose first two moments match those of the target distribution. It should also be pointed out that the saddlepoint approximation and Imhof's formula, which incidentally is not closed form, need to be recalculated at each point of the distribution. Moreover, as can be seen for instance from Huzurbazar (1999), Figure 2, the saddlepoint approximation may not be accurate throughout the entire range of the distribution.

As will be explained, the results also apply to ratios of certain quadratic forms. Such ratios arise for example in regression theory, linear models, analysis of variance and time series.

A representation of non-central indefinite quadratic forms, which relies on the spectral decomposition theorem, is derived in Section 2; a formula for determining their moments in terms of their cumulants is provided as well. A so-called Laguerre polynomial approximation of the density function of a positive definite quadratic form, which is expressed as the product of a gamma density function and a single polynomial, is introduced in Section 3; explicit representations of the resulting density and distribution functions of an indefinite quadratic form are also given. We note that the expansions are expressed in terms of Laguerre polynomials (or their coefficients) since their associated weight functions are proportional to gamma density functions, which are suitable for approximating the distribution of positive linear combinations of chi-square random variables. An algorithm describing the methodology is provided in Section 4.

Several numerical examples, including an application of the proposed technique to the Durbin–Watson statistic, are presented in Section 5. Finally, certain computational considerations are discussed in Section 6.

2. THE MOMENTS OF A LINEAR COMBINATION OF CHI-SQUARE RANDOM VARIABLES

Since linear combinations of possibly non-central chi-square random variables can be expressed in terms of quadratic forms, we shall provide a representation of the moments of the latter in this section. These moments are required in order to implement the proposed density approximation methodology.

Indefinite quadratic forms in normal random variables can be expressed in terms of standard normal variables as follows. Let $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ where Σ is a positive definite covariance matrix. On letting $\mathbf{Z} \sim \mathcal{N}_p(\mathbf{0}, I)$, where I is a $p \times p$ identity matrix, one has $\mathbf{X} = \Sigma^{\frac{1}{2}}\mathbf{Z} + \boldsymbol{\mu}$ where $\Sigma^{\frac{1}{2}}$ denotes the symmetric square root of Σ. Then, the quadratic form $Q = \mathbf{X}' A \mathbf{X}$ where A is a $p \times p$ real symmetric matrix and $\mathbf{X}'$ denotes the transpose of $\mathbf{X}$ can be expressed as follows:

$$
\begin{aligned}
Q &= \big(\mathbf{Z} + \Sigma^{-\frac{1}{2}}\boldsymbol{\mu}\big)' \,\Sigma^{\frac{1}{2}} A\, \Sigma^{\frac{1}{2}} \big(\mathbf{Z} + \Sigma^{-\frac{1}{2}}\boldsymbol{\mu}\big) \\
&= \big(\mathbf{Z} + \Sigma^{-\frac{1}{2}}\boldsymbol{\mu}\big)' P P' \,\Sigma^{\frac{1}{2}} A\, \Sigma^{\frac{1}{2}} P P' \big(\mathbf{Z} + \Sigma^{-\frac{1}{2}}\boldsymbol{\mu}\big)
\end{aligned}
\tag{2.1}
$$

where P is an orthogonal matrix that diagonalizes $\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$, that is, $P' \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} P = \text{diag}(\lambda_1, ..., \lambda_p)$, $\lambda_1, ..., \lambda_p$ being the eigenvalues of $\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$ (or equivalently those of $A\Sigma$) in *decreasing* order. Let $\mathbf{v}_i$ denote the *normalized* eigenvector of $\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$ *corresponding to* λ_i (such that $\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} \mathbf{v}_i = \lambda_i \mathbf{v}_i$ *and* $\mathbf{v}_i{}' \mathbf{v}_i = 1$), $i = 1, ..., p$, and $P = (\mathbf{v}_1, ..., \mathbf{v}_p)$. Letting $\mathbf{U} = P'\mathbf{Z}$, one has $\mathbf{U} \sim \mathcal{N}_p(\mathbf{0}, I)$ since P is an orthogonal matrix, and then, according to the spectral decomposition theorem,

$$
\begin{aligned}
Q &= (\mathbf{U} + \mathbf{b})' \,\text{diag}(\lambda_1, ..., \lambda_p)\, (\mathbf{U} + \mathbf{b}) \\
&= \sum_{j=1}^{p} \lambda_j (U_j + b_j)^2
\end{aligned}
\tag{2.2}
$$

where $\text{diag}(\lambda_1, ..., \lambda_p)$ is a diagonal matrix whose diagonal elements are $\lambda_1, ..., \lambda_p$, $\mathbf{b} = P' \Sigma^{-\frac{1}{2}} \boldsymbol{\mu}$ with $\mathbf{b} = (b_1, ..., b_p)'$, $\mathbf{U} = (U_1, ..., U_p)'$, and $U_j + b_j$ are independently distributed $\mathcal{N}(b_j, 1)$ random variables, $j = 1, ..., p$. Thus,

$$
\begin{aligned}
Q &= \sum_{j=1}^{r} \lambda_j\, (U_j + b_j)^2 \;-\; \sum_{j=r+\theta+1}^{p} |\lambda_j|\, (U_j + b_j)^2 \\
&\equiv Q_1 - Q_2 \;,
\end{aligned}
\tag{2.3}
$$

where r is the number of positive eigenvalues of $A\Sigma$ and $p - r - \theta$ is the number of negative eigenvalues of $A\Sigma$, θ being the number of null eigenvalues. Consequently, a non-central indefinite quadratic form, Q, can be expressed as a difference of independently distributed linear combinations of independent non-central chi-square random variables having one degree of freedom each. This will be referred to as a general linear combination of such variables. It should be noted that the chi-square random variables are central whenever $\boldsymbol{\mu} = \mathbf{0}$. When $A \geq 0$, Q is a positive semidefinite quadratic form, and $Q \sim Q_1$ as defined in Equation (2.3). We note that if A is not symmetric, it suffices to replace this matrix by $(A+A')/2$, which results in the same quadratic form. Accordingly, it will be assumed without any loss of generality that the matrices of the quadratic forms being considered are symmetric.

As shown in Mathai and Provost (1992), the s^{th} cumulant of $\mathbf{X}'A\mathbf{X}$ where $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ is

$$k(s) = 2^{s-1} s! \Big(\mathrm{tr}(A\Sigma)^s / s + \boldsymbol{\mu}'(A\Sigma)^{s-1} A \boldsymbol{\mu} \Big) , \tag{2.4}$$

$\mathrm{tr}(\cdot)$ denoting the trace of $(\cdot)$. It should be noted that $\mathrm{tr}(A\Sigma)^s = \sum_{j=1}^{p} \lambda_j^s$ where the λ_j's, $j=1, ..., p$, are the eigenvalues of $A\Sigma$. The moments of a random variable can be obtained from its cumulants by means of a recursive relationship that is derived for instance in Smith (1995). Accordingly, the h^{th} moment of $\mathbf{X}'A\mathbf{X}$ is given by

$$\mu(h) = \sum_{i=0}^{h-1} \frac{(h-1)!}{(h-1-i)!\, i!} \, k(h-i) \, \mu(i) , \tag{2.5}$$

where $k(s)$ is as specified by Equation (2.4).

One can make use of Equation (2.5) to determine the moments of the positive definite quadratic forms, $Q_1 \equiv \mathbf{W}_1' A_1 \mathbf{W}_1$ and $Q_2 \equiv \mathbf{W}_2' A_2 \mathbf{W}_2$, appearing in Equation (3) where $A_1 = \mathrm{diag}(\lambda_1, ..., \lambda_r)$, $A_2 = \mathrm{diag}(|\lambda_{r+\theta+1}|, ..., |\lambda_p|)$, $\mathbf{W}_1 \sim \mathcal{N}_r(\boldsymbol{b_1}, I)$ with $\boldsymbol{b_1} = (b_1, ..., b_r)'$, and $\mathbf{W}_2 \sim \mathcal{N}_{p-r-\theta}(\boldsymbol{b_2}, I)$ with $\boldsymbol{b_2} = (b_{r+\theta+1}, ..., b_p)'$, the b_j's being as defined in Equation (2.2).

Since an indefinite quadratic form is distributed as the difference of two positive definite quadratic forms, its density function can be obtained via the transformation of variables technique. For the problem at hand, letting $h_Q(q)$, $f_{Q_1}(q_1)$ and $f_{Q_2}(q_2)$ respectively denote the approximate densities of Q, Q_1 and Q_2, the approximate density function of the indefinite quadratic form Q is given by

$$h_Q(q) = \begin{cases} h_P(q) & \text{for } q \geq 0 , \\ h_N(q) & \text{for } q < 0 , \end{cases} \tag{2.6}$$

where

$$h_P(q) = \int_0^{\infty} f_{Q_1}(q+x) \, f_{Q_2}(x) \, \mathrm{d}x , \tag{2.7}$$

$$h_N(q) = \int_{-q}^{\infty} f_{Q_1}(q+x)\, f_{Q_2}(x)\, \mathrm{d}x \ , \tag{2.8}$$

and $h_P(q)$ and $h_N(q)$ are explicitly given in the next section.

3. LAGUERRE POLYNOMIAL DENSITY APPROXIMANTS

In order to approximate the distribution of a positive definite quadratic form, it is appropriate to make use of an approximation that is based on Laguerre polynomials since their associated weight function is proportional to a gamma density function with parameters $\alpha \equiv \nu+1$ and $\beta = 1$. Accordingly, letting Y be a gamma-type random variable whose exact raw moments are denoted by $\mu_Y(h)$, $h = 0, 1, ..., d$, we first approximate the distribution of $X = Y/\beta$ where β, the second parameter of the gamma approximation, which can be easily obtained by matching moments, is given by

$$\beta = \frac{\mu_Y(2)}{\mu_Y(1)} - \mu_Y(1) \ . \tag{3.1}$$

Similarly, the shape parameter ν in the weight function is determined as follows:

$$\nu = \frac{\mu_Y^2(1)}{\mu_Y(2) - \mu_Y^2(1)} - 1 \ . \tag{3.2}$$

Let $L_i^{\nu}(x)$ denote the i^{th} degree Laguerre polynomial with parameter ν, that is,

$$L_i^{\nu}(x) = \sum_{k=0}^{i} d_{i,k}^{\nu}\, x^k \tag{3.3}$$

where

$$d_{i,k}^{\nu} = \frac{(-1)^{i-k}\, \Gamma(i+\nu+1)}{(i-k)!\, k!\, \Gamma(\nu+k+1)} \ . \tag{3.4}$$

As explained in Provost and Ha (2009), on equating $\int_0^\infty L_i^{\nu}(y)\, f(y)\, \mathrm{d}y$ to $\int_0^\infty L_i^{\nu}(y)$ $f_{Y_d}(y)\, \mathrm{d}y$ for $h = 0, 1, ..., d$, where $f(y)$ is the exact density function being approximated and $f_{Y_d}(x)$ denotes the representation of the approximate density function given in Equation (3.19) (which, incidentally, is equivalent to assuming that the first d moments of the approximate distribution coincide with those of the target distribution), one can determine the coefficients of the Laguerre polynomials by making use of their orthogonality property. Then, by collecting the coefficients of each monomial x^k in the resulting representation, one can express the d^{th} degree Laguerre polynomial density approximant as

$$g_{X_d}(x) = c_\nu\, w_\nu(x) \sum_{k=0}^{d} \xi_{\nu,k}\, x^k \ , \tag{3.5}$$

where

$$c_\nu = 1/\Gamma(\nu+1) , \tag{3.6}$$

$$w_\nu(x) = x^\nu e^{-x} , \tag{3.7}$$

and the coefficients $\xi_{\nu,k}$ can be obtained as

$$\xi_{\nu,k} = \begin{cases} 1 + \sum_{i=2}^{d} \eta_i^\nu d_{i,k}^\nu , & \text{for } k = 0 ; \\ \sum_{i=2}^{d} \eta_i^\nu d_{i,k}^\nu , & \text{for } k = 1 , \\ \sum_{i=k}^{d} \eta_i^\nu d_{i,k}^\nu , & \text{for } k = 2, ..., d , \end{cases} \tag{3.8}$$

with

$$\eta_i^\nu = \frac{i!}{\Gamma(\nu+i+1)} \sum_{k=0}^{i} d_{i,k}^\nu \, \mu_X(k) \tag{3.9}$$

and

$$\mu_X(k) = \mu_Y(k)/\beta^k . \tag{3.10}$$

Thus, the representation of the approximate density function given in Equation (3.5) can be viewed as a mixture of $d+1$ gamma densities with parameters $\nu+k+1$ and 1. The density function of the random variable Y can then be obtained from $g_{X_d}(x)$ as specified in Equation (3.5) via the transformation $Y = \beta X$ as

$$f_{Y_d}(y) = g_{X_d}(y/\beta)/\beta . \tag{3.11}$$

This form of the approximate density function lends itself more readily to algebraic manipulations than that specified in Equation (3.19), which may be somewhat simpler to evaluate.

The corresponding approximate cumulative distribution function of Y evaluated at $c_0 > 0$ is then

$$\begin{aligned} F_{Y_d}(c_0) &= \int_0^{c_0} g_{X_d}(y/\beta)/\beta \, \mathrm{d}y \\ &= \int_0^{c_0/\beta} g_{X_d}(x) \, \mathrm{d}x \\ &= \int_0^{c_0/\beta} c_\nu \, w_\nu(x) \sum_{k=0}^{d} \xi_{\nu,k} \, x^k \, \mathrm{d}t \\ &= \sum_{i=0}^{d} \xi_{\nu,i} \frac{\Gamma(\nu+i+1) - \Gamma(\nu+i+1, c_0/\beta)}{\Gamma(\nu+1)} , \end{aligned} \tag{3.12}$$

where

$$\Gamma(a,\theta) = \int_\theta^\infty t^{a-1} e^{-t}\,\mathrm{d}t \tag{3.13}$$

denotes the incomplete gamma function. Conditions ensuring that the proposed approximants, whether applied to quadratic forms or random variables having an asymptotic chi-square distribution, will converge to their exact density functions, are available in Alexits (1961, p. 304).

The density functions of Q_1 and Q_2 as defined in Equation (2.3) can be approximated from their respective moments which can be determined in Equation (2.5). The density of an indefinite quadratic form $Q = Q_1 - Q_2$, where Q_1 and Q_2 are positive definite quadratic forms, can then be approximated by making use of Equation (2.6) where $f_{Q_1}(\cdot)$ and $f_{Q_2}(\cdot)$ respectively denote the Laguerre polynomial density approximants of Q_1 and Q_2, which are available from Equation (3.11). Explicit representations of $h_P(q)$ and $h_N(q)$ as specified by Equations (2.7) and (2.8), respectively, can be obtained as follows. When q is positive, the probability density function of Q is given by

$$\begin{aligned} h_P(q) &= \int_0^\infty f_{Q_1}(q+y)\, f_{Q_2}(y)\,\mathrm{d}y \\ &= \int_0^\infty \Big(\gamma_{\nu_1,\beta_1}(q+y) \sum_{i=0}^d \xi_{\nu_1,i} \Big(\frac{q+y}{\beta_1}\Big)^i\Big) \Big(\gamma_{\nu_2,\beta_2}(y) \sum_{j=0}^d \xi_{\nu_2,j} \Big(\frac{y}{\beta_2}\Big)^j\Big)\,\mathrm{d}y \end{aligned} \tag{3.14}$$

with $\gamma_{\nu_\ell,\beta_\ell}(z) = z^{\nu_\ell}\, e^{-z/\beta_\ell}/(\beta_\ell^{\nu_\ell+1}\,\Gamma(\nu_\ell+1))$, $\ell = 1,2$; ν_ℓ and β_ℓ determined from Equations (3.1) and (3.2), respectively, $\ell = 1,2$, the coefficients $\xi_{\nu_1,i}$ and $\xi_{\nu_2,i}$ being as defined in Equation (3.8). Identities 3.384 3 and 9.220 4 from Gradshteyn and Ryzhik (1980) yield

$$\begin{aligned} h_P(q) &= \sum_{i=0}^d \sum_{j=0}^d \xi_{\nu_1,i}\,\xi_{\nu_2,j} \int_0^\infty \Big(\frac{q+y}{\beta_1}\Big)^i \Big(\frac{y}{\beta_2}\Big)^j \gamma_{\nu_1,\beta_1}(q+y)\,\gamma_{\nu_2,\beta_2}(y)\,\mathrm{d}y \\ &= \sum_{i=0}^d \sum_{j=0}^d \frac{\xi_{\nu_1,i}\,\xi_{\nu_2,j}\, e^{-q/\beta_1}}{\beta_1^{\nu_1+i+1}\,\beta_2^{\nu_2+j+1}\,\Gamma(\nu_1+1)\,\Gamma(\nu_2+1)} \bigg(\Big(\frac{\beta_1+\beta_2}{\beta_1\beta_2}\Big)^{-1-i-j-\nu_1-\nu_2} \\ &\quad \times \Gamma\big(i+j+\nu_1+\nu_2+1\big)\,{}_1F_1\big(-i-\nu_1, -i-j-\nu_1-\nu_2, q(\beta_1+\beta_2)/(\beta_1\beta_2)\big) \\ &\quad + \frac{\Gamma\big(-1-i-j-\nu_1-\nu_2\big)\,\Gamma(j+\nu_2+1)}{\Gamma(-i-\nu_1)}\, q^{i+j+\nu_1+\nu_2+1} \\ &\quad \times {}_1F_1\big(j+\nu_2+1,\, i+j+\nu_1+\nu_2+2,\, q(\beta_1+\beta_2)/(\beta_1\beta_2)\big)\bigg)\,, \end{aligned} \tag{3.15}$$

where

$${}_1F_1(a,b,z) = \sum_{k=0}^\infty \frac{\Gamma(a+k)\,\Gamma(b)\,z^k}{\Gamma(a)\,\Gamma(b+k)\,k!}$$

is Kummer's confluent hypergeometric function. Similarly, when q is negative, one has

$$
\begin{aligned}
h_N(q) &= \int_{-q}^{\infty}\Big(\gamma_{\nu_1,\beta_1}(q+y)\sum_{i=0}^{d}\xi_{\nu_1,i}\Big(\frac{q+y}{\beta_1}\Big)^i\Big)\Big(\gamma_{\nu_2,\beta_2}(y)\sum_{j=0}^{d}\xi_{\nu_2,j}\Big(\frac{y}{\beta_2}\Big)^j\Big)\,\mathrm{d}y\\
&= \int_{0}^{\infty}\Big(\gamma_{\nu_1,\beta_1}(w)\sum_{i=0}^{d}\xi_{\nu_1,i}\Big(\frac{w}{\beta_1}\Big)^i\Big)\Big(\gamma_{\nu_2,\beta_2}(w-q)\sum_{j=0}^{d}\xi_{\nu_2,j}\Big(\frac{w-q}{\beta_2}\Big)^j\Big)\,\mathrm{d}w\\
&= \sum_{i=0}^{d}\sum_{j=0}^{d}\frac{\xi_{\nu_1,i}\,\xi_{\nu_2,j}}{\beta_1^i\,\beta_2^j}\int_0^\infty w^i\,(w-q)^j\,\gamma_{\nu_1,\beta_1}(w)\,\gamma_{\nu_2,\beta_2}(w-q)\,\mathrm{d}w\\
&= \sum_{i=0}^{d}\sum_{j=0}^{d}\frac{\xi_{\nu_1,i}\,\xi_{\nu_2,j}\,e^{q/\beta_2}}{\beta_1^{\nu_1+i+1}\,\beta_2^{\nu_2+j+1}\,\Gamma(\nu_1+1)\,\Gamma(\nu_2+1)}\Big(\Big(\frac{\beta_1+\beta_2}{\beta_1\,\beta_2}\Big)^{-1-i-j-\nu_1-\nu_2}\\
&\quad\times\Gamma\big(i+j+\nu_1+\nu_2+1\big)\,{}_1F_1\big(-j-\nu_2,-i-j-\nu_1-\nu_2,-q(\beta_1+\beta_2)/(\beta_1\,\beta_2)\big)\\
&\quad+\frac{\Gamma(-1-i-j-\nu_1-\nu_2)\,\Gamma(i+\nu_1+1)}{\Gamma(-j-\nu_2)}\,(-q)^{i+j+\nu_1+\nu_2+1}\\
&\quad\times{}_1F_1\big(i+\nu_1+1,\,i+j+\nu_1+\nu_2+2,\,-q(\beta_1+\beta_2)/(\beta_1\,\beta_2)\big)\Big)\,. \qquad (3.16)
\end{aligned}
$$

We note that the series representations given in Equations (3.15) and (3.16) do not converge when ν_1 or ν_2 are integer-valued. In this case, one would have to evaluate the integral representations by numerical integration.

The corresponding cumulative distribution function is then obtained by integration. The approximate cumulative distribution function for the negative part of Q is given by

$$
\begin{aligned}
H_N(y) &= \int_{-\infty}^{y} h_N(q)\,\mathrm{d}q\\
&= \sum_{i=0}^{d}\sum_{j=0}^{d}\frac{\xi_{\nu_1,i}\,\xi_{\nu_2,j}}{\beta_1^{\nu_1+i+1}\,\beta_2^{\nu_2+j+1}\,\Gamma(\nu_1+1)\,\Gamma(\nu_2+1)}\Big(\Big(\frac{\beta_1+\beta_2}{\beta_1\,\beta_2}\Big)^{-1-i-j-\nu_1-\nu_2}\\
&\quad\times\Gamma(i+j+\nu_1+\nu_2+1)\\
&\quad\times\int_{-\infty}^{y} e^{q/\beta_2}\,{}_1F_1\big(-j-\nu_2,-i-j-\nu_1-\nu_2,-q(\beta_1+\beta_2)/(\beta_1\,\beta_2)\big)\,\mathrm{d}q \qquad (3.17)\\
&\quad+\frac{\Gamma(-1-i-j-\nu_1-\nu_2)\,\Gamma(i+\nu_1+1)}{\Gamma(-j-\nu_2)}\\
&\quad\times\int_{-\infty}^{y}(-q)^{i+j+\nu_1+\nu_2+1}\,e^{q/\beta_2}\\
&\qquad\times{}_1F_1\big(i+\nu_1+1,\,i+j+\nu_1+\nu_2+2,\,-q(\beta_1+\beta_2)/(\beta_1\,\beta_2)\big)\,\mathrm{d}q\Big) =
\end{aligned}
$$

$$= \sum_{i=0}^{d}\sum_{j=0}^{d}\sum_{k=0}^{\infty} \frac{\xi_{\nu_1,i}\,\xi_{\nu_2,j}\,(\beta_1+\beta_2)^k}{\beta_1^{\nu_1+k+i+1}\,\beta_2^{\nu_2+k+j+1}\,\Gamma(\nu_1+1)\,\Gamma(\nu_2+1)} \Bigg(\Big(\frac{\beta_1+\beta_2}{\beta_1\,\beta_2}\Big)^{-1-i-j-\nu_1-\nu_2}$$
$$\times\ \Gamma(1+i+j+\nu_1+\nu_2)\,\frac{\Gamma(-j+k-\nu_2)\,\Gamma(-i-j-\nu_1-\nu_2)}{\Gamma(-j-\nu_2)\,\Gamma(-i-j+k-\nu_1-\nu_2)\,k!}$$
$$\times \int_{-\infty}^{y}(-q)^k\,e^{q/\beta_2}\mathrm{d}q + \frac{\Gamma(-1-i-j-\nu_1-\nu_2)}{\Gamma(-j-\nu_2)}$$
$$\times\ \frac{\Gamma(i+k+\nu_1+1)\,\Gamma(i+j+\nu_1+\nu_2+2)}{\Gamma(i+j+k+\nu_1+\nu_2+2)\,k!} \int_{-\infty}^{y}(-q)^{i+j+k+\nu_1+\nu_2+1}\,e^{q/\beta_2}\,\mathrm{d}q\Bigg)$$

(3.17)

$$= \sum_{i=0}^{d}\sum_{j=0}^{d}\sum_{k=0}^{\infty} \frac{\xi_{\nu_1,i}\,\xi_{\nu_2,j}\,(\beta_1+\beta_2)^k}{\beta_1^{\nu_1+k+i+1}\,\beta_2^{\nu_2+k+j+1}\,\Gamma(\nu_1+1)\,\Gamma(\nu_2+1)} \Bigg(\Big(\frac{\beta_1+\beta_2}{\beta_1\,\beta_2}\Big)^{-1-i-j-\nu_1-\nu_2}$$
$$\times\ \Gamma(1+i+j+\nu_1+\nu_2)\frac{\Gamma(-j+k-\nu_2)\,\Gamma(-i-j-\nu_1-\nu_2)\,\beta_2^{k+1}\,\Gamma(k+1,-y/\beta_2)}{\Gamma(-j-\nu_2)\,\Gamma(-i-j+k-\nu_1-\nu_2)\,k!}$$
$$+\ \frac{\Gamma(-1-i-j-\nu_1-\nu_2)}{\Gamma(-j-\nu_2)}\ \frac{\Gamma(i+k+\nu_1+1)\,\Gamma(i+j+\nu_1+\nu_2+2)}{\Gamma(i+j+k+\nu_1+\nu_2+2)\,k!}$$
$$\times\ \beta_2^{i+j+k+\nu_1+\nu_2+2}\,\Gamma(i+j+k+\nu_1+\nu_2+2,-y/\beta_2)\Bigg)\,.$$

Similarly, the approximate cumulative distribution function for the positive part of Q can be expressed as follows:

$$H_P(y)\ =\ H_N(0)+\int_0^y h_P(q)\,\mathrm{d}q$$
$$=\ H_N(0)+\sum_{i=0}^{d}\sum_{j=0}^{d}\ \frac{\xi_{\nu_1,i}\,\xi_{\nu_2,j}}{\beta_1^{\nu_1+i+1}\,\beta_2^{\nu_2+j+1}\,\Gamma(\nu_1+1)\,\Gamma(\nu_2+2)}$$
$$\times\ \Bigg(\Big(\frac{\beta_1+\beta_2}{\beta_1\,\beta_2}\Big)^{-1-i-j-\nu_1-\nu_2}\ \Gamma(i+j+\nu_1+\nu_2+1)$$
$$\times \int_0^y e^{-q/\beta_1}\,{}_1F_1\big(-i-\nu_1,\ -i-j-\nu_1-\nu_2,\ q(\beta_1+\beta_2)/(\beta_1\,\beta_2)\big)\,\mathrm{d}q$$
$$+\ \frac{\Gamma(-1-i-j-\nu_1-\nu_2)\,\Gamma(j+\nu_2+1)}{\Gamma(-i-\nu_1)}$$
$$\times \int_0^y q^{i+j+\nu_1+\nu_2+1}\,e^{-q/\beta_1}\,{}_1F_1\big(j+\nu_2+1, i+j+\nu_1+\nu_2+2, q(\beta_1+\beta_2)/(\beta_1\,\beta_2)\big)\,\mathrm{d}q\Bigg)$$

(3.18)

$$=\ H_N(0)+\sum_{i=0}^{d}\sum_{j=0}^{d}\sum_{k=0}^{\infty}\ \frac{\xi_{\nu_1,i}\,\xi_{\nu_2,j}}{\beta_1^{\nu_1+i+1}\,\beta_2^{\nu_2+j+1}\,\Gamma(\nu_1+1)\,\Gamma(\nu_2+1)}$$
$$\times\Bigg(\Big(\frac{\beta_1+\beta_2}{\beta_1\,\beta_2}\Big)^{-1-i-j+k-\nu_1-\nu_2}\ \frac{\Gamma(1+i+j+\nu_1+\nu_2)\,\Gamma(-i-\nu_1+k)\,\Gamma(-i-j-\nu_1-\nu_2)}{\Gamma(-i-\nu_1)\,\Gamma(-i-j+k-\nu_1-\nu_2)\,k!}$$
$$\times\ \beta_1^{k+1}\Big(\Gamma(1+k)-\Gamma(1+k,y/\beta_1)\Big)$$
$$+\ \frac{\Gamma(-1-i-j-\nu_1-\nu_2)}{\Gamma(-i-\nu_1)}\ \frac{\Gamma(j+k+\nu_2+1)\,\Gamma(i+j+\nu_1+\nu_2+2)}{\Gamma(i+j+k+\nu_1+\nu_2+2)\,k!}$$
$$\times\ \beta_1^{i+j+k+\nu_1+\nu_2+2}\Big(\Gamma(i+j+k+\nu_1+\nu_2+2)-\Gamma(i+j+k+\nu_1+\nu_2+2,y/\beta_1)\Big)\Bigg)$$

where

$$\begin{aligned} H_N(0) \;=\; & \sum_{i=0}^{d}\sum_{j=0}^{d}\sum_{k=0}^{\infty} \frac{\xi_{\nu_1,i}\,\xi_{\nu_2,j}\,(\beta_1+\beta_2)^k}{\beta_1^{\nu_1+k+i+1}\,\beta_2^{\nu_2+k+j+1}\,\Gamma(\nu_1+1)\,\Gamma(\nu_2+1)} \\ & \times \left(\left(\frac{\beta_1+\beta_2}{\beta_1\,\beta_2}\right)^{-1-i-j-\nu_1-\nu_2} \Gamma(1+i+j+\nu_1+\nu_2) \right. \\ & \times \frac{\Gamma(-j+k-\nu_2)\,\Gamma(-i-j-\nu_1-\nu_2)\,\beta_2^{k+1}\,\Gamma(k+1)}{\Gamma(-j-\nu_2)\,\Gamma(-i-j+k-\nu_1-\nu_2)\,k!} \\ & + \frac{\Gamma(-1-i-j-\nu_1-\nu_2)}{\Gamma(-j-\nu_2)}\,\frac{\Gamma(i+k+\nu_1+1)\,\Gamma(i+j+\nu_1+\nu_2+2)}{k!} \\ & \left. \times\, \beta_2^{i+j+k+\nu_1+\nu_2+2} \right). \end{aligned}$$

Even though the sum over k has infinitely many summands, we observed that fifty terms provide sufficient accuracy. In most cases of interest, a suitable degree for a density approximation can be determined by a *de visu* inspection of the density plots of approximants of successive degrees. More specifically, one might be satisfied that an approximant of degree $d+1$ is adequate if no noticeable differences are observed when comparing the plots of approximants of degrees d and $d+2$. This criterion was applied to all the examples presented in Section 5. Equivalently, one may wish to set a tolerance for the integrated absolute difference of approximants of successive degrees and select the number of terms to be used in the approximation accordingly. If one wishes to determine the number of terms required to obtain a satisfactory approximation for a specific percentile, one could evaluate the percentile approximations for successive values of d until convergence is observed or a preset tolerance value exceeds the difference of two successive approximations. Since we are dealing with a sequence of approximants converging to the exact density function, the close proximity of successive approximants indicates that convergence is nearly attained. Bounds for the integrated absolute and squared truncation errors are obtained in the remainder of this section. In light of Equation (3.19) of Provost (2005), the truncated density function corresponding to that given in Equation (3.11) can be expressed as

$$f_{Y_d}(y) \;=\; \frac{y^{\nu}e^{-y/\beta}}{\beta^{\nu+1}}\sum_{j=0}^{d}\delta_j^{\nu}L_j^{\nu}(y/\beta) \tag{3.19}$$

with $L_j^{\nu}(\cdot)$ as defined in Equation (11) and

$$\delta_j^{\nu} \;=\; \sum_{k=0}^{j}\frac{(-1)^k\,j!\,\mu_X(j-k)}{k!\,(j-k)!\,\Gamma(\nu+j-k+1)}\;.$$

Let $F_{Y_d}(y)$ and $F_Y(y)$ respectively denote the cumulative distribution functions of Y_d and Y and $f_Y(y)$ denote the density function being approximated.

Letting

$$\delta_j^\nu = \frac{j!}{\Gamma(\nu+j+1)}\;\psi_j^\nu$$

where

$$\psi_j^\nu = \sum_{k=0}^{j} \frac{(-1)^k\,\Gamma(\nu+j+1)\;\mu_X(j-k)}{k!\;(j-k)!\;\Gamma(\nu+j-k+1)}\;,$$

a bound for the truncation error with respect to the probability density function of Y can be determined as follows:

$$\begin{aligned} \mathcal{E}_d(y) &= \bigl|f_Y(y) - f_{Y_d}(y)\bigr| \\ &= \frac{y^\nu e^{-y/c}}{c^{\nu+1}} \sum_{j=d+1}^{\infty} \frac{j!}{\Gamma(\nu+j+1)}\;|\psi_j^\nu|\,|L_j(\nu,y/c)|\;, \end{aligned} \tag{3.20}$$

where according to Szegö (1975),

$$\begin{aligned} L_j(\nu,y/c) &\le \frac{(\nu+1)_j}{j!}\;e^{y/(2c)} \\ &= \frac{\Gamma(\nu+1+j)}{\Gamma(\nu+1)\,j!}\;e^{y/(2c)}\;. \end{aligned} \tag{3.21}$$

Thus,

$$\mathcal{E}_d(y) \le \frac{y^\nu e^{-y/(2c)}}{c^{\nu+1}\,\Gamma(\nu+1)} \sum_{j=d+1}^{\infty} |\psi_j^\nu|\;, \tag{3.22}$$

and letting $\lambda_d = \sum_{j=d+1}^{\infty} |\psi_j^\nu|$, a bound for e_d, the integrated absolute truncation error, can be obtained as follows:

$$\begin{aligned} e_d &= \int_0^\infty \mathcal{E}_d(y)\,\mathrm{d}y \\ &\le \int_0^\infty \lambda_d\,\frac{y^\nu e^{-y/(2c)}}{c^{\nu+1}\,\Gamma(\nu+1)}\;\mathrm{d}y \\ &= 2^{\nu+1}\lambda_d \\ &= 2^{\nu+1} \sum_{j=d+1}^{\infty} \left| \sum_{k=0}^{j} \frac{(-1)^k\,\Gamma(\nu+j+1)\;\mu_X(j-k)}{k!\;(j-k)!\;\Gamma(\nu+j-k+1)} \right|. \end{aligned} \tag{3.23}$$

This result yields a bound for the distribution function integrated absolute error:

$$\begin{aligned} |F_Y(y) - F_{Y_d}(y)| &= \left| \int_0^y \bigl(f_Y(y) - f_{Y_d}(y)\bigr)\,\mathrm{d}x \right| \\ &\le \int_0^\infty \bigl|f_Y(y) - f_{Y_d}(y)\bigr|\,\mathrm{d}x \\ &\le 2^{\nu+1}\lambda_d\;. \end{aligned} \tag{3.24}$$

A bound for the density function integrated squared error can be similarly obtained:

$$
\begin{aligned}
e_d^* &= \int_0^\infty \mathcal{E}_d^2(y)\,\mathrm{d}y \\
&\le \int_0^\infty \lambda_d^2 \,\frac{y^{2\nu} e^{-y/c}}{c^{2(\nu+1)}\,\Gamma^2(\nu+1)}\,\mathrm{d}y \qquad (3.25)\\
&= \frac{\lambda_d^2\,\Gamma(2\nu+1)}{c\,\Gamma^2(\nu+1)}\,.
\end{aligned}
$$

Admittedly, these bounds are not very tight. Moreover, a precise order of convergence cannot be determined since these error bounds depend on the moments of the distribution being approximated.

4. THE ALGORITHM

The following algorithm can be utilized to approximate the density function of the quadratic form $Q = \mathbf{X}'A\mathbf{X}$ where $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu},\, \Sigma)$, $\Sigma > 0$ and A is an indefinite symmetric real matrix.

1. The eigenvalues of $A\Sigma$ denoted by $\lambda_1 \ge \cdots \ge \lambda_r > 0 > \lambda_{r+\theta+1} \ge \cdots \ge \lambda_p$, and the *corresponding normalized* eigenvectors, $\boldsymbol{\nu}_1, ..., \boldsymbol{\nu}_p$, are determined.

