

Víctor Leiva

"Chilean Journal of Statistics"

Ten years after its launch:

A message from the new Editor-in-Chief

1

Jhonnata B. de Carvalho, Murilo C. Silva, George F. von Borries,
André L.S. de Pinho and Ricardo F. von Borries

**A combined Fourier analysis and support vector machine
for EEG classification**

3

Luis Benites, Rocío Maehara, Victor H. Lachos and Heleno Bolfarine

**Linear regression models using finite mixtures of skew
heavy-tailed distributions**

21

Ednário Mendonça, Michelli Barros and Joelson Campos

**Goodness-of-fit test for the Birnbaum-Saunders distribution
based on the Kullback-Leibler information**

41

Nathalia L. Chaves, Caio L. N. Azevedo, Filidor Vilca-Labra
and Juvêncio S. Nobre

**A new Birnbaum-Saunders type distribution based on the
skew-normal model under a centered parameterization**

55

Gauss M. Cordeiro, M. Mansoor and Serge B. Provost

**The Harris extended Lindley distribution for modeling
hydrological data**

77

CHILEAN JOURNAL OF STATISTICS

Edited by Víctor Leiva

Volume 10 Number 1
April 2019

ISSN: 0718-7912 (print)
ISSN: 0718-7920 (online)

Published by the
Chilean Statistical Society

S O C H E 

SOCIEDAD CHILENA DE ESTADÍSTICA

AIMS

The Chilean Journal of Statistics (ChJS) is an official publication of the Chilean Statistical Society (www.soche.cl). The ChJS takes the place of *Revista de la Sociedad Chilena de Estadística*, which was published from 1984 to 2000.

The ChJS covers a broad range of topics in statistics, including research, survey and teaching articles, reviews, and material for statistical discussion. In particular, the ChJS considers timely articles organized into the following sections: Theory and methods, computation, simulation, applications and case studies, education and teaching, development, evaluation, review, and validation of statistical software and algorithms, review articles, letters to the editor.

The ChJS editorial board plans to publish one volume per year, with two issues in each volume. On some occasions, certain events or topics may be published in one or more special issues prepared by a guest editor.

EDITOR-IN-CHIEF

Víctor Leiva *Pontificia Universidad Católica de Valparaíso, Chile.*

EDITORS

Héctor Allende Cid	<i>Pontificia Universidad Católica de Valparaíso, Chile</i>
José M. Angulo	<i>Universidad de Granada, Spain</i>
Roberto G. Aykkroyd	<i>University of Leeds, UK</i>
Narayanaswamy Balakrishnan	<i>McMaster University, Canada</i>
Carmen Batanero	<i>Universidad de Granada, Spain</i>
Ionut Bebu	<i>The George Washington University, US</i>
Marcelo Bourguignon	<i>Universidade Federal do Rio Grande do Norte, Brazil</i>
Márcia Branco	<i>Universidade de São Paulo, Brazil</i>
Oscar Bustos	<i>Universidad Nacional de Córdoba, Argentina</i>
Luis M. Castro	<i>Pontificia Universidad Católica de Chile</i>
George Christakos	<i>San Diego State University, US</i>
Enrico Colosimo	<i>Universidade Federal de Minas Gerais, Brazil</i>
Gauss Cordeiro	<i>Universidade Federal de Pernambuco, Brazil</i>
Francisco Cribari-Neto	<i>Universidade Federal de Pernambuco, Brazil</i>
Francisco Cysneiros	<i>Universidade Federal de Pernambuco, Brazil</i>
Mario de Castro	<i>Universidade de São Paulo, São Carlos, Brazil</i>
José A. Díaz-García	<i>Universidad Autónoma Agraria Antonio Narro, Mexico</i>
Raul Fierro	<i>Universidad de Valparaíso, Chile</i>
Jorge Figueroa	<i>Universidad de Concepción, Chile</i>
Isabel Fraga	<i>Universidade de Lisboa, Portugal</i>
Manuel Galea	<i>Pontificia Universidad Católica de Chile</i>
Christian Genest	<i>McGill University, Canada</i>
Marc G. Genton	<i>King Abdullah University of Science and Technology, Saudi Arabia</i>
Patricia Giménez	<i>Universidad Nacional de Mar del Plata, Argentina</i>
Hector Gómez	<i>Universidad de Antofagasta, Chile</i>
Daniel Griffith	<i>University of Texas at Dallas, US</i>
Eduardo Gutiérrez-Peña	<i>Universidad Nacional Autónoma de Mexico</i>
Nikolai Kolev	<i>Universidade de São Paulo, Brazil</i>
Eduardo Lalla	<i>University of Twente, Netherlands</i>
Shuangzhe Liu	<i>University of Canberra, Australia</i>
Jesús López-Fidalgo	<i>Universidad de Navarra, Spain</i>
Liliana López-Kleine	<i>Universidad Nacional de Colombia</i>
Rosangela H. Loschi	<i>Universidade Federal de Minas Gerais, Brazil</i>
Carolina Marchant	<i>Universidad Católica del Maule, Chile</i>
Manuel Mendoza	<i>Instituto Tecnológico Autónomo de Mexico</i>
Orietta Nicolis	<i>Universidad Andrés Bello, Chile</i>
Felipe Osorio	<i>Universidad Técnica Federico Santa María, Chile</i>
Carlos D. Paulino	<i>Instituto Superior Técnico, Portugal</i>
Fernando Quintana	<i>Pontificia Universidad Católica de Chile</i>
Nalini Ravishanker	<i>University of Connecticut, US</i>
Fabrizio Ruggeri	<i>Consiglio Nazionale delle Ricerche, Italy</i>
José M. Sarabia	<i>Universidad de Cantabria, Spain</i>
Helton Saulo	<i>Universidade de Brasília, Brazil</i>
Pranab K. Sen	<i>University of North Carolina at Chapel Hill, US</i>
Julio Singer	<i>Universidade de São Paulo, Brazil</i>
Milan Stehlik	<i>Johannes Kepler University, Austria</i>
M. Dolores Ugarte	<i>Universidad Pública de Navarra, Spain</i>
Andrei Volodin	<i>University of Regina, Canada</i>

MANAGING EDITOR

Marcelo Rodríguez *Universidad Católica del Maule, Chile*

FOUNDING EDITOR

Guido del Pino *Pontificia Universidad Católica de Chile.*

Chilean Journal of Statistics

VOLUME 10, NUMBER 1

APRIL 2019

CONTENTS

Víctor Leiva <i>“Chilean Journal of Statistics” Ten years after its launch: A message from the new Editor-in-Chief</i>	1
Jhonnata B. de Carvalho, Murilo C. Silva, George F. von Borries, André L.S. de Pinho and Ricardo F. von Borries <i>A combined Fourier analysis and support vector machine for EEG classification</i>	3
Luis Benites, Rocío Maehara, Victor H. Lachos and Heleno Bolfarine <i>Linear regression models using finite mixtures of skew heavy-tailed distributions</i>	21
Ednário Mendonça, Michelli Barros and Joelson Campos <i>Goodness-of-fit test for the Birnbaum-Saunders distribution based on the Kullback-Leibler information</i>	41
Nathalia L. Chaves, Caio L. N. Azevedo, Filidor Vilca-Labra and Juvêncio S. Nobre <i>A new Birnbaum-Saunders type distribution based on the skew-normal model under a centered parameterization</i>	55
Gauss M. Cordeiro, M. Mansoor and Serge B. Provost <i>The Harris extended Lindley distribution for modeling hydrological data</i>	77

TENTH VOLUME – FIRST NUMBER
EDITORIAL PAPER

**“Chilean Journal of Statistics”
Ten years after its launch:
A message from the new Editor-in-Chief**

Welcome to the first issue of the tenth volume of the Chilean Journal of Statistics (ChJS). Today, April 29, 2019, the ChJS celebrates ten years of life and begins leaving its childhood, walking quickly to become a teenager. I remember perfectly well when the baby ChJS was born. I was present at that birth and I accompanied the baby during its first three years of life as its Executive Editor. The first volume of the ChJS had two issues, published in April and September 2010, which paid tribute to Dr. Pilar Iglesias, a beloved Chilean statistician. Pilar was the main motivation for the Chilean editorial board to launch this journal, which has as its ancestor the *Revista de la Sociedad Chilena de Estadística*, published in Spanish from 1984 to 2000. I would like to name its dearest uncles who helped the ChJS to survive. Among them are Marcia Branco and Rosangela Loschi from Brazil, Eduardo Gutiérrez-Peña and Manuel Mendoza from Mexico, Marc Genton from Switzerland, as well as Guido del Pino, Manuel Galea, Ronny Vallejos, and Reinaldo Arellano from Chile. I would like to take this opportunity to congratulate Reinaldo, who has honored the Chilean statistical community as the recent winner of the “Mahalanobis Prize 2019” awarded by the International Statistical Institute. Obviously, the ChJS would be nothing without the valuable contributions of renowned international researchers who have honored us by publishing their interesting works in our journal; all of these papers are available for free at <http://chjs.mat.utfsm.cl/issues.html>. We also thank all the anonymous reviewers who have contributed to keeping the top quality standards of the ChJS.

Although the ChJS is published by the Chilean Statistical Society (www.soche.cl) and belongs to the Chilean statistical community, our journal can be recognized as an international publication since its editorial board is composed of colleagues from practically the five continents. Our Editors are from Argentina, Australia, Austria, Bulgaria, Brazil, Canada, Chile, China, Colombia, Greece, India, Italy, Mexico, Netherlands, Peru, Portugal, Romania, Saudi Arabia, Spain, Switzerland, UK, and US. Our current Editorial Board, presented at <http://chjs.mat.utfsm.cl/board.html>, is a mixture of experienced editors and talented young researchers, the latter mainly from Chile and Brazil, who with great interest and enthusiasm have honored us by accepting to be part of the ChJS. They are having their first editorial experiences, although they all have extensive experience as researchers as well as reviewers for prestigious international journals.

I would also like to thank the members of the Directory of the Chilean Statistical Society (<https://soche.cl/quienes-somos>) headed by its President, Dr. Mauricio Castro, for the trust placed in me to be the new Editor-in-Chief of the ChJS. They can rest assured that, just as I did in the past as its Executive Editor, I will make my best effort to bring the ChJS to the highest standards of professionalism, impartiality and quality that all scientific journals must strive for.

In addition to this presentation note, the first issue of the tenth volume of the ChJS comprises five papers. Jhonnata B. de Carvalho, Murilo C. Silva, George F. von Borries, André L.S. de Pinho, and Ricardo F. von Borries, from Brazil and US, combined Fourier analysis and support vector machines to conduct an interesting work for classification of electroencephalograms, a relevant current theme related to data science. Luis Benites, Rocío Maehara, Víctor H. Lachos, and Heleno Bolfarine, from Peru, US and Brazil, proposed a regression model based on a finite mixture of skew heavy-tailed distributions, a widely studied topic by Brazilian and Chilean researchers within the context of statistical modeling. Ednário Mendonça, Michelli Barros, and Joelson Campos, from Brazil, derived goodness-of-fit tests based on the Kullback-Leibler information for the Birnbaum-Saunders model, a distribution which has had some of its more important developments in Chile and Brazil. Nathalia L. Chaves, Caio L.N. Azevedo, Filidor Vilca, and Juvêncio S. Nobre, from Brazil and Chile, introduced a new distribution to describe data with positive support and asymmetry by combining the Birnbaum-Saunders and centered skew-normal models, providing different statistical and mathematical features for this new model. Finally, our fifth paper is presented by Gauss M. Cordeiro, M. Mansoor, and Serge B. Provost, from Brazil, Pakistan and Canada, who derived their work in the setting of distribution theory, an area of wide development around the world, connecting the Harris and Lindley distributions to perform an interesting study which was applied to the modeling of hydrological data.

As a final comment, I would like the Chilean statistical community, as well as the international statistical community, our prestigious Editorial Board and past authors to champion ChJS as an emerging international journal and to encourage others to submit new works to the ChJS. Currently, we are indexed by several international systems, including the Institute for Scientific Information (ISI) Web of Science in the Emerging Sources Citation Index. The ChJS faces important challenges for the near future, such as reaching the Science Citation Index and looking for partnerships with prestigious publishers, societies and associations. However, just as with statistics itself, our success will depend on a team effort. Each one of us is important in meeting these challenges. We need you all.

Víctor Leiva
Editor-in-Chief
Chilean Journal of Statistics
<http://www.victorleiva.cl>

DATA SCIENCE
RESEARCH PAPER

A combined Fourier analysis and support vector machine for EEG classification

JHONNATA B. DE CARVALHO^{1,*}, MURILO C. SILVA², GEORGE F. VON BORRIES²,
ANDRÉ L.S. DE PINHO¹ and RICARDO F. VON BORRIES³

¹Department of Statistics, Universidade Federal do Rio Grande do Norte, Natal, Brazil,

²Department of Statistics, Universidade de Brasília, Brasília, Brazil,

³Department of Electrical and Computer Engineering, University of Texas at El Paso, El Paso, United States of America

(Received: 30 September 2018 · Accepted in final form: 10 March 2019)

Abstract

This paper introduces a method for the classification of electroencephalogram (EEG) data combining Fourier analysis, support vector machine (SVM) and a weighting system, called WFF-SVM, that provides high correct classification rates (accuracy) using a small training data set. Basically, an SVM classifier is calculated for each frequency in the periodogram and a proposed weighting system, based on the error rate of each SVM classifier, is used to obtain a final decision value. Also, it is shown that principal component analysis can be used to identify the best group of EEG channels to apply to the classification method, improving the correct classification rate. Two applications with real data are presented. The first application uses a public data set of epileptic patients and compares the proposed method with other methods presented in the literature. In this case, the correct classification rate obtained was 100%. The second application consists of EEG data collected from a subject submitted to 10 visual stimuli and the correct classification rate obtained was 95.31%. The classifier WFF-SVM combines multiple existing techniques, each one of them widely used in time series and dimensionality reduction problems. Our paper combines standard signal processing techniques to obtain high classification rates of EEG data.

Keywords: Epilepsy data · Periodogram · Principal components analysis · Simple moving averages · Supervised learning.

Mathematics Subject Classification: Primary 62H25 · Secondary 68Q32.

1. INTRODUCTION

Machine learning (ML) techniques have been gaining prominence due to real-world problems as well as large databases. Basically, one can divide ML methods into two classes, supervised learning and unsupervised learning. In unsupervised learning, the method has to recognize the groups by existing standards with a certain criterion. This type of learning tries to gain some understanding of the process that generated the data, e.g., the K-means method applied in DNA gene expression and Internet newsgroups (Ding and He, 2004),

*Corresponding author. Email: jhon_dbz@yahoo.com.br

clustering with hill-climbing optimization method applied to bee species (Friedman and Rubin, 1967), botanical data (Rubin, 1967) and in clustering of plants, wines and heart diseases (Souza et al., 2017). In supervised learning, groups (or classes) are known a priori and it is necessary to provide examples for method training. These methods are often used in classification and regression problems, e.g., logistic regression in the prediction of a financial crisis in Latin American companies (Giampaoli et al., 2016), in the fault diagnosis in chemical processes using Fisher discriminant analysis (Chiang et al., 2000), SVM classification in validation of cancer tissue samples (Furey et al., 2000). However, our interest is in the classification of electroencephalography signals.

An EEG are recordings of the electrical potentials produced by the brain (Bronzino, 1999; Buzsaki, 2006). Basically, the digital EEG is a time series containing information of the electrical activity generated by the brain. EEG has vast application in areas such as epilepsy detection (Andrzejak et al., 2001), emotion regulation using neurofeedback (Ruiz et al., 2014), affective neuroscience (Sitaram et al., 2011), and brain computer interface (Kübler et al., 2001; Wolpaw et al., 2002). For an efficient classification of EEG, an algorithm should address two main problems: feature extraction and classification method. Several methods have been used to extract features of EEG data, such as discrete wavelet transforms (DWT) (Jahankhani et al., 2006; Subasi, 2007; Subasi and Gursoy, 2010), amplitude values (Kaper et al., 2004), clustering techniques (Li and Wen, 2011), autoregressive and adaptive autoregressive parameters (Penny et al., 2000; Pfurtscheller et al., 1998), wavelet packet decomposition and extracted eigenvalues from the resultant wavelet coefficients using principal component analysis (PCA) (Acharya et al., 2012), continuous wavelet transform (CWT), higher order spectra (Acharya et al., 2013), approximate entropy and DWT (Ocak, 2009), analytic time-frequency flexible wavelet transform and fractal dimension (Sharma et al., 2017).

In order to classify a set of extracted features, several pattern recognition methods have been used, such as artificial neural network (Guo et al., 2009; Jahankhani et al., 2006; Nigam and Graupe, 2004; Subasi, 2007), mixture of expert model (Subasi, 2007), linear discriminant analysis (Subasi and Gursoy, 2010), SVM (Chandaka et al., 2009; Subasi and Gursoy, 2010), decision trees (Polat and Günes, 2007), least squares SVM (Li and Wen, 2011; Übeyli, 2010) and hidden markov models (Chiappa and Bengio, 2004). For a more complete review refer to Lotte et al. (2007).

Recently several algorithms have been developed to classify EEG in a variety of applications, such as in Zhang et al. (2016), which proposed a linear Bayesian discriminant with a Laplace prior, named sparse Bayesian method by exploiting a Laplace prior. A major advantage of this method is that it estimates automatically all the parameters of the classifier, without the need to use cross-validation. However, we point out that any Bayesian procedure needs a suitable prior distribution and although the Laplace distribution has been suggested it is conceivable that for a particular application a better prior distribution can be found. Wang et al. (2016) introduces a new approach that utilizes spatiotemporal feature extraction with multivariate linear regression (MLR) to learn discriminative of steady-state visual evoked potentials (SSVEP) features, for improving the detection accuracy. SSVEP are signals that are natural responses to visual stimulation at specific frequencies. MLR is implemented on dimensionality reduced EEG training data and a constructed label matrix to find optimally discriminative subspaces. Jiao et al. (2017) proposed a method that is an extension of multiset canonical correlation analysis (MsetCCA), called multilayer correlation maximization (MCM) model for further improving SSVEP recognition accuracy. MCM combines advantages of both Canonical Correlation Analysis and MsetCCA by carrying out three layers of correlation maximization processes. Zhang et al. (2018) introduced a new method, called multi-kernel extreme learning machine (MKELM) to EEG classification. Basically, this method transforms the EEG through the common

spatial pattern (CSP) and inserts a kernel function in the extreme learning machine (ELM). The MKELM provides a way to circumvent calculation of the hidden layer outputs and inherently encode it in a kernel matrix.

The proposed WFF-SVM is a classifier based on the SVM and the Fourier transform, providing the periodogram as feature extraction. In addition, it uses a weighting system based on the error rate. Thus, we call this classifier weighted Fourier frequencies and SVM, WFF-SVM for short. The WFF-SVM classifier differs from the other methods because it requires just one data transformation (Fourier), which leads to a good capacity to discriminate among groups. The PCA is used to identify the most active regions of the brain, providing the use of fewer electrodes and reducing the complexity of the data, since some electrodes pick up only noises, whereas the other methods ended up losing information by reducing the dimension based on the application of CSP or PCA. In relation to the Fourier transform, we observed that analyzing the signals in the frequency domain (periodogram), as shown in Figure 1, allows us to discriminate the signals for some frequencies. Our classifier takes into account the most distinct frequencies for classification through the weighting system. However, we point out that the choice of the kernel function is not unique, but for our applications the results are virtually the same by considering different kernels, suggesting a robust procedure.

Visual stimuli are commonly used to understand different components, such as color, texture, motion, objects, readability (text versus nontext), and others (Thomas and Vinod, 2017). Moreover, visual stimuli are also used in biometric authentication (Zuquete et al., 2010), emotion classification (Wang et al., 2014), person identification (Das et al., 2009), and others. We tested our classification method using real-world EEG data of two main applications: epilepsy and vision. The first application (described in Subsection 4.1) uses a publicly available data set described in Andrzejak et al. (2001), already used in previous works on EEG classification, and it allows a direct comparison of our classification method to other methods presented in the literature. In this application, the proposed method achieved a correct classification rate of 100.00% under a relatively simple model, showing that the proposed method performs well compared to other methods in the literature. The second application (described in Subsection 4.2) uses a data set collected in an experiment conducted at the University of Texas at El Paso in which the EEG data are acquired while the subject is submitted to visual stimuli. The proposed method showed a high correct classification rate of 95.31% using only three signals from each class in the training phase.

This paper is organized as follows. Section 2 provides a brief review of the SVM classifier relevant for our work and presents the periodogram, which is used for feature extraction. Section 3 presents our classification method integrating Fourier data analysis, SVM and a weighting system. Section 4 reports the performance of our method using real-world data of two applications and compares it with concurrent methods found in the literature. Section 5 provides some discussions, conclusions and recommendations for future work.

2. BACKGROUND

In this section, the methods used in the WFF-SVM classifier are described. The first method is the SVM and it includes three main blocks: the basic classifier, parameters estimation and SVM with nonlinear functions. The other methods are the Fourier analysis, periodogram, and the technique of simple moving averages.

2.1 SUPPORT VECTOR MACHINE

The SVM is a pattern recognition technique that has been widely used in problems like regression and classification (Hastie et al., 2008; Hornik et al., 2006; Theodoridis

and Koutroumbas, 2008; Vapnik, 1996). In classification problems the SVM technique separates two classes (say W_1 and W_{-1}) by a hyperplane $\langle \boldsymbol{\beta}, \mathbf{x} \rangle + \beta_0 = 0$, where $\langle \cdot, \cdot \rangle$ is the inner product, $\mathbf{x}, \boldsymbol{\beta} \in \mathbb{R}^D$ and $\beta_0 \in \mathbb{R}$, corresponding to the decision function

$$f(\mathbf{x}) = \text{sign}(\langle \boldsymbol{\beta}, \mathbf{x} \rangle + \beta_0). \quad (1)$$

The optimal hyperplane is defined as the one maximizing the margin of separation between classes. Note that the optimal hyperplane does not necessarily guarantee a complete separation of points from the two classes. This hyperplane can be constructed using Lagrange multipliers and then solving a constrained convex optimization problem.

Consider a set of training samples \mathbf{x}_i with $i = 1, 2, \dots, N$, then the primal optimization problem along with the soft margin method (Cortes and Vapnik, 1995) is given by

$$\begin{aligned} \min_{\boldsymbol{\beta}, \beta_0, \xi_i} \quad & \frac{1}{2} \|\boldsymbol{\beta}\|^2 + c \sum_{i=1}^N \xi_i, \\ \text{subject to} \quad & \begin{cases} y_i (\langle \boldsymbol{\beta}, \mathbf{x}_i \rangle + \beta_0) \geq 1 - \xi_i, \\ \xi_i \geq 0, \text{ for } i = 1, \dots, N, \end{cases} \end{aligned} \quad (2)$$

where the constant c is previously chosen and determines the influence of the two terms in the minimization problem. The variables ξ_i are known as slack variables measuring the proportional amount of predictions that fall on the wrong side of the margin, and y_i is an indicator variable defined by

$$y_i = \begin{cases} +1, & \text{if } \mathbf{x}_i \in W_1, \\ -1, & \text{if } \mathbf{x}_i \in W_{-1}. \end{cases}$$

Using Lagrange multipliers (Hastie et al., 2008), one can obtain the Wolfe dual function given by

$$L_D = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^N \alpha_i \alpha_k y_i y_k \langle \mathbf{x}_i, \mathbf{x}_k \rangle. \quad (3)$$

The solution is obtained by maximizing L_D , a simple convex optimization problem which must satisfy the conditions $0 \leq \alpha_i \leq c$ and $\sum_{i=1}^N \alpha_i y_i = 0$.