2. Letting $P = (\boldsymbol{\nu}_1, ..., \boldsymbol{\nu}_p)$, $\gamma_1, ..., \gamma_p$ be the eigenvalues of Σ, $\mathbf{t}_1, ..., \mathbf{t}_p$ be the *normalized* eigenvectors of Σ corresponding to $\gamma_1, ..., \gamma_p$, $T = (\mathbf{t}_1, ..., \mathbf{t}_p)$, $\Sigma^{-1/2} = T\,\mathrm{diag}(\gamma_1^{-1/2}, ..., \gamma_p^{-1/2})\;\; T'$, $\boldsymbol{b} = (b_1, \cdots, b_p)' = P'\,\Sigma^{-1/2}\,\boldsymbol{\mu}$ and the U_j's denote independently distributed standard normal variables, one has the decomposition $Q = \sum_{j=1}^r \lambda_j (U_j + b_j)^2 - \sum_{j=r+\theta+1}^p |\lambda_j| (U_j + b_j)^2 \equiv Q_1 - Q_2$, where $Q_1 \equiv \mathbf{W}_1' A_1 \mathbf{W}_1$, $\mathbf{W}_1 \sim \mathcal{N}_r(\boldsymbol{b_1}, I)$, $\boldsymbol{b_1} = (b_1, ..., b_r)'$, $A_1 = \mathrm{diag}(\lambda_1, ..., \lambda_r)$, and $Q_2 \equiv \mathbf{W}_2' A_2 \mathbf{W}_2$, $\mathbf{W}_2 \sim \mathcal{N}_{p-r-\theta}(\boldsymbol{b_2}, I)$, $\boldsymbol{b_2} = (b_{r+\theta+1}, ..., b_p)'$, $A_2 = \mathrm{diag}(|\lambda_{r+\theta+1}|, ..., |\lambda_p|)$. Clearly, $\mathbf{b} = \mathbf{0}$ whenever $\boldsymbol{\mu} = \mathbf{0}$ and, in that case, there is no need to determine the matrices P or T.

3. The cumulants and the moments of Q_1 and Q_2 are obtained from Equations (2.4) and (2.5), respectively.

4. Laguerre polynomial density approximants, as specified by Equation (3.11), are obtained for each of the positive definite quadratic form Q_1 and Q_2 on the basis of their respective moments and denoted by $f_{Q_1}(\cdot)$ and $f_{Q_2}(\cdot)$. This requires the determination of β_i and ν_i from Equations (3.1) and (3.2) for each Q_i, $i = 1, 2$. The degree d of a given approximant can initially be set equal to 6 and then progressively increased until convergence is observed (graphically or with respect to certain percentiles of interest).

5. Given $f_{Q_1}(\cdot)$ and $f_{Q_2}(\cdot)$, the approximate density of Q is obtained from Equation (2.6) where $h_P(\cdot)$ and $h_N(\cdot)$ are respectively specified by Equation (3.15) and (3.16).
6. The corresponding cumulative distribution function can then be evaluated from Equations (3.17) and (3.18).

Remarks. In the case of a nonnegative definite quadratic form, that is, $Q = \mathbf{X}'A\mathbf{X}$ where $A = A'$ and $A \geq 0$, all the eigenvalues of A are nonnegative and one has $Q = Q_1$ whose approximate density and distribution functions are directly obtained from Equations (3.11) and (3.12), respectively.

5. NUMERICAL EXAMPLES

In this section, the proposed Laguerre polynomial approximation methodology is applied to positive definite and indefinite quadratic forms as well as the Durbin–Watson statistic. In each case, the approximated distribution will either be compared with the exact or simulated distributions. It should be noted that Equations (3.15) and (3.16) can be viewed as closed form representations since the ${}_1F_1$ hypergeometric function can be readily evaluated by most mathematical or statistical computing packages. More precision can be obtained by increasing the degree d of the polynomial adjustment appearing in Equation (13). However, when several successive approximations are seen to be nearly identical, the gain in accuracy becomes minimal. Percentage points were obtained by equating the distribution functions to a given probability and solving the resulting equations numerically. The simulated distribution functions were generated by making use of the Monte Carlo technique.

Example 1.

We first consider the case of a positive definite central quadratic form in independently distributed standard normal variables, which, according to Equation (2.2), can be expressed as

$$Q^I = \mathbf{X}'A\mathbf{X} = \sum_{j=1}^{r} \lambda_j Y_j \ , \tag{5.1}$$

where $A > 0$, $\mathbf{X} \sim \mathcal{N}_p(\mathbf{0}, I)$, λ_j, $j = 1, ..., r$, are the positive eigenvalues of A, the Y_j's, $j = 1, ..., r$ are independently distributed central chi-square random variables, each having one degree of freedom.

In this first example, $\lambda_1 = \lambda_2 = 1$, $\lambda_3 = \lambda_4 = 2.5$, and $\lambda_5 = \lambda_6 = 9$. Since the eigenvalues occur in pairs, the exact density function can be determined from the positive part of Equation (3.23) wherein $\lambda'_k = \lambda_{k/2}$, $s = t = r/2$, $\rho = 0$ and an empty product is interpreted as 1. In this case, with $\nu = 0.77054$ and $\beta = 14.12$, the density function of Q^I can be directly approximated by means of Equation (3.11) in conjunction with Equations (3.1), (3.2) and (3.5). Certain quantiles determined from the exact distribution, the gamma density and the sixth and fourteenth-degree Laguerre polynomial approximant specified by Equation (3.12) are included in Table 1.

The 95$^{\text{th}}$ percentiles obtained from approximants of degrees 4, 6, 8, 10, 12 and 14 are respectively 60.5291, 62.5418, 62.3713, 61.8045, 61.7053 and 61.8384. This sequence suggests that a fourteenth-degree approximant might be sufficiently accurate. The exact 95$^{\text{th}}$ percentile is in fact 61.8999. Certain extreme tail quantiles obtained from the exact density function and the fourteenth-degree Laguerre polynomial approximants are presented in Table 2. Bounds for the integrated absolute and squared errors are plotted in Figure 1 for various values of d. Figure 2 shows exact integrated absolute (left panel) and squared (right panel) differences between the exact and approximate cumulative distribution function versus d.

Table 1: Certain quantiles of Q^I.

CDF	Gamma	Laguerre ($d = 6$)	Laguerre ($d = 14$)	Exact
0.01	1.43483	1.92384	2.51869	2.5795
0.05	3.77669	4.63033	5.04397	5.04193
0.10	5.88517	6.83939	7.03708	7.00919
0.50	20.4832	20.3014	20.0027	20.0400
0.90	50.0482	49.0916	49.3561	49.4183
0.95	61.6596	62.5418	61.8384	61.8999
0.99	87.6053	91.4214	90.9503	90.8707

Table 2: Certain extreme tail quantiles of Q^I.

CDF	Gamma	Laguerre ($d = 6$)	Laguerre ($d = 14$)	Exact
0.0001	0.102918	0.149769	0.403458	0.491026
0.001	0.380511	0.542588	1.00356	1.09778
0.999	123.408	127.632	132.491	132.317
0.9999	158.391	183.558	173.364	173.764

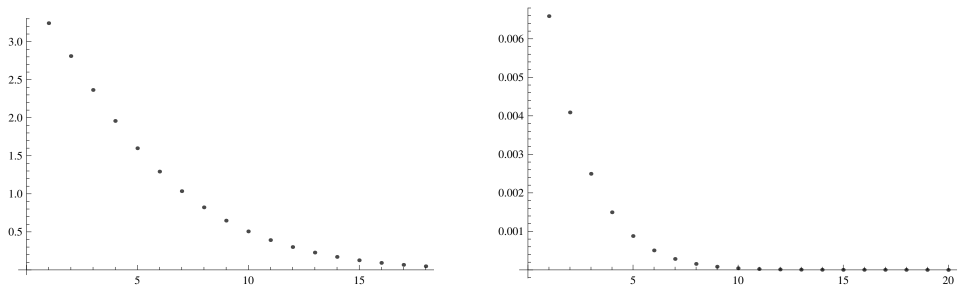

Figure 1: Bounds for the Integrated Absolute (left panel) and Squared (right panel) Truncation Errors with Respect to the Truncation Order.

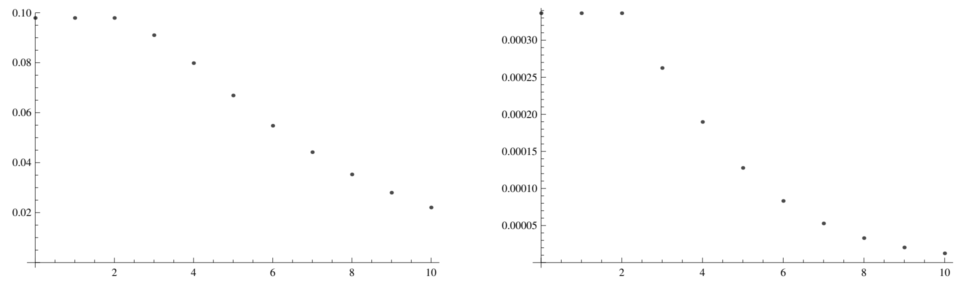

Figure 2: Integrated Absolute Difference (left panel); Integrated Squared Difference (right panel).

Example 2.

We now consider the general case of a non-central indefinite quadratic form, $Q^{II} = \mathbf{X}'A\mathbf{X}$ where

$$A = \begin{pmatrix} 1 & 1 & 2 & 6 \\ 1 & 8 & 0 & 0 \\ 2 & 0 & -1/2 & 1 \\ 6 & 0 & 1 & -2 \end{pmatrix},$$

$\mathbf{X} \sim \mathcal{N}_4(\boldsymbol{\mu}, \Sigma)$ with $\boldsymbol{\mu} = (1, 2, 3, 4)'$ and

$$\Sigma = \begin{pmatrix} 1 & 4/5 & -1/5 & 0 \\ 4/5 & 1 & 1/3 & 1/4 \\ -1/5 & 1/3 & 1 & 0 \\ 0 & 1/4 & 0 & 1 \end{pmatrix}.$$

In light of Equation (2.3), Q^{II} can be re-expressed as

$$Q^{II} = Q_1 - Q_2 = \sum_{i=1}^{2} \lambda_i (U_i + b_i)^2 - \sum_{j=3}^{4} |\lambda_j| \, (U_j + b_j)^2 \tag{5.2}$$

where the U_i's, $i = 1, 2, 3, 4$, are standard normal random variables, $\lambda_1 = 14.487$,

$\lambda_2 = 0.175399$, $\lambda_3 = -1.05353$, $\lambda_4 = -6.30884$, $b_1 = 3.04567$, $b_2 = 7.26373$, $b_3 = -2.10575$, and $b_4 = -2.93822$. Clearly Q^{II} can also be regarded as a general linear combination of non-central chi-square random variables. In this case, the matrices P and $\Sigma^{1/2}$ are respectively

$$P = \begin{pmatrix} 0.552559 & 0.72748 & 0.200698 & 0.353796 \\ 0.537095 & 0.0528413 & -0.257096 & -0.801647 \\ 0.119867 & -0.0192312 & -0.919478 & 0.373928 \\ 0.62597 & -0.683821 & 0.219504 & 0.303922 \end{pmatrix}$$

and

$$\Sigma^{1/2} = \begin{pmatrix} 0.829443 & 0.524395 & -0.186334 & -0.048092 \\ 0.524395 & 0.798945 & 0.248679 & 0.157654 \\ -0.186334 & 0.248679 & 0.950168 & -0.0248771 \\ -0.048092 & 0.157654 & -0.0248771 & 0.986009 \end{pmatrix}.$$

The approximate density functions of Q_1 and Q_2 were obtained by making use of sixth-degree Laguerre polynomial approximants. The resulting approximations of the density and distribution functions of Q^{II} as evaluated from Equations (3.15) and (3.16) and Equations (3.17) and (3.18) with $\nu_1 = 2.05092$, $\beta_1 = 51.8858$, $\nu_2 = 1.99611$ and $\beta_2 = 22.1952$, are plotted in Figure 3. The right panel of Figure 3 also shows the simulated distribution function, which was obtained on the basis of 100,000 replications. Accordingly, the standard error is at most $1/633 \approx 0.0016$.

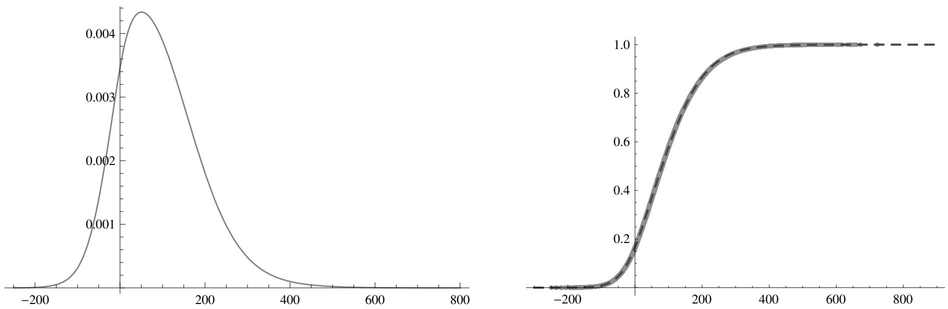

Figure 3: Approximated PDF (left panel);
Simulated and Approximated CDF (right panel).

Example 3.

Consider the following general linear combination of independently distributed central chi-square random variables:

$$Q^{III} = Q_1 - Q_2 = \sum_{i=1}^{r} \lambda_i Y_i - \sum_{j=r+\theta+1}^{p} |\lambda_j|\, Y_j \ , \tag{5.3}$$

where $\theta = 0$, the Y_j's, $j = 1, ..., 16$ are independently distributed central chi-square random variables having one degree of freedom and $\lambda_1 = \lambda_2 = 2$, $\lambda_3 = \lambda_4 = 4$, $\lambda_5 = \lambda_6 = 6$, $\lambda_7 = \lambda_8 = 8$, $\lambda_9 = \lambda_{10} = 10$, $\lambda_{11} = \lambda_{12} = -20$, $\lambda_{13} = \lambda_{14} = -30$ and $\lambda_{15} = \lambda_{16} = -40$.

Since the eigenvalues occur in pairs in the right-hand side of Equation (3.21), Q^{III} can be expressed as

$$Q^{III} = \sum_{i=1}^{s} \lambda'_i \, T_i \; - \sum_{j=s+1}^{t} |\lambda'_j| \, T_j \; , \tag{5.4}$$

where $s = r/2$, $t = p/2$, $\lambda'_k = \lambda_{k/2}$, $k = 1, ..., t$, and the T_i's and T_j's are independently distributed chi-square random variables, each one having *two* degrees of freedom. Imhof (1961) derived the following representation of the exact density function of Q^{III}

$$g(q) = \begin{cases} \sum_{j=1}^{s} \dfrac{\lambda_j'^{\,t-2} e^{-2q/(2\lambda'_j)}}{2\Big(\prod_{k=1,k\neq j}^{s}(\lambda'_j - \lambda'_k)\Big)\Big(\prod_{k=s+1}^{t}(|\lambda'_j| + |\lambda'_k|)\Big)} \; , & q \geq 0 \; , \\[2ex] \sum_{j=s+1}^{t} \dfrac{|\lambda'_j|^{t-2} \, e^{2q/(2|\lambda'_j|)}}{2\Big(\prod_{k=s+1,k\neq j}^{t}(|\lambda'_j| - |\lambda'_k|)\Big)\Big(\prod_{k=1}^{s}(\lambda'_j + \lambda'_k)\Big)} \; , & q < 0 \; . \end{cases} \tag{5.5}$$

The sixth-degree Laguerre polynomial density approximant of Q^{III} as determined from Equations (3.14) and (3.15) is shown in Figure 4, superimposed on the exact density.

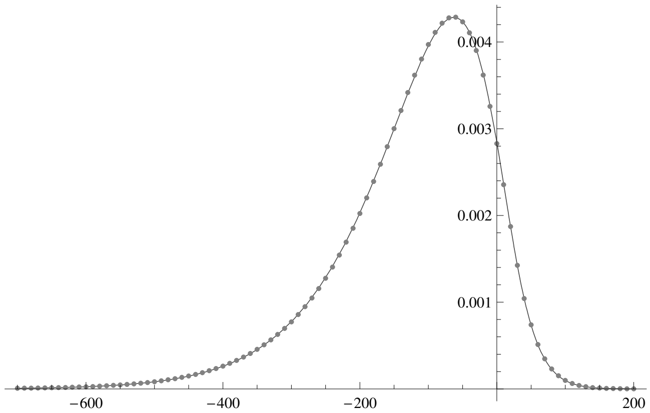

Figure 4: Exact density and Laguerre Polynomial Approximant (dotted line).

Example 4.

The statistic proposed by Durbin and Watson (1950), which in fact assesses whether the errors in the linear regression model

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{5.6}$$

are uncorrelated, can be expressed as

$$D = \frac{\hat{\boldsymbol{\epsilon}}' A^* \hat{\boldsymbol{\epsilon}}}{\hat{\boldsymbol{\epsilon}}' \hat{\boldsymbol{\epsilon}}} \tag{5.7}$$

where

$$\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - X\hat{\boldsymbol{\beta}} \tag{5.8}$$

is the vector of residuals,

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{Y} \tag{5.9}$$

being the ordinary least-squares estimator of $\boldsymbol{\beta}$, and $A^* = (a^*_{ij})$ is a symmetric tridiagonal matrix with $a^*_{11} = a^*_{pp} = 1$; $a^*_{ii} = 2$, for $i = 2, ..., p-1$; $a^*_{ij} = -1$ if $|i-j| = 1$; and $a^*_{ij} = 0$ if $|i-j| \geq 2$. Assuming that the error vector is normally distributed, one has $\boldsymbol{\epsilon} \sim \mathcal{N}_p(\mathbf{0}, I)$ under the null hypothesis.

Then, on writing $\hat{\boldsymbol{\epsilon}}$ as $M\mathbf{Y}$ where $M_{p\times p} = I - X(X'X)^{-1}X' = M'$ is an idempotent matrix of rank $p-k$, the test statistic can be expressed as the following ratio of quadratic forms:

$$D = \frac{\mathbf{Z}'MA^*M\mathbf{Z}}{\mathbf{Z}'M\mathbf{Z}}\,, \tag{5.10}$$

where $\mathbf{Z} \sim \mathcal{N}_p(\mathbf{0}, I)$; this can be seen from the fact that $M\mathbf{Y}$ and $M\mathbf{Z}$ are identically distributed singular normal vectors with mean vector $\mathbf{0}$ and covariance matrix MM'. We note that the distribution function of D (and, in general, ratios of quadratic forms of the form $(\mathbf{X}'B\mathbf{X})/(\mathbf{X}'C\mathbf{X})$) at the point t_0 can be determined as follows:

$$\begin{aligned} \Pr(D < t_0) &= \Pr\Big(\mathbf{Z}'MA^*M\mathbf{Z} < t_0\mathbf{Z}'M\mathbf{Z}\Big) \\ &= \Pr\Big(\mathbf{Z}'M(A^*M - t_0 I)\mathbf{Z} < 0\Big)\,. \end{aligned} \tag{5.11}$$

On letting $U = \mathbf{Z}'M(A^*M - t_0 I)\mathbf{Z}$, U can be re-expressed as a difference of two positive quadratic forms by applying Steps 1 and 2 of the algorithm provided in Section 4, with $A = M(A^*M - t_0 I)$, $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I$. The moments and the Laguerre polynomial approximant of the density function of U are then obtained from Steps 3, 4 and 5.

We make use of a data set that is provided in Hildreth and Lu (1960, p. 58). In this case, there are $k = 5$ independent variables, $p = 18$, the observed value of D is 0.96, and the 13 non-zero eigenvalues of $M(A^*M - t_0 I)$ are those of MA^*M minus t_0. The non-zero eigenvalues of MA^*M are $3.92807, 3.82025$, 3.68089, 3.38335, 3.22043, 2.95724, 2.35303, 2.25696, 1.79483, 1.48804, 0.948635, 0.742294 and 0.378736. For instance, when $t_0 = 1.80977$, which corresponds to the 10^{th} percentile of the simulated cumulative distribution function resulting from 1,000,000 replications, the eigenvalues of the positive definite quadratic form Q_1 are 2.11817, 2.01035, 1.87099, 1.57345, 1.41053, 1.14734, 0.54313 and 0.44706, while those of Q_2 are 0.01507, 0.32186, 0.861265, 1.06761 and 1.43116. The approximate cumulative distribution function of D based on ten moments was evaluated from Equations (3.17) and (3.18) at certain percentiles of the distribution obtained by simulation. The results reported in Table 3 indicate that the empirical and approximate distribution functions are in close agreement for the given simulated percentiles.

Table 3: Approximate CDF evaluated at certain empirical percentile of D.

CDF	Simulated	Approximate CDF
0.01	1.36069	0.010435
0.025	1.51197	0.025476
0.05	1.64792	0.050280
0.1	1.80977	0.099761
0.25	2.08536	0.247875
0.5	2.39014	0.495934
0.75	2.6861	0.748343
0.9	2.93742	0.902156
0.95	3.07679	0.952783
0.975	3.18896	0.977276
0.99	3.31005	0.991466
1	3.83768	1

6. COMPUTATIONAL CONSIDERATIONS AND CONCLUDING REMARKS

Laguerre polynomials for which an explicit representation is provided in this paper are readily available from numerous mathematical packages including *Mathematica* and *Maple*. It should be pointed out that, after the determination of the parameters ν and β, the only remaining step for obtaining a density approximant is the evaluation of the polynomial coefficients, which are easily determined from Equation (3.8). Quantiles can then be obtained by numerical integration or from the explicit representation of the cumulative distribution function given in Equations (3.17) and (3.18) for indefinite quadratic forms. Conveniently, the requisite calculations can be handled by most mathematical or statistical packages. The symbolic computational package *Mathematica* was used for evaluating the approximants and plotting the graphs, the code being available from the authors upon request.

The proposed density approximation methodology is conceptually simple since it is essentially based on a moment-matching technique. Moreover, it is easy to program and consistently yields remarkably accurate percentage points. Although most applications require relatively few moments, the proposed approximation can accommodate a large number of moments, if need be. The applicability of the results is not restricted to quadratic forms since this methodology can also be utilized to approximate the density functions of random variables that are approximately or asymptotically distributed as gamma random variables, such as those that are proportional to the logarithm of the inverse of certain likelihood ratio test statistics or those that can be expressed as general linear combinations of independently distributed non-central chi-square random variables, which occur in asymptotic theory.

ACKNOWLEDGMENTS

We would like to express our gratitude to an Associate Editor and two reviewers for valuable suggestions that improved the presentation of the results. Thanks are also due to Deli Li for supplying useful references. The financial support of the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged. This research was also supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0013791).

REFERENCES

[1] Alexits, G. (1961). *Convergence Problems of Orthogonal Series*, Pergamon Press, New York.

[2] Anderson, O. D. (1990). Moments of the sampled autocovariances and autocorrelations for a Gaussian white-noise process, *Canad. J. Statist.*, **18**, 271–284.

[3] Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification, *Ann. Math. Statist.*, **25**, 290–302.

[4] Davis, R. B. (1973). Numerical inversion of a characteristic function, *Biometrika*, **60**, 415–417.

[5] De Gooijer, J. G. and MacNeill, I. B. (1999). Lagged regression residuals and serial correlation tests, *Journal of Business and Economic Statistics*, **17**, 236–247.

[6] Dempster, A. P.; Schatzoff, M. and Wermouth, N. (1977). A simulation study of alternatives to ordinary least squares, *Journal of the American Statistical Association*, **72**, 77–106.

[7] Devroye, L. (1989). On random variate generation when only moments or Fourier coefficients are known, *Mathematics and Computers in Simulation*, **31**, 71–89.

[8] Dreze, J. (1976). Bayesian limited information analysis of the simultaneous equations model, *Econometrica*, **44**, 1045–1075.

[9] Dwivedi, T. and Srivastava, V. (1979). Estimation of seemingly unrelated regression equations, *Journal of Econometrics*, **10**, 15–32.

[10] Gradshteyn, I. S. and Ryzhik, I. M. (1980). *Table of Integrals, Series, and Products, Corrected and Enlarged Edition*, Academic Press, New York.

[11] Gurland, J. (1948). Distribution of quadratic forms and ratios of quadratic forms, *Ann. Math. Statist*, **19**, 228–237.

[12] Gurland, J. (1953). Distribution of quadratic forms and ratios of quadratic forms, *Ann. Math. Statist.*, **24**, 416–427.

[13] Gurland, J. (1956). Quadratic forms in normally distributed random variables, *Sankhy ā*, **17**, 37–50.

[14] Hendry, D. F. (1979). The behaviour of inconsistent instrumental variables estimators in dynamic systems with autocorrelated errors, *Journal of Econometrics*, **9**, 295–314.

[15] Hendry, D. F. (1979). The behaviour of inconsistent instrumental variables estimators in dynamic systems with autocorrelated errors, *Journal of Econometrics*, **9**, 295–314.

[16] Hendry, D. F. and Harrison, R. W. (1974). Monte Carlo methodology and the small sample behaviour of ordinary and two-stage least squares, *Journal of Econometrics*, **2**, 151–174.

[17] Hendry, D. F. and Harrison, R. W. (1980). An empirical application and Monte Carlo analysis of tests of dynamic specification, *Review of Economic Studies*, **47**, 21–45.

[18] Hildreth, C. and Lu, J. Y. (1960). *Demand Relations with Auto-Correlated Disturbances*, East Lansing, Michigan: Michigan State University, Agricultural Experiment Station, Department of Agricultural Economics, Technical Bulletin 276.

[19] Hillier, G. H. (1980). The density of a quadratic form in a vector uniformly distributed on the n-sphere, *Econometric Theory*, **17**, 1–28.

[20] Huzurbazar, S. (1999). Practical saddlepoint approximations, *The American Statistician*, **53**, 225–232.

[21] Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables, *Biometrika*, **48**, 419–426.

[22] Johnson, N. L.; Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Vol. 1, John Wiley & Sons, New York.

[23] Kotz, S.; Johnson, N. L. and Boyd, D. W. (1967a). Series representation of distribution of quadratic forms in normal variables I. Central case, *Ann. Math. Statist.*, **38**, 823–837.

[24] Kotz, S.; Johnson, N. L. and Boyd, D. W. (1967b). Series representation of distribution of quadratic forms in normal variables II. Non-central case, *Ann. Math. Statist.*, **38**, 838–848.

[25] MacNeill, I. B. (1978). Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times, *The Annals of Statistics*, **6**, 422–433.

[26] Mathai, A. M. and Provost, S. B. (1992). *Quadratic Forms in Random Variables, Theory and Applications*, Marcel Dekker Inc., New York.

[27] Mathai, A. M. and Saxena, R. K. (1978). *The H-function with Applications in Statistics and Other Disciplines*, John Wiley and Sons, New York.

REVSTAT – Statistical Journal
Volume 11, Number 3, November 2013, 255–276

MULTIPLICATIVE CENSORING: ESTIMATION OF A DENSITY AND ITS DERIVATIVES UNDER THE L_p-RISK

Authors: Mohammad Abbaszadeh
– Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran
abbaszadeh.mo@stu-mail.um.ac.ir

Christophe Chesneau
– Université de Caen, LMNO, Campus II, Science 3, 14032, Caen, France
christophe.chesneau@unicaen.fr

Hassan Doosti
– Department of Mathematics, Kharazmi University, Tehran, Iran
and
Department of Mathematics and Statistics, The University of Melbourne, Melbourne, Australia
hassan.doosti@unimelb.edu.au

Received: May 2012 Revised: January 2013 Accepted: January 2013

Abstract:

- We consider the problem of estimating a density and its derivatives for a sample of multiplicatively censored random variables. The purpose of this paper is to present an approach to this problem based on wavelets methods. Two different estimators are developed: a linear based on projections and a nonlinear using a term-by-term selection of the estimated wavelet coefficients. We explore their performances under the L_p-risk with $p \geq 1$ and over a wide class of functions: the Besov balls. Fast rates of convergence are obtained. Finite sample properties of the estimation procedure are studied on a simulated data example.

Key-Words:

- *density estimation; multiplicative censoring; inverse problem; wavelets; Besov balls; L_p-risk.*

AMS Subject Classification:

- 62G07, 62G20.

1. INTRODUCTION

The multiplicative censoring density model can be described as follows. We observe n *i.i.d.* random variables $Y_1, ..., Y_n$ where, for any $i \in \{1, ..., n\}$,

$$Y_i = U_i X_i \ , \tag{1.1}$$

$U_1, ..., U_n$ are n unobserved *i.i.d.* random variables having the common uniform distribution on $[0, 1]$ and $X_1, ..., X_n$ are n unobserved *i.i.d.* random variables with common unknown density $f : [0, 1] \to [0, \infty]$. For any $i \in \{1, ..., n\}$, we suppose that U_i and X_i are independent. Our aim is to estimate f (or a transformation of f) from $Y_1, ..., Y_n$. Details, applications and results of this model can be found in, e.g., [37], [38], [2] and [1]. For recent applications in the field of signal processing, we refer to [7] and references therein for further readings.

In this paper, we investigate the estimation of $f^{(m)}$ (including f for $m = 0$). This is particularly of interest to detect possible bumps, concavity or convexity properties of f. The estimation of the derivatives of a density have been investigated by several authors. The pioneers are [4], [35] and [36]. Recent studies can be found in [31], [9, 10], [32] and [8].

In recent years, wavelet methods in nonparametric function estimation have become a powerful technique. The major advantages of these methods are their spatial adaptivity and asymptotic optimality properties over large function spaces. We refer to, e.g., [3], [23] and [39]. These facts motivate the estimation of $f^{(m)}$ via wavelet methods. To the best of our knowledge, this has never been investigated before for (1.1). Combning the approaches of [1] and [31], we construct two different wavelet estimators: a linear one and a nonlinear adaptive one based on a hard thresholding rule introduced by [18]. The latter method has the advantage to be adaptive; it does not depend on the knowledge of the smoothness of $f^{(m)}$ in its construction. We explore their performances via the L_p-risk with $p \geq 1$ (including the Mean Integrated Squared Error (MISE) which corresponds to $p = 2$) over a "standard" wide class of unknown functions: the Besov balls $B^s_{r,q}(M)$. Our main result proves that the considered adaptive wavelet estimator achieves a fast rate of convergence. Then we show the finite sample properties of the considered estimators by a simulated data.

The rest of the paper is organized as follows. Section 2 briefly describes the wavelet basis and the Besov balls. Assumptions on the model and the wavelet estimators are presented in Section 3. The theoretical results are given in Section 4. A simulation study is done in Section 5. The proofs are gathered in Section 6.

2. WAVELETS AND BESOV BALLS

This section is devoted to bascics on wavelets and Besov balls.

2.1. Wavelets

Let N be a positive integer such that $N > 10(m+1)$ (where m refers to the estimation of $f^{(m)}$).

Throughout the paper, we work within an orthonormal multiresolution analysis of $L_2([0,1]) = \left\{h\colon [0,1] \to \mathbb{R};\ \int_0^1 h(x)^2\, dx < \infty\right\}$, associated with the initial wavelet functions ϕ and ψ of the Daubechies wavelets $db2N$. The features of these functions are to be compactly supported and $\mathcal{C}^{m+1}$.

Set

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)\ , \qquad \psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)\ .$$

Then, with an appropriate treatment at the boundaries, there exists an integer τ satisfying $2^\tau \geq 2N$ such that, for any $\ell \geq \tau$, the system

$$\mathcal{S} = \left\{\phi_{\ell,k};\ k \in \{0,...,2^\ell - 1\};\ \psi_{j,k};\ j \in \mathbb{N} - \{0,...,\ell - 1\},\ k \in \{0,...,2^j - 1\}\right\}$$

is an orthonormal basis of $L_2([0,1])$.

For any integer $\ell \geq \tau$, any $h \in L_2([0,1])$ can be expanded on $\mathcal{S}$ as

$$h(x) = \sum_{k=0}^{2^\ell - 1} c_{\ell,k}\, \phi_{\ell,k}(x) + \sum_{j=\ell}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}\, \psi_{j,k}(x)\ , \qquad x \in [0,1]\ , \tag{2.1}$$

where

$$c_{j,k} = \int_0^1 h(x)\, \phi_{j,k}(x)\, dx\ , \qquad d_{j,k} = \int_0^1 h(x)\, \psi_{j,k}(x)\, dx\ . \tag{2.2}$$

See, e.g., [14] and [27].

As usual in nonparametric statistics via wavelets, we will suppose that the unknown function $f^{(m)}$ belongs to Besov balls defined below.

2.2. Besov balls

Let $M > 0, s > 0, r \geq 1, q \geq 1$ and $L_r([0,1]) = \left\{h\colon [0,1] \to \mathbb{R};\ \int_0^1 |h(x)|^r dx < \infty\right\}$. Set, for every measurable function h on [0,1] and $\epsilon \geq 0$, $\Delta_\epsilon(h)(x) = h(x+\epsilon) - h(x)$, $\Delta_\epsilon^2(h)(x) = \Delta_\epsilon(\Delta_\epsilon(h))(x)$ and, identically, $\Delta_\epsilon^N(h)(x) = \Delta_\epsilon^{N-1}(\Delta_\epsilon(h))(x)$.

Let

$$\rho^N(t,h,r) \;=\; \sup_{|\epsilon|\le t} \left(\int_0^1 |\Delta_\epsilon^N(h)(x)|^r \, dx \right)^{1/r} .$$

Then, for $s \in [0,N)$, we define the Besov ball $B^s_{r,q}(M)$ by

$$B^s_{r,q}(M) \;=\; \left\{ h \in L_r([0,1]) ;\; \left(\int_0^1 \left(\frac{\rho^N(t,h,r)}{t^s} \right)^q \frac{dt}{t} \right)^{1/q} \le M \right\} ,$$

with the usual modifications if $r=\infty$ or $q=\infty$.

We have the equivalence: $h \in B^s_{r,q}(M)$ if and only if there exists a constant $M^* > 0$ (depending on M) such that (2.2) satisfy

$$\left(\sum_{k=0}^{2^\tau-1} |c_{\tau,k}|^r \right)^{1/r} + \left(\sum_{j=\tau}^{\infty} \left(2^{j(s+1/2-1/r)} \left(\sum_{k=0}^{2^j-1} |d_{j,k}|^r \right)^{1/r} \right)^q \right)^{1/q} \le\; M^* \;,$$

with the usual modifications if $r=\infty$ or $q=\infty$.

In this expression, s is a smoothness parameter and r and q are norm parameters. Details on Besov balls can be found in [28] and [23, Chapter 9].

3. ESTIMATORS

This section describes our wavelet estimation approach.

3.1. Wavelet methodology

Suppose that, for any $v \in \{0,...,m\}$, $f^{(v)} \in L_2([0,1])$. Then we have the wavelet series expansion:

$$f^{(m)}(x) \;=\; \sum_{k=0}^{2^\ell-1} c^{(m)}_{\ell,k}\, \phi_{\ell,k}(x) + \sum_{j=\ell}^{\infty} \sum_{k=0}^{2^j-1} d^{(m)}_{j,k}\, \psi_{j,k}(x) \;, \qquad x \in [0,1] \;,$$

where $c^{(m)}_{j,k} = \int_0^1 f^{(m)}(x)\, \phi_{j,k}(x)\, dx$ and $d^{(m)}_{j,k} = \int_0^1 f^{(m)}(x)\, \psi_{j,k}(x)\, dx$ and m is the order of the density derivative to be estimated.

We now aim to construct natural estimators for these unknown wavelet coefficients. Combining the approaches of [1] and [31], let us investigate a more arranging expression for $c^{(m)}_{j,k}$ (the same development holds for $d^{(m)}_{j,k}$).

Suppose that, for any $v \in \{0, ..., m\}$, $f^{(v)}(0) = f^{(v)}(1) = 0$. It follows from m-fold integration by parts that

$$c_{j,k}^{(m)} = (-1)^m \int_0^1 f(x)\,(\phi_{j,k})^{(m)}(x)\,dx \ .$$

Note that, since $U_1 \sim \mathcal{U}([0,1])$ and U_1 and X_1 are independent, the density of Y_1 is

$$g(x) = \int_x^1 \frac{f(y)}{y}\,dy \ , \qquad x \in [0,1] \ . \tag{3.1}$$

Hence $f(x) = -xg'(x)$, $x \in [0,1]$.