One can also generalize the SVM technique using a non-linear discriminant (unlike the hyperplane). In this case, a mapping is used in a larger number of dimensions. It can be shown (Theodoridis and Koutroumbas, 2008) that this mapping in a larger number of dimensions can be implemented without increasing the computational demand by replacing the inner product $\langle \mathbf{x}_i, \mathbf{x}_k \rangle$ in Equation (3) by a kernel $K(\mathbf{x}_i, \mathbf{x}_k)$ to compute the inner product in a higher dimensional space. In this study, we consider two popularly used kernels:

- Gaussian kernel: $K_1(\mathbf{x}_i, \mathbf{x}_j) = \exp \{-\sigma \|\mathbf{x}_i - \mathbf{x}_j\|^2\}$;
- Polynomial kernel: $K_2(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^d$;

where σ and d are kernel width and polynomial degree, respectively. Note when $d = 1$, the polynomial kernel is called linear kernel.

2.2 FOURIER ANALYSIS

Fourier frequency analysis is a very important tool in signal processing and the periodogram is one of its subproducts (Fuller, 1996). The periodogram shows how the covariance of a time series is distributed in frequency. Any stationary time series can be represented as a sum of sines and cosines (Fuller, 1996), that is, a discrete stationary time series $\{X_t\}$, where $t = 1, \dots, n$, (n being odd) can be represented by

$$X_t = \frac{a_0}{2} + \sum_{j=1}^{\lfloor n/2 \rfloor} a_j \cos(\omega_j t) + b_j \sin(\omega_j t),$$

where $\lfloor n/2 \rfloor$ is the largest integer less than or equal to $n/2$, a_k and b_k are parameters to be estimated. Also, the Fourier frequencies are defined by

$$\omega_k = \frac{2\pi k}{n}, \quad k = 0, \dots, \left\lfloor \frac{n}{2} \right\rfloor.$$

The periodogram can be defined as the sequence $\{J_k\}$, where

$$J_k = \frac{n}{2} (a_k^2 + b_k^2), \quad (4)$$

and the sum of squares removed by $\cos(\omega_k t)$ and $\sin(\omega_k t)$ is

$$J_k = \frac{2}{n} \left[\left(\sum_{t=1}^n X_t \cos(\omega_k t) \right)^2 + \left(\sum_{t=1}^n X_t \sin(\omega_k t) \right)^2 \right].$$

Thus, the value of the periodogram at frequency ω_k is the contribution from this frequency to the sum of squares of $\{X_t\}$ or, equivalently, its energy.

Some periodograms shown in this paper are smoothed using a moving average technique (Brockwell and Davis, 2002). Considering $\{J_k\}$ a sequence of points in the periodogram, for some $\alpha \in \mathbb{N}$, we define the smoothing by

$$J_k^\alpha = \frac{1}{\alpha} \sum_{j=1}^{\alpha} J_{k+j-1}, \quad k = 0, 1, \dots, \left\lfloor \frac{n}{2} \right\rfloor + 1 - \alpha, \quad (5)$$

where J_k^α is the average of α terms in sequence starting at the point J_k , meaning that each point J_k^α is the average contribution of α frequencies for the total energy of the series.

Let $\mathbf{X}_{i,1}$ and $\mathbf{X}_{i,2} \in \mathbb{R}^{P \times C}$ EEG samples of two classes from the i -trial with C and P being the number of channels and samples, respectively. The application of the Fourier transform will be in each column (channel) of $\mathbf{X}_{i,1}$ and $\mathbf{X}_{i,2}$ from the i -trial, building a vector

$$\mathbf{J}_{\ell,k}^\alpha = (J_{\ell,k_{i,g}}^\alpha)^\top, \quad (6)$$

with $i = 1, \dots, N_g$, N_g being the number of trials belonging to class g ($g = 1, 2$) and $\ell = 1, \dots, C$. These vectors together with the vector of labels $\mathbf{y} = (y_1, y_2, \dots, y_{N_1+N_2})^\top$ are the inputs of the classifier WFF-SVM.

3. NEW METHOD FOR EEG CLASSIFICATION

The classification of EEG data is a difficult task, with the analysis disturbed because most of the EEG channels may not be relevant to the classification at hand. Usually, traditional classification techniques alone do not provide good results when applied to EEG data. Therefore, it is important to construct a new method able to distinguish important brain regions and to capture the essential information contained in the data.

3.1 MOTIVATION

The Fourier analysis, especially the periodogram, can reveal hidden patterns in signals. Figure 1 has a set of 4 plots, all of which represent the signals generated by two different stimuli and captured by a channel of the EEG data (red for W_1 class and black for W_{-1} class) for a visual stimuli study (see Section 4.2). The top-left graph represents the superimposed plots of the original EEG signals. Note that it is difficult to visually distinguish two different classes in the time-domain plots presented in this graph. The top-right and bottom-left graphs represent the periodogram and the smoothed periodogram (J_k^4 in Equation (5)) of the signals, respectively. Now, it is easier to notice hidden patterns revealed by the periodograms of the data.

The plots indicate that the periodograms of W_1 have higher values at central frequencies than the periodograms of W_{-1} . In fact, the bottom-right graph in Figure 1 shows a possible discriminant (the dashed line) for these periodograms. Note that the periodograms of W_1 always have values above this hypothetical discrimination line for the central frequencies of the periodogram. However, it should be noted that this type of pattern does not occur for all the channels nor in all regions of the brain. It is necessary to use methods that identify both the relevant channels and the relevant frequencies in a set of periodograms, so that in an application, such as epilepsy detection of signals can be automatically classified into one of the expected classes.

3.2 CALCULATING THE DISCRIMINANT

The graphs in Figure 1 are revealing. It is easy to discriminate the periodograms for certain frequencies, but this separation is not so clear for other frequencies. It is noticeable that each frequency has its own importance and, therefore, could be evaluated individually and not as a whole. Thus, this paper describes a method in which a different discriminant is calculated for each frequency using the SVM classifier.

Considering the set of training $\mathbf{J}_{\ell,k}^\alpha$ of Equation (6) and the label vector \mathbf{y} with C channels, $\ell = 1, 2, \dots, C$ and a set of F frequencies, $k = 0, 1, \dots, F$ (k -th point of the smoothed periodogram and $F = \lfloor n/2 \rfloor$), define $\text{SVM}_{\ell,k}[j_{\ell,k}^\alpha]$ as the discriminant function generated by SVM, given by Equation (1), that classifies a new value $j_{\ell,k}^\alpha$ of the periodogram for a test signal into one of two classes, W_1 or W_{-1} , according to

$$\text{SVM}_{\ell,k}[j_{\ell,k}^\alpha] = \begin{cases} +1, & \text{if } j_{\ell,k}^\alpha \text{ is classified in } W_1, \\ -1, & \text{if } j_{\ell,k}^\alpha \text{ is classified in } W_{-1}. \end{cases} \quad (7)$$

Then, each discriminant will classify a new signal between two classes depending on whether the periodogram has higher or lower value at a particular frequency. Figure 2 shows an example of these discriminants. Note that each discriminant function $\text{SVM}_{\ell,k}[\cdot]$ could present a different decision. Thus, in order to unify these decisions, the next two sections present a weighting system that generates a single answer to the decision problem.

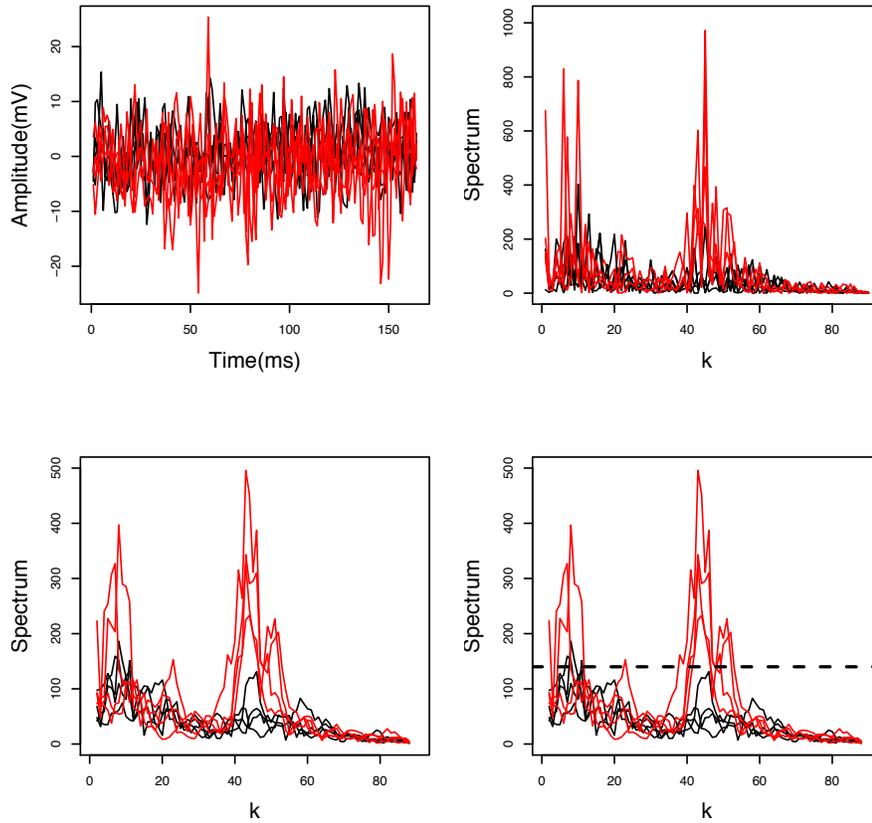


Figure 1. Representations of a set of signals generated by two stimuli. Each line is a signal from the W_1 class (red/lighter lines) or W_{-1} class (black/darker lines). Top-left: original signals. Top-right: periodogram of the signals. Bottom-left: smoothed periodogram of the signals. Bottom-right: smoothed periodogram of the signals with a possible naive discriminant (dashed line). These data are obtained at the Multi-Sensing-Processing and Learning Laboratory (MSPL) at the University of Texas at El Paso (UTEP).

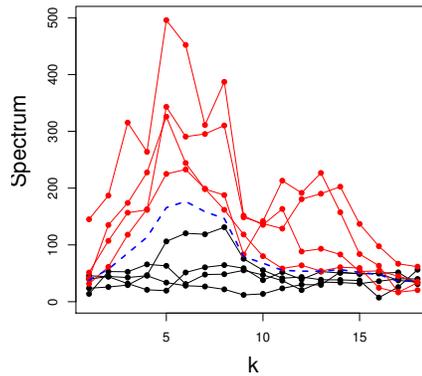


Figure 2. Some discriminating points (dashed line) for some Fourier frequencies ω_k for classes W_1 and W_{-1} . Red (lighter lines) represents class W_1 and black (darker lines) represents class W_{-1} .

3.3 WEIGHTING SYSTEM

Now, we have several discriminant functions, one for each EEG channel and each point in frequency, with discriminant functions producing different decisions. However, it is clear

that there are some discriminants more reliable than others and this reliability is determined by the incorrect classification rate (or error rate) on the training phase of the classification problem. For example, if for some channel ℓ and frequency k the discriminant function $\text{SVM}_{\ell,k}[\cdot]$ provides a low error rate on the training phase, then it is considered more reliable than another discriminant function with a higher error rate. Having this in mind, we introduce a weighting system based on the error rate for each discriminant.

The weight for channel ℓ and frequency k is defined as

$$\Psi_{\ell,k} = [1 - 2 \cdot \min(\text{Error Rate}, 0.5)]^{\hat{\rho}_{\ell,k}}, \quad (8)$$

where $\text{Error Rate} \in [0, 1]$ and $\hat{\rho}_{\ell,k} \geq 1$ is a constant given by

$$\hat{\rho}_{\ell,k} = \frac{\text{SS}_{\text{Total}}}{\text{SS}_{\text{Treatment}}}, \quad (9)$$

where $\text{SS}_{\text{Total}} = \sum_{i=1}^{n_c} \sum_{j=1}^{N_i} (J_{i,j}^\alpha - \bar{J})^2$ and $\text{SS}_{\text{Treatment}} = \sum_{i=1}^{n_c} \sum_{j=1}^{N_i} (\bar{J}_i - \bar{J})^2$, with n_c representing the number of classes (in this case we have $n_c = 2$), N_i is the number of frequencies of the smoothed periodogram of the i -th class, $J_{i,j}^\alpha$ is the j -th smoothed periodogram of the i -th class, \bar{J}_i is the arithmetic mean of the i -th class and \bar{J} is the mean of all smoothed periodograms. The basic concept of our truncated weighting system is to allocate 0 to the ones that have at least a 50% error rate, since $\min\{0, 0.5\} = 0$ implies zero weight. This is so because, based on our experience, it does not make sense to consider classifiers that provide over 50% error rate. On the other hand, the weighting system is an increasing function as the error rate tends to zero, achieving its maximum value when the error rate is zero. Finally, the power $\hat{\rho}_{\ell,k}$ is used to penalize the classifiers that have an error rate between 0 and 50%.

There are several advantages in the use of the exponent $\hat{\rho}_{\ell,k}$ in Equation (9) for the weighting system. It only involves sums, is easy to implement, does not involve optimization, has computational cost almost zero, it uses the data for calculation, it measures the distance between the groups taking into account the variability between and within the groups, and each frequency will have its own weight for SVM.

It is very important to use this kind of information to classify EEG data because much of the data contain non-relevant information of non-activated brain regions such as artifacts in EEG or noise. The next section will show how to use these weights to produce a single decision between one of the two classes W_1 or W_{-1} for new signals.

The implementation of the WFF-SVM method is presented in Algorithm 1. In Figure 3 we display a flowchart of the SVM framework that summarizes all the steps proposed. This classifier is denominated weighted Fourier and support vector machine (WFF-SVM).

Algorithm 1 Training WFF-SVM algorithm.

- 1:** Let $\mathbf{X}_{1,i} \in \mathbb{R}^{P \times C}$ and $\mathbf{X}_{2,i} \in \mathbb{R}^{P \times C}$ denote EEG samples of two classes recorded from the i -th trial. Choose the SVM kernel, the value of c and α smoothing parameter of Equation (5);
 - 2:** Apply the Fourier transform of Equation (4) in each column (channel) of $\mathbf{X}_{i,1}$ and $\mathbf{X}_{i,2}$ from the i -trial and use the moving average technique of Equation (5);
 - 3:** Use the SVM in the smoothed periodograms in step 2, totalizing $C \times F$ models;
 - 4:** Calculate the training error rate to each model in step 3 and the respective weight of Equation (8).
-

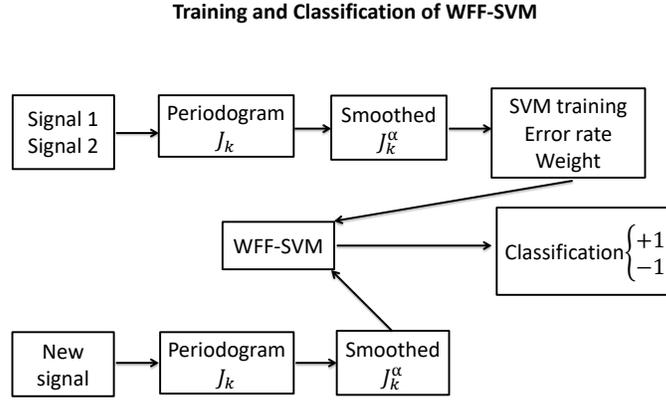


Figure 3. Flowchart for the training and classification phase of a new signal.

3.4 TEST PHASE FOR PRACTICAL APPLICATION

On the test phase for a practical application, we have a new set of signals (one signal per channel) to be classified as class W_1 or W_{-1} . This is done in two different ways (which will be compared later in this paper) using the discriminant function of Equation (7) associated with the weight of Equation (8).

The proposed classification method comprises the following main steps: first, consider a new stimulus $\mathbf{X} \in \mathbb{R}^{P \times C}$ and for each channel $\ell (\ell = 1, \dots, C)$ calculate the periodogram $\{J_{\ell,k}^\alpha\}$. Then, for each channel ℓ and frequency k of the periodogram use the discriminant function $\text{SVM}_{\ell,k}[J_{\ell,k}^\alpha]$ given by Equation (7) to obtain a particular decision (+1 or -1). Finally, using the weights, two decision methods are devised to classify the EEG signals.

In the first decision method, which we label as D_1 , each decision $\text{SVM}_{\ell,k}[J_{\ell,k}^\alpha]$ is weighted by $\Psi_{\ell,k}$ and each channel has its own decision weighting as in

$$D_1 = \text{sign} \left\{ \sum_{\ell=1}^C \text{sign} \left\{ \sum_{k=0}^F \Psi_{\ell,k} \times \text{SVM}_{\ell,k}[J_{\ell,k}^\alpha] \right\} \right\}. \quad (10)$$

In the second decision method, which we label as D_2 , each channel has its own decision weighting $\text{SVM}_{\ell,k}[J_{\ell,k}^\alpha]$ by $\Psi_{\ell,k}$, and the final decision is a pool between channels. Thus, we define

$$D_2 = \text{sign} \left\{ \frac{\sum_{\ell=1}^C \sum_{k=0}^F \text{SVM}_{\ell,k}[J_{\ell,k}^\alpha] \times \Psi_{\ell,k}}{\sum_{k=0}^F \sum_{\ell=1}^C \Psi_{\ell,k}} \right\}. \quad (11)$$

Basically, this decision system takes into account the performance of the channel in the training phase, because if there is a considerable disagreement regarding the classifiers in a given channel, the contribution of this channel to the final classification will not have a

great influence. Then, for both decision methods, we apply the criteria

$$\text{Decision} = \begin{cases} W_1, & \text{if } D_j = +1, \\ W_{-1}, & \text{if } D_j = -1, \\ \text{None}, & \text{if } D_j = 0, \end{cases} \quad (12)$$

for $j = 1, 2$. The implementation of the classification of a new signal is presented in Algorithm 2.

Algorithm 2 Classification of a new signal in WFF-SVM algorithm.

- 1:** Let $\mathbf{X}_{\text{new}} \in \mathbb{R}^{P \times C}$ denote EEG sample of a new recorded;
 - 2:** Apply the Fourier transform of Equation (4) in each column (channel) of \mathbf{X}_{new} and use the moving average technique of Equation (5);
 - 3:** Apply the $C \times F$ SVM models calculated by Algorithm 1 in the smoothed periodograms of step 2, totalizing $C \times F$ of values of Equation (7);
 - 4:** Use the $C \times F$ values calculated in the step 3 and use the decision weighting of Equations (10) or (11).
-

The following sections present two applications with real EEG data. First the proposed method is compared to other methods proposed in the literature, then we use it with a new data set.

4. APPLICATIONS AND RESULTS

This section presents two applications of our classification method. The first application uses a publicly available data set described in Andrzejak et al. (2001) which is used in several papers and is very useful to compare the proposed classification method with other methods. The second application uses a data set collected in an experiment conducted by the MSPL at UTEP. The classifier is implemented in the R software and to have access to the respective code, visit <https://carvalhomyssearches.weebly.com>; see R (2018).

4.1 EPILEPSY DATA CLASSIFICATION

The epilepsy data consists of five distinct sets each containing 100 single-channel EEG segments (Andrzejak et al., 2001). Two of these sets, denoted A and B, are obtained from EEG recordings from five healthy volunteers in an awake state with eyes open and eyes closed, respectively. Sets C, D, and E originated from an EEG archive of pre-surgical diagnosis. Segments in set D are recorded from within the epileptogenic zone, and those in set C from the hippocampal formation of the opposite hemisphere of the brain. While sets C and D contained only activity measured during seizure free intervals, set E only contained seizure activity (for more details about these data sets see Andrzejak et al. (2001)). As in previous studies (Nigam and Graupe, 2004; Subasi, 2007; Subasi and Gursoy, 2010), we used only two datasets (A and E) to test the classifier.

Both sets A and E have 100 signals each, one signal for each channel and each signal corresponding to 4097 samples. To perform the classification it is cut out the beginning and the end of the signals and subsampled them into 20 signals (components) of 200 samples each. Then, for each set A and E, we randomly selected 10 of the corresponding 20 signals to use in the training phase. In the test phase we repeated this same subsampling process to all the signals in both sets A and E. Thus, it is generated 2000 signals to use in the test phase.

Many authors also proposed methods for the classification of EEG data using data sets A and E to test their classifiers. Table 1 has a summary of the overall results and also the result with the application of the proposed method, named WFF-SVM. In WFF-SVM is used the linear kernel, $c = 1$, $\alpha = 5$ and D_2 as described in Equations (2), (5) and (11), respectively.

According to Zhang et al. (2018), the MKELM is more efficient than the following methods: multilayer perceptron with a single hidden layer; the conventional SVM; SVM with Gaussian and polynomial kernel; multi-kernel SVM using both Gaussian and polynomial kernels; the conventional ELM; ELM with Gaussian kernel; ELM with polynomial kernel, and finally, the multi-kernel ELM using both Gaussian and polynomial kernels. Therefore, we also considered in the comparison the new classifier proposed by Zhang et al. (2018), called MKELM, in both applications.

Table 1. Comparison of results for epilepsy data.

Reference	% Accuracy	Method
Subasi (2007)	94.50	ME
	93.20	MLPNN
	98.75	DWT, PCA and SVM
Subasi and Gursoy (2010)	99.50	DWT, ICA and SVM
	100.00	DWT, LDA and SVM
Jahankhani et al. (2006)	98.00	NN
Guo et al. (2009)	95.00	RWE and NN
Nigam and Graupe (2004)	97.20	NN
Polat and Günes (2007)	98.72	TRF
Li and Wen (2011)	99.90	LS-SVM
Chandaka et al. (2009)	95.96	SVM
Übeyli (2010)	99.56	LS-SVM
Zhang et al. (2018)	100.00	MKELM
Proposed method	100.00	WFF-SVM

where ME is mixed of experts; MLPNN is multi-layer perceptron neural network; DWT is discrete wavelet transform; LDA is linear discriminant analysis; ICA is independent component analysis; NN is neural networks; RWE is relative wavelet energy; LS-SVM is least square support vector machine; MKELM is multi-kernel extreme learning machine using both Gaussian and polynomial kernels with CSP feature.

Note that the proposed method is as efficient as (or more efficient than) the other methods. A possible reason for this improvement is the weighting system capturing the most important regions for classification, strengthening the process.

Despite the greater efficiency of the proposed method, it can be noted that all methods are very efficient for this problem. The main reason for this result is that it is relatively easy to classify the epilepsy data; in fact, neurologists can visually distinguish the EEG patterns of epileptic patients and non-epileptics patients. For this reason, the following example presents a more complex application that uses EEG data collected in an experiment based on visual stimuli with a set of tasks to classify.

4.2 CLASSIFICATION OF VISUAL STIMULI

In the visual stimuli application, the objective is to calculate the discriminant function so that, given a new visual stimulus event, our classification method is capable of identifying the slide presented to the subject from the EEG data recordings only. To do this, the proposed method is used after a selection of activated channels using PCA.

Experimental Design The data set used in this application is acquired at the MSPL at UTEP. The EEG data are recorded from a volunteer test subject using a Biosemi EEG acquisition system with 128 channels. The acquisition system recorded EEG signals

corresponding to 10 different visual stimuli, each one presented multiple times in random order and during a regular interval of time. The visual stimuli used correspond to the slides shown in Figure 4. Each stimulus is shown on a computer monitor screen 4 times (in random order) with a five seconds break between each slide, corresponding to a blank screen. An audible tone alerted the subject each time a new slide is about to be displayed. Thus, the EEG data set of the second experiment comprised of 4 EEG signals for each one of the 10 visual stimuli, acquired by 128 channels.