One integration by parts yields

$$\begin{aligned} c_{j,k}^{(m)} &= (-1)^m \left(-\int_0^1 g'(x)\,x(\phi_{j,k})^{(m)}(x)\,dx \right) \\ &= (-1)^m \int_0^1 \Big((\phi_{j,k})^{(m)}(x) + x(\phi_{j,k})^{(m+1)}(x) \Big)\, g(x)\,dx \\ &= \mathbb{E}\bigg((-1)^m \Big((\phi_{j,k})^{(m)}(Y_1) + Y_1(\phi_{j,k})^{(m+1)}(Y_1) \Big) \bigg) \ . \end{aligned}$$

The method of moments gives the following unbiased estimator for $c_{j,k}^{(m)}$:

$$\hat{c}_{j,k}^{(m)} = \frac{(-1)^m}{n} \sum_{i=1}^{n} \Big((\phi_{j,k})^{(m)}(Y_i) + Y_i(\phi_{j,k})^{(m+1)}(Y_i) \Big) \tag{3.2}$$

and, similarly, an unbiased estimator for $d_{j,k}^{(m)}$ is

$$\hat{d}_{j,k}^{(m)} = \frac{(-1)^m}{n} \sum_{i=1}^{n} \Big((\psi_{j,k})^{(m)}(Y_i) + Y_i(\psi_{j,k})^{(m+1)}(Y_i) \Big) \ . \tag{3.3}$$

Further properties of these wavelet coefficients estimators are explored in Propositions 6.1 and 6.2 below. We are now in the position to present the considered estimators for $f^{(m)}$.

3.2. Main estimators

We define the linear estimator $\hat{f}_{lin}^{(m)}$ by

$$\hat{f}_{lin}^{(m)}(x) = \sum_{k=0}^{2^{j_0}-1} \hat{c}_{j_0,k}^{(m)} \phi_{j_0,k}(x) \ , \qquad x \in [0,1] \ , \tag{3.4}$$

where $\hat{c}_{j,k}^{(m)}$ is defined by (3.2) and j_0 is an integer which will be properly chosen later.

Recent developments on the linear wavelet estimators for various density estimation problems can be found in [11].

We define the hard thresholding estimator $\hat{f}^{(m)}_{hard}$ by

$$\hat{f}^{(m)}_{hard}(x) = \sum_{k=0}^{2^\tau-1} \hat{c}^{(m)}_{\tau,k}\,\phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1}\sum_{k=0}^{2^j-1} \hat{d}^{(m)}_{j,k}\,\mathbf{1}_{\left\{|\hat{d}^{(m)}_{j,k}|\ge\kappa\delta^{(m)}_j\right\}}\,\psi_{j,k}(x)\ , \qquad x\in[0,1]\ , \tag{3.5}$$

where $\hat{c}^{(m)}_{j,k}$ and $\hat{d}^{(m)}_{j,k}$ are defined by (3.2) and (3.3), $\mathbf{1}$ is the indicator function, j_1 is the integer satisfying

$$\left(\frac{n}{\ln n}\right)^{1/(2m+3)} < 2^{j_1+1} \le 2\left(\frac{n}{\ln n}\right)^{1/(2m+3)},$$

$\delta^{(m)}_j$ is the threshold:

$$\delta^{(m)}_j = 2^{j(m+1)}\sqrt{\frac{\ln n}{n}}$$

and κ is a large enough constant (see Remark 4.2 and Proposition 6.2).

The major difference between $\hat{f}^{(m)}_{lin}$ and $\hat{f}^{(m)}_{hard}$ is the term-by-term selection of the wavelet coefficients estimators which makes $\hat{f}^{(m)}_{hard}$ adaptive. Discussions on hard thresholding estimators in nonparametric function estimation can be found in, e.g., [18], [23], [16] and [39].

Remark 3.1. A preliminary idea is to rewrite the model (1.1) as: $-\ln Y_i = -\ln X_i - \ln U_i$. In this form, it becomes the standard density deconvolution model where $-\ln U_1, ..., -\ln U_n$ are n unobserved *i.i.d.* random variables having the common exponential distribution with parameter 1 and $-\ln X_1, ..., -\ln X_n$ are n unobserved *i.i.d.* random variables with unknown density

$$q(x) = e^{-x} f(e^{-x})\ , \qquad x\in(0,\infty)\ .$$

Then there exist a wide variety of methods to estimate q. See, e.g., [19], [22], [29], [5], [15] and [26]. Results on the estimation of $q^{(m)}$ via kernel methods can be found in [19]. However, due to the definition of q, it seems difficult to deduce results on the estimation of $f^{(m)}$ from $q^{(m)}$ under the L_p-risk.

Remark 3.2. Another possible approach to estimate $f^{(m)}$ is described below. Since $f(x) = -xg'(x)$, $x\in[0,1]$, we have

$$f^{(m)}(x) = -\big(mg^{(m)}(x) + xg^{(m+1)}(x)\big)\ , \qquad x\in[0,1]\ . \tag{3.6}$$

Then a plug-in approach to estimate $f^{(m)}$ consists in estimating $g^{(m)}$ by $\hat{g}^{(m)}$ and $g^{(m+1)}$ by $\hat{g}^{(m+1)}$, and to inject them in (3.6). This yields the estimator

$$\hat{f}^{(m)}_*(x) = -\big(m\hat{g}^{(m)}(x) + x\hat{g}^{(m+1)}(x)\big)\ , \qquad x\in[0,1]\ .$$

However, there are at least two disadvantages to this approach.

- Firstly, since two different estimators are required, more errors are injected in $\hat{f}_*^{(m)}$ in comparison to (3.5).
- Secondly, the choices of $\hat{g}^{(m)}$ and $\hat{g}^{(m+1)}$ are not so clear. If we focus our attention on wavelet estimators, one can chose hard thresholding versions as in [31]. However, the presence of x in front of $\hat{g}^{(m+1)}(x)$ implies that we work with the nonorthonormal basis $\mathcal{S}_* = \{x\phi_{\ell,k}(x),\ k \in \{0,...,2^\ell-1\};\ x\psi_{j,k}(x);\ j \in \mathbb{N} - \{0,...,\ell-1\},\ k \in \{0,...,2^j-1\}\}$. And it is not immediately clear how we can manipulate it in the context of the L_p-risk.

4. RESULTS

Before presenting the main results, let us formulate the following assumptions:

(**A1**) for any $v \in \{0,...,m\}$, $f^{(v)}(0) = f^{(v)}(1) = 0$,

(**A2**) there exists a known constant $C > 0$ such that, for any $v \in \{0,...,m\}$,

$$\int_0^1 \left(f^{(v)}(x)\right)^2 dx \le C\ ,$$

(**A3**) there exists a known constant $C > 0$ such that

$$\sup_{x\in[0,1]} g(x) \le C\ ,$$

where g is as in (3.1).

Theorems 4.1 and 4.2 below explore the performance of our estimators under the L_p-risk over Besov balls.

Theorem 4.1 (L_p-risk for $\hat{f}_{lin}^{(m)}$). *Consider* (1.1) *under (A1), (A2) and (A3). Let $p \ge 1$. Suppose that $f^{(m)} \in B^s_{r,q}(M)$ with $s > 0$, $r \ge 1$ and $q \ge 1$. Set $s_* = \min(s, s-1/r+1/p)$ and let $\hat{f}_{lin}^{(m)}$ be as in* (3.4) *with j_0 being the integer such that*

$$n^{1/(2s_*+2m+3)} < 2^{j_0+1} \le 2n^{1/(2s_*+2m+3)}\ .$$

Then there exists a constant $C > 0$ such that

$$\mathbb{E}\left(\int_0^1 \left(\hat{f}_{lin}^{(m)}(x) - f^{(m)}(x)\right)^p dx\right) \le C\,n^{-s_*p/(2s_*+2m+3)}\ .$$

Remark 4.1. As usual in linear wavelet estimation, we distinguish in Theorem 4.1 two different zones: the homogeneous zone corresponding to $r \geq p$, and the inhomogeneous zone corresponding to $p > r$ (following the classification of [23, Remark 10.4]). For the homogeneous zone, we obtain the rate of convergence $u_{m,n} = n^{-sp/(2s+2m+3)}$ whereas for the inhomogeneous zone, $u_{m,n} = n^{-(s-1/r+1/p)p/(2(s-1/r+1/p)+2m+3)}$ which is slower than the previous one. Observe that these rates of convergence are similar to those attained by wavelet estimators for some inverse problems (see, e.g., [29], [24] and [12] for deconvolution problems).

Theorem 4.2 (L_p-risk for $\hat{f}^{(m)}_{hard}$)**.** *Consider* (1.1) *under (A1), (A2) and (A3). Let* $\hat{f}^{(m)}_{hard}$ *be* (3.5)*. Suppose that* $f^{(m)} \in B^s_{r,q}(M)$ *with* $s > 0$, $r \geq 1$ *and* $q \geq 1$. *Then there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left(\int_0^1 \left(\hat{f}^{(m)}_{hard}(x) - f^{(m)}(x)\right)^p dx\right) \leq C\,\varphi_{n,m}\ ,$$

where

$$\varphi_{n,m} = \begin{cases} \left(\dfrac{\ln n}{n}\right)^{sp/(2s+2m+3)}, & \text{for } rs > (m+3/2)(p-r), \\ \left(\dfrac{\ln n}{n}\right)^{(s-1/r+1/p)p/(2s-2/r+2m+3)}, & \text{for } rs < (m+3/2)(p-r), \\ \left(\dfrac{\ln n}{n}\right)^{(s-1/r+1/p)p/(2s-2/r+2m+3)} (\ln n)^{(p-r/q)_+}, & \text{for } rs = (m+3/2)(p-r). \end{cases}$$

We see in Theorems 4.1 and 4.2 that

- over the homogeneous zone (i.e., $r \geq p$), $\hat{f}^{(m)}_{hard}$ attains a rate of convergence close to the one of $\hat{f}^{(m)}_{lin}$, i.e., $n^{-sp/(2s+2m+3)}$ (the only difference is a logarithmic term).
- over the inhomogeneous zone (i.e., $p > r$), $\hat{f}^{(m)}_{hard}$ attains a better rate of convergence than the one of $\hat{f}^{(m)}_{lin}$. From an asymptotic point of view, the difference is really significant.

Naturally, taking into account that $\hat{f}^{(m)}_{hard}$ is adaptive, it is preferable to $\hat{f}^{(m)}_{lin}$ in the estimation of $f^{(m)}$.

Remark 4.2. The optimal choice of the threshold κ is difficult to explicit because it depends on numerous constants including those in (A2) and (A3), some norms of the elements of the wavelet basis and the universal constants appearing in Bernstein inequality (see Proposition 6.2). The knowledge of these constants is however determinant for the knowledge of κ and, a fortiori, for the adaptivity of $\hat{f}^{(m)}_{hard}$.

Remark 4.3. Note that Theorem 4.2 taken with $p = 2$ and $m = 0$ coincides with [1, Theorem 4.2] taken with $w(x) = 1$.

Perspectives. A possible extension of this work will be to consider more complex thresholding technique as the block thresholding one (see, e.g., [6] and [13]). Moreover, to ensure that $n^{-sp/(2s+2m+3)}$ is the optimal in the minimax sense, lower bounds must be proved. However, important technical difficulties related to the estimation of $f^{(m)}$ (not only f) appear. All these aspects need further investigations that we leave for a future work.

5. SIMULATION STUDY

We investigate the performances of three wavelets estimators: the linear wavelet estimator (3.4) defined with $j_0 = 7$ (which is an arbitrary choice since s is unknown), the hard thresholding wavelet estimator 3.5 defined with the "universal threshold constant" $\kappa = \hat{\sigma}\sqrt{2}$, where $\hat{\sigma}$ is the standard deviation of the estimated wavelet coefficients (see [17]) and a linear wavelet estimator after local linear smoothing.

Remark 5.1. As noticed in [33], the smooth linear wavelet estimator is motivated by the fact that, when $f^{(m)}$ is smoother than the decomposing wavelet (or the sample size is small), the wavelet shrinkage estimators may contain abusive peaks and artifacts. A possible solution is to consider another smoothing method such as the local linear regression smoother introduced by [20, 21] which enjoys good sampling properties and high minimax efficiency. The construction of the considered estimator is based on [21, eq (2.1)-(2.4)], where Y_j is the wavelet linear estimator (3.4) with $j_0 = j$, $X_j = j/n$, K denotes the Gaussian kernel and $h = 0.08$. Note that we do not claim any theoretical properties of this estimator in this study.

The quality of the estimated density is measured by ANorm which are obtained by following formula

$$ANorm = \frac{1}{N}\sum_{l=1}^{N}\left(\sum_{i=1}^{n}\left(\hat{f}_l^{(m)}(i/n) - f_l^{(m)}(i/n)\right)^2\right)^{1/2},$$

where N is the number of replications and $\hat{f}_l^{(m)}$ is estimator of $f_l^{(m)}$ in three state linear, hard threshold and smoothing methods. We select $N = 100$ and $m \in \{0, 1\}$ at (ANorm) formula. The codes were written in MATLAB software and use Daubechies-Lagarias algorithm for calculating various orthonormal wavelets.

In two examples, we consider samples from a Beta distribution and from a mixture of two Beta distributions.

In both of these examples, the smooth linear wavelet estimators is better than others. On the other hand, hard thresholding estimator (see (3.5)) works better than the linear estimator (see (3.4)).

Example 1. We generate samples $X_1, ..., X_n$ from a Beta distribution $Beta(\alpha, \beta)$ with parameters $\alpha = 3$ and $\beta = 3$ with size $n = 1000$. Also we generate $n = 1000$ samples from uniform distribution on $[0, 1]$ that are independent of the X_i's to produce multiplicative censoring. Then we estimate original density using various wavelet methods for derivatives of order $m \in \{0, 1\}$. Fig. 1 shows the original density and Fig. 2 its derivative.

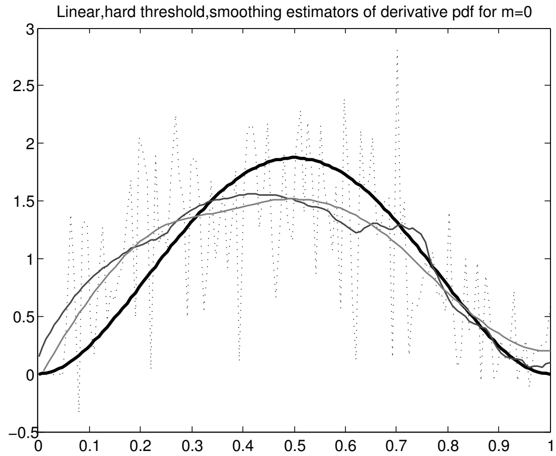

Figure 1: The original density with black line, linear estimator with dotted line, hard threshold estimator with blue line and smoothing estimator is red line. (Density estimation for Beta distribution).

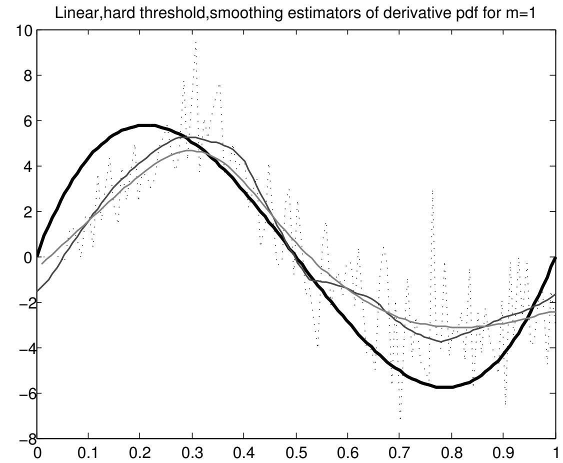

Figure 2: The original density with black line, linear estimator with dotted line, hard threshold estimator with blue line and smoothing estimator is red line. (Derivative estimation for Beta distribution).

The original density with black line, linear estimator with dotted line, hard threshold estimator with blue line and smoothing estimator is red line.

With obvious ANorm and standard deviation in Tables 1 and 2, we conclude when sample size increases, ANorm is smaller and we have better performance.

Table 1: Computed values for ANorm and (Standard deviation) with Beta distribution for $m = 0$.

Estimation Methods	ANorm and (Standard deviation)			
	$n = 256$	$n = 512$	$n = 1024$	$n = 2048$
Linear	12.7357 (0.9611)	9.2113 (0.7728)	6.8497 (0.5323)	5.2332 (0.4201)
Hard Thresholding	6.7080 (1.6041)	5.0373 (1.1540)	3.8189 (0.9575)	3.2546 (0.6949)
Smoothing	3.0102 (1.0213)	2.5472 (0.7237)	2.4273 (0.4855)	2.3779 (0.4057)

Table 2: Computed values for ANorm and (Standard deviation) with Beta distribution for $m = 1$.

Estimation Methods	ANorm and (Standard deviation)			
	$n = 256$	$n = 512$	$n = 1024$	$n = 2048$
Linear	40.4498 (3.3822)	31.5682 (2.5285)	25.8549 (1.6848)	22.2075 (1.0914)
Hard Thresholding	23.1572 (4.4096)	20.8959 (3.2689)	19.3508 (1.7315)	18.6700 (1.2586)
Smoothing	19.0204 (3.1333)	18.6047 (1.9025)	18.3701 (1.5284)	18.1202 (1.0132)

Example 2. In this example, we consider mixture Beta distribution. We generate $n = 1000$ samples $X_1, ..., X_n$ such that $f \sim (1/3)\, Beta(4,6) + (2/3)\, Beta(3,4)$ and proceed as the previous example. Fig. 3 and Fig. 4 show plot from defined estimators.

We calculated ANorm and standard deviation in Tables 3 and 4 for different values of n.

Table 3: Computed values for ANorm and (Standard deviation) with Beta mixture distribution for $m = 0$.

Estimation Methods	ANorm and (Standard deviation)			
	$n = 256$	$n = 512$	$n = 1024$	$n = 2048$
Linear	13.0786 (1.1155)	9.2829 (0.9413)	6.8973 (0.6469)	5.1470 (0.4214)
Hard Thresholding	9.0021 (1.8353)	6.0055 (1.4657)	4.5009 (0.9155)	3.7527 (0.5689)
Smoothing	3.0145 (1.0037)	2.7518 (0.8104)	2.7084 (0.6871)	2.5294 (0.3554)

Table 4: Computed values for ANorm and (Standard deviation) with Beta distribution for $m = 1$.

Estimation Methods	ANorm and (Standard deviation)			
	$n = 256$	$n = 512$	$n = 1024$	$n = 2048$
Linear	64.1098 (7.1914)	48.2251 (4.7379)	36.6836 (2.8821)	29.3822 (2.0521)
Hard Thresholding	32.9540 (10.2685)	26.2743 (7.7777)	25.0095 (4.1568)	21.9653 (2.2657)
Smoothing	20.5346 (4.5341)	20.2737 (4.0706)	19.6435 (2.3666)	19.3836 (1.9703)

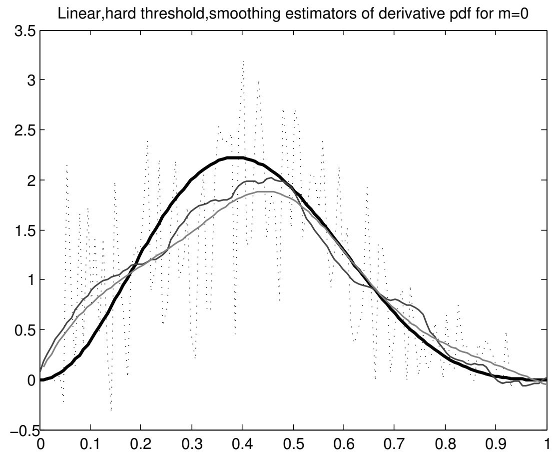

Figure 3: The original density with black line, linear estimator with dotted line, hard threshold estimator with blue line and smoothing estimator is red line. (Density estimation for Beta mixture distribution).

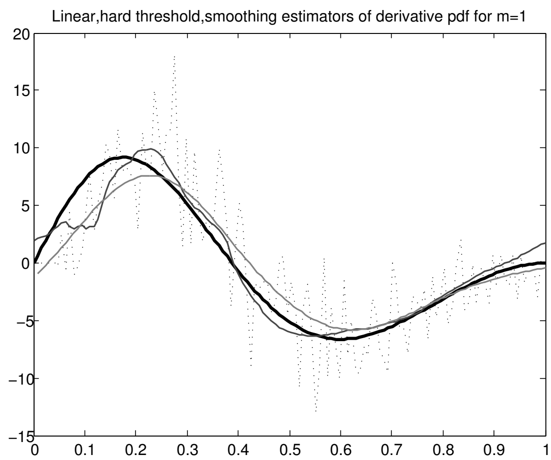

Figure 4: The original density with black line, linear estimator with dotted line, hard threshold estimator with blue line and smoothing estimator is red line. (Derivative estimation for Beta mixture distribution).

6. PROOFS

In this section, C denotes any constant that does not depend on j, k and n. Its value may change from one term to another and may depend on ϕ or ψ.

This section is organized as follows. Firstly, we introduce two auxiliary results on some properties of (3.2) and (3.3) at the heart of the proofs of our main theorems.

Proposition 6.1. *Let $p \geq 1$. For any integer $j \geq \tau$ such that $2^j \leq n$ and any $k \in \{0, ..., 2^j - 1\}$, let $\hat{c}_{j,k}^{(m)}$ be (3.2), $\hat{d}_{j,k}^{(m)}$ be (3.3), $c_{j,k}^{(m)} = \int_0^1 f^{(m)}(x)\phi_{j,k}(x)dx$ and $d_{j,k}^{(m)} = \int_0^1 f^{(m)}(x)\psi_{j,k}(x)dx$. Then, under (A1), (A2) and (A3), there exists a constant $C > 0$ such that*

$$\mathbb{E}\left(\left(\hat{c}_{j,k}^{(m)} - c_{j,k}^{(m)}\right)^{2p}\right) \leq C\, 2^{j(2m+2)p} \frac{1}{n^p}$$

and

$$\mathbb{E}\left(\left(\hat{d}_{j,k}^{(m)} - d_{j,k}^{(m)}\right)^{2p}\right) \leq C\, 2^{j(2m+2)p} \frac{1}{n^p}\ .$$

Proposition 6.2. *Let $p \geq 1$. For any integer $j \geq \tau$ such that $2^j \leq n/\ln n$ and any $k \in \{0, ..., 2^j - 1\}$, let $\hat{d}_{j,k}^{(m)}$ be (3.3) and $d_{j,k}^{(m)} = \int_0^1 f^{(m)}(x)\psi_{j,k}(x)dx$. Then, under (A1), (A2) and (A3), there exists a constant $\kappa > 0$ such that*

$$\mathbb{P}\left(\left|\hat{d}_{j,k}^{(m)} - d_{j,k}^{(m)}\right| \geq \frac{\kappa}{2}\, 2^{j(m+1)}\sqrt{\frac{\ln n}{n}}\right) \leq 2\left(\frac{\ln n}{n}\right)^p.$$

Proof of Proposition 6.1: For convenience, let us prove the second inequality, the proof of the first one is identical.

For the sake of simplicity, for any $i \in \{1, ..., n\}$, set

$$Q_{i,j,k}^{(m)} = (-1)^m\left((\psi_{j,k})^{(m)}(Y_i) + Y_i(\psi_{j,k})^{(m+1)}(Y_i)\right)$$

and

$$U_i = Q_{i,j,k}^{(m)} - d_{j,k}^{(m)}\ .$$

Then we can write

$$\mathbb{E}\left((\hat{d}_{j,k}^{(m)} - d_{j,k}^{(m)})^{2p}\right) = \frac{1}{n^{2p}}\,\mathbb{E}\left(\left(\sum_{i=1}^{n} U_i\right)^{2p}\right). \tag{6.1}$$

Let us now investigate the bound of this expectation via the Rosenthal inequality presented below (see [34]).

Lemma 6.1 (Rosenthal's inequality)**.** *Let n be a positive integer, $\gamma \geq 2$ and $U_1, ..., U_n$ be n zero mean i.i.d. random variables such that $\mathbb{E}(|U_1|^\gamma) < \infty$. Then there exists a constant $C > 0$ such that*

$$\mathbb{E}\left(\left|\sum_{i=1}^{n} U_i\right|^\gamma\right) \leq C \max\Big(n\,\mathbb{E}\big(|U_1|^\gamma\big),\, n^{\gamma/2}\left(\mathbb{E}(U_1^2)\right)^{\gamma/2}\Big).$$

Observe that $U_1, ..., U_n$ are *i.i.d.* and, since $\mathbb{E}(Q^{(m)}_{i,j,k}) = d^{(m)}_{j,k}$, $\mathbb{E}(U_1) = 0$.

Since $Y_1(\Omega) = [0,1]$, we have

$$\begin{aligned} \left|Q^{(m)}_{i,j,k}\right| &\leq \left|(\psi_{j,k})^{(m)}(Y_i)\right| + \left|Y_i(\psi_{j,k})^{(m+1)}(Y_i)\right| \\ &\leq \left|(\psi_{j,k})^{(m)}(Y_i)\right| + \left|(\psi_{j,k})^{(m+1)}(Y_i)\right|. \end{aligned} \tag{6.2}$$

Let $\upsilon \geq 1$. It follows from $\mathbb{E}(Q^{(m)}_{i,j,k}) = d^{(m)}_{j,k}$, the Hölder inequality and (6.2) that

$$\begin{aligned} \mathbb{E}\big(|U_1|^\upsilon\big) &\leq C\,\mathbb{E}\big(|Q^{(m)}_{1,j,k}|^\upsilon\big) \\ &\leq C\left(\mathbb{E}\big(\left|(\psi_{j,k})^{(m)}(Y_1)\right|^\upsilon\big) + \mathbb{E}\big(\left|(\psi_{j,k})^{(m+1)}(Y_1)\right|^\upsilon\big)\right). \end{aligned} \tag{6.3}$$

Using (A3), $(\psi_{j,k})^{(m)}(x) = 2^{j(2m+1)/2}\psi^{(m)}(2^j x - k)$ and doing the change of variables $y = 2^j x - k$, we have

$$\begin{aligned} \mathbb{E}\big(\left|(\psi_{j,k})^{(m)}(Y_1)\right|^\upsilon\big) &= \int_0^1 \left|(\psi_{j,k})^{(m)}(x)\right|^\upsilon g(x)\,dx \;\leq C\int_0^1 \left|(\psi_{j,k})^{(m)}(x)\right|^\upsilon dx \\ &= C\,2^{j\upsilon(2m+1)/2}\int_0^1 \left|\psi^{(m)}(2^j x - k)\right|^\upsilon dx \\ &= C\,2^{j(\upsilon(2m+1)/2-1)}\int_{-k}^{2^j-k} \left|\psi^{(m)}(y)\right|^\upsilon dy \;\leq\; C\,2^{j(\upsilon(2m+1)/2-1)}. \end{aligned} \tag{6.4}$$

In a similar way, we prove that

$$\mathbb{E}\big(\left|(\psi_{j,k})^{(m+1)}(Y_1)\right|^\upsilon\big) \leq C\,2^{j(\upsilon(2m+3)/2-1)}. \tag{6.5}$$

Putting (6.3), (6.4) and (6.5) together, we obtain

$$\mathbb{E}\big(|U_1|^\upsilon\big) \leq C\,2^{j(\upsilon(2m+3)/2-1)}. \tag{6.6}$$

Using the Rosenthal inequality with $U_1, ..., U_n$, $\gamma = 2p$ and $2^j \leq n$, we have

$$\begin{aligned} \mathbb{E}\left(\left(\sum_{i=1}^{n} U_i\right)^{2p}\right) &\leq C\max\Big(n\,\mathbb{E}(U_1^{2p}),\, n^p\left(\mathbb{E}(U_1^2)\right)^p\Big) \\ &\leq C\max\Big(n\,2^{j((2m+3)p-1)},\, n^p\,2^{j(2m+2)p}\Big) \\ &\leq C\,n^p\,2^{j(2m+2)p}. \end{aligned} \tag{6.7}$$

By (6.1) and (6.7), we have

$$\mathbb{E}\left(\left(\hat{d}_{j,k}^{(m)} - d_{j,k}^{(m)}\right)^{2p}\right) \le C\,\frac{1}{n^{2p}}\,n^p\,2^{j(2m+2)p} \le C\,2^{j(2m+2)p}\,\frac{1}{n^p}\,.$$

Similarly, we prove that

$$\mathbb{E}\left(\left(\hat{c}_{j,k}^{(m)} - c_{j,k}^{(m)}\right)^{2p}\right) \le C\,2^{j(2m+2)p}\,\frac{1}{n^p}\,.$$

The proof of Proposition 6.1 is complete. □

Proof of Proposition 6.2: For the sake of simplicity, for any $i \in \{1,...,n\}$, set

$$Q_{i,j,k}^{(m)} = (-1)^m \left((\psi_{j,k})^{(m)}(Y_i) + Y_i(\psi_{j,k})^{(m+1)}(Y_i)\right)$$

and

$$U_i = Q_{i,j,k}^{(m)} - d_{j,k}^{(m)}\,.$$

Then, for any $\kappa > 0$, we can write

(6.8)
$$\mathbb{P}\left(\left|\hat{d}_{j,k}^{(m)} - d_{j,k}^{(m)}\right| \ge \frac{\kappa}{2}\,2^{j(m+1)}\sqrt{\frac{\ln n}{n}}\right) = \mathbb{P}\left(\left|\sum_{i=1}^{n} U_i\right| \ge C\,\frac{\kappa}{2}\,2^{j(m+1)}\sqrt{n\ln n}\right).$$

Let us now explore the bound of this probability via the Bernstein inequality described below (see [30]).

Lemma 6.2 (Bernstein's inequality)**.** *Let n be a positive integer and $U_1,...,U_n$ be n i.i.d. zero mean independent random variables such that there exists a constant $M > 0$ satisfying $|U_1| \le M < \infty$. Then, for any $\upsilon > 0$,*

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} U_i\right| \ge \upsilon\right) \le 2\exp\left(-\frac{\upsilon^2}{2\big(n\,\mathbb{E}(U_1^2) + \upsilon M/3\big)}\right).$$

Observe that $U_1,...,U_n$ are *i.i.d.* and, since $\mathbb{E}(Q_{i,j,k}^{(m)}) = d_{j,k}^{(m)}$, $\mathbb{E}(U_1) = 0$.

Since $Y_1(\Omega) = [0,1]$, $(\psi_{j,k})^{(m)}(x) = 2^{j(2m+1)/2}\psi^{(m)}(2^j x - k)$, $\sup_{y\in[0,1]} |(\psi_{j,k})^{(m)}(y)| \le C2^{j(2m+1)/2}$ and $\sup_{y\in[0,1]} |(\psi_{j,k})^{(m+1)}(y)| \le C2^{j(2m+3)/2}$, we have

$$\begin{aligned}|Q_{1,j,k}^{(m)}| &\le \left|(\psi_{j,k})^{(m)}(Y_1)\right| + \left|Y_1(\psi_{j,k})^{(m+1)}(Y_1)\right| \\ &\le C\left(\sup_{y\in[0,1]}\left|(\psi_{j,k})^{(m)}(y)\right| + \sup_{y\in[0,1]}\left|(\psi_{j,k})^{(m+1)}(y)\right|\right) \le C\,2^{j(2m+3)/2}\,.\end{aligned}$$

Observe that, thanks to (A2) and the Cauchy–Schwarz inequality,

$$|d_{j,k}^{(m)}| \le \left(\int_0^1 \big(f^{(m)}(x)\big)^2 dx\right)^{1/2} \left(\int_0^1 \big(\psi_{j,k}(x)\big)^2 dx\right)^{1/2} \le C\,.$$

Using $2^j \leq n/\ln n$, we have

$$\begin{aligned}|U_1| \;&\leq\; C\big(\big|Q^{(m)}_{1,j,k}\big| + \big|d^{(m)}_{j,k}\big|\big) \leq C\big(2^{j(2m+3)/2} + C\big)\\ &=\; C\,2^{j(2m+3)/2} \leq C\,2^{j(m+1)}\sqrt{\frac{n}{\ln n}}\ .\end{aligned}$$

It follows from (6.6) that

$$\mathbb{E}(U_1^2) \;\leq\; C\,2^{j(2m+2)}\ .$$

The Bernstein inequality applied with $U_1, ..., U_n$ and $\upsilon = (\kappa/2)\,2^{j(m+1)}\sqrt{n\ln n}$ gives

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{i=1}^{n} U_i\right| \geq \upsilon\right) &\leq 2\,\exp\!\left(-\frac{\upsilon^2}{2\big(n\,\mathbb{E}(U_1^2) + \upsilon M/3\big)}\right)\\ &\leq\; 2\,\exp\!\left(-\frac{(\kappa/2)^2\, 2^{j(2m+2)}\, n\,\ln n}{C n 2^{j(2m+2)} + C(\kappa/2)\,2^{j(m+1)}\sqrt{n\ln n}\;2^{j(m+1)}\sqrt{n/\ln n}}\right) \qquad (6.9)\\ &=\; 2\,n^{-C\frac{\kappa^2}{1+\kappa}}\ .\end{aligned}$$

By (6.8) and (6.9), there exists a constant $\kappa > 0$ such that

$$\mathbb{P}\left(\big|\hat{d}^{(m)}_{j,k} - d^{(m)}_{j,k}\big| \geq \frac{\kappa}{2}\,2^{j(m+1)}\sqrt{\frac{\ln n}{n}}\right) \;\leq\; 2\,n^{-C\frac{\kappa^2}{1+\kappa}} \;\leq\; 2\left(\frac{\ln n}{n}\right)^{p}.$$

Proposition 6.2 is proved. □

Proof of Theorem 4.1: We expand the function $f^{(m)}$ on $\mathcal{S}$ as

$$f^{(m)}(x) \;=\; \sum_{k=0}^{2^{j_0}-1} c^{(m)}_{j_0,k}\,\phi_{j_0,k}(x) + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} d^{(m)}_{j,k}\,\psi_{j,k}(x)\ ,$$

where $c^{(m)}_{j_0,k} = \int_0^1 f^{(m)}(x)\,\phi_{j_0,k}(x)\,dx$ and $d^{(m)}_{j,k} = \int_0^1 f^{(m)}(x)\,\psi_{j,k}(x)\,dx$.

We have

$$\mathbb{E}\left(\int_0^1 \Big(\hat{f}^{(m)}_{lin}(x) - f^{(m)}(x)\Big)^p dx\right) \leq 2^{p-1}(A+B)\ , \qquad (6.10)$$

where

$$A \;=\; \mathbb{E}\left(\int_0^1 \left(\sum_{k=0}^{2^{j_0}-1}\big(\hat{c}^{(m)}_{j_0,k} - c^{(m)}_{j_0,k}\big)\,\phi_{j_0,k}(x)\right)^p dx\right)$$

and

$$B \;=\; \int_0^1 \left(\sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} d^{(m)}_{j,k}\,\psi_{j,k}(x)\right)^p dx\ .$$

Let us now introduce a L_p-norm result for wavelets.

Lemma 6.3. *Let $p \geq 1$. For any sequence of real number $(\theta_{j,k})_{j,k}$, there exists a constant $C > 0$ such that*

$$\int_0^1 \left(\sum_{k=0}^{2^j-1} \theta_{j,k}\, \phi_{j,k}(x) \right)^p dx \;\leq\; C\, 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\theta_{j,k}|^p .$$

The proof can be found in, e.g., [23, Proposition 8.3].