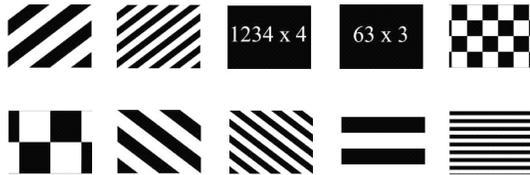


Figure 4. Ten visual stimuli shown to the test subject during EEG signal acquisition.

Using PCA for source localization The PCA is used to explain the variance-covariance structure of a set of variables by a smaller set of variables formed by linear combinations of the original ones (Johnson and Wichern, 2007). Generally, in databases that contain strongly correlated variables (as in EEG data) the PCA is very useful to reduce the dimensionality of the problem. In PCA, the first principal component is the linear combination with the highest possible variance. This means, in the case of EEG data, that the most important channels for the composition of the first principal component are the channels that capture signals with higher variance (the channels corresponding to the activated brain regions) as described in von Borries et al. (2013). Figure 5 shows contours obtained for the first principal component when PCA is applied to EEG signals from 128 channels of the visual stimuli experiment. One can observe that most of the variability in this experiment is present in the channels located on the brain’s frontal lobe. The next sections show that, in fact, this region is the most important for classification and the other regions basically do not bring relevant information to the classification problem at hand. Actually, our results show that the correct classification rate increases when the signals from those regions are not included in classification.

Data analysis First, we train the classifier. Since the proposed method is a binary classifier and we have 10 apparently different visual stimuli, the classification process is implemented sequentially by pairs of visual stimuli. Moreover, as many images are very similar, the classification is performed only with abstract images against images with arithmetic operations, making a total of 16 discriminants (or 16 pairs). Cross-validation is used to approximate the correct classification rate of this method, as follows: for each pair of images analyzed, the first repetition of each image (independent of the others) is excluded in the training phase to be used in the testing phase. Then, the second repetition of each image (independent of the others) is excluded in the training phase to be used in the testing phase, and so on. Thus, $4 \times 16 \times 2 = 128$ signals are used in the test phase. Note that the signals used in the test phase are not used to build the discriminant, resulting in a reliable analysis. The first test is done using the periodogram with the configurations $\alpha = 1$ and 4, linear kernel and using $c = 1$. Note in Table 2, the classification rates for each configuration. There is an increase of around 10% for all configurations when the smoothed periodogram ($\alpha = 4$) is used, indicating that smoothing is a good option to improve the classification rate. Furthermore, D_1 method is better than the D_2 , but not having a very large difference between the rates. Figure 6 shows a contour plot of the accuracy of each brain region. It should be noted that the EEG signals located at the brain’s frontal lobe

had the best correct classification rates. The similarity between Figures 5 and 6 is remarkable, indicating that the regions identified using PCA actually correspond to the regions of higher correct classification rates. Therefore, one might think that the non-activated regions contain non-relevant information that actually disturbs the classification. Thus, the cross-validation process is repeated using 53 channels with the highest hit rates, where most are from the front of the brain, with parameters $c = 1, 10, 100$. The results presented in Table 3 indicate that the correct classification rates increase when using the smoothed periodograms and specially when selecting only the most relevant channels. Therefore, it appears to be extremely important, in a classification analysis of EEG data, to remove from the analysis the channels that appear basically to capture non-relevant information. However, the cost value does not seem to influence much on the results and the classification rates are very similar for all values of c , so, for the analyzes that will be done from now on, will be used $c = 1$.

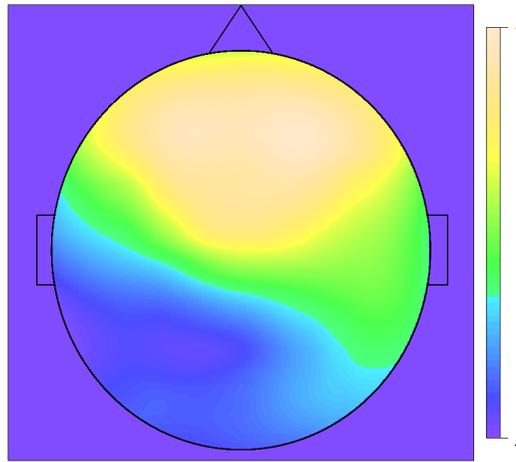


Figure 5. Variability of signals through the Brain. Contours for the first principal component when PCA is applied to EEG signals from 128 channels of the visual stimuli experiment. The front of the brain presents most of the signal variability.

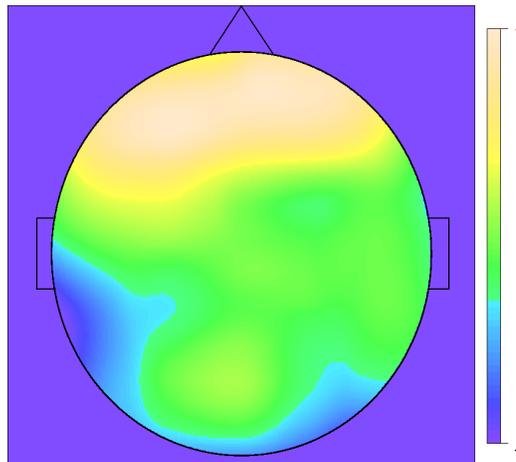


Figure 6. Contour lines for the correct classification rates by channel: new method with the smoothed periodogram, $\alpha = 4$ and $c = 1$.

Table 2. Accuracy for some settings of smoothing parameter in the WFF-SVM algorithm.

Classifier	α	Method	% accuracy
WFF-SVM	1	D_1	75.00
		D_2	73.44
	4	D_1	87.50
		D_2	84.37

Table 3. Results using $\alpha = 4$ for the accuracy using some values of cost (c), number of channels and type of decision.

Cost	Channel	Method	% accuracy
$c = 1$	128	D_1	87.50
		D_2	84.37
	53	D_1	92.97
		D_2	92.97
$c = 10$	128	D_1	85.94
		D_2	85.16
	53	D_1	92.97
		D_2	92.97
$c = 100$	128	D_1	85.95
		D_2	85.16
	53	D_1	92.97
		D_2	92.97

After some α variations, we obtained a classification rate of 95.31% with $c = 1$, $\alpha = 5$, using D_2 with 53 channels, and 73.44% to MKELM using all the channels with CSP feature. These are the best results found in this study. The non-requirement of an extensive training data set constitutes an important characteristic of the proposed classification method since in real-world applications the collection of signals available to train the classifier can be limited to only a few cases.

5. DISCUSSION AND CONCLUSIONS

EEG technique is employed to help in a variety of diagnosis, such as posttraumatic stress, human emotions and epilepsy. Regarding the latter one, there is a special interest to detect as early as possible epilepsy in order to initiate the proper treatment and mitigate this neurological disorder effects. Several studies were conducted with this objective, such as [Fergus et al. \(2015\)](#) who uses machine learning, whereas [Thodoroff et al. \(2016\)](#) and [Acharya et al. \(2018\)](#) have used the deep learning (DL) approach. The DL method has been used in several problems as in image recognition ([Krizhevsky et al., 2012](#)), diagnosis of Alzheimer’s disease ([Ortiz et al., 2016](#)), prediction of sale prices of real estate units ([Rafiei and Adeli, 2015](#)) and in the estimation of concrete compressive strength [Rafiei et al. \(2017\)](#). There are examples in the literature that use SVM and DL, such as in [Tang \(2013\)](#), who developed an approach in DL replacing the softmax layer by a linear SVM. [Erfani et al. \(2016\)](#) used a hybrid model where an unsupervised deep belief networks is trained to extract generic underlying features, and one class SVM is trained from the features learned by the deep belief networks. Therefore, these works show that the use of SVM in DL is not new and suggests that in future works WFF-SVM in DL can also be contemplated in order to search for more efficient methods. The WFF-SVM can be used in any type of signal, EEG, electrocardiogram, electromyogram, etc. In order to accomplish

that, it is sufficient to represent the data as a time series or in a certain proper order. This proposed paper is based on a broader study found in [Carvalho \(2016\)](#), in which electromyogram data were also considered. Furthermore, this classifier can be used in clinical application or any other application. Regarding the computational intensive aspect, with the rapidly increasing performance of new computers, including parallel programming and the promising quantum programming the tendency is to be feasible. The application using epilepsy data showed that the proposed method has no better competitor among other methods presented in the literature. This paper presents a second and more complete application. This application using EEG data captured during an experiment involving visual stimuli showed a number of specific features for the classification of EEG data. In particular, this application showed that the brain region identified using PCA was similar to the region where the channels had the best individual correct classification rates. In fact, the correct classification rate increased significantly by discarding the EEG channels that had non-relevant information. The proposed method of using smoothed periodograms and assigning weights to the channels based on their individual error rates resulted in higher correct classification rates than other methods reported in the literature. It should be noted that the proposed method showed a high correct classification rate of 95.31% using only three signals from each class in the training phase. Thus, a topic for future research is to extend the WFF-SVM to accept more than two groups for training and classification. In addition, it would be important to propose some sort of threshold for decision-making, in guiding the decision Equation (12) on how far it must be from zero to have a more objective classification.

This paper presented a new method for classification of EEG data that uses Fourier analysis and SVM. The proposed method employs a specific SVM decision value for each frequency of the periodogram. In addition, a simple weighting system based on the performance of the classifier, obtained in the training phase, is applied to the classification phase. We used two data sets to test the performance of the proposed classifier. The first data set referred to EEG of an epilepsy study and the second to EEG of a visual stimulation study. Finally, one point for improvement include the extension of our classification method to more than two classes and to expand the performance comparison with other methods.

ACKNOWLEDGMENT

The authors thank and acknowledge financial support from, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Universidade Federal do Rio Grande do Norte (UFRN), Universidade de Brasília (UnB) and University of Texas at El Paso (UTEP). We also thank the Department of Statistics at UFRN for inviting the fourth author to a visit that generate a partnership that led to this research. Furthermore, we thank the anonymous reviewers for their careful reading of our manuscript and their many insightful comments and suggestions.

REFERENCES

- Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., and Adeli, H., 2018. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Computers in Biology and Medicine*, 100, 270-278.
- Acharya, U.R., Sree, S.V., Alvin, A.P.C., and Suri, J.S., 2012. Use of principal component analysis for automatic classification of epileptic EEG activities in wavelet framework. *Expert Systems with Applications*, 39, 9072-9078.
- Acharya, U.R., Yanti, R., Zheng, J.W., Krishnan, M.M.R., Tan, J.H., Martis, R.J., and

- Lim, C.M., 2013. Automated diagnosis of epilepsy using CWT, HOS and texture parameters. *International Journal of Neural Systems*, 23, 1350009.
- Andrzejak, R.G., Lehnertz, K., Mormann, F., Rieke, C., David, P., and Elger, C.E., 2001. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, 64, 061907.
- Brockwell, P.J. and Davis, R.A., 2002. *Introduction to Time Series and Forecasting*. Springer, New York.
- Bronzino, J.D., 1999. *Biomedical engineering handbook*. CRC press, Boca Raton, US.
- Buzsaki, G., 2006. *Rhythms of the Brain*. Oxford University Press, Oxford, UK.
- Carvalho, J.B., 2016. *Classificador Máquina de Suporte Vetorial com Análise de Fourier Aplicada em Dados de EEG e EMG*. Master's thesis. Federal University of Rio Grande do Norte, available at <https://repositorio.ufrn.br/jspui/handle/123456789/20964>.
- Chandaka, S., Chatterjee, A., and Munshi, S., 2009. Cross-correlation aided support vector machine classifier for classification of EEG signals. *Expert Systems with Applications*, 36, 1329-1336.
- Chiang, L.H., Russell, E.L., and Braatz, R.D., 2000. Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 50, 243-252.
- Chiappa, S., and Bengio, S., 2004. HMM and IOHMM modeling of EEG rhythms for asynchronous BCI systems. *IDIAP*.
- Cortes, C., and Vapnik, V., 1995. Support-vector networks. *Machine Learning*, 20, 273-297.
- Das, K., Zhang, S., Giesbrecht, B., and Eckstein, M.P., 2009. Using rapid visually evoked EEG activity for person identification. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 2490-2493). IEEE.
- Ding, C. and He, X., 2004. K-means clustering via principal component analysis. In *Proceedings of the Twenty-first International Conference on Machine Learning* (p. 29). ACM.
- Erfani, S.M., Rajasegarar, S., Karunasekera, S., and Leckie, C., 2016. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. *Pattern Recognition*, 58, 121-134.
- Fergus, P., Hussain, A., Hignett, D., Al-Jumeily, D., Abdel-Aziz, K., and Hamdan, H., 2015. A machine learning system for automated whole-brain seizure detection. *Applied Computing and Informatics*, 12, 70-89.
- Friedman, H.P., and Rubin, J., 1967. On some invariant criteria for grouping data. *Journal of the American Statistical Association*, 62, 1159-1178.
- Fuller, W.A., 1996. *Introduction to Statistical Time Series*. Wiley, New York
- Furey, T.S., Cristianini, N., Duffy, N., Bednarski, D.W., Schummer, M., and Haussler, D., 2000. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 16, 906-914.
- Giampaoli, V., Tamura, K.A., Caro, N.P., and de Araujo, L.J.S., 2016. Prediction of a financial crisis in Latin American companies using the mixed logistic regression model. *Chilean Journal of Statistics*, 7, 31-41.
- Guo, L., Rivero, D., Seoane, J.A., and Pazos, A., 2009. Classification of EEG signals using relative wavelet energy and artificial neural networks. In *Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation* (pp. 177-184). ACM.
- Hastie, T., Tibshirani, R., and Friedman, J., 2008. *The Elements of Statistical Learning*. Springer, New York.
- Hornik, K., Meyer, D., and Karatzoglou, A., 2006. Support vector machines in R. *Journal of statistical software*, 15, 1-28.

- Jahankhani, P., Kodogiannis, V., and Revett, K., 2006. EEG signal classification using wavelet feature extraction and neural networks. In *Modern Computing, 2006. JVA'06. IEEE John Vincent Atanasoff 2006 International Symposium on* (pp. 120-124). IEEE.
- Jiao, Y., Zhang, Y., Wang, Y., Wang, B., Jin, J., and Wang, X., 2018. A novel multi-layer correlation maximization model for improving CCA-based frequency recognition in SSVEP braincomputer interface. *International Journal of Neural Systems*, 28, 1750039.
- Johnson, R.A. and Wichern D.W., 2007. *Applied Multivariate Statistical Analysis*. Prentice-Hall, New Jersey.
- Kaper, M., Meinicke, P., Grossekhoefer, U., Lingner, T., and Ritter, H., 2004. BCI competition 2003-data set IIb: support vector machines for the P300 speller paradigm. *IEEE Transactions on Biomedical Engineering*, 51, 1073-1076.
- Krizhevsky, A., Sutskever, I., and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (pp. 1097-1105).
- Kübler, A., Kotchoubey, B., Kaiser, J., Wolpaw, J.R., and Birbaumer, N., 2001. Brain-computer communication: Unlocking the locked in. *Psychological Bulletin*, 127, 358.
- Li, Y. and Wen, P.P., 2011. Clustering technique-based least square support vector machine for EEG signal classification. *Computer Methods and Programs in Biomedicine*, 104, 358-372.
- Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., and Arnaldi, B., 2007. A review of classification algorithms for EEG-based braincomputer interfaces. *Journal of Neural Engineering*, 4, R1.
- Nigam, V.P., and Graupe, D., 2004. A neural-network-based detection of epilepsy. *Neurological Research*, 26, 55-60.
- Ocak, H., 2009. Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy. *Expert Systems with Applications*, 36, 2027-2036.
- Ortiz, A., Munilla, J., Gorriz, J.M., and Ramirez, J., 2016. Ensembles of deep learning architectures for the early diagnosis of the Alzheimers disease. *International Journal of Neural Systems*, 26, 1650025.
- Penny, W.D., Roberts, S.J., Curran, E.A., and Stokes, M.J., 2000. EEG-based communication: a pattern recognition approach. *IEEE Transactions on Rehabilitation Engineering*, 8, 214-215.
- Pfurtscheller G., Neuper C., Schlogl A., and Lugger K., 1998. Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters. *IEEE Transactions on Rehabilitation Engineering*, 6, 316-325.
- Polat, K. and Günes, S., 2007. Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. *Applied Mathematics and Computation*, 187, 1017-1026.
- R Core Team, 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rafiei, M.H. and Adeli, H., 2015. A novel machine learning model for estimation of sale prices of real estate units. *Journal of Construction Engineering and Management*, 142, 04015066.
- Rafiei, M.H., Khushefati, W.H., Demirboga, R., and Adeli, H., 2017. Supervised deep restricted Boltzmann machine for estimation of concrete. *ACI Materials Journal*, 114, 237-244.
- Rubin, J., 1967. Optimal classification into groups: an approach for solving the taxonomy problem. *Journal of Theoretical Biology*, 15, 103-144.
- Ruiz, S., Buyukturkoglu, K., Rana, M., Birbaumer, N., and Sitaram, R., 2014. Real-time fMRI brain computer interfaces: self-regulation of single brain regions to networks. *Biological Psychology*, 95, 4-20.

- Sharma, M., Pachori, R.B., and Acharya, U.R., 2017. A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension. *Pattern Recognition Letters*, 94, 172-179.
- Sitaram, R., Lee, S., Ruiz, S., Rana, M., Veit, R., and Birbaumer, N., 2011. Real-time support vector classification and feedback of multiple emotional brain states. *Neuroimage*, 56, 753-765.
- Souza, M.P.S., Filho, T.M.S., Amaral, G.J.A., and Souza, R.M.C.R., 2019. Investigating different fitness criteria for swarm-based clustering. *International Journal of Business Intelligence and Data Mining*, pages in press, available at <http://dx.doi.org/10.1504/IJBIDM.2017.10007517>.
- Subasi, A., 2007. EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications*, 32, 1084-1093.
- Subasi, A. and Gursoy, M.I., 2010. EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Systems with Applications*, 37, 8659-8666.
- Tang, Y., 2013. Deep learning using linear support vector machines. Working paper, available at <https://arxiv.org/abs/1306.0239>.
- Theodoridis, S. and Koutroumbas, K., 2008. *Pattern Recognition*. Academic Press, New York.
- Thodoroff, P., Pineau, J., and Lim, A., 2016. Learning robust features using deep learning for automatic seizure detection. In *Machine Learning for Healthcare Conference* (pp. 178-190).
- Thomas, K.P. and Vinod, A.P., 2017. Toward EEG-based biometric systems: the great potential of brain-wave-based biometrics. *IEEE Systems, Man, and Cybernetics Magazine*, 3, 6-15.
- Übeyli, E.D., 2010. Least squares support vector machine employing model-based methods coefficients for analysis of EEG signals. *Expert Systems with Applications*, 37, 233-239.
- Vapnik, V., 1996. *Statistical Learning Theory*. Wiley, New York.
- von Borries, G.F., Rocha, L.C.S., Coutinho, M.S., and von Borries, R.F., 2013. Identification of brain-generated activity with EEG using visual stimuli. *Revista Brasileira de Biometria*, 31, 28-42.
- Wang, H., Zhang, Y., Waytowich, N.R., Krusienski, D.J., Zhou, G., Jin, J., Wang, X., and Cichocki, A., 2016. Discriminative feature extraction via multivariate linear regression for SSVEP-based BCI. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24, 532-541.
- Wang, X.W., Nie, D., and Lu, B.L., 2014. Emotional state classification from EEG data using machine learning approach. *Neurocomputing*, 129, 94-106.
- Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., and Vaughan, T.M., 2002. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113, 767-791.
- Zhang, Y., Wang, Y., Zhou, G., Jin, J., Wang, B., Wang, X., and Cichocki, A., 2018. Multi-kernel extreme learning machine for EEG classification in brain-computer interfaces. *Expert Systems with Applications*, 96, 302-310.
- Zhang, Y., Zhou, G., Jin, J., Zhao, Q., Wang, X., and Cichocki, A., 2016. Sparse Bayesian classification of EEG for braincomputer interface. *IEEE Transactions on Neural Networks and Learning Systems*, 27, 2256-2267.
- Zuquete, A., Quintela, B., and Cunha, J.P.S., 2010. Biometric authentication using electroencephalograms: a practical study using visual evoked potentials. *Electrónica e Telecomunicações*, 5, 185-194.

STATISTICAL MODELING
RESEARCH PAPER

Linear regression models using finite mixtures of skew heavy-tailed distributions

LUIS BENITES^{1,*}, ROCÍO MAEHARA², VÍCTOR H. LACHOS³, and HELENO BOLFARINE⁴

¹Departamento de Ciencias, Pontificia Universidad Católica del Perú, Perú,

²Departamento de Ingeniería, Universidad del Pacífico, Lima, Perú,

³Department of Statistics, University of Connecticut, Storrs, Connecticut, USA,

⁴Departamento de Estatística, Universidade de São Paulo, Brazil

(Received: 15 March 2019 · Accepted in final form: 15 April 2019)

Abstract

In this paper, we propose a regression model based on the assumption that the error term follows a mixture of normal distributions. Specifically, we consider a finite scale mixture of skew-normal distributions, a rich family that contains the skew-normal, skew-t, skew-slash and skew-contaminated normal distributions as members. This model allows us to describe data with high flexibility, simultaneously accommodating multimodality, skewness and heavy tails. We develop a simple EM-type algorithm to perform maximum likelihood inference of the parameters of the proposed model with closed-form expressions for both E- and M-steps. Furthermore, the observed information matrix is derived analytically to account for the corresponding standard errors and a bootstrap procedure is implemented to test the number of components in the mixture. The practical utility of the new model is illustrated with a real dataset and several simulation studies. The proposed algorithm and methods are implemented in an R package named `FMsmnReg`.

Keywords: ECME algorithm · Mixture model · Non-normal error distribution · Scale mixtures of skew-normal distributions

Mathematics Subject Classification: Primary 62J05 · Secondary 62J99

1. BIBLIOGRAPHICAL REVIEW AND MOTIVATING EXAMPLE

1.1 INTRODUCTION

A basic assumption of the linear regression (LR) model is that the error term follows a normal distribution. However, it is well known that data from some phenomena do not always satisfy this assumption, instead having a distribution with heavy tails, skewness or multimodality. Many extensions of this classic model have been proposed to broaden the applicability of Gaussian linear regression (N-LR) analysis to situations where the Gaussian error term assumption may be inadequate, such as, the use of the Student-t distribution (Lange et al., 1989), which is appropriate for datasets involving errors with

*Corresponding author. Email: lbenitess@pucp.edu.pe

longer than normal tails. Other extensions include the use of the symmetrical class of scale mixtures of normal (SMN) distributions (Andrews and Mallows, 1974; Lange and Sinsheimer, 1993), as discussed in Galea et al. (1997), the asymmetrical class of skew-normal (SMSN) distributions proposed by Branco and Dey (2001) or the unified skew-elliptical distributions proposed by Arellano and Genton (2010). However, in practice when nothing is known about the true distribution of the error terms, a risk exists that linear regression analysis based on any of the above models will be performed using an incorrectly specified model. There can also be situations where a single parametric family is unable to provide a satisfactory model for local variations in the observed data.

To overcome these problems, solutions that use finite mixture (FM-LR) models have been recently proposed. For instance, Bartolucci and Scaccia (2005), Soffritti and Galimberti (2011) and Galimberti and Soffritti (2014) developed methods for linear regression analysis by assuming a finite mixture of Gaussian (FM-N-LR) and Student-t (FM-T-LR) components for the error terms.

The classic approach to finite mixture modeling has several challenging aspects. There are nontrivial issues, like non-identifiability and an unbounded likelihood. In this context, Holzmann and Munk (2006) established the identifiability of finite mixtures of elliptical distributions under conditions of the characteristic or probability density function (PDF) generators. More recently, Otianiano et al. (2015) established the identifiability of finite mixture of skew-normal (Azzalini, 1985) and skew-t (Azzalini and Genton, 2008) distributions.