Lemma 6.3, Proposition 6.1 and the Cauchy–Schwarz inequality yield

$$\begin{aligned} A \;&\leq\; C\, 2^{j_0(p/2-1)} \sum_{k=0}^{2^{j_0}-1} \mathbb{E}\left((\hat{c}^{(m)}_{j_0,k} - c^{(m)}_{j_0,k})^p \right) \\ &\leq\; C\, 2^{j_0(p/2-1)} \sum_{k=0}^{2^{j_0}-1} \left(\mathbb{E}\left((\hat{c}^{(m)}_{j_0,k} - c^{(m)}_{j_0,k})^{2p} \right) \right)^{1/2} \qquad (6.11) \\ &\leq\; C\, 2^{j_0(p/2-1)}\, 2^{j_0}\, 2^{j_0(m+1)p} \frac{1}{n^{p/2}} \;=\; C \left(\frac{2^{j_0(2m+3)}}{n} \right)^{p/2} . \end{aligned}$$

On the other hand, using $f^{(m)} \in B^s_{r,q}(M)$ and proceeding as in [18, eq (24)], we have

$$B \;\leq\; C\, 2^{-j_0 s_* p} . \qquad (6.12)$$

It follows from (6.10), (6.11), (6.12) and the definition of j_0 that

$$\begin{aligned} \mathbb{E}\left(\int_0^1 \left(\hat{f}^{(m)}_{lin}(x) - f^{(m)}(x) \right)^p dx \right) \;&\leq\; C \left(\left(\frac{2^{j_0(2m+3)}}{n} \right)^{p/2} + 2^{-j_0 s_* p} \right) \\ &\leq\; C\, n^{-s_* p/(2s_*+2m+3)} . \end{aligned}$$

This ends the proof of Theorem 4.1. □

Proof of Theorem 4.2: Theorem 4.2 is a consequence of Theorem 6.1 below by taking with $\nu = m+1$ and using Propositions 6.1 and 6.2 above. □

Theorem 6.1. *Let $h \in L_2([0,1])$ be an unknown function to be estimated from n observations and (2.1) its wavelet decomposition. Let $\hat{c}_{j,k}$ and $\hat{d}_{j,k}$ be estimators of $c_{j,k}$ and $d_{j,k}$ respectively such that there exist three constants $\nu > 0$, $C > 0$ and $\kappa > 0$ satisfying*

Moments inequalities: *for any $j \geq \tau$ such that $2^j \leq n$ and $k \in \{0, ..., 2^j - 1\}$,*

$$\mathbb{E}\left((\hat{c}_{j,k} - c_{j,k})^{2p} \right) \;\leq\; C\, 2^{2\nu j p} \left(\frac{\ln n}{n} \right)^p$$

and

$$\mathbb{E}\left(\left(\hat{d}_{j,k}-d_{j,k}\right)^{2p}\right) \;\leq\; C\,2^{2\nu jp}\left(\frac{\ln n}{n}\right)^{p}.$$

Concentration inequality: for any $j\geq\tau$ *such that* $2^j\leq n/\ln n$ *and* $k\in\{0,...,2^j-1\}$,

$$\mathbb{P}\left(\left|\hat{d}_{j,k}-d_{j,k}\right|\geq\frac{\kappa}{2}\,2^{\nu j}\sqrt{\frac{\ln n}{n}}\right)\leq C\left(\frac{\ln n}{n}\right)^{p}.$$

Let us define the hard thresholding wavelet estimator of h *by*

$$\hat{h}(x)\;=\;\sum_{k=0}^{2^\tau-1}\hat{c}_{\tau,k}\,\phi_{\tau,k}(x)+\sum_{j=\tau}^{j_1}\sum_{k=0}^{2^j-1}\hat{d}_{j,k}\,\mathbf{1}_{\left\{|\hat{d}_{j,k}|\geq\kappa 2^{\nu j}\sqrt{\ln n/n}\right\}}\psi_{j,k}(x)\;,\qquad x\in[0,1]\;,$$

where j_1 *is the integer satisfying* $(n/\ln n)^{1/(2\nu+1)}<2^{j_1+1}\leq 2\,(n/\ln n)^{1/(2\nu+1)}$.

Suppose that $h\in B^s_{r,q}(M)$ *with* $s>0$, $r\geq 1$ *and* $q\geq 1$. *Then there exists a constant* $C>0$ *such that*

$$\mathbb{E}\left(\int_0^1\left(\hat{h}(x)-h(x)\right)^{p}dx\right)\leq C\,\Theta_{n,\nu}\;,$$

where

$$\Theta_{n,\nu}=\begin{cases}\left(\dfrac{\ln n}{n}\right)^{sp/(2s+2\nu+1)}, & \text{for}\quad rs>(\nu+1/2)(p-r),\\[2ex] \left(\dfrac{\ln n}{n}\right)^{(s-1/r+1/p)p/(2s-2/r+2\nu+1)}, & \text{for}\quad rs<(\nu+1/2)(p-r),\\[2ex] \left(\dfrac{\ln n}{n}\right)^{(s-1/r+1/p)p/(2s-2/r+2\nu+1)}(\ln n)^{(p-r/q)_+}, & \text{for}\quad rs=(\nu+1/2)(p-r).\end{cases}$$

Theorem 6.1 does not appear in this form in the literature but can be proved using similar arguments to [25, Theorem 5.1] for a bound of the L_p-risk and [12, Theorem 4.2] for the determination of the rates of convergence.

ACKNOWLEDGMENTS

The authors thank two anonymous referees for their thorough and useful comments.

REFERENCES

[1] Abbaszadeh, M.; Chesneau, C. and Doosti, H. (2012). Nonparametric estimation of density under bias and multiplicative censoring via wavelet methods, *Statistics and Probability Letters*, **82**, 932–941.

[2] Andersen, K. and Hansen, M. (2001). Multiplicative censoring: density estimation by a series expansion approach, *Journal of Statistical Planning and Inference*, **98**, 137–155.

[3] Antoniadis, A. (1997). Wavelets in statistics: a review (with discussion), *Journal of the Italian Statistical Society, Series B*, **6**, 97–144.

[4] Bhattacharya, P.K. (1967). Estimation of a probability density function and its derivatives, *Sankhya, Serie A*, **29**, 373–382.

[5] Butucea, C. and Matias, C. (2005). Minimax estimation of the noise level and of the signal density in a semiparametric convolution model, *Bernoulli*, **11**(2), 309–340.

[6] Cai, T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach, *The Annals of Statistics*, **27**, 898–924.

[7] Chabert, M.; Ruiz, D. and Tourneret, J.-Y. (2004). Optimal wavelet for abrupt change detection in multiplicative noise. In "IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004", Proceedings, (ICASSP '04), vol. 2, pp. ii 1089-92. IEEE, May 2004.

[8] Chacón, J.E.; Duong, T. and Wand, M.P. (2011). Asymptotics for general multivariate kernel density derivative estimators, *Statistica Sinica*, to appear.

[9] Chaubey, Y.P.; Doosti, H. and Prakasa Rao, B.L.S. (2006). Wavelet based estimation of the derivatives of a density with associated variables, *International Journal of Pure and Applied Mathematics*, **27**(1), 97–106.

[10] Chaubey, Y.P.; Doosti, H. and Prakasa Rao, B.L.S. (2008). Wavelet based estimation of the derivatives of a density for a negatively associated process, *Journal of Statistical Theory and Practice*, **2**(3), 453–463.

[11] Chaubey, Y.P.; Chesneau, C. and Doosti, H. (2011). On linear wavelet density estimation: some recent developments, *Journal of the Indian Society of Agricultural Statistics*, **65**(2), 169–179.

[12] Chesneau, C. (2008). Wavelet estimation via block thresholding: a minimax study under $\mathbb{L}^p$ risk, *Statistica Sinica*, **18**(3), 1007–1024.

[13] Chesneau, C.; Fadili, M.J. and Starck, J.-L. (2010). Stein Block Thresholding For Image Denoising, *Applied and Computational Harmonic Analysis*, **28**(1), 67–88.

[14] Cohen, A.; Daubechies, I.; Jawerth, B. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms, *Applied and Computational Harmonic Analysis*, **24**(1), 54–81.

[15] Delaigle, A. and Gijbels, I. (2006). Estimation of boundary and discontinuity points in deconvolution problems, *Statistica Sinica*, **16**, 773–788.

[16] Delyon, B. and Juditsky, A. (1996). On minimax wavelet estimators, *Applied Computational Harmonic Analysis*, **3**, 215–228.

[17] Donoho, D.L. and Johnstone, I.M. (1994). Ideal spatial adaptation by wavelet shrinkage, *Biometrica*, **81**, 425–455.

[18] Donoho, D.L.; Johnstone, I.M.; Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding, *The Annals of Statistics*, **24**, 508–539.

[19] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problem, *Ann. Statist.*, **19**, 1257–1272.

[20] Fan, J. (1992). Design-adaptive nonparametric regression, *Journal of American Statistical Association*, **87**, 998–1004.

[21] Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies, *Annals of Statistics*, **21**, 196–216.

[22] Fan, J. and Liu, Y. (1997). A note on asymptotic normality for deconvolution kernel density estimators, *Sankhya*, **59**, 138–141.

[23] Härdle, W.; Kerkyacharian, G.; Picard, D. and Tsybakov, A. (1998). *Wavelet, Approximation and Statistical Applications*, Lectures Notes in Statistics, 129, New York, Springer Verlag.

[24] Johnstone, I.; Kerkyacharian, G.; Picard, D. and Raimondo, M. (2004). Wavelet deconvolution in a periodic setting, *Journal of the Royal Statistical Society, Serie B*, **66**(3), 547–573.

[25] Kerkyacharian, G. and Picard, D. (2000). Thresholding algorithms, maxisets and well concentrated bases (with discussion and a rejoinder by the authors), *Test*, **9**(2), 283–345.

[26] Lacour, C. (2006). Rates of convergence for nonparametric deconvolution, *C. R. Acad. Sci. Paris Ser. I Math.*, **342**(11), 877–882.

[27] Mallat, S. (2009). *A wavelet tour of signal processing*, Elsevier/Academic Press, Amsterdam, third edition. The sparse way, With contributions from Gabriel Peyré.

[28] Meyer, Y. (1992). *Wavelets and Operators*, Cambridge University Press, Cambridge.

[29] Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution, *The Annals of Statistics*, **27**, 2033–2053.

[30] Petrov, V.V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*, Oxford: Clarendon Press.

[31] Prakasa Rao, B.L.S. (1996). Nonparametric estimation of the derivatives of a density by the method of wavelets, *Bull. Inform. Cyb.*, **28**, 91–100.

[32] Prakasa Rao, B.L.S. (2010). Wavelet linear estimation for derivatives of a density from observations of mixtures with varying mixing proportions, *Indian J. Pure Appl. Math.*, **41**(1), 275–291.

[33] Ramirez, P. and Vidakovic, B. (2010). Wavelet density estimation for stratified size-biased sample, *Journal of Statistical planning and Inference*, **140**, 419–432.

[34] Rosenthal, H.P. (1970). On the subspaces of L_p ($p \geq 2$) spanned by sequences of independent random variables, *Israel Journal of Mathematics*, **8**, 273–303.

[35] Schuster, E.F. (1969). Estimation of a probability function and its derivatives, *Ann. Math. Statist.*, **40**, 1187–1195.

[36] Singh, R.S. (1977). Applications of estimators of a density and its derivatives to certain statistical problems, *J. Roy. Statist. Soc. B*, **39**, 357–363.

[37] Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation, *Biometrika*, **76**, 751–761.

[38] Vardi, Y. and Zhang, C.H. (1992). Large sample study of empirical distributions in a random multiplicative censoring model, *The Annals of Statistics*, **20**, 1022–1039.

[39] Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, John Wiley & Sons, Inc., New York, 384 pp.

REVSTAT – Statistical Journal
Volume 11, Number 3, November 2013, 277–299

ON THE ESTIMATION OF THE SECOND ORDER PARAMETER FOR HEAVY-TAILED DISTRIBUTIONS

Authors: El hadji Deme
– LERSTAD, Université Gaston Berger,
Saint-Louis, Sénégal
eltinoyaw@yahoo.fr

Laurent Gardes
– Université de Strasbourg & CNRS,
IRMA UMR 7501, France
gardes@unistra.fr

Stéphane Girard
– Mistis, INRIA Rhône-Alpes & Laboratoire Jean Kuntzmann,
France
stephane.girard@inria.fr

Received: May 2012 Revised: November 2012 Accepted: January 2013

Abstract:

- The extreme-value index γ is an important parameter in extreme-value theory since it controls the first order behavior of the distribution tail. In the literature, numerous estimators of this parameter have been proposed especially in the case of heavy-tailed distributions, which is the situation considered here. Most of these estimators depend on the k largest observations of the underlying sample. Their bias is controlled by the second order parameter ρ. In order to reduce the bias of γ's estimators or to select the best number k of observations to use, the knowledge of ρ is essential. In this paper, we propose a simple approach to estimate the second order parameter ρ leading to both existing and new estimators. We establish a general result that can be used to easily prove the asymptotic normality of a large number of estimators proposed in the literature or to compare different estimators within a given family. Some illustrations on simulations are also provided.

Key-Words:

- *extreme-value theory; heavy-tailed distribution; extreme-value index; second order parameter; asymptotic properties.*

AMS Subject Classification:

- 62G32, 62G30, 60G70, 62E20.

1. INTRODUCTION

Extreme-value theory establishes the asymptotic behavior of the largest observations in a sample. It provides methods for extending the empirical distribution function beyond the observed data. It is thus possible to estimate quantities related to the tail of a distribution such as small exceedance probabilities or extreme quantiles. We refer to [11, 25] for general accounts on extreme-value theory. More specifically, let $X_1, ..., X_n$ be a sequence of random variables (rv), independent and identically distributed from a cumulative distribution function (cdf) F. Extreme-value theory establishes that the asymptotic distribution of the maximum $X_{n,n} = \max\{X_1, ..., X_n\}$ properly rescaled is the extreme-value distribution with cdf

$$G_\gamma(x) = \exp\bigl(-(1+\gamma x)_+\bigr)^{-1/\gamma}$$

where $y_+ = \max(y, 0)$. The parameter $\gamma \in \mathbb{R}$ is referred to as the extreme-value index. Here, we focus on the case where $\gamma > 0$. In such a situation, F is said to belong to the maximum domain of attraction of the Fréchet distribution. In this domain of attraction, a simple characterization of distributions is available: the quantile function $U(x) := F^{\leftarrow}(1 - 1/x)$ can be written as

$$U(x) = x^\gamma \ell(x) ,$$

where ℓ is a slowly varying function at infinity *i.e.* for all $\lambda > 0$,

$$\lim_{x\to\infty} \frac{\ell(\lambda x)}{\ell(x)} = 1 . \tag{1.1}$$

The distribution F is said to be heavy tailed and the extreme-value parameter γ governs the heaviness of the tail. The estimation of γ is a central topic in the analysis of such distributions. Several estimators have thus been proposed in the statistical literature and their asymptotic distributions established under a second order condition: There exist a function $A(x) \to 0$ of constant sign for large values of x and a second order parameter $\rho < 0$ such that, for every $\lambda > 0$,

$$\lim_{x\to\infty} \frac{1}{A(x)} \log\left(\frac{\ell(\lambda x)}{\ell(x)}\right) = K_\rho(\lambda) := \int_1^\lambda u^{\rho-1} du . \tag{1.2}$$

Let us highlight that (1.2) implies that $|A|$ is regularly varying with index ρ, see [16]. Hence, as the second order parameter ρ decreases, the rate of convergence in (1.1) increases. Thus, the knowledge of ρ can be of high interest in real problems. For example, the second order parameter is of primordial importance in the adaptive choice of the best number of upper order statistics to be considered in the estimation of the extreme-value index [24]. The estimation of ρ can also be used to propose bias reduced estimators of the extreme value index (see for instance [4, 21, 23]) or of the Weibull tail-coefficient [9, 10], even though some bias reduction can be achieved with the canonical choice $\rho = -1$ as suggested in

[12, 22]. For the above mentioned reasons, the estimation of the second order parameter ρ has received a lot of attention in the extreme-value literature, see for instance [3, 6, 13, 14, 17, 19, 26, 30, 31].

In this paper, we propose a simple and general approach to estimate ρ. Let $\mathbb{I} = {}^t(1, ..., 1) \in \mathbb{R}^d$. The two main ingredients of our approach are a random variable $T_n = T_n(X_1, ..., X_n) \in \mathbb{R}^d$ verifying the following three assumptions:

(T1) There exists rvs ω_n, χ_n and a function $f\colon \mathbb{R}^- \to \mathbb{R}^d$ such that $\omega_n^{-1}(T_n - \chi_n\mathbb{I}) \xrightarrow{\mathbb{P}} f(\rho)$,

and a function $\psi\colon \mathbb{R}^d \to \mathbb{R}$ such that

(Ψ1) $\psi(x + \lambda\mathbb{I}) = \psi(x)$ for all $x \in \mathbb{R}^d$ and $\lambda \in \mathbb{R}$,

(Ψ2) $\psi(\lambda x) = \psi(x)$ for all $\lambda \in \mathbb{R}\backslash\{0\}$.

Note that **(T1)** imposes that T_n properly normalized converges in probability to some function of ρ, while **(Ψ1)** and **(Ψ2)** mean that ψ is both location and shift invariant. Starting from these three assumptions, we straightforwardly obtain that

$$\psi\big(\omega_n^{-1}(T_n - \chi_n\mathbb{I})\big) \;=\; \psi(T_n) \;\xrightarrow{\mathbb{P}}\; \psi(f(\rho)) \;,$$

under a continuity condition on ψ. Denoting by $Z_n := \psi(T_n)$ and by $\varphi := \psi \circ f\colon \mathbb{R}^- \to \mathbb{R}$, we obtain $Z_n \xrightarrow{\mathbb{P}} \varphi(\rho)$. It is thus clear that, under an additional regularity assumption and assuming that both Z_n and φ are known, ρ can be consistently estimated thanks to $\varphi^{-1}(Z_n)$. This estimation principle is described more precisely in Section 2. The consistency and asymptotic normality of the proposed estimator is also established. Examples of T_n random variables are presented in Section 3. Some functions ψ are proposed in Section 4 and it is shown that the above mentioned estimators [6, 13, 14, 17, 19] can be read as particular cases of our approach. As a consequence, this remark permits to establish their asymptotic properties in a simple and unified way. We illustrate how several asymptotically Gaussian estimators can be derived from this framework. Finally, some estimators are compared in Section 5 both from the asymptotic and finite sample size performances points of view.

2. MAIN RESULTS

Recall that T_n is a $\mathbb{R}^d$- random vector verifying **(T1)** and ψ is a function $\mathbb{R}^d \to \mathbb{R}$ verifying **(Ψ1)** and **(Ψ2)**. We further assume that:

(Ψ3) There exist $J_0 \subseteq \mathbb{R}^-$ and an open interval $J \subset \mathbb{R}$ such that $\varphi = \psi \circ f$ is a bijection $J_0 \to J$.

Under this assumption, the following estimator of ρ may be considered:

$$\hat{\rho}_n = \left| \begin{array}{ll} \varphi^{-1}(Z_n) & \text{if } Z_n \in J\,, \\ 0 & \text{otherwise}\,. \end{array} \right. \tag{2.1}$$

To derive the consistency of $\hat{\rho}_n$, an additional regularity assumption is introduced:

(Ψ4) ψ is continuous in a neighborhood of $f(\rho)$ and f is continuous in a neighborhood of ρ.

The proof of the next result is based on the heuristic consideration of Section 1 and is detailed in Section 6.

Theorem 2.1. *If* **(T1)** *and* **(Ψ1)**–**(Ψ4)** *hold then* $\hat{\rho}_n \xrightarrow{\mathbb{P}} \rho$ *as* $n \to \infty$.

The asymptotic normality of $\hat{\rho}_n$ can be established under a stronger version of **(Ψ4)**:

(Ψ5) ψ is continuously differentiable in a neighborhood of $f(\rho)$ and f is continuously differentiable in a neighborhood of ρ,

and the assumption that a normalized version of T_n is itself asymptotically Gaussian:

(T2) There exists two rvs ω_n, χ_n, a sequence $v_n \to \infty$, two functions $f, m\colon \mathbb{R}^- \to \mathbb{R}^d$ and a $d \times d$ matrix Σ such that $v_n\big(\omega_n^{-1}(T_n - \chi_n \mathbb{I}) - f(\rho)\big) \xrightarrow{d} \mathcal{N}_d\big(m(\rho), \gamma^2 \Sigma\big)$.

Theorem 2.2. *Suppose* **(T2)**, **(Ψ1)**–**(Ψ3)** *and* **(Ψ5)** *hold. If* $\rho \in J_0$ *and* $\varphi'(\rho) \neq 0$, *then*

$$v_n(\hat{\rho}_n - \rho) \xrightarrow{d} \mathcal{N}\left(\frac{m_\psi(\rho)}{\varphi'(\rho)}, \frac{\gamma^2 \sigma_\psi^2(\rho)}{(\varphi'(\rho))^2}\right),$$

with $\varphi'(\rho) = {}^t f'(\rho)\, \nabla\psi(f(\rho))$ *and where we have defined*

$$m_\psi(\rho) := {}^t m(\rho)\, \nabla\psi(f(\rho))\,,$$
$$\sigma_\psi^2(\rho) := {}^t\nabla\psi(f(\rho))\, \Sigma\, \nabla\psi(f(\rho))\,.$$

3. EXAMPLES OF T_n RANDOM VARIABLES

Let $X_{1,n} \leq ... \leq X_{n,n}$ be the sample of ascending order statistics and $k = k_n$ be an intermediate sequence *i.e.* such that $k \to \infty$ and $k/n \to 0$ as $n \to \infty$. Most extreme-value estimators are based either on the log-excesses $(\log X_{n-j+1,n} -$

$\log X_{n-k,n}$) or on the rescaled log-spacings $j(\log X_{n-j+1,n} - \log X_{n-j,n})$ defined for $j = 1, ..., k$. In the following, two examples of T_n random variables are presented based on weighted means of the log-excesses and of the rescaled log-spacings.

The first example is based on

$$R_k(\tau) = \frac{1}{k}\sum_{j=1}^{k} H_\tau\left(\frac{j}{k+1}\right) j\left(\log X_{n-j+1,n} - \log X_{n-j,n}\right) , \tag{3.1}$$

where $H_\tau\colon [0,1] \to \mathbb{R}$ is a weight function indexed by a parameter $\tau \in (0,\infty)$. Without loss of generality, one can assume that H_τ integrates to one. This random variable is used for instance in [1] to estimate the extreme-value index γ, in [17, 26, 30] to estimate the second order parameter ρ and in [18] to estimate the third order parameter, see condition **(C2)** below. It is a particular case of the kernel statistic introduced in [7]. Let us also note that, in the case where $H_\tau(u) = 1$ for all $u \in [0,1]$, $R_k(\tau)$ reduces to the well-known Hill estimator [27]. The asymptotic properties of $R_k(\tau)$ require some technical condition (denoted by **(C1)**) on the weight function H_τ. It has been first introduced in [1] and it is recalled hereafter. Introducing the operator

$$\mu\colon \quad h \in L_2([0,1]) \longrightarrow \mu(h) = \int_0^1 h(u)\,du \in \mathbb{R}$$

and $I_t(u) = u^{-t}$ for $t \le 0$ and $u \in (0,1]$, the condition can be written as

(C1) $H_\tau \in L_2([0,1])$, $\mu(|H_\tau| I_{\rho+1+\varepsilon}) < \infty$ and

$$H_\tau(t) = \frac{1}{t}\int_0^t u(\nu)\,d\nu$$

for some $\varepsilon > 0$ and for some function u satisfying for all $j = 1, ..., k$

$$\left|(k+1)\int_{(j-1)/(k+1)}^{j/(k+1)} u(t)\,dt\right| \le g\left(\frac{j}{k+1}\right),$$

where g is a positive continuous and integrable function defined on $(0,1)$. Furthermore, for $\eta \in \{0,1\}$, and $k \to \infty$:

$$\frac{1}{k}\sum_{j=1}^{k} H_\tau\left(\frac{j}{k+1}\right)\left(\frac{j}{k+1}\right)^{-\eta\rho} = \mu(H_\tau I_{\eta\rho}) + o(k^{-1/2}) ,$$

$$\max_{j\in\{1,...,k\}}\left|H_\tau\left(\frac{j}{k+1}\right)\right| = o(k^{1/2}) .$$

It is then possible to define $T_n^{(R)}$ on the basis of $R_k(\tau)$, given in (3.1), as

$$T_n^{(R)} = \left(T_{n,i}^{(R)} = \left(R_k(\tau_i)/\gamma\right)^{\theta_i},\ i = 1, ..., d\right) , \tag{3.2}$$

where θ_i, $i=1,...,d$ are positive parameters. In the next lemma, it is proved that $T_n^{(R)}$ satisfies condition **(T2)** under a third order condition, which is a refinement of (1.2):

(C2) There exist functions $A(x)\to 0$ and $B(x)\to 0$ both of constant sign for large values of x, a second order parameter $\rho<0$ and a third order parameter $\beta<0$ such that, for every $\lambda>0$,

$$\lim_{x\to\infty} \frac{\big(\log\ell(\lambda x)-\log\ell(x)\big)/A(x)-K_\rho(\lambda)}{B(x)} \;=\; L_{(\rho,\beta)}(\lambda)$$

where

$$L_{(\rho,\beta)}(\lambda) \;:=\; \int_1^\lambda s^{\rho-1}\int_1^s u^{\beta-1}\,du\,ds\ ,$$

and the functions $|A|$ and $|B|$ are regularly varying functions with index ρ and β respectively.

This condition is the cornerstone for establishing the asymptotic normality of estimators of ρ. Let us denote by $Y_{n-k,n}$ the $n-k$ largest order statistics from a n-sample of standard Pareto rv.

Lemma 3.1. *Suppose* **(C1)**, **(C2)** *hold and let* $k=k_n$ *be an intermediate sequence* k *such that*

$$\begin{aligned} &k\to\infty\,, \quad n/k\to\infty\,, \quad k^{1/2}A(n/k)\to\infty\ ,\\ &k^{1/2}A^2(n/k)\to\lambda_A\,, \quad k^{1/2}A(n/k)\,B(n/k)\to\lambda_B\ , \end{aligned} \tag{3.3}$$

for $\lambda_A\in\mathbb{R}$ *and* $\lambda_B\in\mathbb{R}$. *Then, the random vector* $T_n^{(R)}$ *satisfies* **(T2)** *with* $\omega_n^{(R)}= A(Y_{n-k,n})/\gamma$, $\chi_n^{(R)}=1$, $v_n=k^{1/2}A(n/k)$,

$$\begin{aligned} f^{(R)}(\rho) &= \Big(\theta_i\,\mu(H_{\tau_i}I_\rho),\ i=1,...,d\Big)\,,\\ m^{(R)}(\rho) &= \left(\lambda_A\frac{\theta_i(\theta_i-1)}{2\gamma}\,\mu^2(H_{\tau_i}I_\rho)-\lambda_B\,\theta_i\,\mu(H_{\tau_i}I_\rho K_{-\beta});\ i=1,...,d\right), \end{aligned}$$

and, for $(i,j)\in\{1,...,d\}^2$, $\Sigma_{i,j}^{(R)}=\theta_i\,\theta_j\,\mu(H_{\tau_i}H_{\tau_j})$.

The proof is a straightforward consequence of Theorem 2 and Appendix A.5 in [17].

The second example requires some additional notations. Let us consider the operator $\vartheta\colon L_2([0,1])\times L_2([0,1])\longrightarrow\mathbb{R}$ defined by

$$\vartheta(h_1,h_2) \;=\; \int_0^1\!\!\int_0^1 h_1(u)\,h_2(v)\,(u\wedge v-uv)\,du\,dv$$

and the two functions $\bar{I}_t(u) = (1-u)^{-t}$ and $J_t(u) = (-\log u)^{-t}$ defined for $t \le 0$ and $u \in (0,1]$. The random variables of interest are

$$S_k(\tau,\alpha) = \frac{1}{k}\sum_{j=1}^{k} G_{\tau,\alpha}\left(\frac{j}{k+1}\right)\left(\log X_{n-j+1,n} - \log X_{n-k,n}\right)^{\alpha}, \tag{3.4}$$

where $G_{\tau,\alpha}$ is a positive function indexed by two positive parameters α and τ. Without loss of generality, it can be assumed that $\mu(G_{\tau,\alpha}J_{-\alpha}) = 1$. In [8, 20, 29] several estimators of γ based on $S_k(\tau,\alpha)$ are introduced in the particular case where G is constant. Most recently, in [6, 14, 26, 28], $S_k(\tau,\alpha)$ is used to estimate the parameters γ and ρ. The asymptotic distribution of these estimators is obtained under the following assumption on the function $G_{\tau,\alpha}$.

(C3) The function $G_{\tau,\alpha}$ is positive, non-increasing and integrable on $(0,1)$. Furthermore, there exists $\delta > 1/2$ such that $0 < \mu(G_{\tau,\alpha}I_\delta) < \infty$ and $0 < \mu(G_{\tau,\alpha}\bar{I}_\delta) < \infty$.

It is then possible to define $T_n^{(S)}$ on the basis of $S_k(\tau,\alpha)$, see (3.4), as

$$T_n^{(S)} = \left(T_{n,i}^{(S)} = \left(S_k(\tau_i,\alpha_i)/\gamma^{\alpha_i}\right)^{\theta_i},\ i=1,\dots,d\right). \tag{3.5}$$

The following result is the analogous of Lemma 3.1 for the above random variables.

Lemma 3.2. *Suppose* **(C2)**, **(C3)** *hold. If the intermediate sequence k satisfy (3.3) then the random vector $T_n^{(S)}$ satisfies* **(T2)** *with $\omega_n^{(S)} = A(n/k)/\gamma$, $\chi_n^{(S)} = 1$, $v_n = k^{1/2}A(n/k)$,*

$$f^{(S)}(\rho) = \left(-\theta_i\,\alpha_i\,\mu\big(G_{\tau_i,\alpha_i}J_{1-\alpha_i}K_{-\rho}\big);\ i=1,\dots,d\right),$$

$$m^{(S)}(\rho) = \left(\lambda_A\frac{\theta_i\alpha_i(\alpha_i-1)}{2\gamma}\,\mu\big(G_{\tau_i\alpha_i}J_{2-\alpha_i}K_{-\rho}^2\big) + \lambda_B\,\alpha_i\,\theta_i\,\mu\big(G_{\tau_i,\alpha_i}J_{1-\alpha_i}L_{(-\rho,-\beta)}\big);\right.$$
$$\left. i=1,\dots,d\right),$$

and, for $(i,j)\in\{1,\dots,d\}^2$, $\Sigma_{i,j}^{(S)} = \theta_i\,\theta_j\,\alpha_i\,\alpha_j\,\vartheta\big(G_{\tau_i,\alpha_i}J_{1-\alpha_i},\,G_{\tau_j,\alpha_j}J_{1-\alpha_j}\big)$.

The proof is a straightforward consequence of Proposition 3 and Lemma 1 in [6]. In the next section, we illustrate how the combination of $T_n^{(R)}$ or $T_n^{(S)}$ with some function ψ following (2.1) can lead to existing or new estimators of ρ.

4. APPLICATIONS

In this section, we propose estimators of ρ based on the random variable $T_n^{(R)}$ (subsection 4.1) and $T_n^{(S)}$ (subsection 4.2). In both cases, $d = 8$ and the following function $\psi_\delta\colon \mathcal{D} \mapsto \mathbb{R}\backslash\{0\}$ is considered

$$\psi_\delta(x_1, ..., x_8) \;=\; \widetilde{\psi}_\delta(x_1 - x_2, x_3 - x_4, x_5 - x_6, x_7 - x_8)\ , \tag{4.1}$$

where $\delta \geq 0$, $\mathcal{D} = \big\{(x_1, ..., x_8) \in \mathbb{R}^8;\ x_1 \neq x_2,\ x_3 \neq x_4, \text{ and } (x_5-x_6)(x_7-x_8) > 0\big\}$, and $\widetilde{\psi}_\delta\colon \mathbb{R}^4 \mapsto \mathbb{R}$ is given by:

$$\widetilde{\psi}_\delta(y_1, ..., y_4) \;=\; \frac{y_1}{y_2}\left(\frac{y_4}{y_3}\right)^{\delta}.$$

Let us highlight that ψ_δ verifies the invariance properties **(Ψ1)** and **(Ψ2)**.

4.1. Estimators based on the random variable $R_k(\tau)$

Since $d = 8$, the random variable $T_n^{(R)}$ defined in (3.2) depends on 16 parameters: $\big\{(\theta_i, \tau_i) \in (0,\infty)^2,\ i = 1, ..., 8\big\}$. The following condition on these parameters is introduced. Let $\tilde\theta = (\tilde\theta_1, ..., \tilde\theta_4) \in (0,\infty)^4$ with $\tilde\theta_3 \neq \tilde\theta_4$.

(C4) $\big\{\theta_i = \tilde\theta_{\lceil i/2\rceil},\ i = 1, ..., 8\big\}$ with $\delta = (\tilde\theta_1 - \tilde\theta_2)/(\tilde\theta_3 - \tilde\theta_4)$. Furthermore, $\tau_1 < \tau_2 \leq \tau_3 < \tau_4,\ \tau_5 < \tau_6 \leq \tau_7 < \tau_8$,

where $\lceil x\rceil = \inf\{n \in \mathbb{N} \mid x \leq n\}$. Under this condition, $T_n^{(R)}$ involves 12 free parameters. We also introduce the following notations: $Z_n^{(R)} = \psi_\delta(T_n^{(R)})$ and $\varphi_\delta^{(R)} = \psi_\delta \circ f^{(R)}$ where $f^{(R)}$ is given in Lemma 3.1. Note that, since $\delta = (\tilde\theta_1 - \tilde\theta_2)/(\tilde\theta_3 - \tilde\theta_4)$, it is easy to check that $Z_n^{(R)}$ does not depend on the unknown parameter γ. We now establish the asymptotic normality of the estimator $\hat\rho_n^{(R)}$ defined by (2.1) when $T_n^{(R)}$ and the function ψ_δ are used:

$$\hat\rho_n^{(R)} = \left|\begin{array}{ll} (\varphi_\delta^{(R)})^{-1}(Z_n^{(R)}) & \text{if } Z_n^{(R)} \in J\,, \\ 0 & \text{otherwise}\,. \end{array}\right. \tag{4.2}$$

The following additional condition is required:

(C5) The function $\nu_\rho(\tau) = \mu(H_\tau I_\rho)$ is differentiable with, for all $\rho < 0$ and all $\tau \in \mathbb{R}$, $\nu'_\rho(\tau) > 0$.

Let us denote for $i \in \{1, ..., 4\}$,

$$m_A^{(R,i)} \;=\; \exp\Big\{(\tilde\theta_i - 1)\,\big(\nu_\rho(\tau_{2i-1}) + \nu_\rho(\tau_{2i})\big)\Big\}\,,$$

$$m_B^{(R,i)} \;=\; \exp\left\{\frac{\mu\big((H_{\tau_{2i-1}} - H_{\tau_{2i}})\, I_\rho\, K_{-\beta}\big)}{\nu_\rho(\tau_{2i-1}) - \nu_\rho(\tau_{2i})}\right\},$$

and for $u \in [0,1]$,

$$v^{(R,i)}(u) \;=\; \exp\left\{\frac{H_{\tau_{2i-1}}(u) - H_{\tau_{2i}}(u)}{\nu_\rho(\tau_{2i-1}) - \nu_\rho(\tau_{2i})}\right\}.$$

For the sake of simplicity, we also introduce $m_A^{(R)} = \big(m_A^{(R,i)},\ i=1,...,4\big)$, $m_B^{(R)} = \big(m_B^{(R,i)},\ i=1,...,4\big)$ and $v^{(R)} = \big(v^{(R,i)},\ i=1,...,4\big)$.

Corollary 4.1. *Suppose* **(C1)**, **(C2)**, **(C4)** *and* **(C5)** *hold. There exist two intervals J and J_0 such that for all $\rho \in J_0$ and for a sequence k satisfying (3.3),*

$$k^{1/2} A(n/k)\,(\hat\rho_n^{(R)} - \rho) \ \xrightarrow{d}\ \mathcal{N}\left(\frac{\lambda_A}{2\gamma}\,\mathcal{AB}_1^{(R)}(\delta,\rho) - \lambda_B\,\mathcal{AB}_2^{(R)}(\delta,\rho,\beta),\ \gamma^2 \mathcal{AV}^{(R)}(\delta,\rho)\right)$$

where

$$\mathcal{AB}_1^{(R)}(\delta,\rho) \;=\; \frac{\varphi_\delta^{(R)}(\rho)}{\big[\varphi_\delta^{(R)}\big]'(\rho)}\ \log\tilde\psi_\delta\big(m_A^{(R)}\big)\,,$$

$$\mathcal{AB}_2^{(R)}(\delta,\rho,\beta) \;=\; \frac{\varphi_\delta^{(R)}(\rho)}{\big[\varphi_\delta^{(R)}\big]'(\rho)}\ \log\tilde\psi_\delta\big(m_B^{(R)}\big)\,,$$

$$\mathcal{AV}^{(R)}(\delta,\rho) \;=\; \left(\frac{\varphi_\delta^{(R)}(\rho)}{\big[\varphi_\delta^{(R)}\big]'(\rho)}\right)^2 \mu\Big(\log^2\tilde\psi_\delta\big(v^{(R)}\big)\Big)\,.$$

Note that this result can be read as an extension of [17], Proposition 3, in two ways. First, we do not limit ourselves to the case $\delta = 1$. Second, we do not assume that the function $\varphi_\delta^{(R)}$ is a bijection, but it is shown to be a consequence of **(C4)**. Besides, the proof of Corollary 4.1 is very simple based on Theorem 2.2 and Lemma 3.1, see Section 6 for details.

As an example, the function $H_\tau \colon u \in [0,1] \mapsto \tau u^{\tau-1}$, $\tau \ge 1$ satisfies conditions **(C1)** and **(C5)** since $\nu_\rho(\tau) = \tau/(\tau-\rho)$. Letting $\tau_1 \le \tau_5$, $\tau_2 = \tau_3$, $\tau_4 = \tau_8$ and $\tau_6 = \tau_7$ leads to a simple expression of $\varphi_\delta^{(R)}$:

$$\varphi_\delta^{(R)}(\rho) \;=\; \omega(\delta,\tilde\theta)\left(\frac{\tau_4-\rho}{\tau_1-\rho}\right)\left(\frac{\tau_5-\rho}{\tau_4-\rho}\right)^{\delta} \tag{4.3}$$

where

$$\omega(\delta,\tilde\theta) = \left(\frac{\tilde\theta_1(\tau_1-\tau_2)}{\tilde\theta_2(\tau_2-\tau_4)}\right)\left(\frac{\tilde\theta_4(\tau_6-\tau_4)}{\tilde\theta_3(\tau_5-\tau_6)}\right)^{\delta}.$$

Moreover, one also has explicit forms for J_0 and J in two situations:

- **(i)** If $0 \le \delta \le \delta_0 := (\tau_4-\tau_1)/(\tau_4-\tau_5)$ then $\varphi_\delta^{(R)}$ is increasing from $J_0 = \mathbb{R}^-$ to $J = \omega(\delta,\tilde\theta) \bullet (1, \widetilde\psi_\delta(\tau_4,\tau_1,\tau_4,\tau_5))$.
- **(ii)** If $\delta \ge \delta_1 := \delta_0\tau_5/\tau_1$ then $\varphi_\delta^{(R)}$ is decreasing from $J_0 = \mathbb{R}^-$ to $J = \omega(\delta,\tilde\theta) \bullet (\widetilde\psi_\delta(\tau_4,\tau_1,\tau_4,\tau_5), 1)$.

Here, $\bullet$ denotes the scaling operator. The case $\delta \in (\delta_0,\delta_1)$ is not considered here, since one can show that, in this situation, $J_0 \subsetneq \mathbb{R}^-$ and thus the condition $\rho \in J_0$

of Corollary 4.1 is not necessarily satisfied. Let us now list some particular cases where the inverse function of $\varphi_\delta^{(R)}$ is explicit.

Example 4.1. Let $\delta = 1$ *i.e.* $\tilde\theta_1 - \tilde\theta_2 = \tilde\theta_3 - \tilde\theta_4$. The rv $Z_n^{(R)}$ is denoted by $Z_{n,1}^{(R)}$. Since $\delta_0 > 1$, we are in situation (i) and

$$\hat\rho_{n,1}^{(R)} = \frac{\tau_5\,\omega(1,\tilde\theta) - \tau_1 Z_{n,1}^{(R)}}{\omega(1,\tilde\theta) - Z_{n,1}^{(R)}}\,\mathbb{I}\Big\{Z_{n,1}^{(R)} \in \omega(1,\tilde\theta) \bullet \big(1, \widetilde\psi_1(\tau_4,\tau_1,\tau_4,\tau_5)\big)\Big\}\,.$$

Remark that this estimator coincides with the one proposed in [17], Lemma 1.

Example 4.2. Let $\delta = 0$ *i.e.* $\tilde\theta_1 = \tilde\theta_2$. The rv $Z_n^{(R)}$ is thus denoted by $Z_{n,2}^{(R)}$. Again, we are in situation (i) and a new estimator of ρ is obtained

$$\hat\rho_{n,2}^{(R)} = \frac{\tau_4\,\omega(0,\tilde\theta) - \tau_1 Z_{n,2}^{(R)}}{\omega(0,\tilde\theta) - Z_{n,2}^{(R)}}\,\mathbb{I}\Big\{Z_{n,2}^{(R)} \in \omega(0,\tilde\theta) \bullet \big(1, \widetilde\psi_0(\tau_4,\tau_1,\tau_4,\tau_5)\big)\Big\}\,.$$

Example 4.3. Let $\tau_1 = \tau_5$. In this case $\delta_0 = \delta_1 = 1$ and thus, we are in situation (i) if $\delta < 1$ and in situation (ii) otherwise. In this case, the rv $Z_n^{(R)}$ is denoted by $Z_{n,3}^{(R)}$. A new estimator of ρ is obtained:

$$\hat\rho_{n,3}^{(R)} = \frac{\tau_4\big(Z_{n,3}^{(R)}/\omega(\delta,\tilde\theta)\big)^{1/(\delta-1)} - \tau_1}{\big(Z_{n,3}^{(R)}/\omega(\delta,\tilde\theta)\big)^{1/(\delta-1)} - 1}\,\mathbb{I}\big\{Z_{n,3}^{(R)} \in J\big\}\,.$$

4.2. Estimators based on the random variable $S_k(\tau,\alpha)$

The random variable $T_n^{(S)}$ defined in (3.5) depends on 24 parameters: $\big\{(\theta_i,\tau_i,\alpha_i) \in (0,\infty)^3,\ i=1,...,8\big\}$. Let $(\zeta_1,...,\zeta_4) \in (0,\infty)^4$ with $\zeta_3 \neq \zeta_4$. In the following, we assume that

(C6) $\big\{\theta_i\alpha_i = \zeta_{\lceil i/2\rceil},\ i=1,...,8\big\}$ with $\delta = (\zeta_1-\zeta_2)/(\zeta_3-\zeta_4)$. Furthermore, $(\tau_{2i-1},\alpha_{2i-1}) \neq (\tau_{2i},\alpha_{2i})$, for $i=1,...,4$ and, for $i=3,4$, $(\tau_{2i-1},\alpha_{2i-1}) < (\tau_{2i},\alpha_{2i})$,

where $(x,y) \neq (s,t)$ means that $x \neq s$ and/or $y \neq t$ and $(x,y) < (s,t)$ means that $x < s$ and $y \le t$ or $x = s$ and $y < t$. We introduce the notations: $Z_n^{(S)} = \psi_\delta(T_n^{(S)})$ and $\varphi_\delta^{(S)} = \psi_\delta \circ f^{(S)}$ where $f^{(S)}$ is given in Lemma 3.2. Under this condition, $T_n^{(S)}$ involves 20 free parameters. Besides, since $\delta = (\zeta_1-\zeta_2)/(\zeta_3-\zeta_4)$, it is easy to check that $Z_n^{(S)}$ does not depend on the unknown parameter γ. To establish the asymptotic distribution of the estimator $\hat\rho_n^{(S)}$, the following condition is required:

(C7) For all $\rho < 0$, the function $\nu_\rho(\tau, \alpha) = \mu(G_{\tau,\alpha} J_{1-\alpha} K_{-\rho})$ is differentiable with $\frac{\partial}{\partial \tau}\nu_\rho(\tau, \alpha) > 0$ and $\frac{\partial}{\partial \alpha}\nu_\rho(\tau, \alpha) > 0$ for all $\alpha > 0$ and all $\tau \in \mathbb{R}$.

For $i = 1, ..., 4$, let $\log m_A^{(S,i)}$ and $\log m_B^{(S,i)}$ denote respectively

$$\left\{ \frac{(\alpha_{2i-1} - 1)\, \mu\big(G_{\tau_{2i-1},\alpha_{2i-1}} J_{2-\alpha_{2i-1}} K^2_{-\rho}\big) - (\alpha_{2i} - 1)\, \mu\big(G_{\tau_{2i},\alpha_{2i}} J_{2-\alpha_{2i}} K^2_{-\rho}\big)}{\nu_\rho(\tau_{2i}, \alpha_{2i}) - \nu_\rho(\tau_{2i-1}, \alpha_{2i-1})} \right\},$$

$$\left\{ \frac{\mu\big(G_{\tau_{2i-1},\alpha_{2i-1}} J_{1-\alpha_{2i-1}} L_{(-\rho,-\beta)}\big) - \mu\big(G_{\tau_{2i},\alpha_{2i}} J_{1-\alpha_{2i}} L_{(-\rho,-\beta)}\big)}{\nu_\rho(\tau_{2i}, \alpha_{2i}) - \nu_\rho(\tau_{2i-1}, \alpha_{2i-1})} \right\},$$

and for $u \in [0, 1]$,

$$v^{(S,i)}(u) = \frac{G_{\tau_{2i-1},\alpha_{2i-1}}(u)\, J_{1-\alpha_{2i-1}}(u) - G_{\tau_{2i},\alpha_{2i}}(u)\, J_{1-\alpha_{2i}}(u)}{\nu_\rho(\tau_{2i}, \alpha_{2i}) - \nu_\rho(\tau_{2i-1}, \alpha_{2i-1})} .$$

Let us also consider $m_A^{(S)} = \big(m_A^{(S,i)},\ i=1, ..., 4\big)$ and $m_B^{(S)} = \big(m_B^{(S,i)},\ i=1, ..., 4\big)$. The next result is a direct consequence of Theorem 2.2 and Lemma 3.2, see Section 6 for a short proof.

Corollary 4.2. *Suppose* **(C2)**, **(C3)**, **(C6)** *and* **(C7)** *hold. There exist two intervals J and J_0 such that for all $\rho \in J_0$ and for a sequence k satisfying (3.3),*

$$k^{1/2} A(n/k)\, \big(\hat\rho_n^{(S)} - \rho\big) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda_A}{2\gamma}\, \mathcal{AB}_1^{(S)}(\delta, \rho) + \lambda_B\, \mathcal{AB}_2^{(S)}(\delta, \rho, \beta),\ \gamma^2 \mathcal{AV}^{(S)}(\delta, \rho)\right)$$

where

$$\mathcal{AB}_1^{(S)}(\delta, \rho) = \frac{\varphi_\delta^{(S)}(\rho)}{\big[\varphi_\delta^{(S)}\big]'(\rho)} \log \tilde\psi_\delta\big(m_A^{(S)}\big),$$

$$\mathcal{AB}_2^{(S)}(\delta, \rho, \beta) = \frac{\varphi_\delta^{(S)}(\rho)}{\big[\varphi_\delta^{(S)}\big]'(\rho)} \log \tilde\psi_\delta\big(m_B^{(S)}\big),$$

$$\mathcal{AV}^{(S)}(\delta, \rho) = \left(\frac{\varphi_\delta^{(S)}(\rho)}{\big[\varphi_\delta^{(S)}\big]'(\rho)}\right)^2 \times \vartheta\Big(v^{(S,1)} - v^{(S,2)} - \delta\big(v^{(S,3)} - v^{(S,4)}\big), v^{(S,1)} - v^{(S,2)} - \delta\big(v^{(S,3)} - v^{(S,4)}\big)\Big).$$

Let us highlight that Proposition 5, Proposition 7 and Proposition 9 of [6] are particular cases of Corollary 4.2 for three different value of δ ($\delta = 2$, $\delta = 1$ and $\delta = 0$ respectively). The asymptotic normality of the estimators proposed in [19] and in [14] can also be easily established with Corollary 4.2.

As an example of function $G_{\tau,\alpha}$, one can consider the function defined on $[0, 1]$ by:

$$G_{\tau,\alpha}(u) = \frac{g_{\tau-1}(u)}{\int_0^1 g_{\tau-1}(x)\, J_{-\alpha}(x) dx} \qquad \text{for} \quad \tau \geq 1 \quad \text{and} \quad \alpha > 0 ,$$

where the function g_τ is given by

$$g_0(x) = 1 \,, \qquad g_{\tau-1}(x) = \frac{\tau}{\tau-1}\,(1 - x^{\tau-1}), \quad \forall\, \tau > 1 \,.$$

Clearly, the function $G_{\tau,\alpha}$ satisfies condition **(C7)** and, under **(C6)**, the expression of $\varphi_\rho^{(S)}$ is

$$\varphi_\delta^{(S)}(\rho) \;=\; \frac{\zeta_1}{\zeta_2}\left(\frac{\zeta_4}{\zeta_3}\right)^{\!\delta} \frac{\nu_\rho(\tau_1,\alpha_1) - \nu_\rho(\tau_2,\alpha_2)}{\nu_\rho(\tau_3,\alpha_3) - \nu_\rho(\tau_4,\alpha_4)} \left[\frac{\nu_\rho(\tau_7,\alpha_7) - \nu_\rho(\tau_8,\alpha_8)}{\nu_\rho(\tau_5,\alpha_5) - \nu_\rho(\tau_6,\alpha_6)}\right]^{\delta}$$

with

$$\nu_\rho(\tau,\alpha) \;=\; \frac{1 - (1-\rho)^{-\alpha} + (\tau-\rho)^{-\alpha} - \tau^{-\alpha}}{\alpha\,\rho\,(1 - \tau^{-\alpha-1})} \qquad \text{if} \quad \tau \neq 1 \,,$$

and

$$\nu_\rho(1,\alpha) \;=\; \frac{1}{\alpha\rho}\,\frac{(1-\rho)^\alpha - 1}{(1-\rho)^\alpha} \,.$$

Even if Corollary 4.2 ensures the existence of intervals J_0 and J, they are impossible to specify in the general case. In the following, we consider several sets of parameters where these intervals can be easily exhibited and for which the inverse function $\varphi_\delta^{(S)}$ admits an explicit form. To this end, it is assumed that $\tau_2 = \tau_3 = \tau_5 = \tau_6 = \tau_7 = \tau_8 = \alpha_7 = 1$, $\alpha_6 = 3$, $\alpha_8 = 2$ and the following notation is introduced:

$$\omega^*(\delta,\zeta) \;=\; \frac{\zeta_1}{\zeta_2}\left(\frac{3\,\zeta_4}{\zeta_3}\right)^{\!\delta} .$$

In all the examples below, $J_0 = \mathbb{R}^-$ and thus the condition $\rho \in J_0$ is always satisfied. The first three examples correspond to existing estimators of the second order parameter while the three last examples give rise to new estimators.

Example 4.4. Let $\delta = 0$ (i.e. $\zeta_1 = \zeta_2$), $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 1$, $\tau_1 = 2$ and $\tau_4 = 3$. Denoting by $Z_{n,4}^{(S)}$ the rv $Z_n^{(S)}$, the estimator of ρ is given by:

$$\hat\rho_{n,4}^{(S)} \;=\; \frac{6\big(Z_{n,4}^{(S)} + 2\big)}{3\,Z_{n,4}^{(S)} + 4}\; \mathbb{I}\Big\{Z_{n,4}^{(S)} \in (-2, -4/3)\Big\} \,.$$

Note that this estimator corresponds to the estimator $\hat\rho_{n,k}^{[2]}$ defined in [6], Section 5.2.

Example 4.5. Let $\delta = 0$, $\alpha_1 = \alpha_3 = \alpha_4 = 1$ and $\tau_1 = \tau_4 = \alpha_2 = 2$. Denoting by $Z_{n,5}^{(S)}$ the rv $Z_n^{(S)}$, we find back the estimator $\hat\rho_{n,k}^{[3]}$ proposed in [6], Section 5.2:

$$\hat\rho_{n,5}^{(S)} \;=\; \frac{2\big(Z_{n,5}^{(S)} - 2\big)}{2\,Z_{n,5}^{(S)} - 1}\; \mathbb{I}\Big\{Z_{n,5}^{(S)} \in (1/2, 2)\Big\} \,.$$

Example 4.6. Let $\alpha_1 = \zeta_1 = 4$, $\alpha_3 = \zeta_2 = \zeta_4 = 2$, $\zeta_3 = 3$ and $\alpha_2 = \alpha_4 = \alpha_5 = \tau_1 = \tau_4 = 1$. These choices entail $\delta = 2$. Denoting by $Z_{n,6}^{(S)}$ the rv $Z_n^{(S)}$, the estimator of ρ given by:

$$\hat{\rho}_{n,6}^{(S)} = \frac{6Z_{n,6}^{(S)} - 4 + \left(3Z_{n,6}^{(S)} - 2\right)^{1/2}}{4Z_{n,6}^{(S)} - 3} \, \mathbb{I}\Big\{ Z_{n,6}^{(S)} \in (2/3, 3/4) \Big\} .$$

corresponds to the one proposed in [19], equation (12).

Example 4.7. Consider the case $\delta = 1$ (i.e. $\zeta_1 - \zeta_2 = \zeta_3 - \zeta_4$), $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 1$, $\tau_1 = \alpha_5 = 2$ and $\tau_4 = 3$. Denoting by $Z_{n,7}^{(S)}$ the rv $Z_n^{(S)}$, a new estimator of ρ is given by:

$$\hat{\rho}_{n,7}^{(S)} = \frac{6Z_{n,7}^{(S)} + 4\omega^*(1,\zeta)}{3Z_{n,7}^{(S)} + 4\omega^*(1,\zeta)} \, \mathbb{I}\Big\{ Z_{n,7}^{(S)} \in \omega^*(1,\zeta) \bullet (-4/3, -2/3) \Big\} .$$

Example 4.8. Let $\delta = 1$, $\alpha_1 = \alpha_3 = \alpha_4 = 1$ and $\tau_1 = \tau_4 = \alpha_2 = \alpha_5 = 2$. Denoting by $Z_{n,8}^{(S)}$ the rv $Z_n^{(S)}$, we obtain a new estimator of ρ:

$$\hat{\rho}_{n,8}^{(S)} = \frac{6Z_{n,8}^{(S)} - 4\omega^*(1,\zeta)}{2Z_{n,8}^{(S)} - \omega^*(1,\zeta)} \, \mathbb{I}\Big\{ Z_{n,8}^{(S)} \in \omega^*(1,\zeta) \bullet (1/2, 2/3) \Big\} .$$

Example 4.9. Let $\tau_1 = \tau_4 = \alpha_1 = 1$, $\alpha_2 = \alpha_3 = \alpha_5 = 2$ and $\alpha_4 = 3$. Denoting by $Z_{n,9}^{(S)}$ the rv $Z_n^{(S)}$, the estimator of ρ is given by:

$$\hat{\rho}_{n,9}^{(S)} = \frac{3\Big(Z_{n,9}^{(S)} / \big(3\omega^*(\delta,\zeta)\big)\Big)^{1/(\delta+1)} - 1}{\Big(Z_{n,9}^{(S)} / \big(3\omega^*(\delta,\zeta)\big)\Big)^{1/(\delta+1)} - 1} \, \mathbb{I}\Big\{ Z_{n,9}^{(S)} \in \omega^*(\delta,\zeta) \bullet (3^{-\delta}, 3) \Big\} .$$

In the particular case where $\delta = 0$, this estimator corresponds to the one proposed in [13].

To summarize, we have illustrated how Theorem 2.2 may be used to prove the asymptotic normality of estimators built on $T_n^{(R)}$ or $T_n^{(S)}$: Corollary 4.1 and Corollary 4.2 cover a large number of estimators proposed in the literature. Five new estimators of ρ have been introduced: $\hat{\rho}_{n,2}^{(R)}$, $\hat{\rho}_{n,3}^{(R)}$, $\hat{\rho}_{n,7}^{(S)}$, $\hat{\rho}_{n,8}^{(S)}$ and $\hat{\rho}_{n,9}^{(S)}$. All of them are explicit and are asymptotically Gaussian. The comparison of their finite sample properties is a huge task since they may depend on their parameters (θ_i, τ_i, α_i) as well as on the simulated distribution. We conclude this study by proposing a method for selecting some "asymptotic optimal" parameters within a family of estimators. The performances and the limits of this technique are illustrated by comparing several estimators on simulated data.

5. COMPARISON OF SOME ESTIMATORS

Some estimators of ρ are now compared on a specific Pareto-type model, namely the Burr distribution with cdf $F(x) = 1 - (\zeta/(\zeta + x^\eta))^\lambda$, $x > 0$, $\zeta, \lambda, \eta > 0$, considered for instance in [2], equation (3). The associated extreme-value index is $\gamma = 1/(\lambda\eta)$ and this model satisfies the third order condition **(C2)** with $\rho = \beta = -1/\lambda$, $A(x) = \gamma x^\rho/(1 - x^\rho)$ and $B(x) = \rho x^\rho/(1 - x^\rho)$. We limit ourselves to the case $\zeta = 1$ and $\lambda = 1/\eta$ so that $\gamma = 1$.

5.1. Estimators based on the random variable $R_k(\tau)$

Let us first focus on the estimators of ρ based on the random variables $R_k(\tau_i)$ considered in Section 4.1 with kernel functions $H_{\tau_i}(u) = \tau_i u^{\tau_i - 1}$, for $i = 1, ..., 8$. The values of the parameters $\tau_1, ..., \tau_8$, $\tilde{\theta}_1$, $\tilde{\theta}_3$ and $\tilde{\theta}_4$ are taken as in [17, 30]: $\tau_1 = 1.25$, $\tau_2 = \tau_3 = 1.75$, $\tau_4 = \tau_8 = 2$, $\tau_5 = 1.5$, $\tau_6 = \tau_7 = 1.75$, $\tilde{\theta}_1 = 0.01$, $\tilde{\theta}_3 = 0.02$ and $\tilde{\theta}_4 = 0.04$. According to the authors, these values yield good results for distributions satisfying the third order condition **(C2)** with $\beta = \rho$. For these parameters, a simple expression of $\varphi_\delta^{(R)}$ is obtained, see (4.3), and we have $\delta_0 = 1.5$ and $\delta_1 = 1.8$. Recall that $\tilde{\theta}_2 = \tilde{\theta}_1 + \delta(\tilde{\theta}_4 - \tilde{\theta}_3)$ for $\delta \geq 0$. In the following, we propose to choose the remaining parameter δ using a method similar to the one proposed in [15]. It consists in minimizing with respect to δ an upper bound on the asymptotic mean-squared error. The method is described in Paragraph 5.1.1 and an example of application is presented in Paragraph 5.1.2.

5.1.1. Controlling the asymptotic mean-squared error

As in [17], we assume that $\rho = \beta$. Following Corollary 4.1, the asymptotic bias components of $\hat{\rho}_n^{(R)}$ are respectively proportional to $\mathcal{AB}_1^{(R)}(\delta, \rho)$ and $\mathcal{AB}_2^{(R)}(\delta, \rho, \rho)$ while its asymptotic variance is proportional to $\mathcal{AV}^{(R)}(\delta, \rho)$. The asymptotic mean-squared error $\mathcal{AMSE}(\delta, \gamma, \rho)$ of $\hat{\rho}_n^{(R)}$ can be defined as

$$(5.1) \qquad \frac{1}{kA^2(n/k)} \left(\left(\frac{\lambda_A}{2\gamma} \mathcal{AB}_1^{(R)}(\delta, \rho) - \lambda_B \mathcal{AB}_2^{(R)}(\delta, \rho, \rho) \right)^2 + \gamma^2 \mathcal{AV}^{(R)}(\delta, \rho) \right).$$

One way to choose the parameter δ could be to minimize the above asymptotic mean-squared error. In practice, the parameters γ, ρ as well as the functions A and B are unknown and thus the asymptotic mean-squared error cannot be evaluated. To overcome this problem, it is possible to introduce an upper bound on $\mathcal{AMSE}(\delta, \gamma, \rho)$. Assuming that $\delta \in [0, \delta_0) \cup (\delta_1, \infty)$ and $\rho \in [\rho_{\min}, \rho_{\max}]$, it is easy to check that $|\mathcal{AB}_1^{(R)}(\delta,\rho)| \geq |\mathcal{AB}_1^{(R)}(\delta_1,\rho_{\max})|$, $|\mathcal{AB}_2^{(R)}(\delta,\rho,\rho)| \geq |\mathcal{AB}_2^{(R)}(\delta_0,\rho_{\min},\rho_{\min})|$. Besides, numerically, one can observe that $\mathcal{AV}^{(R)}(\delta,\rho) \geq \mathcal{AV}^{(R)}(1.32,-0.46)$. We thus have:

$$\mathcal{AMSE}(\delta,\gamma,\rho) \leq \frac{C\pi(\delta,\rho)}{k\,A^2(n/k)}\ ,$$

with $\pi(\delta,\rho) = (\mathcal{AB}_1^{(R)}(\delta,\rho)\,\mathcal{AB}_2^{(R)}(\delta,\rho,\rho))^2\,\mathcal{AV}^{(R)}(\delta,\rho)$ and where the constant C does not depend on δ and ρ. We thus consider for $\rho<0$ the parameter δ minimizing the function $\pi(\delta,\rho)$. For instance, when ρ is in the neighborhood of 0, one can show that the optimal value is $\delta=\delta_0=1.5$.

5.1.2. Illustration on the Burr distribution

Three estimators are compared:

- the estimator $\hat{\rho}_{n,1}^{(R)}$ proposed in [17], and which corresponds to the case $\delta=1$, see Example 4.1,
- the new explicit estimator $\hat{\rho}_{n,2}^{(R)}$ introduced in Example 4.2 which corresponds to the case $\delta=0$,
- the new implicit estimator defined by $\hat{\rho}_{n,0}^{(R)} := \hat{\rho}_n^{(R)}$ with $\delta=\delta_0=1.5$, see equation (4.2).

First, the estimators are compared on the basis of their asymptotic mean-squared errors. Taking $\lambda_A = k^{1/2}A^2(n/k)$ and $\lambda_B = \rho\lambda_A/\gamma$, the asymptotic mean-squared errors are plotted on the left panel of Figure 1 as a function of $k \in \{1500,...,4999\}$ with $n=5000$ and for $\rho \in \{-1,-0.25\}$. It appears that $\hat{\rho}_{n,0}^{(R)}$ yields the best results for $\rho=-1$. This is in accordance with the results from the previous paragraph: $\delta=1.5$ is the "optimal" when ρ is close to 0. As a preliminary conclusion, the criterion $\pi(\cdot)$ seems to be well-adapted for tuning the estimator parameters. At the opposite, when $\rho=-0.25$, the best estimator from the asymptotic mean-squared error point of view is $\hat{\rho}_{n,2}^{(R)}$.

Second, the estimators are compared on their finite sample size performances. For each estimator, and for each value of $k \in \{1500,...,4999\}$, the empirical mean-squared error is computed on 500 replications of the sample of size $n=5000$. The results are displayed on the right panel of Figure 1. The conclusions are qualitatively the same: $\hat{\rho}_{n,0}^{(R)}$ yields the best results in the case $\rho \geq -1$ where as $\hat{\rho}_{n,2}^{(R)}$ yields the best results in the case $\rho < -1$. Let us note that, consequently, $\hat{\rho}_{n,1}^{(R)}$ is never the best estimator in the situation considered here. In practice, the case $\rho \geq -1$ is the more interesting one, since it corresponds to a strong bias. For this reason, it seems to us that $\hat{\rho}_{n,0}^{(R)}$ should be preferred.

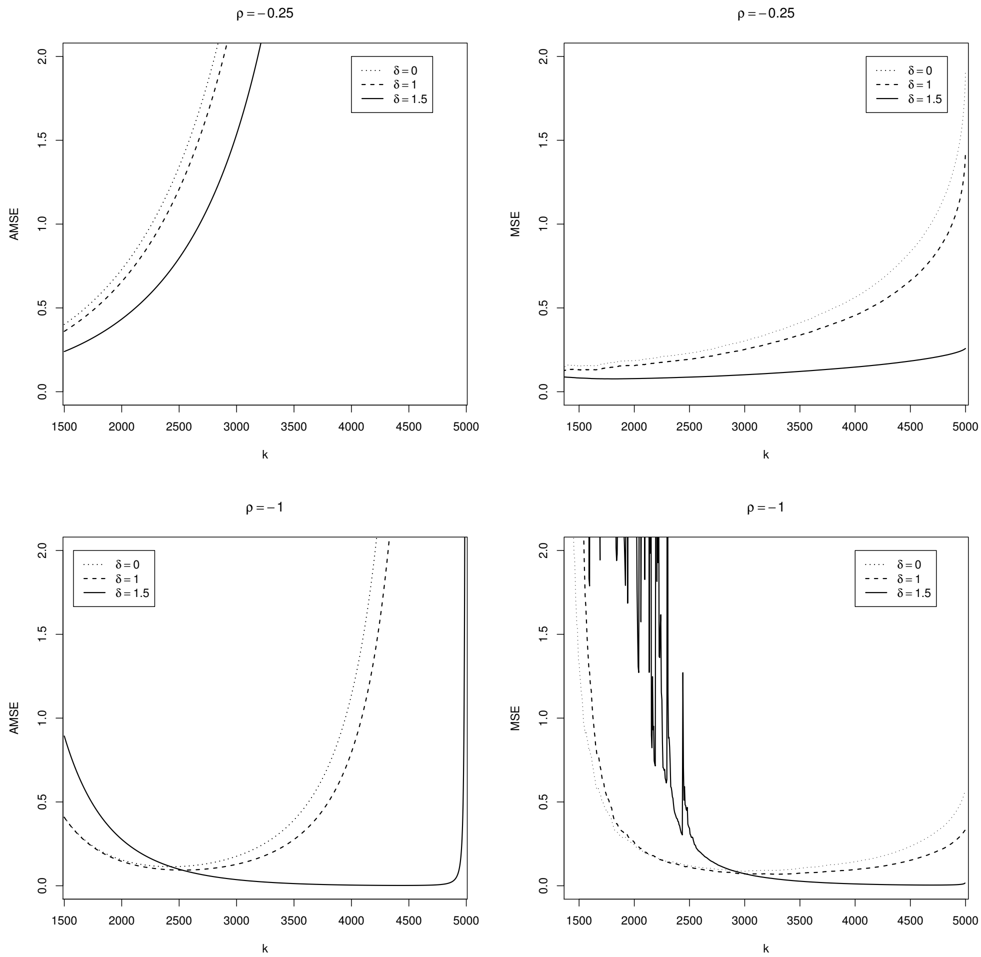

Figure 1: Asymptotic mean-squared errors (left) and empirical mean-squared errors (right) of $\hat{\rho}^{(R)}_{n,0}$, $\hat{\rho}^{(R)}_{n,1}$ and $\hat{\rho}^{(R)}_{n,2}$ as a function of k for a Burr distribution.

5.2. Estimators based on the random variable $S_k(\tau,\alpha)$

Let us now consider the estimators of ρ based on the random variables $S_k(\tau_i,\alpha_i)$ for $i=1,...,8$ considered in Section 4.2 in the case where $(\tau_1,\alpha_1)=(\tau_7,\alpha_7)$, $(\tau_2,\alpha_2)=(\tau_8,\alpha_8)$, $(\tau_3,\alpha_3)=(\tau_5,\alpha_5)$ and $(\tau_4,\alpha_4)=(\tau_6,\alpha_6)$. In Paragraph 5.2.1, we show that the asymptotic mean-squared error is independent of δ. In contrast, Paragraph 5.2.2 illustrates the finite sample behavior of the estimators when δ varies.

5.2.1. Comparison in terms of asymptotic mean-squared error

From Corollary 4.2, the asymptotic bias and variance components of $\hat{\rho}_n^{(S)}$ are respectively proportional to

$$\mathcal{AB}_1^{(S)}(\delta,\rho) = \frac{g^{(S)}(\rho)}{\left(g^{(S)}\right)'(\rho)}\left(\log m_A^{(S,1)} - \log m_A^{(S,2)}\right),$$

$$\mathcal{AB}_2^{(S)}(\delta,\rho) = \frac{g^{(S)}(\rho)}{\left(g^{(S)}\right)'(\rho)}\left(\log m_B^{(S,1)} - \log m_B^{(S,2)}\right),$$

$$\mathcal{AV}^{(S)}(\delta,\rho) = \left(\frac{g^{(S)}(\rho)}{\left(g^{(S)}\right)'(\rho)}\right)^2 \vartheta\left(v^{(S,1)} - v^{(S,2)},\, v^{(S,1)} - v^{(S,2)}\right),$$

where

$$g^{(S)}(\rho) = \frac{\zeta_1}{\zeta_2}\,\frac{\mu(G_{\tau_1,\alpha_1}J_{1-\alpha_1}K_{-\rho}) - \mu(G_{\tau_2,\alpha_2}J_{1-\alpha_2}K_{-\rho})}{\mu(G_{\tau_3,\alpha_3}J_{1-\alpha_3}K_{-\rho}) - \mu(G_{\tau_4,\alpha_4}J_{1-\alpha_4}K_{-\rho})}\,.$$

It thus appears that the asymptotic mean-squared error (defined similarly to (5.1)) does not depend on δ. From the asymptotic point of view, all the estimators $\hat{\rho}_n^{(S)}$ such that $(\tau_1,\alpha_1) = (\tau_7,\alpha_7)$, $(\tau_2,\alpha_2) = (\tau_8,\alpha_8)$, $(\tau_3,\alpha_3) = (\tau_5,\alpha_5)$ and $(\tau_4,\alpha_4) = (\tau_6,\alpha_6)$ are thus equivalent.

5.2.2. Comparison on the simulated Burr distribution

For the sake of simplicity, we fix $\alpha_1 = \alpha_7 = \theta_5 = \theta_7 = \tau_1 = ... = \tau_8 = 1$, $\alpha_2 = \alpha_3 = \alpha_5 = \alpha_8 = 2$, $\alpha_4 = \alpha_6 = 3$, $\theta_3 = \theta_8 = 1/2$, $\theta_4 = 1/3$, $\theta_6 = 2/3$, $\theta_1 = \delta + 1$ and $\theta_2 = (\delta+1)/2$ so that δ is the unique free parameter. The resulting estimator is $\hat{\rho}_{n,9}^{(S)}$, it coincides with the one proposed in [13] when $\delta = 0$. For each value of $k \in \{500, ..., 4999\}$, the empirical mean-squared error associated to $\hat{\rho}_{n,9}^{(S)}$ is computed on 500 replications of the sample of size $n = 5000$ for $\delta \in \{0,1,2\}$ and for $\rho \in \{-0.25, -1\}$. The results are displayed on Figure 2. It appears that $\delta = 0$ yields the best results for both values of ρ: the empirical mean-squared error is smaller than these associated to $\delta = 1$ or $\delta = 2$. This hierarchy cannot be observed on the asymptotic mean-squared error.

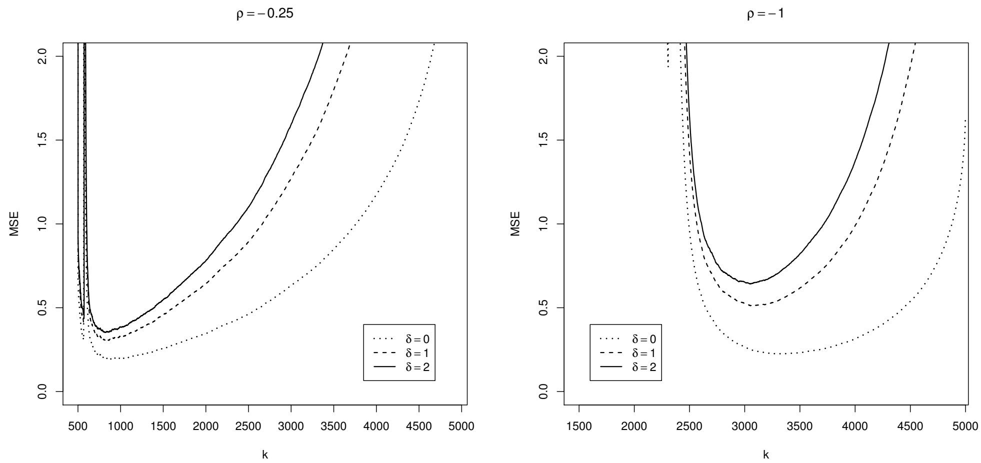

Figure 2: Empirical mean-squared errors of $\hat{\rho}_{n,9}^{(S)}$ as a function of k for a Burr distribution.

5.3. Tentative conclusion

The families of estimators of the second order parameter usually depend on a large set, say Θ, of parameters (12 parameters for estimators based on the random variables $R_k(\tau)$ and 20 parameters for $S_k(\tau,\alpha)$). The methodology proposed in Paragraph 5.1.1 permits to compute an upper bound $\pi(\cdot)$ on the asymptotic mean-squared error $\mathcal{AMSE}$ associated to the estimators. This requires to show that the quantities $\mathcal{AB}_1$, $\mathcal{AB}_2$ and $\mathcal{AV}$ are lower bounded when Θ varies in some region R_Θ. Thus, it may be possible, for some well chosen region R_Θ, to find an "optimal" set of parameters minimizing $\pi(\cdot)$. Unfortunately, the $\mathcal{AMSE}$ may not depend on all the parameters in Θ (see Paragraph 5.2.1) whereas the finite sample performances of the estimator does (see Paragraph 5.2.2). In such a case, the definition of a criterion for selecting an optimal Θ is an open question.