The class of SMSN distributions, proposed by Branco and Dey (2001), is attractive since it simultaneously models skewness with heavy tails (Prates et al., 2012) and contains as proper elements distributions such as the skew-normal, skew-t, skew-slash, skew-contaminated normal and all the symmetric class of scale mixtures of normal (SMN) distributions defined by Andrews and Mallows (1974). Besides this, it has a stochastic representation for easy implementation of the Expectation-Maximization (EM) algorithm (Dempster et al., 1977) and it also facilitates the study of many useful properties. Thus, this extension results in a flexible class of models for robust estimation and inference in FM-LR models.

The objective of this paper is to propose a mixture regression model (and associated likelihood inference) based on the mixtures of the class of scale mixtures of skew-normal (SMSN) distributions, by extending the mixture model based on symmetrical distributions. An advantage of this model is the possibility of fitting multimodality, heavy tails and skewness simultaneously. We derive a mixture model for the random errors based on the class of SMSN distributions (FM-SMSN-LR model) and evaluate the performance of the FM-SMSN-LR model by simulations. In order to motivate our research, we describe the following example with a dataset from the Australian Institute of Sport data (AIS).

1.2 MOTIVATING EXAMPLE

Before discussing the goal of this work, we present a motivating example. More specifically, we extend the linear regression model proposed by Bartolucci and Scaccia (2005), which is defined as

$$Y_i = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad f(\varepsilon_i) = \sum_{j=1}^g p_j \phi(\varepsilon_i | \mu_j, \sigma_j^2), \quad i = 1, \dots, n,$$

where Y_i is the response of case i , $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ is a vector of explanatory variable values, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ is a vector of unknown linear parameters, p_j are positive weights summing to 1, the μ_j terms satisfy the constraint $\sum_{j=1}^g p_j \mu_j = 0$, $\phi(\cdot; \mu_j, \sigma_j^2)$ denotes

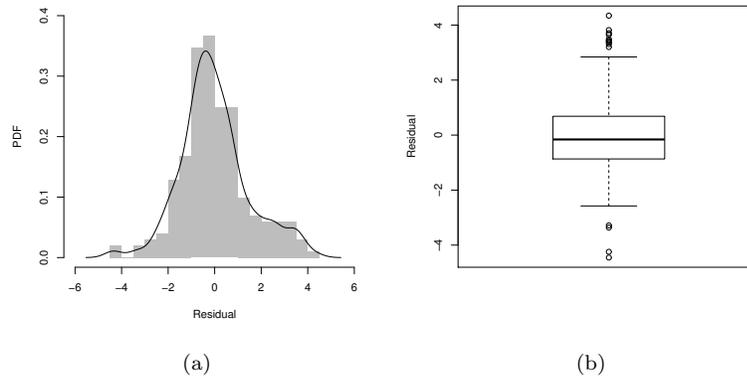


Figure 1. Histogram with a kernel PDF estimate superimposed (a) and the boxplot of ordinary residuals (b) with AIS data.

the PDF of the normal distribution, by assuming that the distribution of the error terms follows a finite mixture of SMSN distribution, so that the FM-SMSN-LR is defined. It is important to stress that our proposal is different from that of Zeller et al. (2016), where the linear regression is modeled with different regression functions, the so-called mixture of regressions or switching regression (Quandt and Ramsey, 1978). An important question that is addressed in this paper is whether a mixture model ($g \geq 2$) is needed instead of a one-component model. Thus, we use the parametric bootstrap log-likelihood ratio statistic, which was proposed by Turner (2000).

To test our proposed model, we use the AIS data available in an R package named `FMsmnReg`. Figure 1 (panels a and b) displays the histogram with a kernel PDF estimate superimposed and the boxplot of ordinary residuals, respectively, obtained by fitting a N-LR model to the AIS data. The plots reveal the existence of multimodal residuals, with evident presence of outliers. In summary, it is necessary to consider a more robust structure in the error. Therefore, this example serves as a motivation for the FM-SMSN-LR model.

1.3 ORGANIZATION OF THE PAPER

The remainder of the paper is organized as follows. In Section 2, we briefly discuss some properties of the univariate SMSN family. In Section 3, we present the FM-SMSN-LR model, including the EM-type algorithm for maximum likelihood (ML) estimation, and derive the empirical information matrix analytically to obtain the standard errors. In Section 4, numerical samples using both simulated and real data are given to illustrate the performance of the proposed model. Finally, some concluding remarks are presented in Section 5.

2. BACKGROUND

2.1 SCALE MIXTURES OF SKEW-NORMAL DISTRIBUTIONS

Next, we start by defining the skew-normal (SN) distribution and then we introduce some useful properties. As defined by Azzalini (1985), a random variable Z has a skew-normal distribution with location parameter μ , scale parameter σ^2 and skewness parameter $\lambda \in \mathbb{R}$, denoted by $Z \sim \text{SN}(\mu, \sigma^2, \lambda)$, if its PDF is given by

$$\phi_{\text{SN}}(z|\mu, \sigma^2, \lambda) = 2\phi(z|\mu, \sigma^2)\Phi(\lambda(z - \mu)/\sigma).$$

The relation between the SMSN class and the SN distribution is provided in the next definition.

DEFINITION 2.1 A random variable Y has an SMSN distribution with location parameter μ , scale parameter σ^2 and skewness parameter λ , denoted by $\text{SMSN}(\mu, \sigma^2, \lambda; H)$, if it has the stochastic representation

$$Y = \mu + \kappa^{1/2}(U)Z, \quad U \perp Z,$$

where $Z \sim \text{SN}(0, \sigma^2, \lambda)$, U is a positive random variable with cumulative distribution function $H(\cdot | \boldsymbol{\nu})$ indexed by a scalar or vector parameter $\boldsymbol{\nu}$ and $\kappa(u)$ is a positive function of u .

The mean and variance of Y are given respectively by

$$\text{E}[Y] = \mu + \sqrt{\frac{2}{\pi}} K_1 \Delta, \quad \text{Var}[Y] = \sigma^2 \left(K_2 - \frac{2}{\pi} K_1^2 \delta^2 \right), \quad (1)$$

where $\Delta = \sigma\delta$, with $\delta = \lambda/\sqrt{1 + \lambda^2}$ and $K_r = \text{E}[U^{-r/2}]$, $r = 1, 2, \dots$. Although we can deal with any $\kappa(\cdot)$ function, in this paper we restrict our attention to the case where $\kappa(u) = 1/u$, since it leads to good mathematical properties. Given $U = u$, we have that $Y|U = u \sim \text{SN}(\mu, u^{-1}\sigma^2, \lambda)$. Thus, the PDF of Y is expressed as

$$f(y) = \phi_{\text{SMSN}}(y|\mu, \sigma^2, \lambda, \boldsymbol{\nu}) = 2 \int_0^\infty \phi(y|\mu, u^{-1}\sigma^2) \Phi\left(u^{1/2}\lambda(y - \mu)/\sigma\right) dH(u|\boldsymbol{\nu}). \quad (2)$$

When H is degenerate, with $u = 1$, we obtain the $\text{SN}(\mu, \sigma^2, \lambda)$ distribution, and when $\lambda = 0$, the SMSN distribution reduces to the class of scale-mixtures of the normal (SMN) distribution represented by the PDF $f_0(y) = \phi_{\text{SMN}}(y|\mu, \sigma^2, \boldsymbol{\nu}) = \int_0^\infty \phi(y|\mu, u^{-1}\sigma^2) dH(u|\boldsymbol{\nu})$.

2.2 SPECIAL CASES OF THE SMSN DISTRIBUTIONS

Some special families of SMSN distributions are the following:

- The skew-t distribution with ν degrees of freedom. In this case, the PDF of Y takes the form

$$\phi_{\text{T}}(y|\mu, \sigma^2, \lambda, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu\sigma}} \left(1 + \frac{d}{\nu}\right)^{-\frac{\nu+1}{2}} T\left(\sqrt{\frac{\nu+1}{d+\nu}} A|\nu+1\right), \quad y \in \mathbb{R},$$

where $d = (y - \mu)^2/\sigma^2$, $A = \lambda(y - \mu)/\sigma$ and $T(\cdot|\nu)$ denotes the distribution function of the standard Student-t distribution, with location zero, scale one and ν degrees of freedom, namely $t(0, 1, \nu)$. We use the notation $Y \sim \text{ST}(\mu, \sigma^2, \lambda, \nu)$.

- The skew-slash distribution. It is denoted by $Y \sim \text{SSL}(\mu, \sigma^2, \lambda, \nu)$ and the associated PDF is given by

$$\phi_{\text{SL}}(y|\mu, \sigma^2, \lambda, \nu) = 2\nu \int_0^1 u^{\nu-1} \phi(y|\mu, u^{-1}\sigma^2) \Phi(u^{1/2}A) du, \quad y \in \mathbb{R}.$$

The skew-slash is a heavy-tailed distribution having as limiting distribution the skew-normal one (when $\nu \rightarrow \infty$).

- The skew contaminated normal distribution. We denote it by $Y \sim \text{SCN}(\mu, \sigma^2, \lambda, \nu, \gamma)$. Its PDF is given by

$$\phi_{\text{SCN}}(y|\mu, \sigma^2, \lambda, \nu) = 2\{\nu\phi(y|\mu, \gamma^{-1}\sigma^2)\Phi(\gamma^{1/2}A) + (1-\nu)\phi(y|\mu, \sigma^2)\Phi(A)\}, \quad \nu, \gamma \in (0, 1].$$

The parameters ν and γ can be interpreted as the proportion of outliers and a scale factor, respectively. The skew contaminated normal distribution reduces to the skew-normal distribution when $\gamma = 1$.

2.3 COMPUTATIONAL FRAMEWORK

The R software (R Core Team, 2016) produces statistical analyses, with its open source codes. This non-commercial computational program may be downloaded from <http://www.r-project.org>. Our method was implemented in R and its codes are available through the FMsmnReg package (Benites et al., 2016). We use the mixmsmn package, which allows the simulation of mixture the class of scale mixture of skew-normal distributions, see Prates et al. (2013). This computational framework is useful for conducting the simulation studies and the empirical illustration carried out in Section 4.

3. THE LINEAR REGRESSION MODEL WITH FM-SMSN ERRORS

3.1 GENERAL CONTEXT

Next, we introduce the linear regression model using finite mixture of skew heavy tailed distributions where the distribution of the error terms follows a finite mixture of scale mixture of skew-normal distributions (FM-SMSN-LR), following a similar setup as that developed by Bartolucci and Scaccia (2005). Consider the linear regression model expressed as

$$Y_i = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, n, \quad (3)$$

where Y_i is the response of case i , $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ is a vector of explanatory variables of dimension $(p+1) \times 1$, and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ is the regression parameter vector. Furthermore, we assume that

$$f(\varepsilon_i) = \sum_{j=1}^g p_j \phi_{\text{SMSN}}(\varepsilon_i | \mu_j + b\Delta_j, \sigma_j^2, \lambda_j, \boldsymbol{\nu}_j), \quad i = 1, \dots, n, \quad (4)$$

where p_j are positive weights summing to 1, the μ_j s satisfy the identifiability constraint $\sum_{j=1}^g p_j \mu_j = 0$, $b = -\sqrt{2/\pi}K_1$, $K_1 = \text{E}[U^{-1/2}]$, $\Delta_j = \sigma_j \delta_j$ with $\delta_j = \lambda_j / \sqrt{1 + \lambda_j^2}$. Then from Equation (1), we have that $\text{E}(\varepsilon_i) = 0$. Thus, for linearity of SMSN distributions, the PDF of Y_i is expressed as

$$f(y_i | \boldsymbol{\theta}) = \sum_{j=1}^g p_j \phi_{\text{SMSN}}(y_i | \mu_{ij} + b\Delta_j, \sigma_j^2, \lambda_j, \boldsymbol{\nu}_j), \quad \mu_{ij} = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + \mu_j = \vartheta_j + \mathbf{x}_i^\top \boldsymbol{\beta}, \quad (5)$$

where $\mu_{ij} = \mathbf{x}_i^\top \boldsymbol{\beta} + \vartheta_j$, $\vartheta_j = \beta_0 + \mu_j$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, (p_1, \dots, p_{g-1})^\top, \vartheta_1, \dots, \vartheta_g, \sigma_1^2, \dots, \sigma_g^2, \lambda_1, \dots, \lambda_g, \nu_1, \dots, \nu_g)^\top$ is the vector with all parameters. Concerning the parameter $\boldsymbol{\nu}_j$ of the mixing distribution $H(\cdot | \boldsymbol{\nu}_j)$, for $j = 1, \dots, g$, it can be a vector of parameters, e.g.,

the contaminated normal distribution. Thus, for computational convenience we assume that $\boldsymbol{\nu}_1 = \dots = \boldsymbol{\nu}_g = \boldsymbol{\nu}$. This strategy works very well in the empirical studies that we have conducted and greatly simplifies the optimization problem. For $U = 1$, Equations (3) and (4) lead to the FM-N-LR defined by [Bartolucci and Scaccia \(2005\)](#). Moreover, when $g = 1$ and a nonlinear function is used instead of $\mathbf{x}_i^\top \boldsymbol{\beta}$, the FM-SMSN-LR framework reduces to the model discussed by [Garay et al. \(2011\)](#). For each i and j , consider the latent indicator variable Z_{ij} , such that

$$Z_{ij} = \begin{cases} 1, & \text{if the } i\text{th subject is from the } j\text{th component;} \\ 0, & \text{otherwise.} \end{cases}$$

Observe that $Z_{ij} = 1$ if and only if $Z_i = j$. Then

$$P(Z_{ij} = 1) = 1 - P(Z_{ij} = 0) = p_j \quad \text{and} \quad y_i | Z_{ij} = 1 \sim \text{SMSN}(\mu_{ij} + b\Delta_j, \sigma_j^2, \lambda_j; H(\boldsymbol{\nu})). \quad (6)$$

Note that by integrating out $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{ig})^\top$, we obtain the marginal PDF presented in Equation (2) and $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ are independent random vectors, each one having a multinomial distribution with PDF defined as $f(\mathbf{z}_i) = p_1^{z_{i1}} p_2^{z_{i2}} \dots (1 - p_1 - \dots - p_{g-1})^{z_{ig}}$, which we denote by $\mathbf{Z}_i \sim \text{M}(1; p_1, \dots, p_g)$. These latent vectors appear in the hierarchical representation given next, which is used to build the Expectation Conditional Maximization Either (ECME) algorithm as proposed by [Liu and Rubin \(1994\)](#), which is a variant of the EM algorithm [Dempster et al. \(1977\)](#). From Equation (6) along with Definition 2.1, the FM-SMSN-LR model can be represented as

$$Y_i | u_i, t_i, Z_{ij} = 1 \stackrel{\text{IND}}{\sim} \text{N}(\mu_{ij} + \Delta_j t_i, u_i^{-1} \Gamma_j), \quad (7)$$

$$T_i | u_i, Z_{ij} = 1 \stackrel{\text{IND}}{\sim} \text{TN}(b, u_i^{-1}, (b, \infty)),$$

$$U_i | Z_{ij} = 1 \stackrel{\text{IND}}{\sim} H(u_i; \boldsymbol{\nu}),$$

$$\mathbf{Z}_i \stackrel{\text{IID}}{\sim} \text{M}(1; p_1, \dots, p_g), \quad i = 1, \dots, n, \quad j = 1, \dots, g, \quad (8)$$

where IND denotes independent, whereas IID stands for independent and identically distributed, with $\Gamma_j = (1 - \delta_j^2)\sigma_j^2$, $\Delta_j = \sigma_j \delta_j$ and $\delta_j = \lambda_j / \sqrt{1 + \lambda_j^2}$.

3.2 PARAMETER ESTIMATION VIA THE ECME ALGORITHM

Next, we show how to implement the ECME algorithm for ML estimation of the parameters of the FM-SMSN-LR model. By using Equations (7) to (8), we have that the complete-data log-likelihood function is given by

$$\begin{aligned} \ell_c(\boldsymbol{\theta} | \mathbf{y}, \mathbf{t}, \mathbf{u}, \mathbf{z}) = c + \sum_{i=1}^n \sum_{j=1}^g Z_{ij} \left\{ \log(p_j) - \frac{1}{2} \log(\Gamma_j) - \frac{u_i}{2\Gamma_j} (y_i - \mu_{ij} - \Delta_j t_i)^2 \right. \\ \left. + \log(h(u_i | \boldsymbol{\nu})) + \log[\phi_{\text{TN}}(t_i | b, u_i^{-1}, (b, \infty))] \right\}, \end{aligned}$$

where c is a constant that is independent of the parameter vector $\boldsymbol{\theta}$. By defining the quantities $\hat{z}_{ij} = \text{E}[Z_{ij} | \hat{\boldsymbol{\theta}}, y_i]$, $\hat{s}_{1ij} = \text{E}[Z_{ij} U_i | \hat{\boldsymbol{\theta}}, y_i]$, $\hat{s}_{2ij} = \text{E}[Z_{ij} U_i T_i | \hat{\boldsymbol{\theta}}, y_i]$ and $\hat{s}_{3ij} =$

$E[Z_{ij}U_iT_i^2|\hat{\boldsymbol{\theta}}, y_i]$, as having known properties of conditional expectation, we obtain

$$\hat{z}_{ij} = \frac{\hat{p}_j \phi_{\text{SMSN}}(y_i|\mu_{ij} + b\Delta_j, \sigma_j^2, \lambda_j, \boldsymbol{\nu})}{\sum_{j=1}^g \hat{p}_j \phi_{\text{SMSN}}(y_i|\mu_{ij} + b\Delta_j, \sigma_j^2, \lambda_j, \boldsymbol{\nu})},$$

$\hat{s}_{1ij} = \hat{z}_{ij}\hat{u}_{ij}$, $\hat{s}_{2ij} = \hat{z}_{ij}(\hat{u}_{ij}\hat{\mu}_{T_{ij}} + \widehat{M}_{T_j}\hat{\tau}_{1_{ij}})$ and $\hat{s}_{3ij} = \hat{z}_{ij}(\hat{u}_{ij}\hat{\mu}_{T_{ij}}^2 + \widehat{M}_{T_j}^2 + \widehat{M}_{T_j}(\hat{\mu}_{T_{ij}} + b)\hat{\tau}_{1_{ij}})$, where

$$\hat{\tau}_{1_{ij}} = E \left[U_i^{1/2} W_{\Phi_1} \left(\frac{U_i^{1/2} \hat{\mu}_{T_{ij}}}{\widehat{M}_{T_j}} \right) \mid \hat{\boldsymbol{\theta}}, y_i, Z_{ij} = 1 \right], \quad i = 1, \dots, n, \quad j = 1, \dots, g,$$

$$\widehat{M}_{T_j}^2 = \frac{\Gamma_j}{\Gamma_j + \Delta_j^2}, \quad \hat{\mu}_{T_{ij}} = b + \frac{\Delta_j}{\Gamma_j + \Delta_j^2} (y_i - \mu_{ij} - \Delta b) \quad \text{and} \quad \hat{u}_{ij} = E[U_j|\hat{\boldsymbol{\theta}}, y_i, Z_{ij} = 1].$$

Once again, at each step the conditional expectations \hat{u}_{ij} and $\hat{\tau}_{1_{ij}}$ can be easily derived from the results given in [Basso et al. \(2010\)](#). Thus, the Q -function is given by

$$\begin{aligned} Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = c + \sum_{i=1}^n \sum_{j=1}^g & \left(\hat{z}_{ij}^{(k)} (\log(p_j) - \frac{1}{2} \log(\Gamma_j) - \frac{1}{2\Gamma_j} (\hat{s}_{1ij}^{(k)} (y_i - \mu_{ij})^2 - 2(y_i - \mu_{ij}) \Delta_j \hat{s}_{2ij}^{(k)} \right. \\ & \left. + \Delta_j^2 \hat{s}_{3ij}^{(k)}) + E[Z_{ij} \log(h(U_i|\boldsymbol{\nu}))|\hat{\boldsymbol{\theta}}^{(k)}, y_i] + E[Z_{ij} \log(\phi_{\text{TN}}(T_i|b, u_i^{-1}, (b, \infty)))|\hat{\boldsymbol{\theta}}^{(k)}, y_i] \right). \end{aligned}$$

In the CML-step we update the estimate of $\boldsymbol{\nu}$ by direct maximization of the marginal log-likelihood, circumventing the computation of the conditional expectations $\hat{s}_{4ij} = E[Z_{ij} \log(h(U_i|\boldsymbol{\nu}))|\hat{\boldsymbol{\theta}}, y_i]$ and $\hat{s}_{5ij} = E[Z_{ij} \log(\phi_{\text{TN}}(T_i|b, u_i^{-1}, (b, \infty)))|\hat{\boldsymbol{\theta}}^{(k)}, y_i]$. Thus, the ECME algorithm for ML estimation of $\boldsymbol{\theta}$ is defined as follows:

E-step: Given a current estimate $\hat{\boldsymbol{\theta}}^{(k)}$, compute \hat{z}_{ij} , \hat{s}_{1ij} , \hat{s}_{2ij} , \hat{s}_{3ij} , for $i = 1, \dots, n$ and $j = 1, \dots, g$.

CM-steps: Update $\hat{\boldsymbol{\theta}}^{(k)}$ by maximizing $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = E[\ell_c(\boldsymbol{\theta})|\mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)}]$ over $\boldsymbol{\theta}$, which leads to the closed-form expressions given by

$$\begin{aligned} \hat{p}_j^{(k+1)} &= n^{-1} \sum_{i=1}^n \hat{z}_{ij}^{(k)}, \\ \hat{\vartheta}_j^{(k+1)} &= \left(\sum_{i=1}^n (\hat{s}_{1ij}^{(k)} (y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}) - \hat{\Delta}_j^{(k)} \hat{s}_{2ij}^{(k)}) \right) / \sum_{i=1}^n \hat{s}_{1ij}^{(k)}, \\ \hat{\boldsymbol{\beta}}^{(k+1)} &= \left(\sum_{i=1}^n \sum_{j=1}^g \frac{\hat{s}_{1ij}^{(k)} \mathbf{x}_i \mathbf{x}_i^\top}{\hat{\Gamma}_j^{(k)}} \right)^{-1} \sum_{i=1}^n \sum_{j=1}^g \frac{1}{\hat{\Gamma}_j^{(k)}} [\hat{s}_{1ij}^{(k)} (y_i - \hat{\vartheta}_j^{(k+1)}) - \hat{\Delta}_j^{(k)} \hat{s}_{2ij}^{(k)}] \mathbf{x}_i, \\ \hat{\Delta}_j^{(k+1)} &= \left(\sum_{i=1}^n (y_i - \hat{\mu}_{ij}^{(k+1)}) \hat{s}_{2ij}^{(k)} \right) / \sum_{i=1}^n \hat{s}_{3ij}^{(k)} \\ \hat{\Gamma}_j^{(k+1)} &= \sum_{i=1}^n \left(\hat{s}_{1ij}^{(k)} (y_i - \hat{\mu}_{ij}^{(k+1)})^2 - 2(y_i - \hat{\mu}_{ij}^{(k+1)}) \hat{\Delta}_j^{(k+1)} \hat{s}_{2ij}^{(k)} + \hat{\Delta}_j^{2(k+1)} \hat{s}_{3ij}^{(k)} \right) / \sum_{i=1}^n \hat{z}_{ij}^{(k)}. \end{aligned}$$

CML-step: Update $\widehat{\boldsymbol{\nu}}^{(k)}$ by maximizing the current marginal log-likelihood function, obtaining

$$\boldsymbol{\nu}^{(k+1)} = \operatorname{argmax}_{\boldsymbol{\nu}} \sum_{i=1}^n \log \left(\sum_{j=1}^g p_j^{(k+1)} \phi_{\text{SMSN}} \left(y_i | \mu_{ij}^{(k+1)} + b(\boldsymbol{\nu}) \Delta_j^{(k+1)}, \sigma_j^{2(k+1)}, \lambda_j^{(k+1)}, \boldsymbol{\nu} \right) \right).$$

Through constraint $\sum_{j=1}^g p_j \mu_j = 0$ (Bartolucci and Scaccia, 2005), we obtain the estimates of β_0 and μ_j as

$$\widehat{\beta}_0^{(k+1)} = \sum_{j=1}^g \widehat{p}_j^{(k+1)} \widehat{\vartheta}_j^{(k+1)} \quad \text{and} \quad \widehat{\mu}_j^{(k+1)} = \widehat{\vartheta}_j^{(k+1)} - \widehat{\beta}_0^{(k+1)},$$

respectively, for $j = 1, \dots, g$. This process is iterated until a suitable stopping criterion is satisfied. To avoid an indication of lack of progress of the algorithm (McNicholas et al., 2010), we adopted the Aitken acceleration method as the stopping criterion. At iteration k , we first compute the Aitken acceleration factor $c^{(k)} = (\ell^{(k+1)} - \ell^{(k)}) / (\ell^{(k)} - \ell^{(k-1)})$, where following Böhning et al. (1994), the asymptotic estimate of the log-likelihood at iteration $k + 1$ is given by

$$\ell_{\infty}^{(k+1)} = \ell^{(k)} + \frac{1}{1 - c^{(k)}} \left[\ell^{(k+1)} - \ell^{(k)} \right]. \quad (9)$$

As pointed out by Lindsay (1995), the algorithm is considered to reach convergence when $\ell_{\infty}^{(k+1)} - \ell^{(k+1)} < \varepsilon$, where ε is the desired tolerance (we use $\varepsilon = 10^{-6}$). A usual criticism is that EM-type procedures tend to get stuck in local modes. A convenient way to avoid this limitation is to try several EM iterations with a variety of starting values. If there are several modes, one can find the global mode by comparing their relative masses and log-likelihood values. We suggest the following strategy: For β_0 and $\boldsymbol{\beta}$ use the ordinary least-squares (OLS) estimate. Initial values for $p_j, \mu_j, \sigma_j^2, \lambda_j$ and $\boldsymbol{\nu}$, $j = 1, \dots, g$, are obtained by fitting the mixture model given in Equation (3) to the OLS residuals (Bartolucci and Scaccia, 2005), which can be done through the FMsmnReg package (Benites et al., 2016).

3.3 MODEL SELECTION AND APPROXIMATE STANDARD ERRORS

Consider the problem of comparing several FM-SMSN-LR models, with different numbers of component PDFs. Here, we use two model selection criteria, the Akaike information criterion plus a bias correction term (Hurvich and Tsai, 1989), denoted by (AIC_c) , and the adjusted Bayesian information criterion (Schlove, 1987), denoted by (BIC_a) . These criteria are defined as

$$\text{AIC}_c = -2\ell(\widehat{\boldsymbol{\theta}}) + \frac{2n\rho}{n - \rho - 1} \quad \text{and} \quad \text{BIC}_a = -2\ell(\widehat{\boldsymbol{\theta}}) + \rho \log \left(\frac{n+2}{2} \right),$$

where $\ell(\boldsymbol{\theta})$ is the actual log-likelihood, ρ is the number of free parameters that have to be estimated in the model, and n is the sample size.

A simple way of obtaining the standard errors of ML estimators of mixture model parameters is to approximate the asymptotic covariance matrix of $\widehat{\boldsymbol{\theta}}$ by the inverse of the observed information matrix. Let $\mathbf{I}_o(\boldsymbol{\theta}) = -\partial^2 \ell(\boldsymbol{\theta} | \mathbf{y}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$ be the observed information matrix, where $\ell(\boldsymbol{\theta} | \mathbf{y})$ is the observed log-likelihood function, which is obtained using Equation (5). In this work we use the alternative method suggested by Basford et al. (1997),

which consists of approximating the inverse of the covariance matrix by

$$\mathbf{I}_o(\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^n \widehat{\mathbf{s}}_i \widehat{\mathbf{s}}_i^\top, \quad \text{where} \quad \widehat{\mathbf{s}}_i = \left. \frac{\partial}{\partial \boldsymbol{\theta}} \log [f(y_i|\boldsymbol{\theta})] \right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}, \quad (10)$$

where $\widehat{\mathbf{s}}_i = (\widehat{s}_{i,\boldsymbol{\beta}}^\top, \widehat{s}_{i,p_1}, \dots, \widehat{s}_{i,p_{g-1}}, \widehat{s}_{i,\vartheta_1}, \dots, \widehat{s}_{i,\vartheta_g}, \widehat{s}_{i,\sigma_1^2}, \dots, \widehat{s}_{i,\sigma_g^2}, \widehat{s}_{i,\lambda_1}, \dots, \widehat{s}_{i,\lambda_g}, \widehat{s}_{i,\boldsymbol{\nu}})^\top$. It is important to stress that the standard error of $\boldsymbol{\nu}$, obtained from $\widehat{s}_{i,\boldsymbol{\nu}}$, depends heavily on the calculation of conditional expectation $E[\log(U_i)|y_{\text{obs}_i}, \widehat{\boldsymbol{\theta}}]$, which relies on computationally intensive Monte Carlo integrations, since no analytical expression for this expected value exists. Therefore, the expressions for the elements $\widehat{s}_{i,\boldsymbol{\beta}}, \widehat{s}_{i,p_j}, \widehat{s}_{i,\vartheta_j}, \widehat{s}_{i,\sigma_j^2}, \widehat{s}_{i,\lambda_j}$, for $j = 1, \dots, g$, are given as

$$\begin{aligned} \widehat{s}_{i,\boldsymbol{\beta}}^\top &= \frac{\sum_{j=1}^G p_j D_{\boldsymbol{\beta}}(y_i; \boldsymbol{\theta}_j)}{f(y_i; \boldsymbol{\theta})}, \quad \widehat{s}_{i,\vartheta_j} = \frac{p_j D_{\vartheta_j}(y_i; \boldsymbol{\theta}_j)}{f(y_i; \boldsymbol{\theta})}, \quad \widehat{s}_{i,\sigma_j^2} = \frac{p_j D_{\sigma_j^2}(y_i; \boldsymbol{\theta}_j)}{f(y_i; \boldsymbol{\theta})}, \quad \widehat{s}_{i,\lambda_j} = \frac{p_j D_{\lambda_j}(y_i; \boldsymbol{\theta}_j)}{f(y_i; \boldsymbol{\theta})}, \\ \widehat{s}_{i,p_j} &= \frac{1}{f(y_i; \boldsymbol{\theta})} [\phi_{\text{SMSN}}(y_i|\mu_{ij} + b\Delta_j, \sigma_j^2, \lambda_j, \boldsymbol{\nu}) - \phi_{\text{SMSN}}(y_i|\mu_{ig} + b\Delta_g, \sigma_g^2, \lambda_g, \boldsymbol{\nu})], \end{aligned}$$

with

$$D_{\vartheta_j}(y_i; \boldsymbol{\theta}_j) = \frac{\partial}{\partial \vartheta_j} \left(\phi_{\text{SMSN}}(y_i|\mu_{ij} + b\Delta_j, \sigma_j^2, \lambda_j, \boldsymbol{\nu}) \right).$$

After some algebraic manipulation, we obtain

$$\begin{aligned} D_{\boldsymbol{\beta}}(y_i; \boldsymbol{\theta}_j) &= \frac{2}{\sqrt{2\pi\sigma_j^2}} \left[\sigma^{-2}(y_i - \mu_{ij} - b\Delta_j) I_{ij}^\Phi(3/2) - \sigma_j^{-1} \lambda_j I_{ij}^\phi(1) \right] \mathbf{x}_i, \\ D_{\vartheta_j}(y_i; \boldsymbol{\theta}_j) &= \frac{2}{\sqrt{2\pi\sigma_j^2}} \left[\sigma_j^{-2}(y_i - \mu_{ij} - b\Delta_j) I_{ij}^\Phi(3/2) - \sigma_j^{-1} \lambda_j I_{ij}^\phi(1) \right], \\ D_{\lambda_j}(y_i; \boldsymbol{\theta}_j) &= \frac{2}{\sqrt{2\pi\sigma_j^2}} \left[\frac{(y_i - \mu_{ij} - b\Delta_j)b}{(1 + \lambda_j^2)^{(3/2)}} I_{ij}^\Phi(3/2) + \left((y_i - \mu_{ij} - b\Delta_j) - \frac{b\Delta_j}{1 + \lambda_j^2} I_{ij}^\phi(1) \right) \right], \\ D_{\sigma_j^2}(y_i; \boldsymbol{\theta}_j) &= \frac{1}{\sqrt{2\pi\sigma_j^2}} \left[-\sigma_j^{-2} I_{ij}^\Phi(1/2) + \sigma_j^{-4}(y_i - \mu_{ij} - b\Delta_j)^2 I_{ij}^\Phi(3/2) \right. \\ &\quad \left. + \sigma_j^{-4}(y_i - \mu_{ij} - b\Delta_j)b\Delta_j I_{ij}^\Phi(3/2) - \lambda_j \sigma_j^{-3}(y_i - \mu_{ij}) I_{ij}^\phi(1) \right] \end{aligned}$$

where the expressions $I_{ij}^\Phi(w)$ and $I_{ij}^\phi(w)$ are given in [Basso et al. \(2010\)](#). The information-based approximation defined in Equation (10) is asymptotically applicable. However, it is less reliable unless the sample size is sufficiently large. Observe that the asymptotic covariance matrix of the ML estimates, that is, the inverse of Equation (10), was obtained using the parametrization $\varphi_j = \beta_0 + \mu_j$, $j = 1, \dots, g$. We can use the traditional delta method (see [Rao, 1973](#), Sec. 6a.2), to obtain standard errors using the original parameterization.

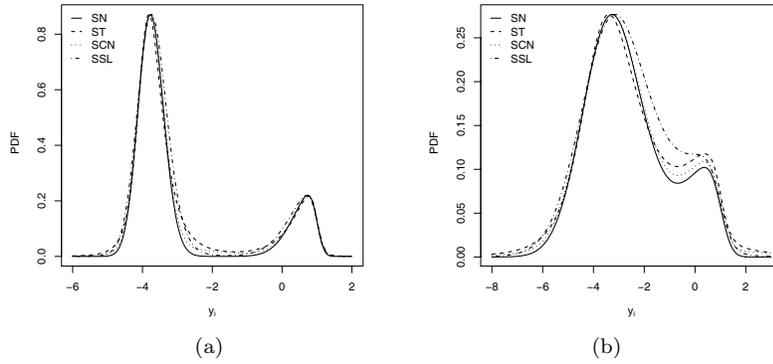


Figure 2. Target mixture PDFs from simulated data in Scenario 1 (a) and Scenario 2 (b).

4. NUMERICAL STUDIES

4.1 PARAMETER RECOVERY (SIMULATION STUDY I)

We conduct three simulation studies to illustrate the performance of our proposed model. The first simulation presented below reports the consistency of the approximate standard errors for the ML estimators of parameters through the EM algorithm with each sample under the stopping criterion in Equation (9), whereas the contents of the second and third simulations are described in the corresponding subsections. In addition, we finish this section of numerical studies with an empirical illustration based on real data.

Here, we consider two scenarios for simulation in order to verify if we can estimate the true parameter values accurately by using the proposed ECME algorithm. This is the first step to ensure that the estimation procedure works satisfactorily. We fit data that were artificially generated from the following model with two components

$$f(y_i|\boldsymbol{\theta}) = \sum_{j=1}^2 p_j \phi_{\text{SMSN}}(y_i | \mu_{ij} + b\Delta_j, \sigma_j^2, \lambda_j, \boldsymbol{\nu}), \quad i = 1, \dots, n,$$

where Z_{ij} is a component indicator of Y_i with $P(Z_{ij} = 1) = p_j$, $j = 1, 2$, $x_i^\top = (x_{i1}, x_{i2})$, such that $x_{i1} \sim U(0, 1)$ and $x_{i2} \sim U(0, 1)$, for $i = 1, \dots, n$, and ε_1 and ε_2 follow a distribution as in the assumption given in Equation (3). We consider the following parameter values: $\beta_0 = -1$, $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top = (-4, -3)^\top$, $\mu_1 = -4$, $\mu_2 = 1$, $\lambda_1 = 1$, $\lambda_2 = -4$ and $p_1 = 0.2$. In addition, we consider the following scenarios (depicted in Figure 2): scenario 1 (well separated components) with $\sigma_1^2 = 0.2$ and $\sigma_2^2 = 0.4$, and scenario 2 (poorly separated components) with $\sigma_1^2 = 2$ and $\sigma_2^2 = 2$. For each combination of parameters, we generated 1000 Monte Carlo samples of size $n = 1000$ from the FM-SMSN-LR models, under four different situations: FM-SN-LR, FM-ST-LR ($\nu = 3$), FM-SSL-LR ($\nu = 3$) and FM-SCN-LR ($\boldsymbol{\nu}^\top = (0.1, 0.1)$). The average values and standard deviations (MC SD) of the estimators across the 1000 Monte Carlo samples were computed, along with the average (IM SE) values of the approximate standard deviations of the estimates obtained through the method described in the Subsection 3.3. Moreover, we compute coverage probability of each parameter (COV), which is defined by $\text{COV}(\hat{\boldsymbol{\theta}}) = (1/m) \sum_{j=1}^m I(\boldsymbol{\theta} \in [\hat{\boldsymbol{\theta}}_L, \hat{\boldsymbol{\theta}}_U])$, where I is the indicator function such that $\boldsymbol{\theta}$ lies in the interval $[\hat{\boldsymbol{\theta}}_L, \hat{\boldsymbol{\theta}}_U]$, with $\hat{\boldsymbol{\theta}}_L$ and $\hat{\boldsymbol{\theta}}_U$ being estimated lower and upper bounds of the 95% CI, respectively. The results are presented in Table 1. Note that under both scenarios (well and poorly separated components), the results suggest that the proposed FM-SMSN-LR model produces satisfactory estimates.

It can be seen from this table that the estimation method of the standard errors provides relatively close results (IM SE and MC SD), indicating that the proposed asymptotic

Table 1. Simulation study I: mean and MC SD are the respective estimated means and standard deviations from fitting a FM-SMSN-LR model based on 1000 samples. IM SE is the average value of the approximate standard error obtained through the information-based method. COV is the coverage probability. True values of parameters are in parentheses.

Parameter		Scenario 1: $\sigma_1^2 = 0.2, \sigma_2^2 = 0.4$				Scenario 2: $\sigma_1^2 = \sigma_2^2 = 2$			
		SN	ST($\nu = 3$)	SCN ($\nu = 0.1$)	SSL($\nu = 3$)	SN	ST($\nu = 3$)	SCN ($\nu = 0.1$)	SSL($\nu = 3$)
$\beta_0(-1)$	Mean	-0.9971	-1.0038	-0.9953	-0.9989	-1.0119	-1.0070	-0.9965	-1.0413
	IM SE	0.0602	0.0859	0.0777	0.0883	0.1928	0.3345	0.2369	0.3238
	MC SD	0.0698	0.0755	0.0713	0.0770	0.0925	0.1214	0.1324	0.1284
	COV	90.6%	96.7%	96.6%	96.0%	99.4%	95.7%	91.8%	95.8%
$\beta_1(-4)$	Mean	-4.0002	-3.9985	-3.9996	-3.9947	-3.9949	-3.9958	-3.9963	-4.0005
	IM SE	0.0368	0.0418	0.0402	0.0423	0.0889	0.1021	0.0974	0.0985
	MC SD	0.0365	0.0426	0.0403	0.0449	0.0899	0.1076	0.0950	0.1031
	COV	94.7%	94.2%	95.5%	95.0%	95.0%	92.9%	95.4%	93.3%
$\beta_2(-3)$	Mean	-3.0012	-2.9998	-3.0014	-2.9938	-2.9994	-2.9989	-2.9967	-3.0013
	IM SE	0.0374	0.0424	0.0410	0.0432	0.0859	0.1005	0.0975	0.1020
	MC SD	0.0370	0.0442	0.0413	0.0430	0.0836	0.1046	0.0977	0.1109
	COV	95.6%	93.7%	94.0%	96.0%	96.2%	94.4%	94.2%	92.0%
$\mu_1(-4)$	Mean	-4.0026	-3.9945	-4.0040	-4.0166	-4.0295	-3.9806	-4.0899	-3.9924
	IM SE	0.0853	0.0800	0.0894	0.0854	0.1396	0.2782	0.1896	0.2531
	MC SD	0.0691	0.0876	0.0744	0.0859	0.1111	0.3161	0.2483	0.2202
	COV	98.2%	99.8%	98.6%	98.6%	97.3%	92.3%	84.5%	94.8%
$\mu_2(1)$	Mean	0.9992	1.0012	1.0007	0.9945	0.9990	1.0103	1.0391	0.9955
	IM SE	0.0837	0.0878	0.0862	0.0873	0.0744	0.1098	0.0861	0.0983
	MC SD	0.0630	0.0625	0.0656	0.0625	0.0692	0.1000	0.1060	0.0813
	COV	98.3%	99.7%	98.4%	99.0%	96.7%	96.7%	86.4%	97.7%
σ_1^2	Mean	0.2097	0.2089	0.2084	0.1946	2.0069	2.2009	1.9385	1.9221
	IM SE	0.0680	0.0575	0.0643	0.0543	1.4238	0.9880	0.7385	1.5234
	MC SD	0.0427	0.0639	0.0644	0.0539	0.5626	1.0118	0.8238	0.9698
	COV	88.7%	89.8%	88.9%	89.0%	99.6%	87.3%	83.3%	89.1%
σ_2^2	Mean	0.3991	0.4026	0.3940	0.3988	2.0452	1.9839	1.8290	2.1521
	IM SE	0.0274	0.0385	0.0343	0.0381	0.1978	0.3796	0.1898	0.2758
	MC SD	0.0283	0.0501	0.0423	0.0463	0.1816	0.2642	0.3309	0.3109
	COV	94.0%	85.9%	85.5%	88.0%	95.9%	93.7%	72.5%	89.2%
$\lambda_1(1)$	Mean	1.0916	1.0534	1.0894	0.9679	1.1614	1.0068	0.6175	0.8514
	IM SE	0.7420	0.4956	0.6466	0.4814	1.4279	1.0923	1.2206	2.7316
	MC SD	0.8216	0.4983	0.6441	0.4385	0.4974	0.7792	1.3124	1.1426
	COV	94.3%	96.3%	95.9%	98.0%	99.6%	96.9%	88.4%	92.4%
$\lambda_2(-4)$	Mean	-4.0874	-4.1108	-4.0739	-4.1418	-4.2153	-4.0168	-3.7773	-4.0682
	IM SE	0.5446	0.5969	0.5971	0.6086	0.6299	0.8950	0.6262	0.6219
	MC SD	0.5406	0.6141	0.6007	0.5477	0.5967	0.6555	0.8671	0.6494
	COV	96.8%	95.5%	94.3%	96.0%	96.8%	94.5%	86.8%	93.6%
$p_1(0.2)$	Mean	0.1998	0.2004	0.1999	0.1985	0.1987	0.2033	0.2028	0.2000
	IM SE	0.0126	0.0131	0.0130	0.0131	0.0146	0.2218	0.0159	0.0204
	MC SD	0.0126	0.0125	0.0129	0.0127	0.0138	0.0235	0.0213	0.0191
	COV	95.3%	95.8%	95.0%	94.0%	96.3%	92.9%	87.3%	94.6%
ν	Mean	-	3.0735	0.1070	2.9791	-	3.2216	0.1342	4.4543
$\gamma(0.1)$	Mean	-	-	0.1098	-	-	-	0.1415	-

approximation for the variances of the ML estimates of Equation (10) is reliable. Note also that the coverage probability (COV) for the regression parameters is quite stable for two scenarios, indicating that the proposed asymptotic approximation for the variance estimates of the ML estimates is reliable.

4.2 ASYMPTOTIC PROPERTIES OF THE EM ESTIMATES (SIMULATION STUDY II)

The main focus in this simulation study is to show the asymptotic properties of the EM estimates. Our strategy is to generate artificial samples from the FM-SMSN-LR model with $x_i^\top = (x_{i1}, x_{i1})$, such that $x_{i1} \sim U(0, 1)$ and $x_{i2} \sim U(0, 1)$, for $i = 1, \dots, n$. We choose sample sizes $n = 100, 250, 500, 1000, 2500$ and 5000 . The true values of the parameters were taken as $\beta_0 = -1, \beta = (\beta_1, \beta_2)^\top = (-4, -3)^\top, \mu_1 = -4, \mu_2 = 1, \sigma_1^2 = 0.2, \sigma_2^2 = 0.4$ and $p_1 = 0.2$. For each combination of parameters and sample sizes, we generated 1000 random samples from the FM-SMSN-LR models, under three different situations: FM-SN-LR, FM-ST-LR ($\nu = 3$), FM-SSL-LR ($\nu = 3$) and FM-SCN-LR ($\nu^\top = (0.1, 0.1)$). In order to analyze asymptotic properties of the EM estimates, we computed the bias and the relative root mean square error (RMSE) for each combination of sample size and

parameter values. For θ_i , they are given by

$$\text{Bias}(\theta_i) = \frac{1}{1000} \sum_{i=1}^{1000} (\theta_i^{(j)} - \theta_i) \text{ and } \text{RMSE}(\theta_i) = \sqrt{\frac{1}{1000} \sum_{i=1}^{1000} (\theta_i^{(j)} - \theta_i)^2},$$

where $\hat{\theta}_i^{(j)}$ is the estimate of θ_i for the j th sample. The results for β_0 , β_1 and β_2 are shown in Figure 3; the results for μ_1 , σ_1 and λ_1 are shown in Figure 4; the results for μ_2 , σ_2 , λ_2 are shown in Figure 5; and the results for p_1 are shown in Figure 6. One can see a pattern of convergence to zero of the bias and RMSE when n increases for all the parameters. As a general rule, we can say that Bias and RMSE tend to approach zero when the sample size increases, indicating that the estimates based on the proposed EM-type algorithm under the FM-SMSN-LR model do provide good asymptotic properties.

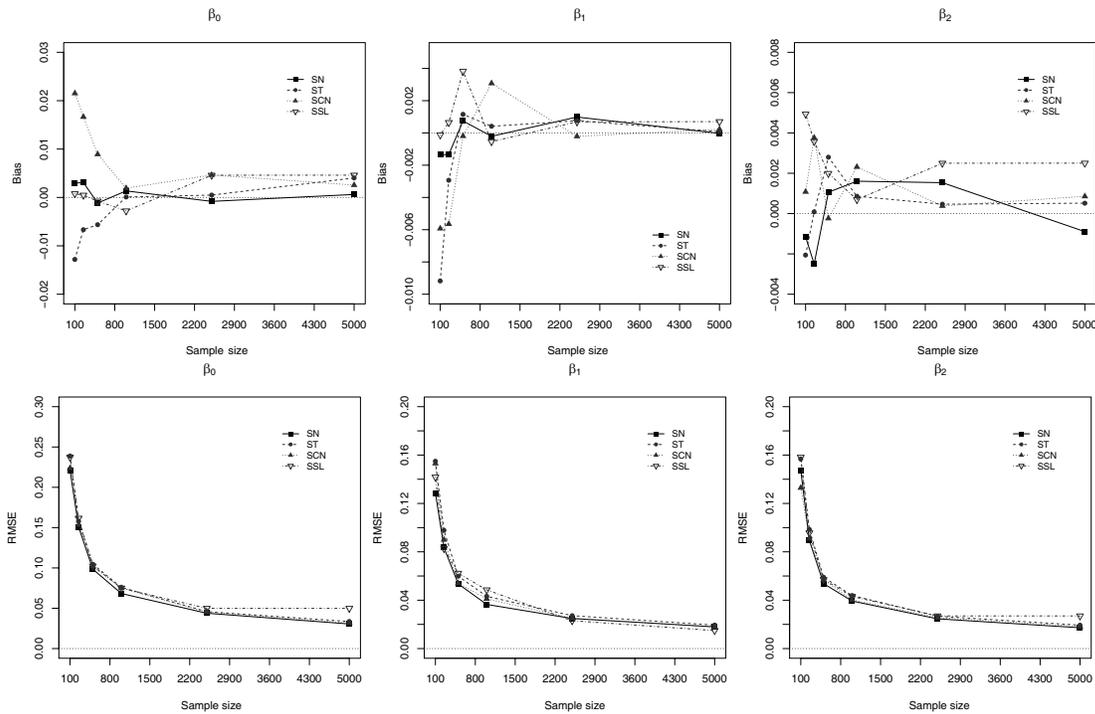


Figure 3. Average bias (1st row) and average RMSE (2nd row) of the estimators of $\beta_0, \beta_1, \beta_2$ for simulation II.