6. PROOFS

Proof of Theorem 2.1: Clearly, **(Ψ1)** and **(Ψ2)** entail $Z_n = \psi(\omega_n^{-1}(T_n - \chi_n \mathbb{I}))$. Moreover, **(T1)** and **(Ψ4)** yield $Z_n \xrightarrow{\mathbb{P}} \psi(f(\rho)) = \varphi(\rho)$. For all $\varepsilon > 0$, we have

$$\begin{aligned}
\mathbb{P}\big(|\hat{\rho}_n - \rho| > \varepsilon\big) &= \mathbb{P}\big(\{|\hat{\rho}_n - \rho| > \varepsilon\} \cap \{Z_n \in J\}\big) + \mathbb{P}\big(\{|\hat{\rho}_n - \rho| > \varepsilon\} \cap \{Z_n \notin J\}\big) \\
&\le \mathbb{P}\big(\{|\hat{\rho}_n - \rho| > \varepsilon\} \cap \{Z_n \in J\}\big) + \mathbb{P}\big(\{Z_n \notin J\}\big) \\
&= \mathbb{P}\big(\{|\varphi^{-1}(Z_n) - \rho| > \varepsilon\} \cap \{Z_n \in J\}\big) + \mathbb{P}\big(\{Z_n \notin J\}\big)\,.
\end{aligned}$$

From **(Ψ3)** and **(Ψ4)**, φ^{-1} is also continuous in a neighborhood of $\varphi(\rho)$. Since $Z_n \xrightarrow{\mathbb{P}} \varphi(\rho)$, it follows that $\mathbb{P}\big(\{|\varphi^{-1}(Z_n)-\rho|>\varepsilon\}\cap\{Z_n\in J\}\big)\to 0$ as $n\to\infty$. Besides, $\rho\in J_0$ yields $\varphi(\rho)\in J$ and thus

$$\mathbb{P}\big(\{Z_n\notin J\}\big)\to 0 \qquad \text{as} \quad n\to\infty\ . \tag{6.1}$$

As a conclusion, $\mathbb{P}\big(|\hat\rho_n-\rho|>\varepsilon\big)\to 0$ as $n\to\infty$ and the result is proved. □

Proof of Theorem 2.2: Recalling that $Z_n=\psi(\omega_n^{-1}(T_n-\chi_n\mathbb{I}))$, a first order Taylor expansion shows that there exists $\varepsilon\in(0,1)$ such that

$$v_n\big(Z_n-\varphi(\rho)\big) = {}^t(v_n\xi_n)\ \nabla\psi\big(f(\rho)+\varepsilon\xi_n\big)\ ,$$

where we have defined $\xi_n=\omega_n^{-1}(T_n-\chi_n\mathbb{I})-f(\rho)$. Therefore, $\xi_n\xrightarrow{\mathbb{P}}0$ and **(Ψ5)** entail that $\nabla\psi(f(\rho)+\varepsilon\xi_n)\xrightarrow{\mathbb{P}}\nabla\psi(f(\rho))$. Thus, taking account of **(T2)**, we obtain that

$$v_n\big(Z_n-\varphi(\rho)\big)\xrightarrow{d}\mathcal{N}\big(m_\psi(\rho),\gamma^2\sigma_\psi^2(\rho)\big)\ . \tag{6.2}$$

Now, $P_n(x):=\mathbb{P}\big(\{v_n(\hat\rho_n-\rho)\le x\}\big)$ can be rewritten as

$$\begin{aligned}
P_n(x) &= \mathbb{P}\big(\{v_n(\hat\rho_n-\rho)\le x\}\cap\{Z_n\in J\}\big)+\mathbb{P}\big(\{v_n(\hat\rho_n-\rho)\le x\}\cap\{Z_n\notin J\}\big)\\
&= \mathbb{P}\big(\{v_n(\varphi^{-1}(Z_n)-\rho)\le x\}\cap\{Z_n\in J\}\big)+\mathbb{P}\big(\{v_n(\hat\rho_n-\rho)\le x\}\cap\{Z_n\notin J\}\big)\\
&=: P_{1,n}(x)+P_{2,n}(x)\ .
\end{aligned}$$

Let us first note that

$$0\le P_{2,n}(x)\le\mathbb{P}\big(\{Z_n\notin J\}\big)\to 0 \qquad \text{as} \quad n\to\infty\ , \tag{6.3}$$

in view of (6.1) in the proof of Theorem 2.1. Focusing on $P_{1,n}(x)$, since φ is continuously differentiable in a neighborhood of ρ and $\varphi'(\rho)\neq 0$, it follows that φ is monotone in a neighborhood of ρ. Let us consider the case where φ is decreasing, the case φ increasing being similar. Writing $J=(a,b)$, it follows that

$$\begin{aligned}
P_{1,n}(x) &= \mathbb{P}\Big(\big\{a\vee\varphi(\rho+x/v_n)\le Z_n\le b\big\}\Big)\\
&= \mathbb{P}\Big(\Big\{v_n\big(a\vee\varphi(\rho+x/v_n)-\varphi(\rho)\big)<v_n\big(Z_n-\varphi(\rho)\big)\le v_n\big(b-\varphi(\rho)\big)\Big\}\Big)\ .
\end{aligned}$$

Introducing G_n the cumulative distribution function of $v_n(Z_n-\varphi(\rho))$, we have

$$\begin{aligned}
1-P_{1,n}(x) &= 1-G_n\Big(v_n\big(b-\varphi(\rho)\big)\Big)+G_n\Big(v_n\big(a\vee\varphi(\rho+x/v_n)-\varphi(\rho)\big)\Big)\\
&= 1-G_n\Big(v_n\big(b-\varphi(\rho)\big)\Big)\\
&\quad + G_n\Big(v_n\big(a-\varphi(\rho)\big)\Big)\vee G_n\Big(v_n\big(\varphi(\rho+x/v_n)-\varphi(\rho)\big)\Big)\\
&=: P_{1,1,n}+P_{1,2,n}\vee P_{1,3,n}(x)\ .
\end{aligned}$$

Let G denote the cumulative distribution function of the $\mathcal{N}\big(m_\psi(\rho), \gamma^2\sigma_\psi^2(\rho)\big)$ distribution. It is straightforward that

$$P_{1,1,n} \le 1 - G\Big(v_n\big(b - \varphi(\rho)\big)\Big) + \sup_{t\in\mathbb{R}}\big|G_n(t) - G(t)\big| \ .$$

Since $\rho \in J_0$, we have $\varphi(\rho) \in J = (a, b)$. In particular, $b > \varphi(\rho)$ yields $1 - G\big(v_n(b - \varphi(\rho))\big) \to 0$ as $n \to \infty$. Besides, (6.2) shows that $G_n(t) \to G(t)$ for all $t \in \mathbb{R}$ and thus $G_n(t) \to G(t)$ uniformly, see for instance [11], p. 552. As a preliminary conclusion $P_{1,1,n} \to 0$ and, similarly, $P_{1,2,n} \to 0$ as $n \to \infty$. Finally,

$$\Big|P_{1,3,n}(x) - G\big(x\varphi'(\rho)\big)\Big| \le \Big|G\Big(v_n\big(\varphi(\rho + x/v_n) - \varphi(\rho)\big)\Big) - G\big(x\varphi'(\rho)\big)\Big| + \sup_{t\in\mathbb{R}}\big|G_n(t) - G(t)\big|$$

and, in view of **(Ψ5)**, $v_n\big(\varphi(\rho + x/v_n) - \varphi(\rho)\big) \to x\varphi'(\rho)$ as $n \to \infty$, which leads to $P_{1,3,n}(x) \to G(x\varphi'(\rho))$ as $n \to \infty$. We thus have shown that

$$P_{1,n}(x) \to 1 - G\big(x\varphi'(\rho)\big) = G(x|\varphi'(\rho)|) \qquad \text{as} \quad n \to \infty \ . \tag{6.4}$$

Collecting (6.3) and (6.4) yields

$$\mathbb{P}\big(\big\{v_n(\hat\rho_n - \rho) \le x\big\}\big) \to G\big(x|\varphi'(\rho)|\big) \qquad \text{as} \quad n \to \infty$$

and concludes the proof. □

Proof of Corollary 4.1: Clearly, ψ_δ given in (4.1) satisfies **(Ψ1)** and **(Ψ2)**. Moreover, Lemma 3.1 shows that **(T2)** holds. To apply Theorem 2.2 it only remains to prove that **(Ψ3)** and **(Ψ5)** are satisfied. First remark that under **(C4)** and **(C5)**, $\varphi_\delta^{(R)}(\rho)$ is well defined for all $\rho \le 0$ since $f^{(R)}(\rho) \in \mathcal{D}$. Furthermore, from Lemma 3.1, we have for $i = 1, ..., 4$,

$$T_{n,2i-1}^{(R)} - T_{n,2i}^{(R)} = \frac{\tilde\theta_i\, A(Y_{n-k,n})}{\gamma}\,\big(\nu_\rho(\tau_{2i-1}) - \nu_\rho(\tau_{2i})\big)\,\big(1 + o_P(1)\big) \ ,$$

as n goes to infinity. Hence, conditions **(C4)** and **(C5)** imply that $T_n^{(R)} \in \mathcal{D}$. Finally, using Lerch's Theorem (see [5], page 345), condition **(C4)** implies that there exists $\rho_0 < 0$ such that the first derivative of $\varphi_\delta^{(R)}$ is non zero at ρ_0. Thus, the inverse function theorem insures the existence of intervals J_0 and J for which the function $\varphi_\delta^{(R)}$ is a continuously differentiable bijection from J_0 to J. In conclusion, conditions **(Ψ3)** and **(Ψ5)** are satisfied and Theorem 2.2 applies. □

Proof of Corollary 4.2: The proof follows the same lines as the one of Corollary 4.1. It consists in remarking that, under **(C6)** and **(C7)**, one has $f^{(S)}(\rho) \in \mathcal{D}$ and $T_n^{(S)} \in \mathcal{D}$ since,

$$T_{n,2i-1}^{(S)} - T_{n,2i}^{(S)} = \frac{\zeta_i\, A(n/k)}{\gamma}\,\big(\nu_\rho(\tau_{2i}, \alpha_{2i}) - \nu_\rho\big(\tau_{2i-1}, \alpha_{2i-1})\big)\,\big(1 + o_P(1)\big) \ ,$$

in view of Lemma 3.2. □

ACKNOWLEDGMENTS

The first author acknowledges support from AIRES-Sud (AIRES-Sud is a program from the French Ministry of Foreign and European Affairs, implemented by the "Institut de Recherche pour le Développement", IRD-DSF) and from the "Ministère de la Recherche Scientifique" of Sénégal.

REFERENCES

[1] Beirlant, J.; Dierckx, G.; Goegebeur, Y. and Matthys, G. (1999). Tail index estimation and an exponential regression model, *Extremes*, **2**, 177–200.

[2] Beirlant, J.; Goegebeur, Y.; Verlaack, R. and Vynckier, P. (1998). Burr regression and portfolio segmentation, *Insurance: Mathematics and Economics*, **23**, 231–250.

[3] Cai, J.; de Haan, L. and Zhou, C. (2012). Bias correction in extreme value statistics with index around zero, *Extremes*, to appear, DOI: 10.1007/s10687-012-0158-x.

[4] Cairo, F.; Gomes, M.I. and Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator, *REVSTAT – Statistical Journal*, **3**(2), 113–136.

[5] Carslaw, H.S. and Jaeger, J.C. (1948). *Operational Methods in Applied Mathematics*, Oxford University Press.

[6] Ciuperca, G. and Mercadier, C. (2010). Semi-parametric estimation for heavy tailed distributions, *Extremes*, **13**, 55–87.

[7] Csörgö, S.; Deheuvels, P. and Mason, D.M. (1985). Kernel estimates of the tail index of a distribution, *The Annals of Statistics*, **13**, 1050–1077.

[8] Dekkers, A.L.M.; Einmahl, J.H.J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution, *The Annals of Statistics*, **17**, 1833–1855.

[9] Diebolt, J.; Gardes, L.; Girard, S. and Guillou, A. (2008). Bias-reduced estimators of the Weibull tail-coefficient, *Test*, **17**(2), 311–331.

[10] Diebolt, J.; Gardes, L.; Girard, S. and Guillou, A. (2008). Bias-reduced extreme quantile estimators of Weibull tail-distributions, *Journal of Statistical Planning and Inference*, **138**(5), 1389–1401.

[11] Embrechts, P.; Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events*, Springer.

[12] Feuerverger, A. and Hall, P. (1999). Estimating a tail exponent by modeling departure from a Pareto distribution, *The Annals of Statistics*, **27**, 760–781.

[13] 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**(2), 193–213.

[14] Fraga Alves, M.I.; de Haan, L. and Lin, T. (2003). Estimation of the parameter controlling the speed of convergence in extreme value theory, *Mathematical Methods of Statistics*, **12**(2), 155–176.

[15] Gardes, L. and Girard, S. (2010). Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels, *Extremes*, **13**(2), 177–204.

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

[17] Goegebeur, Y.; Beirlant, J. and de Wet, T. (2010). Kernel estimators for the second order parameter in extreme value statistics, *Journal of Statistical Planning and Inference*, **140**, 2632–2652.

[18] Goegebeur, Y. and de Wet, T. (2012). Estimation of the third-order parameter in extreme value statistics, *Test*, **21**(2), 330–354.

[19] Gomes, M.I.; de Haan, L. and Peng, L. (2002). Semi-parametric estimation of the second order parameter in statistics of extreme, *Extremes*, **5**, 387–414.

[20] Gomes, M.I. and Martins, M.J. (2001). Generalizations of the Hill estimator – asymptotic versus finite sample behaviour, *Journal of Statistical Planning and Inference*, **93**, 161–180.

[21] 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.

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

[23] Gomes, M.I.; Martins, M.J. and Neves, M. (2007). Improving second order reduced bias extreme value index estimator, *REVSTAT – Statistical Journal*, **5**(2), 177–207.

[24] 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.

[25] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer Series in Operations Research and Financial Engineering, Springer.

[26] Hall, P. and Welsh, A.H. (1985). Adaptative estimates of parameters of regular variation, *The Annals of Statistics*, **13**, 331–341.

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

[28] Peng, L. (1998). Asymptotic unbiased estimators for the extreme value index, *Statistics and Probability Letters*, **38**, 107–115.

[29] Segers, J. (2001). Residual estimators, *Journal of Statistical Planning and Inference*, **98**, 15–27.

[30] de Wet, T.; Goegebeur, Y. and Munch, M.R. (2012). Asymptotically unbiased estimation of the second order tail parameter, *Statistics and Probability Letters*, **82**, 565–573.

[31] Worms, J. and Worms, R. (2012). Estimation of second order parameters using probability weighted moments, *ESAIM: Probability and Statistics*, **16**, 97–113.

REVSTAT – Statistical Journal
Volume 11, Number 3, November 2013, 301–315

TWO NONPARAMETRIC ESTIMATORS OF THE MEAN RESIDUAL LIFE

Authors: Abdel-Razzaq Mugdadi
– Department of Mathematics and Statistics,
Jordan University of Science and Technology,
Irbid, Jordan
aamugdadi@just.edu.jo

Amanuel Teweldemedhin
– Department of Mathematics,
University of Wisconsin – Richland,
Richland Center, WI 53581, U.S.A.
amanuel.teweldemedhi@uwc.edu

Received: March 2012 Revised: January 2013 Accepted: February 2013

Abstract:

- The mean residual life function $L(t)$ can be written based on the vitality function $V(t)$. In this article we propose two methods to estimate $V(t)$. The two methods are based on both the kernel density estimation and the empirical function. In addition, we evaluate the mean square error of the two estimators and we study the consistency for both of them.

Key-Words:

- *mean residual life; kernel; empirical; estimation; mean square error; bandwidth; consistent.*

AMS Subject Classification:

- 62G05, 62N02.

1. INTRODUCTION

The mean residual life MRL is the expected remaining life, $T-t$, given that the item has survived to time t. The unconditional mean of the distribution, $E(T)$, is a special case given by $L(0)$. To determine a formula for this expectation, the conditional probability density function is needed

$$f_{T|T\geq t}(\tau) = \frac{f(\tau)}{P\big[T\geq t\big]} = \frac{f(\tau)}{R(t)}\,, \qquad \tau\geq t\,. \tag{1.1}$$

This conditional probability density function is actually a family of probability density functions (one for each value of t), each of which has an associated mean

$$E\big[T\,|\,T\geq t\big] = \int_t^\infty \tau f_{T|T\geq t}(\tau)\,d\tau = \int_t^\infty \tau\frac{f(\tau)}{R(t)}\,d\tau\,.$$

Thus, in life testing situations, the expected additional lifetime given that a component has survived until time t is called the MRL. Since the MRL function is the expected *remaining* life, t must be subtracted, yielding

$$L(t) = E\big[T-t\,|\,T\geq t\big] = \frac{1}{R(t)}\int_t^\infty \tau f(\tau)\,d\tau \,-\, t\,. \tag{1.2}$$

Thus $L(t)$ can be written as

$$L(t) = V(t)-t\,,$$

where

$$V(t) = \frac{\int_t^\infty s f(s)\,ds}{R(t)} = \frac{M(t)}{R(t)}\,. \tag{1.3}$$

We study the vitality function estimator when $R(t)>0$, since the vitality function estimator generates the mean residual life function estimator directly by the above equation.

Ratio functions for which nonparametric estimators have been considered include the MRL function and hazard rate among others. One estimation method involves individual estimates of the numerator and denominator. An alternative estimator is to estimate the entire function not the separate pieces. For a discussion of ratio functions estimates see Patil *et al.* [13]. In many reliability studies, the MRL function (corresponding to a lifetime distribution with density $f(t)$, and survival function $R(t)$), is of prime importance. A problem of considerable interest, therefore, is the estimation of mean residual life function. The kernel density estimation is the most popular technique to estimate the probability density function, which is basically can be define as follow: Let $X_1, ..., X_n$ be a random sample from unknown continuous probability density function $f(x)$. The kernel density estimate with appropriate kernel function $k(t)$ and smoothing parameter h is

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right) . \tag{1.4}$$

Kernel type estimators of ratio functions, such as the density under random censoring and the hazard rate have been studied by several authors (e.g. Watson and Leadbetter [18], [19], Marron and Padgett [10], Lo *et al.* [9], Sarda and Vieu [16], and Hollander and Proschan [8]).

The basic estimator for $L(t)$ is $\hat{L}(t) = \hat{V}_0(t) - t$, where $\hat{V}_0(t) = \frac{M_e(t)}{R_e(t)}$. $R_e(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{(X_i>t)}$ and $M_e(t) = \frac{1}{n}\sum_{i=1}^{n} X_i 1_{(X_i>t)}$, but Abdous and Berred[1] discussed that $\hat{V}_0(t)$ does not take into account the smoothness of $V(t)$. Guillamon *et al.* [6] studied the estimator $\hat{V}_3(t) = \frac{M_n(t)}{R_n(t)}$ for $V(t)$, where $M_n(t) = \int_t^\infty s f_n(s)\, ds$, $f_n(t)$ is the kernel density estimation defined in (1.4) and $V_n(t)$ is the kernel reliability estimator (see Section 3). Other estimators or cases proposed by Mitra and Basu [11], Ruiz and Guillamon [14], Chaubey and Sen [3], and Abdous and Berred [1]. In this paper we propose and study two new estimators for the MRL both based on the kernel estimator and the empirical function. Also, we propose new techniques to select the bandwidth for the estimators. From the simulations, we can conclude that the new estimator is competitive with the basic one but we can't say it is a better one.

2. THE FIRST ESTIMATOR $\widehat{V}_1(t)$

In the first propose estimator we use kernel estimate for the numerator function and empirical estimator of the survival function in the denominator. Thus,

$$\widehat{V}_1(t) = \frac{\int_t^\infty s f_n(s)\, ds}{R_e(t)} = \frac{M_n(t)}{R_e(t)} , \tag{2.1}$$

where

$$f_n(t) = \frac{1}{n} \sum_{i=1}^{n} k_h(t - X_i) , \tag{2.2}$$

and

$$M_n(t) = \int_t^\infty s f_n(s)\, ds \tag{2.3}$$

is the kernel estimate of the numerator, and

$$R_e(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{(X_i>t)} \;, \tag{2.4}$$

is the frequency count of a set divided by n. Thus, we can estimate the MRL $L(t)$ by:

$$\widehat{L}_1(t) = \widehat{V}_1(t) - t \;. \tag{2.5}$$

2.1. Properties of $\widehat{V}_1(t)$

In this section, we evaluate the Bias, the variance and the Mean Square Error (MSE) of $\widehat{V}_1(t)$. In addition, we derive the optimal bandwidth that minimizes the Asymptotic Mean Square Error (AMSE) and we study the consistency of $\widehat{V}_1(t)$.

Proposition 2.1. *For any t with $R(t) > 0$,*

$$\widehat{V}_1(t)-V(t) = \frac{1}{nR(t)}\left(\sum_{i=1}^{n}\left(\int_t^{\infty} s\,k_h(s-X_i)\,ds - V(t)\,1_{(X_i>t)}\right)\right)\Big(1+o(1)\Big) \,. \tag{2.6}$$

Proof:

$$\begin{aligned}
\widehat{V}_1(t) - V(t) &= \frac{M_n(t)}{R_e(t)} - V(t)\\
&= \left(\frac{M_n(t) - V(t)\,R_e(t)}{R(t)}\right)\left(1 + \frac{R(t) - R_e(t)}{R_e(t)}\right)\\
&= \frac{1}{nR(t)}\left(\sum_{i=1}^{n}\left(\int_t^{\infty} sk_h(s - X_i)\,ds \;-\; V(t)\,1_{(X_i>t)}\right)\right)\Big(1 + o(1)\Big) \,. \;\Box
\end{aligned}$$

Lemma 2.1. *Let $\widehat{V}_1(t)$ be as (2.1), then*

$$\begin{aligned}
E\Big(\widehat{V}_1(t)\Big) &= V(t) + \frac{h^2}{2R(t)}\,\mu_2(k)\int_t^{\infty} sf''(s)\,ds + o(h^2) \;,\\
Var\Big(\widehat{V}_1(t)\Big) &= \frac{1}{nR^2(t)}\Big(\Gamma_2(t) + h\,\gamma_2(t)\,\alpha(k)\Big) + o\big(h^2\big) + o\Big(\frac{h}{n}\Big)\,,\\
MSE\Big(\widehat{V}_1(t)\Big) &= \frac{1}{n}\frac{\Gamma_2(t)}{R^2(t)} + h\frac{\gamma_2(t)\,\alpha(k)}{nR^2(t)} + h^4\,\frac{\mu_2^2(k)}{4R^2(t)}\left(\int_t^{\infty} sf''(s)\,ds\right)^2 + o\Big(h^4 + \frac{h}{n} + h\Big),
\end{aligned}$$

where $\Gamma_i(t) = \int_t^{\infty} \gamma_i(s)\,ds$, $\gamma_i(t) = t^i f(t)$, $\mu_2(k) = \int_{-\infty}^{\infty} s^2 k(s)\,ds$, *and* $\alpha(k) = \int_{-\infty}^{\infty} 2\,s\,W(s)\,k(s)\,ds < \infty$.

Proof: Using (2.6)

$$E\Big(\widehat{V}_1(t)-V(t)\Big) = \frac{1}{R(t)}\left(E\bigg(\int_t^\infty s k_h(s-X)\,ds\bigg) - V(t)\,E\Big(1_{(X>t)}\Big)\right)$$

$$= \frac{1}{R(t)}\left(\Gamma_1(t)+\frac{h^2}{2}\mu_2(k)\int_t^\infty s f''(s)\,ds - V(t)\int_t^\infty f(s)\,ds\right) + o(h^2)$$

$$= \frac{h^2}{2R(t)}\,\mu_2(k)\int_t^\infty s f''(s)\,ds + o(h^2)\ .$$

Thus,

$$E\Big(\widehat{V}_1(t)\Big) = V(t) + \frac{h^2}{2R(t)}\,\mu_2(k)\int_t^\infty s f''(s)\,ds + o(h^2)\ .$$

Also, from (2.6), and after some reduction,

$$Var\Big(\widehat{V}_1(t)-V(t)\Big) =$$

$$= \frac{1}{nR^2(t)}\left(E\bigg(\int_t^\infty s k_h(s-X)\,ds - V(t)\,1_{(X>t)}\bigg)^2 - \left[E\bigg(\int_t^\infty s k_h(s-X)\,ds - V(t)\,1_{(X>t)}\bigg)\right]^2\right)$$

$$= \frac{1}{nR^2(t)}\left(E\bigg(\int_t^\infty s k_h(s-X)\,ds\bigg)^2 - 2V(t)\,E\bigg(1_{X>t}\int_t^\infty s k_h(s-X)\,ds\bigg) + V^2(t)\,E\Big(1_{X>t}\Big)\right) + o\Big(\frac{h}{n}\Big)$$

$$= \frac{1}{nR^2(t)}\Big(\Gamma_2(t) + h\gamma_2(t)\,\alpha(k) - 2V(t)\,A + V(t)\,\Gamma_1(t)\Big) + o\Big(\frac{h}{n}\Big)\,.$$

But

$$A = E\bigg(1_{(X>t)}\int_t^\infty s k_h(s-X)\,ds\bigg)$$

$$= \int_t^\infty\int_{\frac{t-y}{h}}^\infty (y+hx)\,k(x)\,dx\ f(y)\,dy$$

$$= \frac{1}{2}\,\Gamma_1(t) + h\,\Gamma_1'(t)\int_0^\infty s k(s)\,ds + o\big(h^2\big)\,.$$

So that

$$
\begin{aligned}
Var\Big(\widehat{V}_1(t) - V(t)\Big) &= \\
&= \frac{1}{nR^2(t)}\Bigg(\Gamma_2(t) + h\gamma_2(t)\,\alpha(k) - 2V(t)\bigg[\frac{1}{2}\,\Gamma_1(t) + h\Gamma_1'(t)\int_0^\infty sk(s)\,ds\bigg] \\
&\quad + V(t)\,\Gamma_1(t)\Bigg) + o\Big(h^2 + \frac{h}{n}\Big) \\
&= \frac{1}{nR^2(t)}\Bigg(\Gamma_2(t) + h\gamma_2(t)\,\alpha(k) - \Gamma_1(t)V(t) - 2hV(t)\,\Gamma_1'(t)\int_0^\infty sk(s)\,ds\;\Gamma_1(t)\Bigg) \\
&\quad + o\Big(h^2 + \frac{h}{n}\Big) \\
&= \frac{1}{nR^2(t)}\Bigg(\Gamma_2(t) + h\gamma_2(t)\,\alpha(k) - 2hV(t)\,\Gamma_1'(t)\int_0^\infty sk(s)\,ds\Bigg) + o\Big(h^2 + \frac{h}{n}\Big).
\end{aligned}
$$

Thus,

$$
Var\Big(\widehat{V}_1(t)\Big) = \frac{1}{nR^2(t)}\Big(\Gamma_2(t) + h\gamma_2(t)\,\alpha(k)\Big) + o\big(h^2\big) + o\Big(\frac{h}{n}\Big).
$$

Therefore,

$$
\begin{aligned}
MSE\Big(\widehat{V}_1(t)\Big) &= Bias^2\Big(\widehat{V}_1(t)\Big) + Var\Big(\widehat{V}_1(t)\Big) \\
&= \frac{1}{n}\frac{\Gamma_2(t)}{R^2(t)} + h\frac{\gamma_2(t)\,\alpha(k)}{nR^2(t)} + h^4\frac{\mu_2^2(k)}{4R^2(t)}\Bigg(\int_t^\infty sf''(s)\,ds\Bigg)^2 + o\Big(h^2 + h^4 + \frac{h}{n}\Big).
\end{aligned}
$$

□

The following corollaries can be obtained directly from the above Lemma.

Corollary 2.1. *The asymptotic mean integrated square error (AMISE) of* $\hat{V}_1(t)$ *is*

$$
\begin{aligned}
AMISE\Big(\widehat{V}_1(t)\Big) &= \frac{1}{n}\int \frac{\Gamma_2(t)}{R^2(t)}\,dt + \frac{h}{n}\,\alpha(k)\int \frac{\gamma_2(t)}{R^2(t)}\,dt \\
&\quad + \frac{h^4}{4}\,\mu_2^2(k)\int \frac{1}{R^2(t)}\Bigg(\int_t^\infty sf''(s)\,ds\Bigg)^2 dt\,.
\end{aligned} \tag{2.7}
$$

Corollary 2.2. *The optimal bandwidth that minimizes the* $AMISE(\hat{V}_1(t))$ *is*

$$
\widehat{h}_{opt1} = n^{-\frac{1}{3}}\left(\frac{-\alpha(k)}{\mu_2^2(k)}\;\frac{\int \frac{\gamma_2(t)}{R^2(t)}\,dt}{\int \frac{1}{R^2(t)}\Big(\int_t^\infty sf''(s)\,ds\Big)^2 dt}\right)^{\frac{1}{3}}.
$$

Corollary 2.3. *The estimator $\widehat{V}_1(t)$ is a asymptotically consistent estimator of the vitality function $V(t)$. That is*

$$\widehat{V}_1(t) \xrightarrow{p} V(t) \ . \tag{2.8}$$

3. THE SECOND ESTIMATOR $\widehat{V}_2(t)$

In this section we use empirical estimate for the numerator and kernel estimate of the survival function in the denominator. Thus,

$$\widehat{V}_2(t) = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i 1_{(X_i>t)}}{R_n(t)} = \frac{M_e(t)}{R_n(t)} \ , \tag{3.1}$$

where $R_n(t)$ is the kernel reliability estimator

$$R_n(t) = \frac{1}{n}\sum_{i=1}^{n} W\left(\frac{t-X_i}{h}\right) \tag{3.2}$$

(see Nadaraya [12], Azzalini [2] and Swanepoel [17]), $k(x)$ is a class-2 symmetric kernel, $k_h(x) = \frac{1}{h}\, k(\frac{x}{h})$, $W(t) = \int_t^\infty k(s)\,ds$, and h is a bandwidth (or smoothing parameter) verifying $h \to 0$ and $nh \to \infty$ when $n \to \infty$; and

$$M_e(t) = \frac{1}{n}\sum_{i=1}^{n} X_i 1_{(X_i>t)} \ , \tag{3.3}$$

where

$$1_T = \begin{cases} 1 & \text{if T is true}\,, \\ 0 & \text{otherwise}\,, \end{cases}$$

is the empirical estimate of the numerator in the definition of V.

In this case, the MRL estimator is

$$\widehat{L}_2(t) = \widehat{V}_2(t) - t \ . \tag{3.4}$$

3.1. Properties of $\widehat{V}_2(t)$

In this section we evaluate the MSE and the AMISE of $\widehat{V}_2(t)$. Also, we derive the optimal bandwidth and study the consistency of $\widehat{V}_2(t)$.

Proposition 3.1. *For any t with $R(t) > 0$,*

$$\widehat{V}_2(t) - V(t) = \frac{1}{nR(t)} \left[\sum_{i=1}^{n} \left(X_i 1_{(X_i > t)} - V(t)\, W\!\left(\frac{t - X_i}{h}\right) \right) \right] \Big(1 + o(1)\Big) \tag{3.5}$$

where $W(t) = \int_t^\infty k(s)\, ds$.

Proof:

$$\begin{aligned} \widehat{V}_2(t) - V(t) &= \frac{M_e(t)}{R_n(t)} - V(t) \\ &= \left(\frac{M_e(t) - V(t)\, R_n(t)}{R(t)} \right) \left(1 + \frac{R(t) - R_n(t)}{R_n(t)} \right) \\ &= \frac{1}{nR(t)} \left[\sum_{i=1}^{n} \left(X_i 1_{(X_i > t)} - V(t)\, W\!\left(\frac{t - X_i}{h}\right) \right) \right] \Big(1 + o(1)\Big) . \quad \square \end{aligned}$$

Lemma 3.1. *Let $\widehat{V}_2(t)$, $\Gamma_i(t)$, $\gamma_i(t)$, $\mu_2(k)$, and $\alpha(k)$ be as defined earlier, then*

$$\begin{aligned} E\Big(\widehat{V}_2(t)\Big) &= V(t) + \frac{1}{2}\, h^2 \frac{V(t)}{R(t)}\, f'(t)\, \mu_2(k) + o(h^2) \ , \\ Var\Big(\widehat{V}_2(t)\Big) &= \frac{1}{nR^2(t)}\, \Gamma_2(t) + \frac{h}{nR^2(t)} V^2(t)\, f(t)\, \alpha(k) + o(h) + o\Big(\frac{h}{n}\Big) \ , \\ MSE\Big(\widehat{V}_2(t)\Big) &= \frac{1}{4}\, h^4 \frac{V^2(t)}{R^2(t)} \Big(f'(t)\Big)^2 \mu_2^2(k) + \frac{1}{nR^2(t)}\, \Gamma_2(t) + \frac{h}{nR^2(t)}\, V^2(t)\, f(t)\, \alpha(k) \\ &\quad + o\Big(h + \frac{h}{n}\Big) . \end{aligned}$$

Proof: Using the result in Proposition (3.1)

$$\begin{aligned} E\Big(\widehat{V}_2(t) - V(t)\Big) &= \frac{1}{R(t)} \left(E\big(y I_{(y>t)}\big) - V(t)\, E\!\left[W\!\left(\frac{t - X_i}{h}\right) \right] \right) \\ &= \frac{h^2 V(t)}{2R(t)}\, f'(t)\, \mu_2(k) + o(h^2) \ . \end{aligned}$$

Thus,

$$E\Big(\widehat{V}_2(t)\Big) = V(t) + \frac{1}{2}\, h^2 \frac{V(t)}{R(t)}\, f'(t)\, \mu_2(k) + o(h^2) \ . \tag{3.6}$$

Now, we want to evaluate the variance.

$$
\begin{aligned}
Var\Big(\widehat{V}_2(t)-V(t)\Big) &= \frac{1}{nR^2(t)}\left(Var\big(yI_{(y>t)}\big)-V(t)\,W\left(\frac{t-y}{h}\right)\right)\\
&= \frac{1}{nR^2(t)}\left(E\big(yI_{(y>t)}\big)^2-2V(t)\,E\left(yI_{(y>t)}-W\left(\frac{t-y}{h}\right)\right)\right.\\
&\quad \left.+\,V^2(t)\,E\left(W^2\left(\frac{t-y}{h}\right)\right)\right)+o\left(\frac{h}{n}\right)\\
&= \frac{1}{nR^2(t)}\left(\Gamma_2(t)-\frac{1}{2}h^2V(t)\,\mu_2(k)\int_t^\infty sf''(s)\,ds+hV^2(t)\,f(t)\,\alpha(k)\right)\\
&\quad +\,o\Big(h+\frac{h}{n}\Big)\,.
\end{aligned}
$$

Thus,

$$
Var\Big(\widehat{V}_2(t)\Big) = \frac{1}{nR^2(t)}\,\Gamma_2(t)+\frac{h}{nR^2(t)}\,V^2(t)\,f(t)\,\alpha(k)+o\left(h+\frac{h}{n}\right).
$$

Therefore,

$$
\begin{aligned}
MSE\Big(\widehat{V}_2(t)\Big) &= \frac{1}{4}\,h^4\frac{V^2(t)}{R^2(t)}\,\big(f'(t)\big)^2\mu_2^2(k)+\frac{1}{nR^2(t)}\,\Gamma_2(t)+\frac{h}{nR^2(t)}V^2(t)\,f(t)\,\alpha(k)\\
&\quad +\,o\left(h+\frac{h}{n}\right).
\end{aligned}
$$

□

Corollary 3.1.