4.3 ROBUSTNESS OF THE EM ESTIMATES (SIMULATION STUDY III)

The purpose of this simulation study is to compare the effect of the robustness of the estimates of the FM-SMSN-LR models in the presence of outliers on the response variable. We compare the FM-SN-LR, FM-ST-LR ($\nu = 3$), FM-SSL-LR ($\nu = 3$) and the FM-CN-LR ($(\nu, \gamma) = (0.1, 0.1)$) models. In this scenario, we generated 500 samples of size $n = 500$ of the FM-SMSN-LR model with $f(\varepsilon_i) = \sum_{j=1}^2 p_j \phi_{\text{SMSN}}(\varepsilon_i | \mu_j + b\Delta_j, \sigma_j^2, \lambda_j, \nu)$. The true values of the parameters were taken as $\beta_0 = -1$, $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top = (-4, -3)^\top$, $\mu_1 = -4$, $\mu_2 = 1$, $\sigma_1^2 = 0.2$, $\sigma_2^2 = 0.4$ and $p_1 = 0.2$. To assess how much the EM estimates are influenced by the presence of outliers, we replaced observation y_{150} by $y_{150}(v) = y_{150} + v$, with $v = 1, 2, \dots, 10$. For each replication, we obtained the parameter estimates with and without outliers, with the three FM-SMSN-LR models.

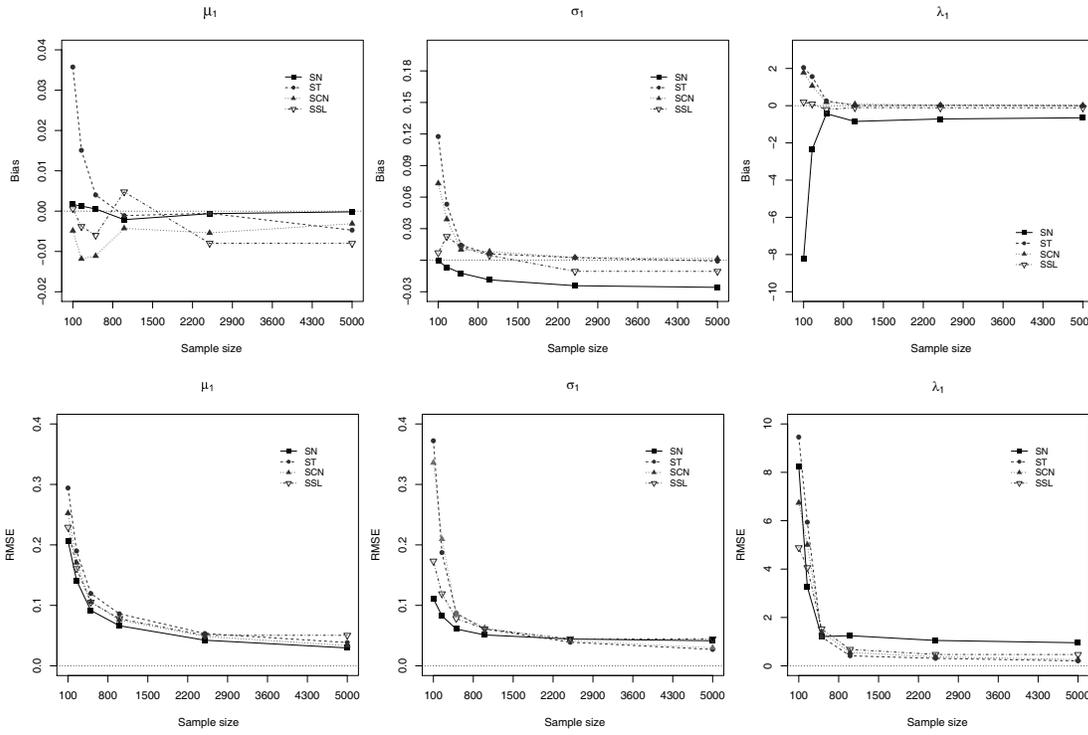


Figure 4. Average bias (1st row) and average RMSE (2nd row) of the estimators of $\mu_1, \sigma_1, \lambda_1$ for simulation II.

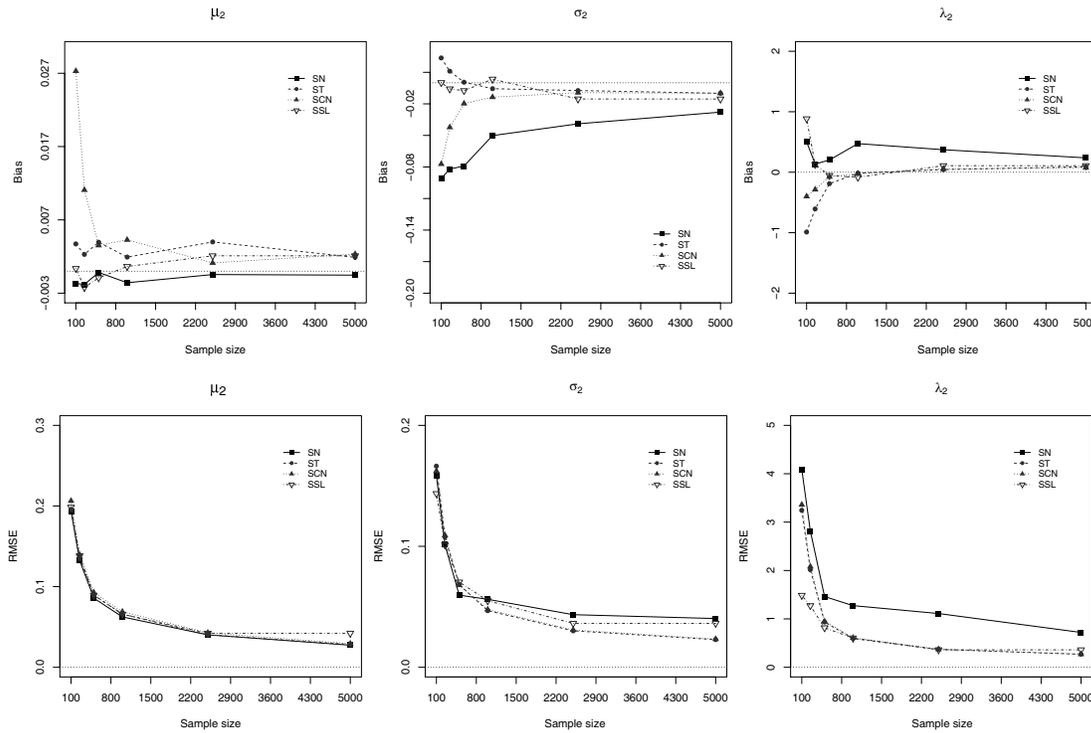


Figure 5. Average bias (1st row) and average RMSE (2nd row) of the estimators of $\mu_2, \sigma_2, \lambda_2$ for simulation II.

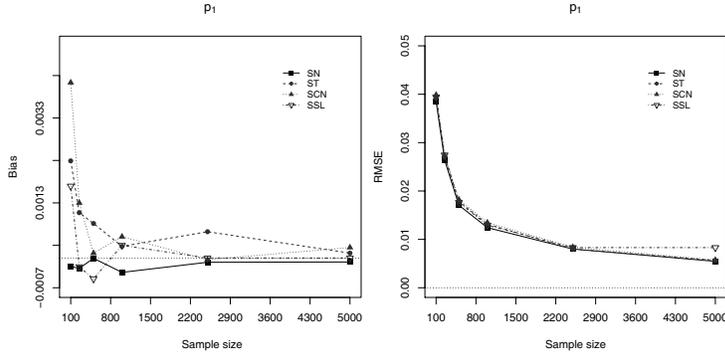


Figure 6. Average bias (1st row) and average RMSE (2nd row) of the estimators of p_1 for simulation II.

We are interested in evaluating the relative change (RC) in the estimates as a function of ν . Given $\Theta = (\beta_1, \beta_2, p_1, p_2, \theta_1, \theta_2)$, with $\theta_j = (\beta_0, \mu_j, \sigma_j^2, \lambda_j)$, $j = 1, 2$, the RC is defined by

$$\text{RC} \left(\hat{\Theta}_i(\nu) \right) = \left| \frac{\hat{\Theta}_i(\nu) - \hat{\Theta}_i}{\hat{\Theta}_i} \right|,$$

where $\hat{\Theta}_i(\nu)$ and $\hat{\Theta}_i$ denote the EM estimates of Θ_i with and without perturbation, respectively.

Figure 7 shows the average values of the relative changes undergone by all the parameters. We note that for all parameters, the average RCs suddenly increase under FM-SN-LR model as the ν value grows. In contrast, for the FM-SMSN-LR models with heavy tails, namely the FM-ST-LR ($\nu = 3$) and FM-SCN-LR ($\nu = (0.1, 0.1)$), the measures vary little, indicating they are more robust than the FM-SN-LR model in the ability to accommodate discrepant observations.

4.4 EMPIRICAL ILLUSTRATION

Next, the proposed techniques are illustrated with the analysis a real dataset, the one previously analyzed by [Cook and Weisberg \(1982\)](#) in a normal regression setting. The dataset comes from the Australian Institute of Sport (AIS) and consists of measurements of 202 athletes. Here, we focus on percent body fat (Bfat), which is assumed to be explained by the sum of skin folds (ssf) and height in cm (Ht). Thus, we consider the FM-SMSN-LR model given by

$$\text{Bfat}_i = \beta_0 + \beta_1 \text{ssf}_i + \beta_2 \text{Ht}_i + \varepsilon_i, \quad i = 1, \dots, 202,$$

where ε_i belongs to the FM-SMSN family.

By using the `FMsmnsnReg` package (see the appendix), we fit the FM-SMSN-LR models as was described in Section 3. Table 2 compares the fit of various mixture models for $g = 1$ to 5 components, using the model selection criteria discussed in Subsection 3.3. Note from this table that, as expected, the heavy-tailed models perform significantly better than the SN model (and the symmetric counterparts such as the normal and Student-t models), with mixtures of two ($g = 2$) components being significantly better in all cases, except for the normal case (FM-N), where a mixture of $g = 3$ is needed.

Moreover, the 2-component FM-ST-LR model fits the data substantially better. This conclusion also is verified through a hypotheses procedure for testing the number of components in the FM-ST-LR model. As reported by [Turner \(2000\)](#), we can use parametric

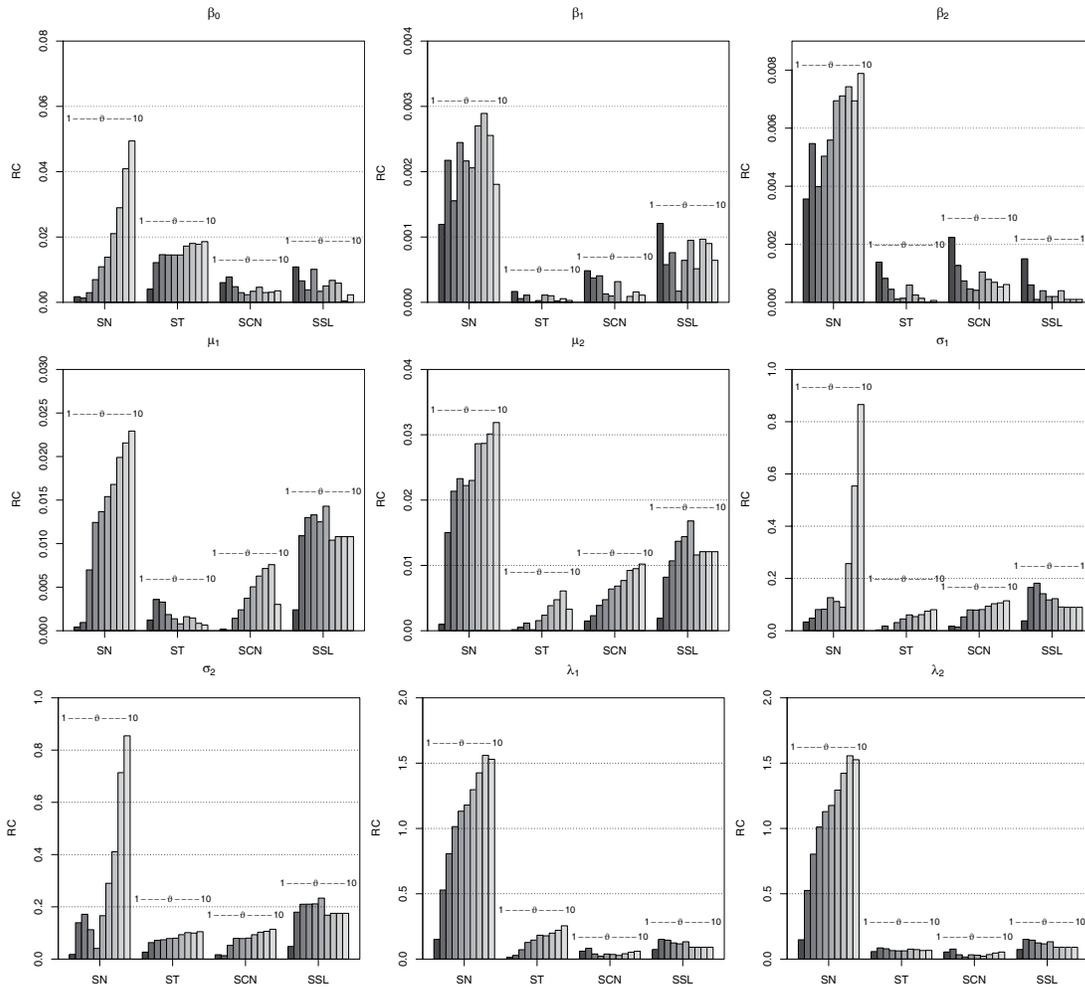


Figure 7. Average RCs of estimates with different perturbations v for simulation study III.

Table 2. Comparison of maximum log-likelihood, AIC_c and BIC_A for fitted FM-SMSN-LR models using the AIS data. The number of parameters is denoted by m .

Model	g	m	log-lik	AIC_c	BIC_a
FM-N	1	5	-367.2395	744.7850	745.1792
FM-N	2	8	-359.2902	735.3265	735.7009
FM-N	3	11	-355.2892	733.9679	734.1192
FM-T	1	6	-363.9525	738.2111	738.6053
FM-T	2	9	-358.2494	733.2449	733.6194
FM-T	3	12	-356.3237	736.0369	736.1881
FM-SN	1	6	-363.0346	738.5001	738.9097
FM-SN	2	10	-356.3079	733.7675	734.0164
FM-SN	3	14	-354.1438	738.5336	738.2486
FM-SN	4	18	-353.1388	746.0152	744.7987
FM-SN	5	22	-352.2579	754.1695	751.5973
FM-ST	1	7	-360.7632	736.1038	736.5070
FM-ST	2	11	-353.9696	731.3286	731.4799
FM-ST	3	15	-353.8492	740.2790	739.7994
FM-ST	4	19	-352.3138	746.8034	745.2888
FM-ST	5	23	-351.7865	755.7752	752.7944
FM-SCN	1	8	-357.0375	738.5001	738.9097
FM-SCN	2	12	-353.7235	733.0978	733.1278
FM-SCN	3	16	-354.1656	743.2717	742.5722
FM-SCN	4	20	-352.0380	748.7169	746.8773
FM-SCN	5	24	-352.8184	760.4164	756.9983
FM-SSL	1	7	-362.3246	739.2264	739.6296
FM-SSL	2	11	-354.1580	731.7054	731.8566
FM-SSL	3	15	-354.1941	740.9689	740.4892
FM-SSL	4	19	-352.2586	746.6930	745.1785
FM-SSL	5	23	-352.3504	756.9031	753.9224

Table 3. AIS data. Parameter estimates of the FM-SMSN- LR models with $g = 2$. SE denotes the corresponding standard errors obtained via the information-based matrix.

Parameter	FM-SN		FM-ST		FM-SCN		FM-SSL	
	ML	SE	ML	SE	ML	SE	ML	SE
β_0	14.7241	0.0001	14.51593	0.00253	14.6622	0.0025	14.7475	0.0025
β_1	0.1799	0.0012	0.17972	0.00850	0.1805	0.0089	0.1796	0.0091
β_2	-0.0757	0.1302	-0.07536	0.19264	-0.0757	0.1458	-0.0754	0.1513
p_1	0.1543	0.9295	0.15418	1.04192	0.1483	1.0841	0.1514	1.0393
μ_1	2.5504	2.2932	1.93244	4.00942	2.3654	3.8355	2.3891	3.9553
μ_2	-0.4652	1.8546	-0.35226	2.94875	-0.4120	2.5091	-0.4263	2.6266
σ_1^2	0.8483	0.5074	3.80681	1.57056	2.2957	1.6255	2.3158	1.6615
σ_2^2	2.2793	0.4021	1.06550	11.56693	1.1240	7.1021	0.9740	7.0029
λ_1	0.1624	0.8467	-5.70438	0.52991	-3.5415	0.4408	-4.8612	0.3724
λ_2	-2.2318	1.7509	-0.62860	9.52263	-1.0111	7.9389	-1.0144	11.9961
ν	-	-	7.45874	-	0.2270	-	2.3036	-
γ	-	-	-	-	0.3075	-	-	-

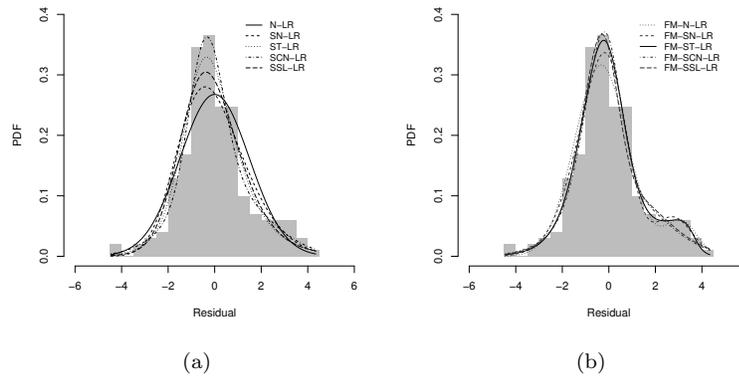


Figure 8. Panels (a) and (b) display the histogram ordinary residuals superimposed on the FM-SMSN-LR residual PDF for $g = 1$ and $g = 2$ components, respectively with AIS dataset.

or semiparametric bootstrap to test hypotheses concerning the number of components in the mixture. Following the method proposed by Turner (2000), we considered 1000 bootstrap statistics for testing $g = 1$ versus $g = 2$, in which case the p -value was 0.027 for the parametric bootstrap. Accordingly, there is strong evidence that at least two components are needed. For testing $g = 2$ versus $g = 3$, the bootstrap p -value was 0.984, so there is no evidence that more than two components are required to model the AIS dataset.

Table 3 presents the ML estimates of the parameters considering the four models with $g = 2$, say FM-SN-LR, FM-ST-LR, FM-SCN-LR and the FM-SSL-LR, along with the corresponding standard errors (SE), obtained via the information-based procedure presented in Subsection 3.3. Notice from Table 3 that the small value of the estimate of ν for the FM-ST-LR and FM-SSL-LR models indicates a lack of adequacy of the SN assumption.

In Figure 8, we plot the histogram of OLS residuals and then display the residual PDFs for the four FM-SMSN-LR models superimposed on a single set of coordinate axes, with $g = 1$ and $g = 2$ components respectively. Additional results related to $g = 3$ and $g = 4$ components are given in Figure 10. Based on this graphical representation, it appears once again that the FM-ST-LR, FT-SCN-LR and the FT-SSL-LR models have quite reasonable and better fit than the FM-SN-LR model with $g = 2$ components.

In order to detect incorrect specification of the error distribution for our best model (FM-ST-LR), we present quantile versus quantile (QQ) plots and simulated envelopes for the residuals $(y - \hat{y})$ in Figure 9. The QQ plots for the other models are given in Figure 11. This figure provides strong evidence that the FM-ST-LR (with $g = 2$ components) yields a better fit to the current data than the ST-LR model (with $g = 1$ component), since there are no observations falling outside the envelope.

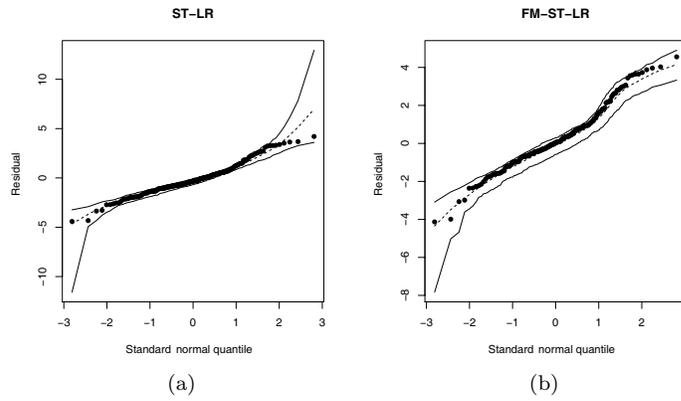


Figure 9. Panels (a) and (b) display the QQ plots and simulated envelopes for the residual $(y - \hat{y})$ with for $g = 1$ and $g = 2$ components, respectively with AIS dataset.

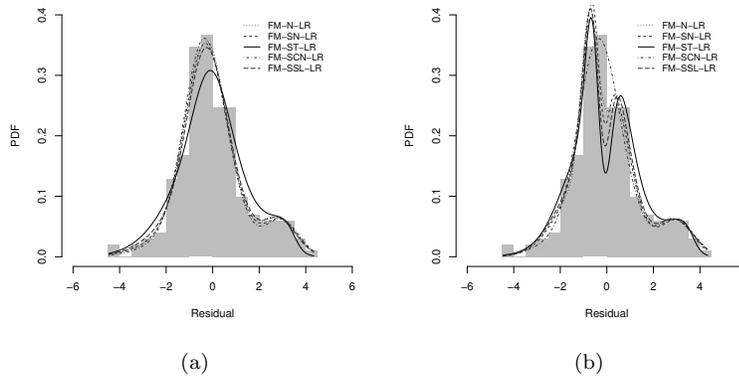


Figure 10. Panels (a) and (b) display the histogram of ordinary residuals with FM-SMSN-LR residual with for $g = 3$ and $g = 4$ components, respectively with AIS dataset.

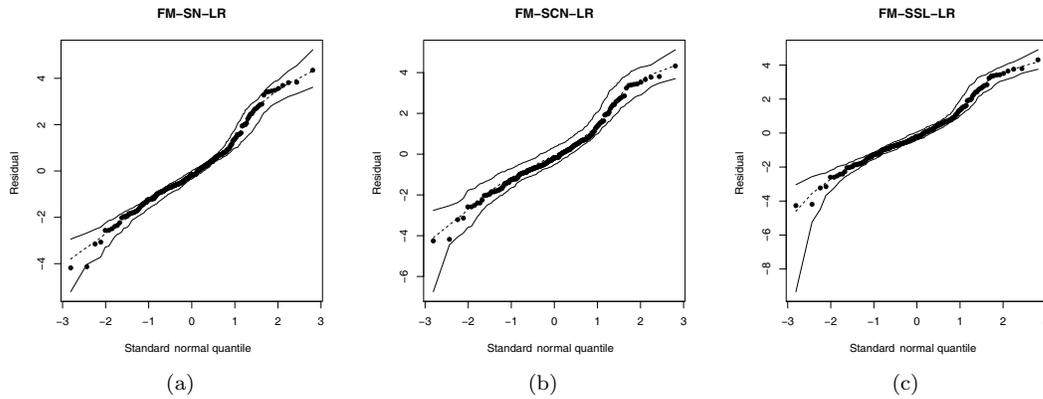


Figure 11. Panels (a), (b) and (c) display the QQ plots and simulated envelopes for the residual $(y - \hat{y})$ for $g = 2$ components based on FM-SN, FM-SCN and FM-SSL distributions, respectively with AIS dataset.

5. CONCLUSIONS

In this paper we consider a regression model whose error term follows a finite mixture of SMSN distributions, which is a rich class of distributions that contains the skew-normal, skew-t, skew-slash and skew-contaminated normal distributions as proper elements. This

approach allows us to model data with great flexibility, simultaneously accommodating multimodality, skewness and heavy tails for the random error in linear regression models. It is important to stress that our proposal is different from that of [Zeller et al. \(2016\)](#), where they use a finite mixture of linear regression models, the so-called switching regression. In this paper, instead of mixtures of regressions, mixtures are exploited as a convenient semiparametric method, which lies between parametric models and kernel PDF estimators, to model the unknown distributional shape of the errors. For this structure we developed a simple EM-type algorithm to perform ML inference of the parameters with closed-form expression at the E-step. The proposed methods are implemented using the `FMsmnReg` package, providing practitioners with a convenient tool for further applications in their domain. The practical utility of the new method is illustrated with the analysis of a real dataset and several simulation studies.

The proposed methods can be extended to multivariate settings using the multivariate SMSN class of distributions ([Cabral et al., 2012](#)), such as the recent proposals of [Soffritti and Galimberti \(2011\)](#) and [Galimberti and Soffritti \(2014\)](#). Due to the popularity of Markov chain Monte Carlo techniques, another potential work is to pursue a fully Bayesian treatment in this context for producing posterior inference. The method can also be extended to mixtures of regressions with skewed and heavy-tailed censored responses, based on recent approaches by [Caudill \(2012\)](#) and [Karlsson and Laitila \(2014\)](#).

APPENDIX: SAMPLE OUTPUT FROM THE `FMsmnReg` PACKAGE

```

-----
Finite Mixture of Scale Mixture Skew Normal Regression Model
-----
Observations = 202
Family = Skew.t

-----
Estimates
-----
      Estimate      SE
beta0  14.51593  0.00253
beta1   0.17972  0.00850
beta2  -0.07536  0.19264
mu1     1.93244  4.00942
mu2    -0.35226  2.94875
sigma1  3.80681  1.57056
sigma2  1.06550 11.5669
shape1 -5.70438  0.52991
shape2 -0.62860  9.52263
pi11    0.15418  1.04192
nu       7.45874    NA

-----
Model selection criteria
-----
Loglik      AIC      BIC      EDC      ICL
Value -357.030 730.235 766.626 739.502 2916.687

-----
Details
-----
Convergence reached? = TRUE
EM iterations = 147 / 500
Criteria = 1e-07
Processing time = 27.11465 secs

```

ACKNOWLEDGMENT

The authors thank the Editors and Reviewers for their constructive comments on an earlier version of this manuscript. Luis Benites and Rocío Maehara acknowledges support from CNPq-Brazil. Victor H. Lachos was supported from CNPq-Brazil (Grant 306334/2015-1). Partial support from CAPES is also acknowledged.

REFERENCES

- Andrews, D.F. and Mallows, C.L., 1974. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society B*, 36, 99-102.
- Arellano-Valle, R.B. and Genton, M.G., 2010. Multivariate unified skew-elliptical distributions. *Chilean Journal of Statistics*, 1, 17-33.
- Azzalini, A., 1985. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171-178.
- Azzalini, A. and Genton, M., 2008. Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review*, 76, 1490-1507.
- Bartolucci, F. and Scaccia, L., 2005. The use of mixtures for dealing with non-normal regression errors. *Computational Statistics and Data Analysis*, 48, 821-834.
- Basford, K., Greenway, D., McLachlan, G., and Peel, D., 1997. Standard errors of fitted component means of normal mixtures. *Computational Statistics*, 12, 1-18.
- Basso, R.M., Lachos, V.H., Cabral, C.R.B., Ghosh, P., 2010. Robust mixture modeling based on scale mixtures of skew-normal distributions. *Computational Statistics and Data Analysis*, 54, 2926-2941.
- Benites, L., Maehara, R., and Lachos, V.H., 2016. FMsmnReg: Regression Models with Finite Mixtures of Skew Heavy-Tailed Errors. R package version 1.0. <http://CRAN.R-project.org/package=FMsmnReg>
- Böhning, D., Dietz, E., Schaub, R., Schlattmann, P., and Lindsay, B.G., 1994. The distribution of the likelihood ratio for mixtures of densities from the one-parameter exponential family. *Annals of the Institute of Statistical Mathematics*, 46, 373-388.
- Branco, M.D. and Dey, D.K., 2001. A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, 79, 99-113.
- Cabral, C.R.B., Lachos, V.H., and Prates, M.O., 2012. Multivariate mixture modeling using skew-normal independent distributions. *Computational Statistics and Data Analysis*, 56, 126-142.
- Caudill, S.B., 2012. A partially adaptive estimator for the censored regression model based on a mixture of normal distributions. *Statistical Methods and Applications*, 21, 121-137.
- Cook, R.D. and Weisberg, S., 1982. *Residuals and Influence in Regression*. Chapman and Hall/CRC, Boca Raton, FL, US.
- Dempster, A., Laird, N., and Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39, 1-38.
- Galea, M., Paula, G.A., and Bolfarine, H., 1997. Local influence in elliptical linear regression models. *Journal of the Royal Statistical Society D*, 46, 71-79.
- Galimberti, G. and Soffritti, G., 2014. A multivariate linear regression analysis using finite mixtures of t distributions. *Computational Statistics and Data Analysis*, 71, 138-150.
- Garay, A.M., Lachos, V.H., and Abanto-Valle, C.A., 2011. Nonlinear regression models based on scale mixtures of skew-normal distributions. *Journal of the Korean Statistical Society*, 40, 115-124.
- Holzmann, H. and Munk, A., 2006. Identifiability of finite mixture of elliptical distributions. *Scandinavian Journal of Statistics*, 33, 753-763.

- Hurvich, C. and Tsai, C.L., 1989. Regression and time series model selection in small samples. *Biometrika*, 76, 297-307.
- Karlsson, M. and Laitila, T., 2014. Finite mixture modeling of censored regression models. *Statistical Papers*, 55, 627-642.
- Lange, K.L., Little, R., and Taylor, J., 1989. Robust statistical modeling using t distribution. *Journal of the American Statistical Association*, 84, 881-896.
- Lange, K.L. and Sinsheimer, J.S., 1993. Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics*, 2, 175-198.
- Lindsay, B., 1995. *Mixture models: Theory, geometry and applications*. Institute of Mathematical Statistics, Hayward, CA, US.
- Liu, C. and Rubin, D.B., 1994. The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika*, 80, 267-278.
- McNicholas, P., Murphy, T., and McDaid, A.D.F., 2010. Serial and parallel implementations of model-based clustering via parsimonious gaussian mixture models. *Computational Statistics and Data Analysis*, 54, 711-723.
- Otianiano, C., Rathie, P., and Ozelim, L., 2015. On the identifiability of finite mixture of skew-normal and skew-t distributions. *Statistics and Probability Letters*, 106, 103-108.
- Prates, M.O., Lachos, V.H., and Cabral, C.B., 2013. mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions. *Journal of Statistical Software*, 12, 1-20.
- Prates, M.O., Dey, D.K., and Lachos, V.H., 2012. A dengue fever study in the state of Rio de Janeiro with the use of generalized skew-normal/independent spatial fields. *Chilean Journal of Statistics*, 3, 143-155.
- Quandt, R.E. and Ramsey, J.B., 1978. Estimating mixtures of normal distributions and switching regressions. *Journal of the American Statistical Association*, 73, 730-738.
- R Core Team, 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rao, C.R., 1973. *Linear Statistical Inference and its Applications*. Wiley, New York.
- Sclove, S., 1987. Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52, 333-343.
- Soffritti, G. and Galimberti, G., 2011. Multivariate linear regression with non-normal errors: a solution based on mixture models. *Statistics and Computing*, 21, 523-536.
- Turner, T.R., 2000. Estimating the propagation rate of a viral infection of potato plants via mixtures of regressions. *Journal of the Royal Statistical Society C*, 49, 371-384.
- Zeller, C.B., Cabral, C.R., Lachos, V.H., 2016. Robust mixture regression modeling based on scale mixtures of skew-normal distributions. *TEST*, 25, 375-396.

GOODNESS-OF-FIT METHODS
RESEARCH PAPER

Goodness-of-fit test for the Birnbaum-Saunders distribution based on the Kullback-Leibler information

EDNÁRIO MENDONÇA¹, MICHELLI BARROS^{2,*}, and JOELSON CAMPOS²

¹Departamento de Estatística, Universidade Estadual da Paraíba, Campina Grande, Brazil,

²Departamento de Estatística, Universidade Federal de Campina Grande, Campina Grande, Brazil

(Received: 08 March 2019 · Accepted in final form: 18 April 2019)

Abstract

In this work, we propose a goodness-of-fit test based on the Kullback-Leibler information for the Birnbaum-Saunders distribution. We use Monte Carlo simulations to evaluate the size and power of the proposed test for several alternative hypotheses under different sample sizes. We compare the powers with standard goodness-of-fit tests based as the Anderson-Darling and Cramér-von Mises tests. Finally, we illustrate the proposed test with a real data set to show its potential applications.

Keywords: Anderson-Darling and Cramér-von Mises tests · Information measures · Maximum likelihood estimation · Monte Carlo method · Power test · R software

Mathematics Subject Classification: Primary 62J20 · Secondary 62J99.

1. INTRODUCTION

The Birnbaum-Saunders (BS) model, proposed by [Birnbaum and Saunders \(1969\)](#), is a life distribution originating from a material fatigue problem, which relates the time to the occurrence of failure with some cumulative damage that is assumed to be Gaussian distributed. The BS model has received much attention in the last decades due to its wide applicability. Based on to its genesis from material fatigue, different cumulative damage processes can be modeled by this distribution, including natural engineering applications, but the BS model can also be applied to other areas as: medicine ([Leiva et al., 2007](#); [Barros et al., 2008](#); [Azevedo et al., 2012](#); [Gomes et al., 2012](#); [Desousa et al., 2018](#); [Leao et al., 2018](#)), atmospheric contamination ([Leiva et al., 2008, 2010, 2015a](#); [Vilca et al., 2011](#); [Ferreira, 2013](#); [Marchant et al., 2018, 2019](#)), water quality ([Leiva et al., 2009](#); [Vilca et al., 2010](#)), neuronal sciences ([Leiva et al., 2015b](#)), human aging ([Leiva and Saunders, 2015](#)), and earthquakes ([Lillo et al., 2018](#)), among others. However, because the BS model is a statistical distribution, we can apply it to several other fields, for example, business, finance, industry, science management, and quality control. For more details about various

*Corresponding author. Email: michelli.karinne@gmail.com

developments on the BS distribution, see [Leiva \(2016\)](#) and references cited therein. The BS model has also been used to construct new more flexible models having heavier and lighter tails than the standard BS distribution, as well as in the construction of models in the unit interval; see [Barros et al. \(2008\)](#), [Azevedo et al. \(2012\)](#), [Mazucheli et al. \(2018\)](#) and [Athayde et al. \(2019\)](#).

In statistics, it is of great interest to determine whether a probabilistic model fits a data set well or not, which could indicate whether these data may have been generated from this model or not. In this sense, several goodness-of-fit tests have been proposed for different probability distributions. Since goodness-of-fit tests measure the discrepancy between a theoretical model and a data set, they can be done in a variety of ways, such as, for example, formulated by chi-squared type tests, by statistics based on the empirical cumulative distribution function or empirical characteristic function. Further details on goodness-of-fit tests can be found in [D'Agostino and Stephens \(1986\)](#), [Castro-Kuriss \(2011\)](#) and [Barros et al. \(2014\)](#).

The Anderson-Darling (AD) and Cramér-von Mises (CM) statistics are often used to test normality. These statistics are based on the distance between the empirical distribution function and the theoretical distribution function. [Chen and Balakrishnam \(1995\)](#) proposed a general purpose approximate goodness-of-fit test based on these statistics which may be used to test the validity of different families of skew distributions. Note that the Kullback-Leibler (KL) criterion is an information measure, which can be used to evaluate the discrepancy between two distribution functions. Such a measure of information has shown good results in testing fitting of models to data sets, in the sense of obtaining more powerful tests than the standard tests; see [Park \(2005\)](#) and [Rad et al. \(2011\)](#). Then, due to the wide applicability of the BS distribution, the objective of this paper is to propose a goodness-of-fit test for the BS distribution based on the KL information and investigate if the proposed test is most powerful than in the case of standard AD and CM tests.

The rest of this paper is organized as follows. In [Section 2](#), we present the methodology with the definitions of entropy, KL information, and a brief review of the BS distribution, as well as an estimation method of its parameters. In addition, in this section, goodness-of-fit test for the BS distribution based on KL information are derived. In [Section 3](#), a simulation study based on the Monte Carlo method is conducted to evaluate the size and power of the proposed test. Also in this section, we illustrate the proposed methodology with a real data set. Finally, [Section 4](#) provides the conclusions of this work and some comments on future research related to this topic.

2. METHODOLOGY

2.1 ENTROPY AND KULLBACK-LEIBLER INFORMATION

In order to quantify the degree of disorder in a physical system the German Rudolph Clausius introduced in [Clausius \(1867\)](#) a new quantity in thermodynamics which he called entropy. Since this concept was introduced in studies of information theory by [Shannon \(1948\)](#). Shannon's idea was to measure the degree of disorder of the occurrence of the values of a random variable (RV) in the sense that the more distinct rare events occur.

Let X be an RV with cumulative distribution function (CDF) F and probability density function (PDF) f . The differential entropy $H(f)$ of X is defined in [Shannon \(1948\)](#) by

$$H(f) = - \int_{-\infty}^{\infty} f(x) \log(f(x)) dx.$$

Let X_1, \dots, X_n , with $n \geq 3$, be a sample from the distribution F , and let $X_{(1)} \leq \dots \leq X_{(n)}$

be their corresponding order statistics. A nonparametric estimator of $H(f)$, proposed by Vasicek (1976), is given by

$$H_{mn} = \frac{1}{n} \sum_{i=1}^n \log \left\{ \frac{n}{2m} (x_{(i+m)} - x_{(i-m)}) \right\}, \quad (1)$$

where the window m is a positive integer less than $n/2$ and $x_{(i-m)} = x_{(1)}$, for $i - m < 1$ and $x_{(i+m)} = x_{(n)}$, for $i + m > n$, such that $x_{(i)}$ is i -th observed value of the corresponding order statistic.

Let $f(x)$ and $g(x)$ be PDFs. The KL information is defined in Kullback and Leibler (1951) as

$$I(f: g) = \int_{-\infty}^{\infty} f(x) \log \left[\frac{f(x)}{g(x)} \right] dx, \quad (2)$$

so that $I(f: g)$ measures the divergence between the PDFs f and g . By using the Gibbs inequality, we can show that $I(f: g) \geq 0$ and $I(f: g) = 0$ if and only if $f(x) = g(x)$. Thus, the sample estimate of the KL information can also be considered for goodness of fit.

2.2 THE BIRNBAUM-SAUNDERS DISTRIBUTION

Let X be a nonnegative RV. Then, X follows a BS distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$, if the CDF of X is given by

$$F(x) = \Phi \left[\frac{1}{\alpha} \left(\sqrt{\frac{x}{\beta}} - \sqrt{\frac{\beta}{x}} \right) \right], \quad x > 0.$$

We use the notation $X \sim \text{BS}(\alpha, \beta)$ for indicating an RV X with BS distribution of shape and scale parameters α and β , respectively. Consequently, the PDF of X is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2\alpha^2} \left(\frac{x}{\beta} + \frac{\beta}{x} - 2 \right) \right] \frac{x^{-3/2}(x + \beta)}{2\alpha\sqrt{\beta}}, \quad x > 0. \quad (3)$$

If $X \sim \text{BS}(\alpha, \beta)$, then the following properties are satisfied:

- (i) The parameter β is also the median of the distribution.
- (ii) If $Z \sim \text{N}(0, 1)$, then X and Z are related by $X = \beta(\alpha Z + (\alpha^2 Z^2 + 4)^{1/2})^2/4$. Thus, $Z = (1/\alpha)[(X/\beta)^{1/2} - (\beta/X)^{1/2}] \sim \text{N}(0, 1)$.
- (iii) $cX \sim \text{BS}(\alpha, c\beta)$, if $c > 0$ and $1/X \sim \text{BS}(\alpha, 1/\beta)$.
- (iv) $E(X) = \beta(1 + \alpha^2/2)$ and $\text{Var}(X) = \beta^2\alpha^2(1 + 5\alpha^2/4)$.
- (v) The q th quantile of X is given by $x_q = \beta(\alpha z_q + (\alpha^2 z_q^2 + 4)^{1/2})^2/4$, where $z_q = \Phi^{-1}(q)$, $\text{N}(0, 1)$ q th quantile.
- (vi) The survival function is expressed as $S(x; \alpha, \beta) = \Phi\{(1/\alpha)[(\beta/x)^{1/2} - (x/\beta)^{1/2}]\}$.

For estimation of the model parameters, we consider the maximum likelihood (ML) method. Let X_1, \dots, X_n be a random sample of size n from $X \sim \text{BS}(\alpha, \beta)$ with PDF given by Equation PDF), so that x_1, \dots, x_n are their respective observed values. Then,

the log-likelihood function for $\boldsymbol{\theta} = (\alpha, \beta)^\top$ is given by

$$\ell(\boldsymbol{\theta}) = K - \frac{1}{2\alpha^2} \sum_{i=1}^n \left(\frac{x_i}{\beta} + \frac{\beta}{x_i} - 2 \right) + \sum_{i=1}^n \log(x_i + \beta) - n \log(\alpha) - \frac{n}{2} \log(\beta),$$

where $K = n(\log(1/\sqrt{2\pi}) - \log(2)) - 3/2 \sum_{i=1}^n \log(x_i)$. The ML estimate of α is defined as

$$\hat{\alpha} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{x_i}{\hat{\beta}} + \frac{\hat{\beta}}{x_i} - 2 \right)}.$$

In the case of the parameter β , the ML estimate do not have closed form requiring the use of a numerical method. Under regularity conditions (see [Cox and Hinkley, 1974](#)), the estimators $\hat{\alpha}$ and $\hat{\beta}$ are consistent and have a bivariate normal joint asymptotic distribution with asymptotic means α and β , respectively, and an asymptotic covariance matrix $\boldsymbol{\Sigma}_{\hat{\theta}}$ that can be obtained from the inverse of the Fisher information matrix given by

$$\mathcal{I}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{2n}{\alpha^2} & 0 \\ 0 & \frac{n}{\beta^2} \left(\frac{1}{4} + \frac{1}{\alpha^2} + I(\alpha) \right) \end{pmatrix},$$

where

$$I(\alpha) = 2 \int_0^\infty \left(\frac{1}{1 + \frac{1}{\xi(az)}} - \frac{1}{2} \right)^2 \phi(z) dz,$$

with ϕ being the PDF of $Z \sim N(0, 1)$ and $\xi(u) = u^{1/2} - u^{-1/2}$. For more details, see [Leiva \(2016\)](#).

2.3 GOODNESS-OF-FIT TESTS FOR THE BS DISTRIBUTION

Given a random sample X_1, \dots, X_n of the RV X , we are interested in testing H_0 : the RV X follows the $BS(\alpha, \beta)$ distribution with PDF given in Equation (3) against H_1 : the RV X does not follow the BS distribution. Note that Equation (2) can be written as

$$\begin{aligned} I(f: g) &= \int_{-\infty}^{\infty} f(x) [\log(f(x)) - \log(g(x))] dx \\ &= -H(f) - \int_{-\infty}^{\infty} f(x) \log(g(x)) dx. \end{aligned} \quad (4)$$

Then, from Equation (4), an estimate of the KL information can be obtained. For doing this, we replace $H(f)$ by its estimate given in Equation (1) and we use the estimated values of the parameters in f . Thus, under the null hypothesis that $f(x) = g(x)$, we can estimate the information of KL using

$$I_{mn} = -H_{mn} - \int_{-\infty}^{\infty} f(x; \hat{\boldsymbol{\theta}}) \log(f(x; \hat{\boldsymbol{\theta}})) dx,$$

where $\hat{\boldsymbol{\theta}}$ is a consistent estimator for $\boldsymbol{\theta}$. Therefore, I_{mn} is a test statistic to verify the suitability of a continuous probabilistic model with PDF given by f to a data set.

For $X \sim \text{BS}(\alpha, \beta)$ and f given in Equation (3), we obtain

$$I_{mn} = -H_{mn} - \log \frac{1}{\sqrt{2\pi}} - \frac{1}{\hat{\alpha}^2} + \log \left(2\hat{\alpha}\sqrt{\hat{\beta}} \right) + \frac{1}{\hat{\alpha}^2} \left(1 + \frac{\hat{\alpha}^2}{2} \right) \\ + \frac{3}{2n} \sum_{i=1}^n \log(x_{(i)}) - \frac{1}{n} \sum_{i=1}^n \log(x_{(i)} + \hat{\beta}),$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the ML estimates of α and β , respectively. Thus, following [Arizono and Ohta \(1989\)](#), we introduce the statistic

$$\text{KL}_{mn} = \frac{1}{\exp(I_{mn})},$$

with $0 \leq \text{KL}_{mn} \leq 1$ since $I_{mn} \in [0, \infty)$. Note that KL_{mn} can be used as test statistic for testing the goodness-of-fit of the BS distribution to a data set. The decision rule is to reject the hypothesis H_0 if $\text{KL}_{mn} \leq \text{KL}_{mn}^*(\rho)$, where $\text{KL}_{mn}^*(\rho)$ is the critical value for a significance level ρ . As we do not have an exact distribution of KL_{mn} , then we obtain $\text{KL}_{mn}^*(\rho)$ through Monte Carlo simulations.

3. NUMERICAL STUDIES

3.1 CRITICAL VALUES FOR THE SIMULATIONS

To obtain the critical values of the proposed test, we conduct Monte Carlo simulation studies with $R = 10,000$ replications each. These studies are based on $n \in \{10, 30, 50, 100\}$, $\alpha \in \{0.5, 1.0, 1.5\}$, and significance level $\rho = 0.05$. In addition, we fix, without loss of generality, $\beta = 1$, since this is a scale parameter. The values considered for the window m are those returned the maximum critical value, according to [Arizono and Ohta \(1989\)](#). This procedure is described in Algorithm 1. All simulations are obtained from implementations in the R statistical software, which is freely distributed from www.R-project.org. For parameters estimation we use the `maxLik` package.

Algorithm 1: Obtaining the critical values of the proposed test.