(3.7)
$$
\begin{aligned}
AMISE\Big(\widehat{V}_2(t)\Big) &= \frac{1}{4}\,h^4\mu_2^2(k)\int\frac{V^2(t)\,\big(f'(t)\big)^2}{R^2(t)}\,dt\;+\;\frac{1}{n}\int\frac{\Gamma_2(t)}{R^2(t)}\,dt\\
&\quad +\;\frac{h\,\alpha(k)}{n}\int\frac{V^2(t)\,f(t)}{R^2(t)}\,dt\;.
\end{aligned}
$$

Corollary 3.2. *The optimal bandwidth that minimizes the* $AMISE(\hat{V}_2(t))$ *is*

$$
\widehat{h}_{opt2} = n^{-\frac{1}{3}}\left(\frac{-\alpha(k)}{\mu_2^2(k)}\;\frac{\int\frac{V^2(t)\,f(t)}{R^2(t)}\,dt}{\int\frac{V^2(t)\,(f'(t))^2}{R^2(t)}\,dt}\right)^{\frac{1}{3}}.
$$

Corollary 3.3. *The estimator* $\widehat{V}_2(t)$ *is a asymptotically consistent estimator of the vitality function* $V(t)$*. That is*

(3.8)
$$
\widehat{V}_2(t)\;\xrightarrow[n\to\infty]{p}\;V(t)\;.
$$

From Corollaries 2.2 and 3.2, we can conclude that the optimal bandwidths decrease at rate $O(n^{-\frac{1}{3}})$, which is the same of rate of convergence for bandwidth for the kernel distribution function estimator.

4. BANDWIDTH SELECTIONS

4.1. Likelihood Cross-Validation

The original cross-validation criterion, proposed by Habbema *et al.* [7] and Duin [4] to select the bandwidth h by minimizing the score function

$$LCV(h) = -\frac{1}{n}\sum_{i=1}^{n}\log \hat{f}_{-i}(X_i)$$

over possible values of h. $\hat{f}_{-i}(X_i)$ is the "leave-one-out" kernel density estimator defined using the data with X_i removed. That is

$$\hat{f}_{-i}(X_i) = \frac{1}{(n-1)\,h}\sum_{j\neq i}^{n} k\left(\frac{X_i - X_j}{h}\right).$$

The method of likelihood cross-validation is a natural development of the idea of using likelihood to judge the adequacy of fit of a statistical model. It is of general applicability beyond choosing h in kernel density estimation, having been used for both parameter estimation and model selection (e.g. Geisser [5]).

Analogous to this we propose this kind of technique to our estimators. That is we will minimize the following function:

$$LCV(h) = -\frac{1}{n}\sum_{i=1}^{n}\log \hat{V}_{j,-i}(X_i) \tag{4.1}$$

where $\hat{V}_{j,-i}(X_i)$, $j = 1, 2$, is the "leave-one-out" vitality function estimators defined using the data with X_i removed. That is

$$\hat{V}_{1,-i}(X_i) = \frac{\frac{1}{n-1}\left[\sum\limits_{j\neq i}^{n} X_i \cdot W\left(\frac{X_i - X_j}{h}\right) + h\sum\limits_{j\neq i}^{n} N_k\left(\frac{X_i - X_j}{h}\right)\right]}{\frac{1}{n-1}\sum\limits_{j\neq i}^{n} 1_{X_j > X_i}} \tag{4.2}$$

and

$$\hat{V}_{2,-i}(X_i) = \frac{\frac{1}{n-1}\sum\limits_{j\neq i}^{n} X_i \cdot 1_{X_j > X_i}}{\frac{1}{n-1}\sum\limits_{j\neq i}^{n} W\left(\frac{X_i - X_j}{h}\right)}. \tag{4.3}$$

4.2. Simulation

We have conducted a numerical study to asses the performance of the estimators that introduced earlier. We simulated repeated samples of size $n = 20, 40, 60, 80$, and 100 from exponential distribution with different means. Thus, the true value being estimated is $L(t)$, when the $f(t)$ is the exponential distribution. The results obtained are based on 1000 repetitions at the sample sizes. We used Epanechnikov kernel, and the likelihood cross-validation for bandwidth selections. The Bias, Variance, and MSE are calculated by repeating the samples 1000 times for each case. Epanechnikov kernel is used for the estimators $\hat{L}_1, \hat{L}_2$, and $\hat{L}_3$. Note that $\hat{L}_0(t) = \frac{M_e(t)}{R_e(t)} - t$ and $\hat{L}_3(t) = \frac{M_n(t)}{R_n(t)} - t$.

Table 1: Simulation from exponential distribution of different means and sample size 20.

n	Mean	Estimators	Bias	Variance	MSE
20	0.5	$\hat{L}_0$	−0.2900	0.3222	0.4063
20	0.5	$\hat{L}_1$	0.3644	0.3034	0.4362
20	0.5	$\hat{L}_2$	0.2643	0.2792	0.3490
20	0.5	$\hat{L}_3$	0.6002	0.2358	0.5960
20	1	$\hat{L}_0$	0.1253	0.0842	0.0999
20	1	$\hat{L}_1$	−0.7525	0.0379	0.6083
20	1	$\hat{L}_2$	−0.2110	0.0575	0.1020
20	1	$\hat{L}_3$	0.0473	0.0542	0.0564
20	5	$\hat{L}_0$	0.1939	0.0478	0.0854
20	5	$\hat{L}_1$	−0.0125	0.0024	0.0026
20	5	$\hat{L}_2$	0.0634	0.0150	0.0191
20	5	$\hat{L}_3$	−0.0714	0.0040	0.0091

Table 2: Simulation from exponential distribution of different means and sample size 40.

n	Mean	Estimators	Bias	Variance	MSE
40	0.5	$\hat{L}_0$	−0.0546	0.3402	0.3432
40	0.5	$\hat{L}_1$	0.1494	0.0834	0.1057
40	0.5	$\hat{L}_2$	−0.9094	0.2316	1.0586
40	0.5	$\hat{L}_3$	0.0260	0.1623	0.1630
40	1	$\hat{L}_0$	0.2041	0.1025	0.1442
40	1	$\hat{L}_1$	0.2277	0.0857	0.1375
40	1	$\hat{L}_2$	−0.1786	0.0370	0.0688
40	1	$\hat{L}_3$	−0.0282	0.0311	0.0319
40	5	$\hat{L}_0$	−0.0406	0.0014	0.0031
40	5	$\hat{L}_1$	0.0632	0.0016	0.0056
40	5	$\hat{L}_2$	−0.0115	0.0036	0.0037
40	5	$\hat{L}_3$	0.0073	0.0019	0.0020

The original purpose of this study was to provide kernel based estimators of mean residual life function. We have found kernel estimation to be a useful tool for nonparametric estimations of reliability functions such as MRL. However, the use of this tool in practice can be hampered by the lack of a suitable bandwidth selection procedure. The likelihood cross-validation proposed in this paper is a suitable technique to select the bandwidth but we can not say it is the optimal one. Also, we can not conclude in the MRL estimators that the smoothing technique is better than the non smoothing technique.

Table 3: Simulation from exponential distribution of different means and sample size 60.

n	Mean	Estimators	Bias	Variance	MSE
60	0.5	$\hat{L}_0$	0.2186	0.1971	0.2449
60	0.5	$\hat{L}_1$	0.1788	0.1156	0.1476
60	0.5	$\hat{L}_2$	0.5544	0.1545	0.4619
60	0.5	$\hat{L}_3$	− 0.2967	0.3261	0.4141
60	1	$\hat{L}_0$	− 0.0105	0.0309	0.0310
60	1	$\hat{L}_1$	− 0.0600	0.0535	0.0571
60	1	$\hat{L}_2$	− 0.1795	0.0174	0.0497
60	1	$\hat{L}_3$	− 0.0730	0.0450	0.0512
60	5	$\hat{L}_0$	− 0.0195	0.0013	0.0017
60	5	$\hat{L}_1$	0.0443	0.0010	0.0030
60	5	$\hat{L}_2$	− 0.0215	0.0006	0.0011
60	5	$\hat{L}_3$	− 0.0309	0.0022	0.0032

Table 4: Simulation from exponential distribution of different means and sample size 80.

n	Mean	Estimators	Bias	Variance	MSE
80	0.5	$\hat{L}_0$	0.0833	0.0600	0.0669
80	0.5	$\hat{L}_1$	0.0319	0.0597	0.0607
80	0.5	$\hat{L}_2$	0.1759	0.2009	0.2317
80	0.5	$\hat{L}_3$	− 0.0866	0.1025	0.1100
80	1	$\hat{L}_0$	0.0392	0.0606	0.0622
80	1	$\hat{L}_1$	0.0536	0.0478	0.0507
80	1	$\hat{L}_2$	0.2112	0.0465	0.0910
80	1	$\hat{L}_3$	− 0.1948	0.0242	0.0622
80	5	$\hat{L}_0$	0.0013	0.0011	0.0012
80	5	$\hat{L}_1$	− 0.0130	0.0011	0.0013
80	5	$\hat{L}_2$	− 0.0516	0.0016	0.0043
80	5	$\hat{L}_3$	− 0.0677	0.0018	0.0064

Table 5: Simulation from exponential distribution of different means and sample size 100.

n	Mean	Estimators	Bias	Variance	MSE
100	0.5	$\hat{L}_0$	0.2901	0.0980	0.1822
100	0.5	$\hat{L}_1$	−0.0888	0.2403	0.2482
100	0.5	$\hat{L}_2$	−0.1093	0.1291	0.1411
100	0.5	$\hat{L}_3$	−0.1420	0.1204	0.1405
100	1	$\hat{L}_0$	−0.0415	0.0293	0.0310
100	1	$\hat{L}_1$	−0.0794	0.0119	0.0182
100	1	$\hat{L}_2$	−0.0278	0.0246	0.0253
100	1	$\hat{L}_3$	0.0736	0.0165	0.0220
100	5	$\hat{L}_0$	0.0131	0.0005	0.0006
100	5	$\hat{L}_1$	−0.0217	0.0020	0.0025
100	5	$\hat{L}_2$	−0.0324	0.0012	0.0022
100	5	$\hat{L}_3$	0.0147	0.0005	0.0007

The MRL estimators proposed in this paper seem natural, reasonable, and intuitively appealing. It is shown that the MRL estimators are asymptotically unbiased and consistent. Note also that the simulation study seem to indicate that the MRL estimators have small variance and MSE. The optimal bandwidth using mean integrated squared error criterion in each MRL estimators is $h = cn^{-1/3}$. It is also proven that the choice of a kernel function is less sensitive to the MRL estimators.

ACKNOWLEDGMENTS

The authors thank the associate editor and the referee for their valuable comments on the earlier three versions of this article.

REFERENCES

[1] Abdous, B. and Berred, A. (2005). Mean residual life estimation, *Journal of Statistical Planning and Inference*, **132**, 3–19.

[2] Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method, *Biometrika*, **68**(1), 326–328.

[3] Chaubey, Yogendra P. and Sen, Pranab K. (1999). On smooth estimation of mean residual life, *Journal of Statistical Planning and Inference*, **75**, 223–236.

[4] Duin, R.P.W. (1976). On the choice of smoothing parameters for Parzen estimators of probability density functions, *IEEE Trans. Comput.*, **C-25**, 1175–1179.

[5] Geisser, S. (1975). The predictive sample reuse method with applications, *J. Amer. Statist. Assoc.*, **70**, 320–328.

[6] Guillamon, A.; Navarro, J. and Ruiz, J.M. (1998). Nonparametric estimator for the mean residual life and vitality function, *Statistical Papers*, **39**, 263–276.

[7] Habbema, J.D.F.; Hermans, J. and van der Broek, K. (1974). *A stepwise discrimination program using density estimation.* In "Compstat" (G. Bruckman, Ed.), Physica Verlag, Vienna.

[8] Hollander, M. and Proschan, F. (1984). *Nonparametric concepts and methods in reliability.* In "Handbook of Statistics" (P.R. Krishnaiah and P.K. Sen, Eds.), **4**, 613-655.

[9] Lo, S.H.; Mack, Y.P. and Wang, J.L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator, *Probab. Theory Rel. Fields*, **80**, 461–473.

[10] Marron, J.S. and Padgett, W.J. (1987): Asymptotically optimal bandwidth selection for kernel density estimators from randomly right censored samples, *Ann. Statist.*, **15**, 1520–1535.

[11] Mitra, M. and Basu, S. (1995). Change point estimation in non-monotonic aging models, *Ann. Inst. Statist. Math.*, **47**(3), 483–491.

[12] Nadaraya, E.A. (1964). Some new estimates for distribution functions, *Theory Prob. Appl.*, **9**, 497–500.

[13] Patil, P.N.; Wells, M.T. and Marron, J.S. (1993). Some heuristics of kernel based estimators of ratio functions, *J. Nonparametric Statistics*, **4**, 203–209.

[14] Ruiz, Jose M. and Guillamon, A. (1999). Nonparametric Recursive Estimator for Mean Residual Life and Vitality Function under dependence conditions, *Commun. Statist. – Theory and Methods*, **25**(9), 1997–2011.

[15] Ruiz, Jose M. and Navarro, J. (1994). Characterization of Distributions By Relationships Between Failure Rate and Mean Residual Life, *IEEE Transactions on Reliability*, **43**(4), 640–644.

[16] Sarda, P. and Vieu, P. (1991). Smoothing parameter selection in hazard estimation, *Statistics & Probability Letters*, **11**, 429–434.

[17] Swanepoel, Jan W.H. (1988). Mean integrated squared error properties and optimal kernels when estimating a distribution function, *Commun. Statist. – Theory Meth.*, **17**(11), 3785–3799.

[18] Watson, G.S. and Leadbetter, M.R. (1964). Hazard analysis I, *Biometrika*, **51**, 175–184.

[19] Watson, G.S. and Leadbetter, M.R. (1964). Hazard analysis II, *Sankhya Ser. A*, **26**, 101–116.

REVSTAT – Statistical Journal
Volume 11, Number 3, November 2013, 317–338

A GENERALIZED SKEW LOGISTIC DISTRIBUTION

Authors: A. Asgharzadeh
– Department of Statistics, University of Mazandaran
Babolsar, Iran
a.asgharzadeh@umz.ac.ir

L. Esmaeili
– Department of Statistics, University of Mazandaran
Babolsar, Iran
l.esmaily@stu.umz.ac.ir

S. Nadarajah
– School of Mathematics, University of Manchester
Manchester M13 9PL, UK
mbbsssn2@manchester.ac.uk

S.H. Shih
– Department of Mathematics and Statistics, American University of Sharjah
Sharjah, UAE
sshih@aus.edu

Received: July 2012 Revised: March 2013 Accepted: March 2013

Abstract:

- In this paper, we introduce a generalized skew logistic distribution that contains the usual skew logistic distribution as a special case. Several mathematical properties of the distribution are discussed like the cumulative distribution function and moments. Furthermore, estimation using the method of maximum likelihood and the Fisher information matrix are investigated. Two real data applications illustrate the performance of the distribution.

Key-Words:

- *estimation; logistic distribution; moments.*

AMS Subject Classification:

- 62E15.

1. INTRODUCTION

Azzalini [2] introduced the skew normal distribution specified by the probability density function (pdf):

$$f_{SN}(x;\lambda) \;=\; 2\,\phi(x)\,\Phi(\lambda x)\;, \qquad -\infty < x < \infty\;, \tag{1.1}$$

where $\lambda \in R$ is the skewness parameter, $\phi(x)$ is the standard normal pdf, and $\Phi(x)$ is the standard normal cumulative distribution function (cdf). Although, Azzalini introduced the skew version (1.1) for the normal distribution, this idea can be applied to any symmetric pdf. Along the same line, the skew logistic distribution with the skewness parameter λ can be proposed as follows. Consider the standard logistic distribution specified by the cdf

$$H(x) \;=\; \frac{1}{1+\exp(-x)}\;, \qquad -\infty < x < \infty\;,$$

and the pdf

$$h(x) \;=\; \frac{\exp(-x)}{\bigl(1+\exp(-x)\bigr)^2}\;, \qquad -\infty < x < \infty\;.$$

Using the idea of Azzalini [2], the pdf of the usual skew logistic distribution is given by

$$f_{SL}(x;\lambda) \;=\; 2\,h(x)\,H(\lambda x) \;=\; \frac{2\exp(-x)}{\bigl(1+\exp(-x)\bigr)^2\,\bigl(1+\exp(-\lambda x)\bigr)} \tag{1.2}$$

for $-\infty < x < \infty$ and $\lambda \in R$. The properties of this distribution have been studied extensively in the literature. See, for example, Nadarajah [12] and Gupta and Kundu [9]. The skew logistic distribution in (1.2) has also received applications; for example, Koessler and Kumar [11] illustrate an application with respect to an adaptive test for the two-sample scale problem based on U-statistics.

Because of the increasing popularity of (1.2), one would like to have generalizations that are more flexible. The aim of this paper is to construct a new generalization of (1.2) using the type III generalized logistic distribution instead of the standard logistic distribution. We study mathematical properties of this new generalization and discuss real data applications.

The type III generalized logistic distribution has the pdf (see Johnson *et al.* [10])

$$g_\alpha(x) = \frac{1}{B(\alpha,\alpha)} \frac{\exp(-\alpha x)}{\big(1+\exp(-x)\big)^{2\alpha}}$$

for $-\infty < x < \infty$ and $\alpha > 0$. This distribution is symmetric for every α. When $\alpha = 1$, the above pdf reduces to the standard logistic pdf. This distribution has the cdf

$$G_\alpha(x) = \frac{B_y(\alpha,\alpha)}{B(\alpha,\alpha)} ,$$

where $y = \big(1+\exp(-x)\big)^{-1}$ and

$$B(\alpha,\alpha) = \frac{\{\Gamma(\alpha)\}^2}{\Gamma(2\alpha)} .$$

Here,

$$\Gamma(a) = \int_0^\infty t^{a-1} \exp(-t)\, dt , \qquad B_x(a,b) = \int_0^x t^{a-1}(1-t)^{b-1}\, dt$$

are the gamma function and the incomplete beta function, respectively.

Now, we define the new skew logistic distribution as follows. If a random variable X has the following pdf

(1.3) $f(x;\alpha,\lambda) = 2\, g_\alpha(x)\, G_\alpha(\lambda x) , \qquad -\infty < x < \infty, \quad \alpha > 0, \quad \lambda \in R ,$

then we say that X has a general skew logistic (GSL) distribution. We write $X \sim$ GSL(α,λ).

From (1.3), some basic properties of GSL(α,λ) can be noted as follows:

- **(i)** When $\alpha = 1$, (1.3) reduces to the usual skew logistic pdf;
- **(ii)** When $\lambda = 0$, (1.3) reduces to the type III generalized logistic pdf;
- **(iii)** If $X \sim$ GSL(α,λ), then $-X \sim$ GSL$(\alpha,-\lambda)$;
- **(iv)** $f(x;\alpha,\lambda) + f(x;\alpha,-\lambda) = 2 g_\alpha(x)$ for all $x \in R$;
- **(v)** $f(x;\alpha,\lambda) \to 2 g_\alpha(x)\, I\{x \geq 0\}$ as $\lambda \to +\infty$ and $f(x;\alpha,\lambda) \to 2 g_\alpha(x)$ $I\{x \leq 0\}$ as $\lambda \to -\infty$ for all α;
- **(vi)** $f(x;\alpha,\lambda) \to 0$ as $x \to \pm\infty$ for all $\alpha > 0$ and $\lambda \in R$.

Numerical investigations show that (1.3) has a single mode. The mode is at x_0, where x_0 is the root of

$$\frac{\lambda}{\alpha} \frac{g_\alpha(x)}{G_\alpha(x)} - \frac{1-\exp(-x)}{1+\exp(-x)} = 0 .$$

Figures 1 and 2 illustrate possible shapes of the pdf (1.3) for $\alpha = 2$ and selected values of λ.

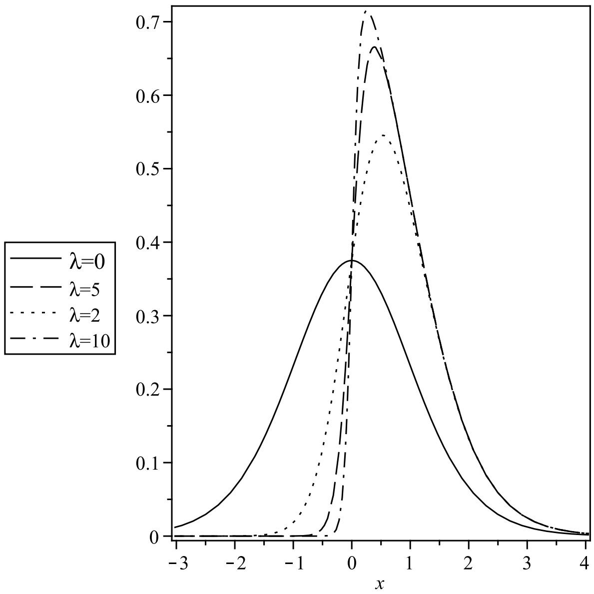

Figure 1: Plots of GSL(α, λ) pdf for $\alpha = 2$ and $\lambda > 0$.

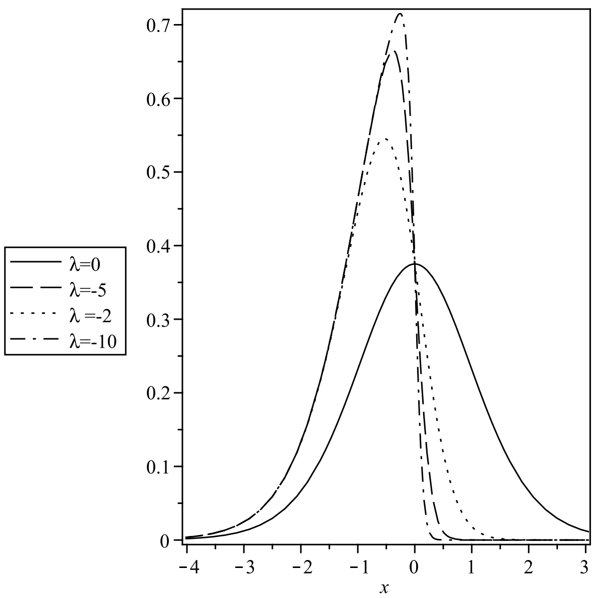

Figure 2: Plots of GSL(α, λ) pdf for $\alpha = 2$ and $\lambda < 0$.

Simulation from (1.3) is straight-forward by using the following representation due to Azzalini [3]:

- $X = S_U U$, where, conditionally on $U = u$, $S_U = +1$ with probability $G_\alpha(\lambda u)$ and $S_U = -1$ with probability $1 - G_\alpha(\lambda u)$;
- $X = S_U |U|$, where, conditionally on $|U| = |u|$, $S_U = +1$ with probability $G_\alpha(\lambda |u|)$ and $S_U = -1$ with probability $1 - G_\alpha(\lambda |u|)$.

Both these representations have physical meanings as explained in Azzalini [3].

In the sequel, we shall use the following functions:

$$\tau_1(b, q) = \sum_{j=0}^{\infty} \frac{\binom{-2q}{j}}{(q+j)^b}\,, \qquad \tau_2(a, b, q, \lambda) = \sum_{j=0}^{\infty} \frac{\binom{-2q}{j}}{(\lambda b + q + j)^a}$$

and, the Gauss hypergeometric function defined by

$${}_2F_1(a, b; c; x) = \sum_{k=0}^{\infty} \frac{(a)_k\,(b)_k}{(c)_k} \frac{x^k}{k!}\,,$$

where $(z)_k = z(z+1) \cdots (z+k-1)$ denotes the ascending factorial.

Throughout the rest of this paper (unless otherwise stated), we shall assume that $\lambda > 0$ since the corresponding results for $\lambda < 0$ can be obtained using the fact that $-X$ has the pdf $2 g_\alpha(x)\, G_\alpha(-\lambda x)$.

Some results of this paper require certain series representations of the general skew logistic pdf (1.3), which we derive now. Using the Taylor series expansion for $[1 + \exp(-\lambda x)]^{-1}$, we can obtain

$$G_\alpha(\lambda x) = \begin{cases} \dfrac{1}{B(\alpha,\alpha)} \displaystyle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \binom{\alpha-1}{i}\binom{-(\alpha+i)}{j} \\ \qquad \cdot \dfrac{(-1)^i \exp(-j\lambda x)}{i+\alpha}\,, & x > 0\,, \\ \dfrac{1}{B(\alpha,\alpha)} \displaystyle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \binom{\alpha-1}{i}\binom{-(\alpha+i)}{j} \\ \qquad \cdot \dfrac{(-1)^i \exp\bigl(\lambda x(i+\alpha+j)\bigr)}{i+\alpha}\,, & x < 0\,. \end{cases}$$

Substituting this into (1.3), a double series representation for the general skew logistic pdf can be obtained as

$$(1.4)\qquad f(x;\alpha,\lambda) = \begin{cases} \dfrac{2}{B(\alpha,\alpha)^2\big(1+\exp(-x)\big)^{2\alpha}} \displaystyle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \binom{\alpha-1}{i}\binom{-(\alpha+i)}{j} \\ \qquad \cdot \dfrac{(-1)^i \exp\big(-x(\lambda j+\alpha)\big)}{i+\alpha}\,, & x>0\,, \\[2ex] \dfrac{2}{B(\alpha,\alpha)^2\big(1+\exp(-x)\big)^{2\alpha}} \displaystyle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \binom{\alpha-1}{i}\binom{-(\alpha+i)}{j} \\ \qquad \cdot \dfrac{(-1)^i \exp\big(x(\lambda i+\lambda\alpha+\lambda j-\alpha)\big)}{i+\alpha}\,, & x<0\,. \end{cases}$$

By expanding the terms in the denominators of (1.4), one can also obtain the triple series representation

$$(1.5)\qquad f(x;\alpha,\lambda) = \begin{cases} \dfrac{2}{B(\alpha,\alpha)^2} \displaystyle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \binom{\alpha-1}{i}\binom{-(\alpha+i)}{j}\binom{-2\alpha}{k} \\ \qquad \cdot \dfrac{(-1)^i \exp\big(-x(\lambda j+k+\alpha)\big)}{i+\alpha}\,, & x>0\,, \\[2ex] \dfrac{2}{B(\alpha,\alpha)^2} \displaystyle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \binom{\alpha-1}{i}\binom{-(\alpha+i)}{j}\binom{-2\alpha}{k} \\ \qquad \cdot \dfrac{(-1)^i \exp\Big(x\big(\lambda(i+\alpha+j)+k+\alpha\big)\Big)}{i+\alpha}\,, & x<0\,. \end{cases}$$

2. CUMULATIVE DISTRIBUTION FUNCTION

Using the double and triple series representations in (1.4) and (1.5), we derive some formulas for the cdf corresponding to (1.3). First, we use the double series representation in (1.4). If $x>0$, then the cdf $F(x)$ can be written as

$$(2.1)\qquad F(x) \;=\; F(0) + \int_0^x \frac{2}{B(\alpha,\alpha)^2} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\binom{\alpha-1}{i}\binom{-(\alpha+i)}{j}(-1)^i}{(i+\alpha)\big(1+\exp(-t)\big)^{2\alpha}} \cdot \exp\big(-t(\lambda j+\alpha)\big)\,dt \;=$$

$$= F(0) + \frac{2}{B(\alpha,\alpha)^2} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\binom{\alpha-1}{i}\binom{-(\alpha+i)}{j}(-1)^i}{i+\alpha} \cdot \int_0^x \frac{\exp\bigl(-t(\lambda j+\alpha)\bigr)}{\bigl(1+\exp(-t)\bigr)^{2\alpha}}\, dt \tag{2.2}$$

$$= F(0) + \frac{2}{B(\alpha,\alpha)^2} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\binom{\alpha-1}{i}\binom{-(\alpha+i)}{j}(-1)^i}{i+\alpha} \cdot \int_{\exp(-x)}^1 \frac{z^{\lambda j+\alpha-1}}{(1+z)^{2\alpha}}\, dz \tag{2.3}$$

$$= F(0) + \frac{2}{B(\alpha,\alpha)^2} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\binom{\alpha-1}{i}\binom{-(\alpha+i)}{j}(-1)^i}{i+\alpha}\, I(x)\ , \tag{2.4}$$

where

$$I(x) = \int_{\exp(-x)}^1 \frac{z^{\lambda j+\alpha-1}}{(1+z)^{2\alpha}}\, dz = \int_0^1 \frac{z^{\lambda j+\alpha-1}}{(1+z)^{2\alpha}}\, dz - \int_0^{\exp(-x)} \frac{z^{\lambda j+\alpha-1}}{(1+z)^{2\alpha}}\, dz = I_1 - I_2\ .$$

By equation (3.194.1) in Gradshteyn and Ryzhik [8], the integrals I_1 and I_2 can be calculated as

$$I_1 = \int_0^1 \frac{z^{\lambda j+\alpha-1}}{(1+z)^{2\alpha}}\, dz = \frac{1}{\lambda j+\alpha}\ {}_2F_1\bigl(2\alpha, \alpha+\lambda j; \alpha+\lambda j+1; -1\bigr)\,, \tag{2.5}$$

and

$$I_2 = \int_0^{\exp(-x)} \frac{z^{\lambda j+\alpha-1}}{(1+z)^{2\alpha}}\, dz = \frac{\exp\bigl(-(\alpha+\lambda j)\,y\bigr)}{\alpha+\lambda j}\ {}_2F_1\bigl(2\alpha, \alpha+\lambda j; \alpha+\lambda j+1; -\exp(-x)\bigr)\,. \tag{2.6}$$

Combining (2.5) and (2.6) and substituting into (2.4), the cdf $F(x)$ for $x > 0$ becomes

$$F(x) = F(0) + \frac{2}{B(\alpha,\alpha)^2} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\binom{\alpha-1}{i}\binom{-(\alpha+i)}{j}(-1)^i}{i+\alpha} \cdot \Biggl\{ \frac{1}{\lambda j+\alpha}\ {}_2F_1\bigl(2\alpha, \alpha+\lambda j; \alpha+\lambda j+1; -1\bigr) - \frac{\exp\bigl(-(\alpha+\lambda j)\,x\bigr)}{\alpha+\lambda j}\ {}_2F_1\bigl(2\alpha, \alpha+\lambda j, \alpha+\lambda j+1; -\exp(-x)\bigr) \Biggr\}\,.$$

Repeating the above argument with $x = 0$ yields the form for $F(0)$ as

$$\begin{aligned} F(0) &= \int_{-\infty}^{0} f(t)\,dt \\ &= \frac{2}{B(\alpha,\alpha)^2} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\binom{\alpha-1}{i}\binom{-(\alpha+i)}{j}(-1)^i}{i+\alpha} \\ &\quad \cdot \int_{-\infty}^{0} \frac{\exp\Big(t\big(\lambda(i+\alpha+j)+\alpha\big)\Big)}{\big(1+\exp(t)\big)^{2\alpha}}\,dt \\ &= \frac{2}{B(\alpha,\alpha)^2} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\binom{\alpha-1}{i}\binom{-(\alpha+i)}{j}(-1)^i}{i+\alpha} \int_0^1 \frac{z^{\lambda(i+\alpha+j)+\alpha-1}}{(1+z)^{2\alpha}}\,dz \\ &= \frac{2}{B(\alpha,\alpha)^2} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\binom{\alpha-1}{i}\binom{-(\alpha+i)}{j}(-1)^i}{(i+\alpha)} \\ &\quad \cdot \frac{{}_2F_1\Big(2\alpha,\ \lambda(i+\alpha+j)+\alpha;\ \lambda(i+\alpha+j)+\alpha+1;\ -1\Big)}{\lambda(i+\alpha+j)+\alpha}\,. \end{aligned}$$

If $x < 0$, then similar arguments by using equation (3.194.1) in Gradshteyn and Ryzhik [8] yields

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \frac{2}{B(\alpha,\alpha)^2} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\binom{\alpha-1}{i}\binom{-(\alpha+i)}{j}(-1)^i}{i+\alpha}\, I\,,$$

where

$$\begin{aligned} I &= \int_{-\infty}^{x} \frac{\exp\Big(t\big(\lambda(j+\alpha+i)+\alpha\big)\Big)}{\big(1+\exp(t)\big)^{2\alpha}} = \int_0^{\exp(x)} \frac{z^{\lambda(i+\alpha+j)+\alpha-1}}{(1+z)^{2\alpha}}\,dz \\ &= \frac{\exp\Big(x\big(\lambda(i+\alpha+j)+\alpha\big)\Big)}{\lambda(i+\alpha+j)+\alpha} \\ &\quad \cdot {}_2F_1\Big(2\alpha,\ \lambda(i+\alpha+j)+\alpha;\ \lambda(i+\alpha+j)+\alpha+1,\ -\exp(x)\Big)\,, \end{aligned}$$

and so the result

$$\begin{aligned} F(x) &= \frac{2}{B(\alpha,\alpha)^2} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\binom{\alpha-1}{i}\binom{-(\alpha+i)}{j}(-1)^i}{i+\alpha} \\ &\quad \cdot \frac{\exp\Big(x\big(\lambda(i+\alpha+j)+\alpha\big)\Big)}{\lambda(i+\alpha+j)+\alpha} \\ &\quad \cdot {}_2F_1\Big(2\alpha,\ \lambda(i+\alpha+j)+\alpha;\ \lambda(i+\alpha+j)+\alpha+1,\ -\exp(x)\Big)\,. \end{aligned}$$

Using the triple series representation, (1.5), the cdf $F(x)$ can be calculated as

$$F(x)=\begin{cases} 1-\dfrac{2}{B(\alpha,\alpha)^2}\displaystyle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{\binom{\alpha-1}{i}\binom{-(\alpha+i)}{j}\binom{-2\alpha}{k}(-1)^i}{(i+\alpha)\,(\lambda j+k+\alpha)} \\ \quad\cdot\ \exp\bigl(-x(\lambda j+k+\alpha)\bigr)\,, & x>0\,, \\ \dfrac{2}{B(\alpha,\alpha)^2}\displaystyle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{\binom{\alpha-1}{i}\binom{-(\alpha+i)}{j}\binom{-2\alpha}{k}(-1)^i}{(i+\alpha)\,\bigl(\lambda(i+\alpha+j)+k+\alpha\bigr)} \\ \quad\cdot\ \exp\Bigl(x\bigl(\lambda(i+\alpha+j)+k+\alpha\bigr)\Bigr)\,, & x<0\,. \end{cases}$$

3. MOMENTS

Many of the interesting characteristics of the general skew logistic distribution can be studied through its moments. Let $X\sim \mathrm{GSL}(\alpha,\lambda)$. In this section, we derive the nth moment of X. It is easy to show that if X follows $\mathrm{GSL}(\alpha,\lambda)$ then $Y=|X|$ has the folded form of the type III generalized logistic distribution specified by the pdf

$$g(y;\alpha,\lambda)\;=\;\frac{2}{B(\alpha,\alpha)}\,\frac{\exp(-\alpha y)}{\bigl(1+\exp(-y)\bigr)^{2\alpha}}$$

for $y>0$. Thus, the even order moments of X are obtained as

$$\begin{aligned} E\left(X^n\right) &= \frac{2}{B(\alpha,\alpha)}\int_0^{\infty}\frac{x^n\exp(-\alpha x)}{\bigl(1+\exp(-x)\bigr)^{2\alpha}}\,dx \;=\; \frac{2}{B(\alpha,\alpha)}\int_0^1\frac{\left(\ln\dfrac{1}{z}\right)^n z^{\alpha-1}}{(1+z)^{2\alpha}}\,dz \\ &= \frac{2}{B(\alpha,\alpha)}\sum_{i=0}^{\infty}\binom{-2\alpha}{i}\int_0^1\left(\ln\frac{1}{z}\right)^n z^{\alpha+i-1}\,dz \\ &= \frac{2n!}{B(\alpha,\alpha)}\sum_{i=0}^{\infty}\frac{\binom{-2\alpha}{i}}{(\alpha+i)^{n+1}} \;=\; \frac{2n!}{B(\alpha,\alpha)}\,\tau_1(n+1,\alpha)\;, \end{aligned}$$

where the penultimate step follows by using equation (4.272.6) in Gradshteyn and Ryzhik [8].

If n is odd then, using the triple series representation, (1.5), one obtains

$$\begin{aligned}
E\left(X^n\right) &= \frac{2}{B^2(\alpha,\alpha)} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \frac{\binom{\alpha-1}{i}\binom{-(\alpha+i)}{j}\binom{-2\alpha}{k}(-1)^i}{i+\alpha} \\
&\quad \cdot \left\{ \int_{-\infty}^{0} x^n \exp\Big(x\big(\lambda(i+\alpha+j)+k+\alpha\big)\Big)\,dx \right. \\
&\quad \left. + \int_{0}^{\infty} x^n \exp\Big(-x\big(\lambda j+k+\alpha\big)\Big)\,dx \right\} \\
&= \frac{2n!}{B^2(\alpha,\alpha)} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \frac{\binom{\alpha-1}{i}\binom{-(\alpha+i)}{j}\binom{-2\alpha}{k}(-1)^i}{i+\alpha} \\
&\quad \cdot \left\{ \frac{1}{(\lambda j+k+\alpha)^{n+1}} - \frac{1}{\big(\lambda(i+\alpha+j)+k+\alpha\big)^{n+1}} \right\} \\
&= \frac{2n!}{B^2(\alpha,\alpha)} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\binom{\alpha-1}{i}\binom{-(\alpha+i)}{j}(-1)^i}{i+\alpha}\, \Delta(n+1,\alpha,\lambda)\ ,
\end{aligned}$$

where $\Delta(n+1,\alpha,\lambda) = \tau_2(n+1,j,\alpha,\lambda) - \tau_2(n+1,i+\alpha+j,\alpha,\lambda)$.

Using these, the first four moments of X can be obtained as

$$E(X) = \frac{2}{B^2(\alpha,\alpha)} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\binom{\alpha-1}{i}\binom{-(\alpha+i)}{j}(-1)^i}{i+\alpha}\, \Delta(2,\alpha,\lambda)\ ,$$

$$E(X^2) = \frac{4}{B(\alpha,\alpha)}\, \tau_1(3,\alpha)\ ,$$

$$E(X^3) = \frac{12}{B^2(\alpha,\alpha)} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\binom{\alpha-1}{i}\binom{-(\alpha+i)}{j}(-1)^i}{i+\alpha}\, \Delta(4,\alpha,\lambda)\ ,$$

and

$$E(X^4) = \frac{48}{B(\alpha,\alpha)}\, \tau_1(5,\alpha)\ .$$

Using the above moments, we can calculate the four measures $E(X)$, $Var(X)$, $Skewness(X)$ and $Kurtosis(X)$. Figures 3 to 6 illustrate the behavior of the four measures for $\lambda = -10, ..., 10$ and $\alpha = 1, 2, 5$. From these figures, we see that:

(i) $E(X)$ increases with increasing λ;

(ii) $E(X)$ decreases with increasing α;

(iii) $Var(X)$ decreases with increasing $|\lambda|$;

(**iv**) $Var(X)$ decreases with increasing α;
(**v**) $Skewness(X)$ increases with increasing λ;
(**vi**) $|Skewness(X)|$ decreases with increasing α;
(**vii**) $Kurtosis(X)$ initially decreases before increasing with increasing $|\lambda|$;
(**viii**) $Kurtosis(X)$ decreases with increasing α.

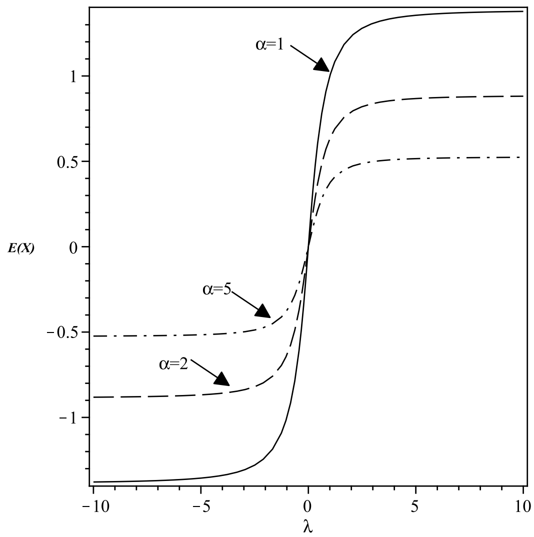

Figure 3: Plot of $E(X)$.

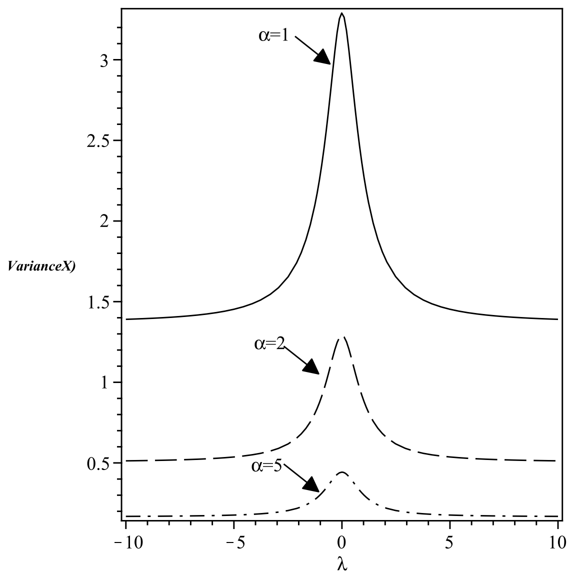

Figure 4: Plot of $Variance(X)$.

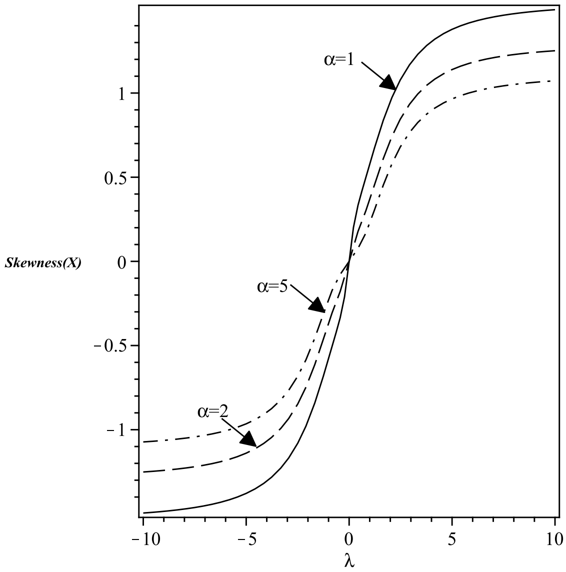

Figure 5: Plot of $Skewness(X)$.

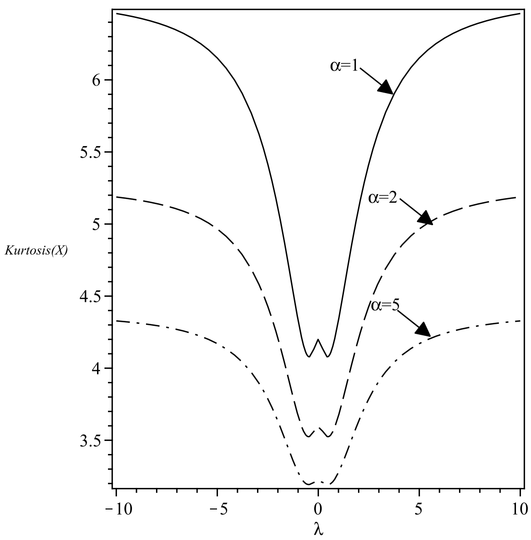

Figure 6: Plot of $Kurtosis(X)$.

4. ESTIMATION

Let us first consider a version of (1.3) with the location parameter $\mu \in R$ and scale parameter $\sigma > 0$, i.e.,

$$f(x;\mu,\sigma,\alpha,\lambda) = \frac{2}{\sigma}\, g_\alpha\!\left(\frac{x-\mu}{\sigma}\right) G_\alpha\!\left[\lambda\left(\frac{x-\mu}{\sigma}\right)\right] \tag{4.1}$$

for $-\infty < x < \infty$, $\alpha > 0$ and $\lambda \in R$. In this section, we consider estimation of the parameters μ, σ, α and λ and provide expressions for the Fisher information matrix. The log-likelihood for a random sample $x_1, ..., x_n$ from (4.1) is:

$$\ell = \ln L(\mu,\sigma,\alpha,\lambda) = -n\ln\sigma + \sum_{i=1}^{n} \ln g_\alpha(y_i) + \sum_{i=1}^{n} \zeta_0(\lambda y_i)\ , \tag{4.2}$$

where $y_i = \frac{x_i-\mu}{\sigma}$ and $\zeta_0(x) = \ln\{2G_\alpha(x)\}$. We also define the derivative $\zeta_m(x) = d^m\zeta_0(x)/dx^m$, $m = 1, 2, 3, ...$ and note that $\zeta_1(x) = g_\alpha(x)/G_\alpha(x)$. All subsequent derivatives can be expressed as functions of $\zeta_1(x)$; in particular, $\zeta_2(x) = -\alpha\left(\frac{1-\exp(-x)}{1+\exp(-x)}\right)\zeta_1(x) - \zeta_1(x)^2$.

By differentiating (4.2) with respect to μ, σ, α and λ, and equating the derivatives to zero, the maximum likelihood estimators are the simultaneous solutions of

$$2\alpha \sum_{i=1}^{n} \frac{\exp(-y_i)}{1+\exp(-y_i)} + \lambda \sum_{i=1}^{n} \zeta_1(\lambda y_i) = n\,\alpha\ , \tag{4.3}$$

$$n + \lambda \sum_{i=1}^{n} y_i\, \zeta_1(\lambda y_i) = \alpha \sum_{i=1}^{n} \frac{y_i\big(1-\exp(-y_i)\big)}{1+\exp(-y_i)}\ , \tag{4.4}$$

$$\sum_{i=1}^{n} y_i + 2\sum_{i=1}^{n} \ln\big\{1+\exp(-y_i)\big\} - \sum_{i=1}^{n} \frac{\partial \ln\big\{2G_\alpha(\lambda y_i)\big\}}{\partial\alpha} = 2n\big(\Psi(2\alpha)-\Psi(\alpha)\big) \tag{4.5}$$

and

$$\sum_{i=1}^{n} y_i\, \zeta_1(\lambda y_i) = 0\ , \tag{4.6}$$

where $\Psi(x) = \ln\Gamma(x)/dx$ is the digamma function. In (4.5), we have

$$\frac{\partial \ln\big\{2G_\alpha(\lambda y)\big\}}{\partial\alpha} = 2\big(\Psi(2\alpha)-\Psi(\alpha)\big) - \frac{\displaystyle\int_{-\infty}^{\lambda y} t\, g_\alpha(t)\, dt}{G_\alpha(\lambda y)} - 2\, \frac{\displaystyle\int_{-\infty}^{\lambda y} \ln\big(1+\exp(-t)\big)\, g_\alpha(t)\, dt}{G_\alpha(\lambda y)}\ .$$

The maximum likelihood estimators $(\widehat{\mu}, \widehat{\sigma}, \widehat{\lambda}, \widehat{\alpha})$ of $(\mu, \sigma, \lambda, \alpha)$ are consistent estimators, and $\sqrt{n}(\widehat{\mu} - \mu, \widehat{\sigma} - \sigma, \widehat{\lambda} - \lambda, \widehat{\alpha} - \alpha)$ is asymptotically normal with zero means and variance covariance matrix $\mathbf{I}^{-1}$, where

$$\mathbf{I} = -\frac{1}{n}\begin{bmatrix} E\left(\frac{\partial^2 \ell}{\partial^2 \mu}\right) & E\left(\frac{\partial^2 \ell}{\partial \mu \partial \sigma}\right) & E\left(\frac{\partial^2 \ell}{\partial \mu \partial \lambda}\right) & E\left(\frac{\partial^2 \ell}{\partial \mu \partial \alpha}\right) \\ E\left(\frac{\partial^2 \ell}{\partial \sigma \partial \mu}\right) & E\left(\frac{\partial^2 \ell}{\partial^2 \sigma}\right) & E\left(\frac{\partial^2 \ell}{\partial \sigma \partial \lambda}\right) & E\left(\frac{\partial^2 \ell}{\partial \sigma \partial \alpha}\right) \\ E\left(\frac{\partial^2 \ell}{\partial \lambda \partial \mu}\right) & E\left(\frac{\partial^2 \ell}{\partial \lambda \partial \sigma}\right) & E\left(\frac{\partial^2 \ell}{\partial^2 \lambda}\right) & E\left(\frac{\partial^2 \ell}{\partial \lambda \partial \alpha}\right) \\ E\left(\frac{\partial^2 \ell}{\partial \alpha \partial \mu}\right) & E\left(\frac{\partial^2 \ell}{\partial \alpha \partial \sigma}\right) & E\left(\frac{\partial^2 \ell}{\partial \alpha \partial \lambda}\right) & E\left(\frac{\partial^2 \ell}{\partial^2 \alpha}\right) \end{bmatrix}.$$

Now, we compute the Fisher information matrix based on the likelihood equations. These enable, for example, construction of confidence intervals based on pivotal quantities using the limiting normal distribution. For simplicity, let us consider interval estimation of (μ, σ, λ) when α is known. In this case, the elements of the Fisher information matrix can be written as

$$\begin{aligned}
-E\left(\frac{\partial^2 \ell}{\partial^2 \mu}\right) &= \frac{2\alpha n}{\sigma^2} I_1 + \frac{n\lambda^2}{\sigma^2} I_2 , \\
-E\left(\frac{\partial^2 \ell}{\partial^2 \sigma}\right) &= -\frac{n}{\sigma^2} + \frac{2n\alpha}{\sigma^2} E(Z) - \frac{4n\alpha}{\sigma^2} I_3 + \frac{2n\alpha}{\sigma^2} I_4 + \frac{n\lambda^2}{\sigma^2} a_{22}(\lambda) , \\
-E\left(\frac{\partial^2 \ell}{\partial^2 \lambda}\right) &= n a_{22}(\lambda) , \\
-E\left(\frac{\partial^2 \ell}{\partial \mu \partial \lambda}\right) &= \frac{n}{\sigma} I_5 - \frac{n\alpha\lambda}{\sigma} I_6 - \frac{n\lambda}{\sigma} a_{12}(\lambda) , \\
-E\left(\frac{\partial^2 \ell}{\partial \sigma \partial \lambda}\right) &= \frac{n\alpha\lambda}{\sigma} I_6 - \frac{n\lambda}{\sigma} a_{12}(\lambda) , \\
-E\left(\frac{\partial^2 \ell}{\partial \mu \partial \sigma}\right) &= \frac{n\alpha}{\sigma^2} + \frac{2n\alpha}{\sigma^2} I_1 - \frac{2n\alpha}{\sigma^2} I_7 - \frac{n\lambda}{\sigma^2} I_5 - \frac{n\lambda^2\alpha}{\sigma^2} I_6 - \frac{n\lambda^2}{\sigma^2} a_{12}(\lambda) , \\
-E\left(\frac{\partial^2 \ell}{\partial^2 \alpha}\right) &= 4n\big(\Psi(1,\alpha) - 2\Psi(1,2\alpha)\big) - nI_8 - 4nI_9 - 4nI_{10} \\
&\quad - nI_{11} + 4nI_{12} + 4nI_{13} , \\
-E\left(\frac{\partial^2 \ell}{\partial \alpha \partial \mu}\right) &= -\frac{n}{\sigma} I_{17} - \frac{2n\lambda}{\sigma} I_{18} + \frac{n\lambda}{\sigma} I_{19} + \frac{n\lambda}{\sigma} I_{20} , \\
-E\left(\frac{\partial^2 \ell}{\partial \alpha \partial \sigma}\right) &= \frac{n}{\sigma} I_{21} - \frac{2n\lambda}{\sigma}\big(\Psi(\alpha) - \Psi(2\alpha)\big) I_5 - \frac{n\lambda^2}{\sigma} a_{21}(\lambda) - \frac{2n\lambda}{\sigma} I_{23} \\
&\quad + \frac{n\lambda}{\sigma} I_{22} + \frac{2n\lambda}{\sigma} I_{16} , \\
-E\left(\frac{\partial^2 \ell}{\partial \alpha \partial \lambda}\right) &= 2nI_{14} - nI_{15} - nI_{16} ,
\end{aligned}$$

where

$$I_1 = E\left(\frac{\exp(-Z)}{1+\exp(-Z)}\right), \quad I_2 = E\big(\zeta_1^2(\lambda Z)\big), \quad I_3 = E\left(\frac{Z\exp(-Z)}{1+\exp(-Z)}\right),$$

$$I_4 = E\left(\frac{Z^2\exp(-Z)}{\big\{1+\exp(-Z)\big\}^2}\right), \quad I_5 = E\big(\zeta_1(\lambda Z)\big),$$

$$I_6 = E\left(\frac{Z\big(1-\exp(-\lambda Z)\big)\,\zeta_1(\lambda Z)}{1+\exp(-\lambda Z)}\right), \quad I_7 = E\left(\frac{Z\exp(-Z)}{\big\{1+\exp(-Z)\big\}^2}\right),$$

$$I_8 = E\left(\frac{b_2(\lambda Z)}{G_\alpha(\lambda Z)}\right), \quad I_9 = E\left(\frac{c_{11}(\lambda Z)}{G_\alpha(\lambda Z)}\right), \quad I_{10} = E\left(\frac{c_{02}(\lambda Z)}{G_\alpha(\lambda Z)}\right),$$

$$I_{11} = E\left(\frac{b_1^2(\lambda Z)}{G_\alpha^2(\lambda Z)}\right), \quad I_{12} = E\left(\frac{c_{01}^2(\lambda Z)}{G_\alpha^2(\lambda Z)}\right), \quad I_{13} = E\left(\frac{b_1(\lambda Z)\,c_{01}(\lambda Z)}{G_\alpha^2(\lambda Z)}\right),$$

$$I_{14} = E\Big(Z\,\zeta_1(\lambda Z)\,\ln\big(1+\exp(-\lambda Z)\big)\Big), \quad I_{15} = E\left(Z\,\zeta_1(\lambda Z)\,\frac{b_1(\lambda Z)}{G_\alpha(\lambda Z)}\right),$$

$$I_{16} = E\left(Z\,\zeta_1(\lambda Z)\,\frac{c_{01}(\lambda Z)}{G_\alpha(\lambda Z)}\right), \quad I_{17} = E\left(\frac{1-\exp(-Z)}{1+\exp(-Z)}\right),$$

$$I_{18} = E\Big(\zeta_1(\lambda Z)\,\ln\big(1+\exp(-\lambda Z)\big)\Big), \quad I_{19} = E\left(\zeta_1(\lambda Z)\,\frac{b_1(\lambda Z)}{G_\alpha(\lambda Z)}\right),$$

$$I_{20} = E\left(\zeta_1(\lambda Z)\,\frac{c_{01}(\lambda Z)}{G_\alpha(\lambda Z)}\right), \quad I_{21} = E\left(Z\,\frac{\exp(-Z)-1}{\exp(-Z)+1}\right),$$

$$I_{22} = E\Big(Z\,\zeta_1(\lambda Z)\,b_1(\lambda Z)\Big), \quad I_{23} = E\Big(Z\,\zeta_1(\lambda Z)\,\ln\big(1+\exp(-Z)\big)\Big),$$

where $Z = (X-\mu)/\sigma$,

$$a_{kh}(\lambda) = E_\lambda\big\{Z^k\,\zeta_1^h(\lambda Z)\big\}, \qquad b_k(x) = \int_{-\infty}^x t^k\,g_\alpha(t)\,dt, \qquad \Psi(n,x) = \frac{d^n}{d^n x}\Psi(x)$$

and

$$c_{kh}(x) = \int_{-\infty}^x t^k\,\ln^h\big\{1+\exp(-t)\big\}\,g_\alpha(t)\,dt\,.$$

Note that $a_{k1}(\lambda)=0$ when k is odd, and that $a_{kh}(\lambda)\geq 0$ when both k and h are even. Also $E(h(Z)\,\zeta_1(\lambda Z))=0$ when $h(x)$ is an odd function and $E(h(Z)\,\zeta_1(\lambda Z))\geq 0$ when $h(x)$ is a even function. In general, these expectations will have to be computed numerically. However, closed-form expressions are possible in some particular cases.

5. REAL DATA APPLICATIONS

In this section, we fit the general skew logistic GSL$(\mu, \sigma, \lambda, \alpha)$ distribution to two real data sets. We compare the fits with those of the usual logistic distribution L(μ, σ), the type III generalized logistic distribution GL(μ, σ, α), the skew logistic distribution SL(μ, σ, λ), Azzalini's [2] skew normal distribution SN(μ, σ, λ), and Azzalini and Capitanio's [5] skew t distribution ST$(\mu, \sigma, \lambda,\ \alpha)$. The parameter λ in the skew normal and skew t distributions is the skewness parameter. The parameter α in the skew t distribution is the degree of freedom parameter. As with Azzalini's [2] skew normal distribution, Azzalini and Capitanio's [5] skew t distribution has been studied by many authors. Two most recent papers are Arellano-Valle and Azzalini [1] and Azzalini and Arellano-Valle [4].

Example 1. The first data set represents the strength data originally reported in Badar and Priest [6]. It represents the strength measured in GPA for single carbon fibers and impregnated 1000-carbon fiber tows. Single fibers were tested under tension at gauge length of 10mm. This data have been analyzed previously by Raqab and Kundu [13] and Gupta and Kundu [9]. The data are as follows:

1.901	2.132	2.203	2.228	2.257	2.350	2.361	2.396	2.397
2.445	2.454	2.474	2.518	2.522	2.525	2.532	2.575	2.614
2.616	2.618	2.624	2.659	2.675	2.738	2.740	2.856	2.917
2.928	2.937	2.937	2.977	2.996	3.030	3.125	3.139	3.145
3.220	3.223	3.235	3.243	3.264	3.272	3.294	3.332	3.346
3.377	3.408	3.435	3.493	3.501	3.537	3.554	3.562	3.628
3.852	3.871	3.886	3.971	4.024	4.027	4.225	4.395	5.020.

We fitted all six distributions to the above data by the method of maximum likelihood. The GSL distribution was fitted by solving (4.3)–(4.6). Table 1 presents the parameter estimates, the log likelihoods (LL), the Kolmogorov–Smirnov (K-S) statistics and respective p-values. Table 2 presents the chi-squared statistics with observed and expected frequencies. Note that the last two columns of Tables 1 and 2 appear identical. This can be explained by the well-known fact that the ST distribution reduces to the SN distribution as α approaches infinity.

Note also that the $\widehat{\alpha}$ for the GSL distribution is very large. Some elementary calculations show that

$$g_\lambda(x) \;\to\; \frac{1}{4^\alpha B(\alpha,\alpha)}\, I\{x=0\}$$

and

$$G_\lambda(x) \;\to\; I\{x \geq 0\}$$

as $\alpha \to \infty$. So, the pdf of the GSL distribution in (1.3) reduces to

$$f(x;\alpha,\lambda) \;\to\; \frac{2^{1-2\alpha}}{B(\alpha,\alpha)}\, I\{x=0\}\, I\{\lambda x \geq 0\}$$

as $\alpha \to \infty$.

Table 1: MLEs, log-likelihoods, Kolmogorov–Smirnov statistics and corresponding p-values for Example 1.

Distribution	L(μ,σ)	SL(μ,σ,λ)	GL(μ,σ,α)	GSL$(\mu,\sigma,\lambda,\alpha)$	SN(μ,σ,λ)	ST$(\mu,\sigma,\lambda,\alpha)$
$\widehat{\mu}$	3.024	2.328	3.048	2.271	2.271	2.271
$\widehat{\sigma}$	0.352	0.550	0.930	195.801	1.000	1.000
$\widehat{\lambda}$	—	3.713	—	4.418	4.419	4.419
$\widehat{\alpha}$	—	—	5.041879	76605.63	—	26491.46
Log-likelihood	−59.330	−56.794	−58.797	−55.902	−55.902	−55.902
KSS	0.094	0.084	0.097	0.073	0.075	0.075
p-value	0.606	0.754	0.571	0.880	0.877	0.877

Table 2: Observed and expected frequencies and chi-squared statistics for Example 1.

Intervals	Observed	L(μ,σ)	SL(μ,σ,λ)	GL(μ,σ,α)	GSL$(\mu,\sigma,\lambda,\alpha)$	SN(μ,σ,λ)	ST$(\mu,\sigma,\lambda,\alpha)$
< 2.5	12	11.61	11.88	11.48	12.28	12.28	12.28
2.5–3.0	20	18.80	22.53	18.00	21.35	21.35	21.35
3.0–3.5	17	19.61	15.25	19.22	15.55	15.55	15.55
3.5–4	9	9.26	7.61	10.51	8.53	8.53	8.53
> 4	5	3.72	5.74	3.79	5.29	5.29	5.29
		$\chi^2 = 0.8832$	$\chi^2 = 0.8345$	$\chi^2 = 1.1055$	$\chi^2 = 0.2682$	$\chi^2 = 0.2684$	$\chi^2 = 0.2684$

From Tables 1 and 2, we see that the GSL distribution provides a better fit for the data than the other five distributions. The GSL distribution takes the smallest chi-squared statistic, the smallest K-S statistic, and the largest p-value. The SN and ST distributions take the second smallest chi-squared statistic, the second smallest K-S statistic, and the second largest p-value. The largest log-likelihood of -55.902 is shared by the GSL, SN and ST distributions. Because of this, one can argue that the SN distribution is a competitor to the GSL distribution (or perhaps that the SN distribution is a better choice than the GSL distribution) since the former has one less parameter.

Figure 7 plots the fitted pdfs on top of the empirical histogram of the data. Figure 8 plots the fitted cdfs on top of the empirical cdf of the data. Both these figures support conclusions based on Tables 1 and 2. In both these figures, the fitted pdfs for the GSL, SN and ST distributions appear almost indistinguishable. Both figures suggest that the GSL distribution captures the tails of the data better than most other distributions.

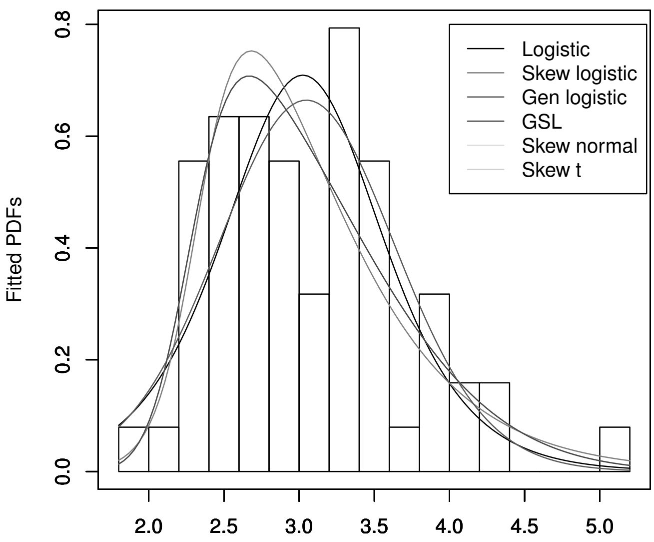

Figure 7: Histogram of the first data set and the fitted pdfs.

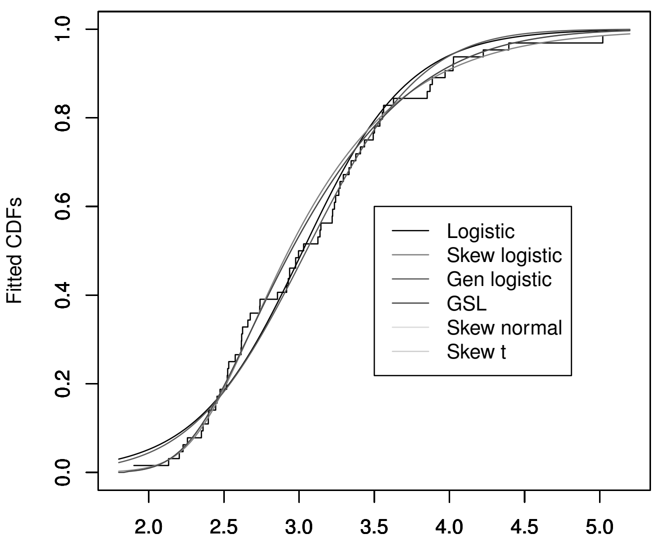

Figure 8: Empirical cdf of the first data set and the fitted cdfs.

Example 2. Here, we analyze the lean body mass of Australian athletes. The data given in Cook and Weisberg [7] are as follows:

```
63.32  58.55  55.36  57.18  53.2   53.77  60.17  48.33  54.57  53.42
68.53  61.85  48.32  66.24  57.92  56.52  54.78  56.31  62.96  56.68
62.39  63.05  56.05  53.65  65.45  64.62  60.05  56.48  41.54  52.78
52.72  61.29  59.59  61.7   62.46  53.14  47.09  53.44  48.78  56.05
56.45  53.11  54.41  55.97  51.62  58.27  57.28  57.3   54.18  42.96
54.46  57.2   54.38  57.58  61.46  53.46  54.11  55.35  55.39  52.23
59.33  61.63  63.39  60.22  55.73  48.57  51.99  51.17  57.54  68.86
63.04  63.03  66.85  59.89  72.98  45.23  55.06  46.96  53.54  47.57
54.63  46.31  49.13  53.71  53.11  46.12  53.41  51.48  53.2   56.58
56.01  46.52  51.75  42.15  48.76  41.93  42.95  38.3   34.36  39.03
```

We fitted all six distributions to the above data by the method of maximum likelihood. Table 3 presents the parameter estimates, the log likelihoods, the Kolmogorov–Smirnov statistics and respective p-values. The corresponding chi-squared statistics with observed and expected frequencies are reported in Table 4.

Table 3: MLEs, log-likelihoods, Kolmogorov–Smirnov statistics and corresponding p-values for Example 2.

Distribution	L(μ,σ)	SL(μ,σ,λ)	GL(μ,σ,α)	GSL($\mu,\sigma,\lambda,\alpha$)	SN(μ,σ,λ)	ST($\mu,\sigma,\lambda,\alpha$)
$\widehat{\mu}$	55.101	57.148	55.036	55.356	54.895	59.085
$\widehat{\sigma}$	3.807	3.990	0.593	0.671	6.887	7.233
$\widehat{\lambda}$	—	−0.389	—	−0.057	8.706×10^{-6}	−0.903
$\widehat{\alpha}$	—	—	0.117	0.133	—	9.924
Log-likelihood	−334.013	−333.557	−333.333	−333.265	−334.865	−333.738
KSS	0.072	0.071	0.070	0.069	0.080	0.072
p-value	0.658	0.712	0.712	0.715	0.642	0.711

Table 4: Observed and expected frequencies and chi-squared statistics for Example 2.

Intervals	Observed	L(μ,σ)	SL(μ,σ,λ)	GL(μ,σ,α)	GSL($\mu,\sigma,\lambda,\alpha$)	SN(μ,σ,λ)	ST($\mu,\sigma,\lambda,\alpha$)
< 42.084	5	3.17	3.85	3.97	4.35	3.14	3.93
42.084–49.808	16	16.77	16.69	14.22	14.47	19.86	16.95
49.808–57.532	48	45.51	44.28	50.67	50.18	41.90	43.57
57.532–65.256	25	28.06	29.66	24.35	24.78	28.47	30.15
> 65.256	6	6.49	5.52	6.79	6.23	6.62	5.40
		$\chi^2=1.5987$	$\chi^2=1.4574$	$\chi^2=0.7444$	$\chi^2=0.3633$	$\chi^2=3.2157$	$\chi^2=1.7422$

From Tables 3 and 4, we can see that the GSL distribution takes the largest log likelihood, the smallest chi-squared statistic, the smallest K-S statistic, and the largest p-value. The GL distribution takes the second largest log likelihood, the second smallest chi-squared statistic, the second smallest K-S statistic, and the second largest p-value. The SN distribution takes the smallest log likelihood, the largest chi-squared statistic, the largest K-S statistic, and the smallest p-value.

Figures 9 and 10 plot the fitted pdfs and fitted cdfs, respectively. Both these figures support conclusions based on Tables 3 and 4. Both figures suggest that the GSL distribution captures the middle part of the data better than most other distributions.

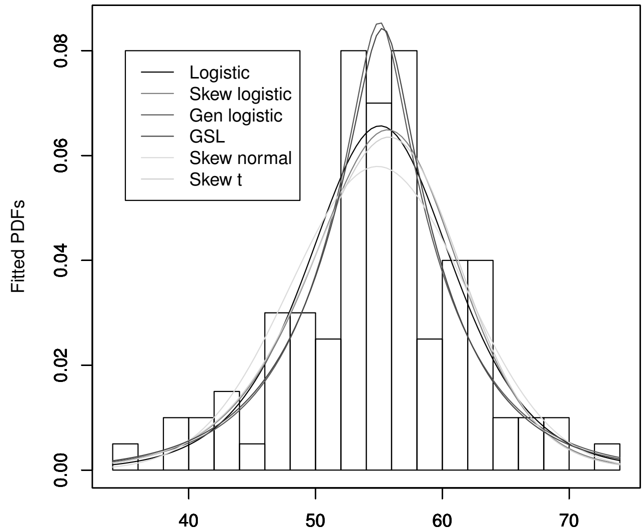

Figure 9: Histogram of the second data set and the fitted pdfs.

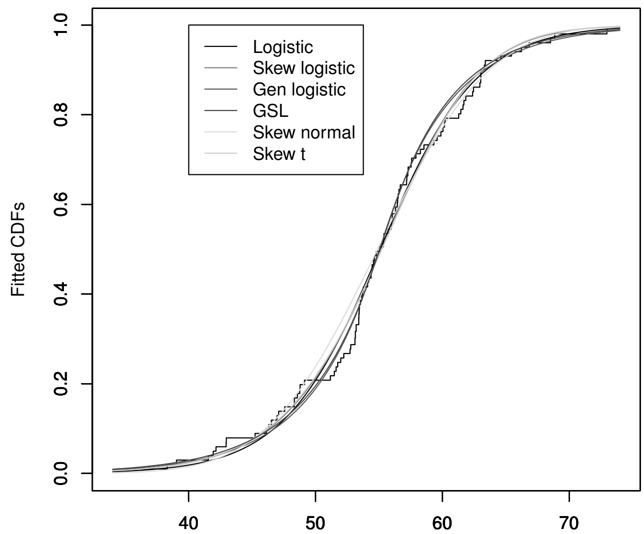

Figure 10: Empirical cdf of the second data set and the fitted cdfs.

ACKNOWLEDGMENTS

The authors would like to thank the Editor and the referee for careful reading and for their comments which greatly improved the paper.

REFERENCES

[1] Arellano-Valle, R.B. and Azzalini, A. (2013). The centred parameterization and related quantities of the skew-t distribution, *Journal of Multivariate Analysis*, **113**, 73–90.

[2] Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, **12**, 171–178.

[3] Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones, *Statistica*, **46**, 199–208.

[4] Azzalini, A. and Arellano-Valle, R.B. (2013). Maximum penalized likelihood estimation for skew-normal and skew-t distributions, *Journal of Statistical Planning and Inference*, **143**, 419–433.

[5] Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution, *Journal of the Royal Statistical Society, B*, **65**, 367–389.

[6] Badar, M.G. and Priest, A.M. (1982). *Statistical aspects of fiber and bundle strength in hybrid composites.* In "Progress in Science and Engineering Composites" (T. Hayashi, K. Kawata and S. Umekawa, Eds.), ICCM-IV, Tokyo, pp. 1129–1136.

[7] Cook, R.D. and Weisberg, S. (1994). *Regression Graphics*, John Wiley and Sons, New York.

[8] Gradshteyn, I.S. and Ryzhik, I.M. (2000). *Table of Integrals, Series and Products*, sixth edition, Academic Press, San Diego.

[9] Gupta, R.D. and Kundu, D. (2010). Generalized logistic distributions, *Journal of Applied Statistical Sciences*, **18**, 51–66.

[10] Johnson, N.L.; Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, volume 2, Wiley, New York.

[11] Koessler, W. and Kumar, N. (2010). An adaptive test for the two-sample scale problem based on U-statistics, *Communications in Statistics—Simulation and Computation*, **39**, 1785–1802.

[12] Nadarajah, S. (2009). The skew logistic distribution, *AStA Advances in Statistical Analysis*, **93**, 187–203.

[13] Raqab, M.Z. and Kundu, D. (2005). Comparison of different estimators of $P(Y < X)$ for a scaled Burr type X distribution, *Communications in Statistics – Simulation and Computation*, **34**, 465–483.

REVSTAT – Statistical Journal

Background

Statistical Institute of Portugal (INE, I.P.), 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.

— Three volumes are scheduled for publication, one in April, 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*, *Statistical Theory and Method Abstracts* and *Zentralblatt für Mathematik*.

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/inicio.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, I.P.
Av. António José de Almeida
1000-043 LISBOA
PORTUGAL

Copyright

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, the INE, I.P., in order to ensure the widest possible dissemination of information, namely through the Statistics Portugal'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.