- 1: Fix n , α and β ;
- 2: Generate 10,000 random samples of size n from $X \sim \text{BS}(\alpha, \beta)$;
- 3: For each sample, estimate the parameter vector $\boldsymbol{\theta} = (\alpha, \beta)^\top$ consistently, through the ML method;
- 4: For each sample, obtain the values of the test statistic KL_{mn} ;
- 5: Sort the test statistic values obtained in the previous step and determine the 5th quantile and then obtain the critical values for the respective significance level.

The critical values obtained, considering the $\text{BS}(0.5, 1)$, $\text{BS}(1, 1)$ and $\text{BS}(1.5, 1)$ distributions are presented in Tables 1-3.

3.2 EVALUATING THE EMPIRICAL SIZE AND POWER OF THE TEST

Next, the empirical size and power of the proposed test are evaluated for different sample sizes based on the Monte Carlo method. We make a comparison among the AD, CM and KL tests, whose statistics are denoted by A^2 , W^2 , KL, and verify in what situations the test based on the KL information is better, in the sense of being most powerful.

Table 1. Critical values for the statistic KL_{mn} considering the $BS(0.5,1)$ distribution and significance level 5%.

n	m									
	1	2	3	4	5	6	7	8	9	10
3	0.2462									
4	0.2577									
5	0.2925	0.4221								
6	0.3256	0.4404								
7	0.3544	0.4620	0.4835							
8	0.3866	0.4935	0.5083							
9	0.4054	0.5102	0.5319	0.5168						
10	0.4250	0.5340	0.5481	0.5401						
12	0.4614	0.5689	0.5840	0.5760	0.5625					
14	0.4911	0.5908	0.6114	0.6072	0.5973	0.5771				
16	0.5159	0.6207	0.6383	0.6354	0.6227	0.6069	0.5880			
18	0.5308	0.6396	0.6597	0.6605	0.6461	0.6331	0.6184	0.5980		
20	0.5499	0.6564	0.6820	0.6796	0.6674	0.6542	0.6428	0.6250	0.6082	
25	0.5754	0.6871	0.7176	0.7194	0.7124	0.7042	0.6905	0.6769	0.6617	0.6489
30	0.5976	0.7132	0.7421	0.7474	0.7481	0.7384	0.7280	0.7153	0.7036	0.6899
35	0.6122	0.7297	0.7593	0.7699	0.7707	0.7655	0.7577	0.7473	0.7352	0.7254
40	0.6243	0.7423	0.7766	0.7904	0.7900	0.7860	0.7789	0.7720	0.7620	0.7527
45	0.6343	0.7547	0.7887	0.8007	0.8053	0.8034	0.7975	0.7917	0.7832	0.7765
50	0.6426	0.7634	0.7982	0.8129	0.8165	0.8142	0.8146	0.8062	0.8027	0.7935
60	0.6568	0.7755	0.8135	0.8291	0.8350	0.8368	0.8355	0.8330	0.8274	0.8235
70	0.6646	0.7854	0.8251	0.8421	0.8498	0.8515	0.8522	0.8501	0.8476	0.8435
80	0.6751	0.7959	0.8349	0.8514	0.8596	0.8641	0.8644	0.8649	0.8628	0.8595
90	0.6804	0.8012	0.8408	0.8598	0.8687	0.8735	0.8758	0.8742	0.8733	0.8718
100	0.6858	0.8075	0.8471	0.8656	0.8760	0.8818	0.8833	0.8841	0.8826	0.8813

Under same the conditions of the obtained critical values, we calculate the empirical size of the test. Algorithm 2 displays this procedure. The results of our simulation study are presented in Table 4. Note that the empirical size is close to the nominal level for all situations considered, indicating that the test is controlled.

Algorithm 2: Obtaining the empirical size of the proposed test.

- 1: Fix n , α and β ;
- 2: Generate 10,000 random samples of size n from $X \sim BS(\alpha, \beta)$;
- 3: For each sample, estimate the parameter vector $\theta = (\alpha, \beta)^\top$ consistently, through the ML method;
- 4: For each sample, obtain the values of the test statistic KL_{mn} ;
- 5: Obtain the empirical size of the test by calculating the proportion of replications that present test statistic value less than the critical value for the corresponding values of n and m .

To determine the empirical power, we consider some probability distributions for the alternative hypothesis. These distributions are chosen and grouped into classes to be analyzed according to the shape of their hazard function: increasing, decreasing and non-monotonous. The probability distributions considered in the evaluation of the power test

Table 2. Critical values for the statistic KL_{mn} considering the BS(1,1) distribution and significance level 5%.

n	m									
	1	2	3	4	5	6	7	8	9	10
3	0.2618									
4	0.2724									
5	0.3066	0.4446								
6	0.3369	0.4698								
7	0.3653	0.4947	0.5095							
8	0.3974	0.5215	0.5398							
9	0.4132	0.5349	0.5650	0.5420						
10	0.4350	0.5575	0.5809	0.5691						
12	0.4681	0.5870	0.6143	0.6133	0.5941					
14	0.4960	0.6066	0.6390	0.6408	0.6338	0.6115				
16	0.5202	0.6338	0.6610	0.6654	0.6591	0.6464	0.6292			
18	0.5342	0.6508	0.6803	0.6871	0.6807	0.6709	0.6592	0.6416		
20	0.5539	0.6671	0.6974	0.7049	0.6989	0.6923	0.6854	0.6683	0.6558	
25	0.5778	0.6946	0.7299	0.7387	0.7371	0.7358	0.7279	0.7177	0.7056	0.6986
30	0.5987	0.7193	0.7525	0.7621	0.7686	0.7639	0.7587	0.7502	0.7436	0.7357
35	0.6140	0.7341	0.7680	0.7823	0.7863	0.7856	0.7827	0.7776	0.7709	0.7651
40	0.6258	0.7460	0.7830	0.7990	0.8038	0.8031	0.7997	0.7968	0.7911	0.7869
45	0.6355	0.7572	0.7942	0.8089	0.8156	0.8169	0.8157	0.8122	0.8085	0.8053
50	0.6436	0.7663	0.8028	0.8196	0.8261	0.8261	0.8289	0.8249	0.8231	0.8175
60	0.6575	0.7776	0.8169	0.8338	0.8422	0.8453	0.8471	0.8463	0.8431	0.8427
70	0.6651	0.7872	0.8279	0.8461	0.8555	0.8583	0.8603	0.8606	0.8596	0.8577
80	0.6753	0.7967	0.8375	0.8546	0.8637	0.8694	0.8717	0.8730	0.8727	0.8716
90	0.6806	0.8024	0.8429	0.8621	0.8722	0.8781	0.8815	0.8809	0.8816	0.8809
100	0.6859	0.8084	0.8484	0.8675	0.8789	0.8850	0.8877	0.8898	0.8895	0.8895

are: gamma, generalized exponential, beta, Pareto type I, Weibull, and half-normal, whose PDFs are the following:

- Gamma($\kappa; \theta$) with PDF

$$f_1(x; \kappa, \theta) = \frac{1}{\Gamma(\kappa)\theta^\kappa} x^{\kappa-1} \exp\left(-\frac{x}{\theta}\right), \quad x > 0, \kappa, \theta > 0$$

and CDF denoted by F_1 .

- GExp($\kappa; \theta$) with PDF

$$f_2(x; \kappa, \theta) = \kappa\theta x \exp\{-\theta x\}[1 - \exp(-\theta x)]^{\kappa-1}, \quad x > 0,$$

$\kappa, \theta > 0$, and CDF denoted by F_2 .

- Beta($\kappa; \theta$), with PDF

$$f_3(x; \kappa, \theta) = \frac{\Gamma(\kappa + \theta)}{\Gamma(\kappa)\Gamma(\theta)} x^{\kappa-1}(1 - x)^{\theta-1}, \quad 0 < x < 1,$$

$\kappa, \theta > 0$, and CDF denoted by F_3 .

Table 3. Critical values for the statistic KL_{mn} considering the BS(1.5,1) distribution and significance level 5%.

n	m									
	1	2	3	4	5	6	7	8	9	10
3	0.2819									
4	0.2911									
5	0.3237	0.4760								
6	0.3514	0.5053								
7	0.3796	0.5303	0.5440							
8	0.4102	0.5547	0.5791							
9	0.4256	0.5665	0.6065	0.5852						
10	0.4449	0.5865	0.6206	0.6114						
12	0.4763	0.6121	0.6529	0.6591	0.6446					
14	0.5036	0.6274	0.6734	0.6850	0.6848	0.6684				
16	0.5270	0.6509	0.6916	0.7056	0.7080	0.7029	0.6938			
18	0.5385	0.6651	0.7067	0.7234	0.7267	0.7247	0.7198	0.7136		
20	0.5564	0.6787	0.7191	0.7371	0.7430	0.7454	0.7462	0.7343	0.7317	
25	0.5806	0.7025	0.7460	0.7630	0.7704	0.7791	0.7791	0.7785	0.7741	0.7770
30	0.6010	0.7260	0.7643	0.7816	0.7940	0.7992	0.8004	0.8024	0.8021	0.8048
35	0.6149	0.7396	0.7777	0.7984	0.8080	0.8131	0.8167	0.8198	0.8209	0.8238
40	0.6273	0.7499	0.7900	0.8116	0.8210	0.8254	0.8293	0.8318	0.8340	0.8349
45	0.6365	0.7602	0.8008	0.8193	0.8298	0.8358	0.8404	0.8425	0.8447	0.8478
50	0.6443	0.7685	0.8080	0.8285	0.8385	0.8425	0.8496	0.8507	0.8541	0.8544
60	0.6581	0.7794	0.8206	0.8402	0.8507	0.8572	0.8614	0.8645	0.8661	0.8708
70	0.6657	0.7884	0.8309	0.8506	0.8618	0.8677	0.8720	0.8749	0.8778	0.8789
80	0.6758	0.7977	0.8397	0.8580	0.8694	0.8761	0.8808	0.8847	0.8867	0.8886
90	0.6807	0.8026	0.8449	0.8652	0.8760	0.8838	0.8891	0.8904	0.8934	0.8946
100	0.6856	0.8089	0.8502	0.8699	0.8824	0.8895	0.8936	0.8972	0.8988	0.9012

Table 4. Empirical size for different sample size and values of the parameter α indicated.

n	m	BS(0.5,1)	BS(1,1)	BS(1.5,1)
10	3	0.0473	0.0588	0.0494
30	5	0.0564	0.0563	0.0513
50	7	0.0520	0.0514	0.0454
100	8	0.0504	0.0515	0.0482

- Pareto($\kappa; \theta$), with PDF

$$f_4(x; \kappa, \theta) = \frac{\kappa \theta^\kappa}{x^{\kappa+1}}, \quad x \in [\theta, \infty), \quad \kappa, \theta > 0,$$

and CDF denoted by F_4 .

- Weibull($\kappa; \theta$), with PDF

$$f_5(x; \kappa, \theta) = \frac{\kappa}{\theta} \left(\frac{x}{\theta}\right)^{\kappa-1} \exp\left\{-\left(\frac{x}{\theta}\right)^\kappa\right\}, \quad x > 0,$$

$\kappa, \theta > 0$ and CDF denoted by F_5 .

- $HN(\theta)$, with PDF

$$f_6(x; \theta) = \frac{2\theta}{\pi} \exp\left(-\frac{x^2\theta^2}{\pi}\right), \quad x \geq 0, \theta > 0,$$

and CDF denoted by F_6 .

The power of the test is calculated based on testing the hypotheses

$$\begin{cases} H_0: X \sim BS(\alpha, \beta), \text{ for some } \alpha > 0 \text{ and } \beta > 0; \\ H_1: X \sim F_i(\boldsymbol{\theta}), \text{ with } \boldsymbol{\theta} > 0 \text{ and } i = 1, \dots, 6. \end{cases}$$

In the procedure, 10,000 Monte Carlo replications and sample sizes $n = 10, 30, 50, 100$ are considered. The powers of the tests are obtained at the significance level $\rho = 0.05$. For each value of n and each distribution in H_1 , with different parameters, the 10,000 samples are generated and the respective values of the test statistic are calculated. Based on the critical values presented in Tables 1-3, we obtain the rejection proportions based on the 10,000 simulated samples. In addition, the power of the test is evaluated based on the CM and AD statistics using the procedure proposed by [Chen and Balakrishnam \(1995\)](#). We make a comparison among the tests and verify in what situations the test based on the KL information is better, in the sense of being most powerful. Tables 5-8 present the powers for the test in question with sample sizes of $n = 10, n = 30, n = 50$ and $n = 100$, respectively.

Table 5. Empirical power for different forms of hazard functions and different distributions considering sample size $n = 10$.

Hazard function	Alternatives	KL_{mn}	W^2	A^2
Increasing	Gamma(3; 1)	0.1534	0.0873	0.0937
	GExp(3; 1)	0.1288	0.0805	0.0825
	Beta(2; 1)	0.6841	0.3970	0.4282
Decreasing	Gamma(0.5; 1)	0.0376	0.0890	0.0959
	GExp(0.5; 1)	0.0428	0.0938	0.1025
Nonmonotone	Pareto(2; 1)	0.4748	0.4070	0.4342
	Weibull(2; 1)	0.2405	0.1507	0.1617
	HN(3)	0.2434	0.2096	0.2298

Table 6. Empirical power for different forms of hazard functions and different distributions considering sample size $n = 30$.

Hazard function	Alternatives	KL_{mn}	W^2	A^2
Increasing	Gamma(3; 1)	0.2656	0.1856	0.2082
	GExp(3; 1)	0.2050	0.1544	0.1719
	Beta(2; 1)	0.9970	0.9053	0.9388
Decreasing	Gamma(0.5; 1)	0.3465	0.3959	0.5343
	GExp(0.5; 1)	0.3638	0.3937	0.5442
Nonmonotone	Pareto(2; 1)	0.9767	0.9039	0.9365
	Weibull(2; 1)	0.5458	0.4172	0.4559
	HN(3)	0.7164	0.6576	0.6987

Table 7. Empirical power for different forms of hazard functions and different distributions considering sample size $n = 50$.

Hazard function	Alternatives	KL_{mn}	W^2	A^2
Increasing	Gamma(3; 1)	0.3575	0.2774	0.3099
	GExp(3; 1)	0.2716	0.2187	0.2498
	Beta(2; 1)	1.0000	0.9905	0.9965
Decreasing	Gamma(0.5; 1)	0.6622	0.7369	0.8711
	GExp(0.5; 1)	0.6779	0.7394	0.8748
Nonmonotone	Pareto(2; 1)	0.9993	0.9922	0.9970
	Weibull(2; 1)	0.7317	0.6190	0.6671
	HN(3)	0.9026	0.8701	0.8999

Table 8. Empirical power for different forms of hazard functions and different distributions considering sample size $n = 100$.

Hazard function	Alternatives	KL_{mn}	W^2	A^2
Increasing	Gamma(3; 1)	0.4937	0.4829	0.5364
	GExp(3; 1)	0.3744	0.3807	0.4276
	Beta(2; 1)	1.0000	1.0000	1.0000
Decreasing	Gamma(0.5; 1)	0.9861	0.9786	0.9969
	GExp(0.5; 1)	0.9882	0.9772	0.9957
Nonmonotone	Pareto(2; 1)	1.0000	1.0000	1.0000
	Weibull(2; 1)	0.9134	0.8922	0.9256
	HN(3)	0.9948	0.9915	0.9958

According to our simulation study, we conclude that the goodness-of-fit test based on the KL information, in general, presents greater powers when compared to standard AD and CM tests, for small sample size. When the hazard function under alternative hypothesis is decreasing, the proposed test has difficulties in discriminating the models, leading to powers close to nominal levels. This is because the hazard functions considered under the alternative hypothesis closely approximate the hazard function of the BS distribution. In addition, as the sample size increases, the power of the test also increases, as expected.

3.3 EMPIRICAL ILLUSTRATION

Next, we consider a set of data related to fatigue life cycles of samples of 6061-T6 aluminum presented in [Birnbaum and Saunders \(1969\)](#). These specimens were cut at an angle parallel to the direction of rotation, oscillating at 18 cycles per second. They were exposed to a pressure with a maximum stress of 26000 psi (pounds per square inch). The data are presented in [Table 9](#).

We want to test the null hypothesis that the sample presented in [Table 9](#) follows the BS distribution. The model parameter estimates are $\hat{\alpha} = 0.1614$ and $\hat{\beta} = 392.7622$. The value observed for the test statistic is $kl_{mn} = 0.9270$, and the critical value for this case is $KL_{mn}^*(\rho) = 0.8834$, at the 5% significance level. Therefore, we do not reject the hypothesis that the data follow the BS distribution. [Figure 1](#) compares the empirical distribution function with the theoretical one. We can observe from this figure that the empirical and theoretical distribution functions are very close, which reinforces the conclusion reached by the test.

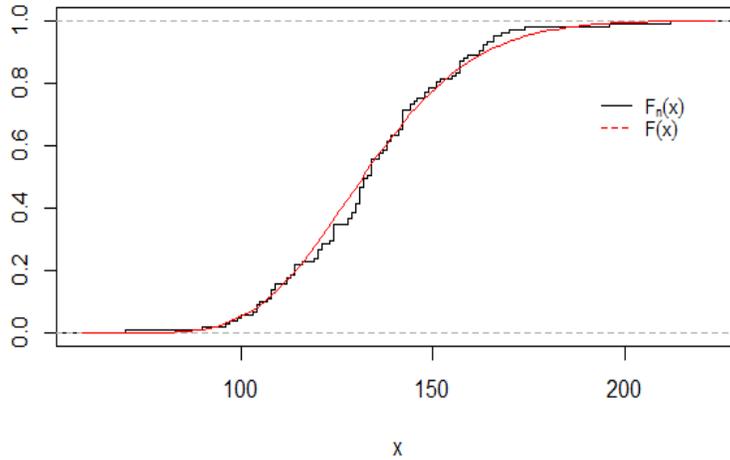


Figure 1. Empirical and theoretical distribution functions BS for aluminum data.

Table 9. Data set of aluminum lifetimes (26.000 psi).

233	258	268	276	290	310	312	315	318	321
321	329	335	336	338	338	342	342	342	344
349	350	350	351	351	352	352	356	358	358
360	362	363	366	367	370	370	372	372	374
375	376	379	379	380	382	389	389	395	396
400	400	400	403	404	406	408	408	410	412
414	416	416	416	420	422	423	426	428	432
432	433	433	437	438	439	439	443	445	445
452	456	456	460	464	466	468	470	470	473
474	476	476	486	488	489	490	491	503	517
540	560								

4. CONCLUSIONS AND FUTURE RESEARCH

In this paper, we proposed a goodness-of-fit test for the Birnbaum-Saunders distribution based on the Kullback-Leibler information. The proposed goodness-of-fit test performed better than the standard Anderson-Darling and Cramér-von Mises tests, in the sense that the proposed test had greater power for the alternatives considered with increasing and nonmonotone hazard functions. When the distribution of the alternative hypothesis had a decreasing hazard function, the test based in KL information presented less power than the Anderson-Darling and Cramér-von Mises tests. In general, the proposed test proved to be a good alternative to the standard Anderson-Darling and Cramér-von Mises tests. As future research, we hope to obtain new tests for the Birnbaum-Saunders distribution based on information measures for censored data, more specifically, for type II and progressively Type-II censored samples.

ACKNOWLEDGMENT

The authors wish to thank the Editors and anonymous referees for their constructive comments on an earlier version of this manuscript, which resulted in this improved version. The study was supported partially by CNPq and CAPES.

REFERENCES

- Arizono, I. and Ohta, H., 1989. A test for normality based on Kullback-Leibler information. *The American Statistician*, 43, 20-22.
- Athayde, E., Azevedo, A., Barros, M., and Leiva, V., 2019. Failure rate of Birnbaum-Saunders distributions: shape, change-point, estimation and robustness. *Brazilian Journal of Probability and Statistics*, 33, 301-328.
- Azevedo, C., Leiva, V., Athayde, E., and Balakrishnan, N., 2012. Shape and change point analyses of the Birnbaum-Saunders-t hazard rate and associated estimation. *Computational Statistics and Data Analysis*, 56, 3887-3897.
- Barros, M., Paula, G.A., and Leiva, V., 2008. A new class of survival regression models with heavy-tailed errors: robustness and diagnostics. *Lifetime Data Analysis*, 14, 316-332.
- Barros, M., Leiva, V., Ospina, R., and Tsuyuguchi, A., 2014. Goodness-of-fit tests for the Birnbaum-Saunders distribution with censored reliability data. *IEEE Transactions on Reliability*, 63, 543-554.
- Birnbaum, Z.W. and Saunders, S.C., 1969. Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability*, 6, 328-347.
- Castro-Kuriss, C., 2011. On a goodness-of-fit test for censored data from a location-scale distribution with application. *Chilean Journal of Statistics*, 37, 15-136.
- Chen, G. and Balakrishnan, N., 1995. A general purpose approximate goodness-of-fit test. *Journal of Quality Technology*, 27, 154-161.
- Clausius, R., 1867. *The mechanical theory of heat: with its applications to the steam-engine and to the physical properties of bodies*. J. van Voorst.
- Cox, D., Hinkley, D., 1974. *Theoretical Statistics*. Chapman and Hall, London, UK.
- D'Agostino, C. and Stephens, M.A., 1986. *Goodness of Fit Techniques*. Marcel Dekker, New York.
- Desousa, M., Saulo, H., Leiva, V., and Scalco, P., 2018. On a tobit-Birnbaum-Saunders model with an application to medical data. *Journal of Applied Statistics*, 45, 932-955.
- Ferreira, M.S., 2013. A study of exponential-type tails applied to Birnbaum-Saunders models. *Chilean Journal of Statistics*, 4, 87-97.
- Gomes, M.I., Ferreira, M., and Leiva, V., 2012. The extreme value Birnbaum-Saunders model and its moments and application in biometry. *Biometrical Letters*, 49, 81-94.
- Kullback, S. and Leibler, R.A., 1951. On information and sufficiency. *The annals of mathematical statistics*, 22, 79-86.
- Leao, J., Leiva, V., Saulo, H., and Tomazella, V., 2018. Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data. *Statistics in Medicine*, 37, 4421-4440.
- Leiva, V., Barros, M., Paula, G.A., and Galea, M., 2007. Influence diagnostics in log Birnbaum-Saunders regression models with censored data. *Computational Statistics and Data Analysis*, 51, 5694-5707.
- Leiva, V., Barros, M., Paula, G.A., and Sanhueza, A., 2008. Generalized Birnbaum-Saunders distributions applied to air pollutant concentration. *Environmetrics*, 19, 235-249.
- Leiva, V., Sanhueza, A., and Angulo, J.M., 2009. A length-biased version of the Birnbaum-Saunders distribution with application in water quality. *Stochastic Environmental Research and Risk Assessment*, 23, 299-307.
- Leiva, V., Vilca, F., Balakrishnan, N., and Sanhueza, A., 2010. A skewed sinhnormal distribution and its properties and application to air pollution. *Communications in Statistics*, 39, 426-443.
- Leiva, V. and Saunders, S.C., 2015. Cumulative damage models. *Wiley StatsRef: Statistics Reference Online* 1-10.

- Leiva, V., Marchant, C., Ruggeri, F., and Saulo, H., 2015a. A criterion for environmental assessment using Birnbaum-Saunders attribute control charts. *Environmetrics*, 26, 463-476.
- Leiva, V., Tejo, M., Guiraud, P., Schmachtenberg, O., Orio, P., and Marmolejo F., 2015b. Modeling neural activity with cumulative damage distributions. *Biological Cybernetics*, 109, 421-433.
- Leiva, V., 2016. *The Birnbaum-Saunders Distribution*. Academic Press, New York.
- Lillo, C., Leiva, V., Nicolis, O., and Aykroyd, R.G., 2018. L-moments of the Birnbaum-Saunders distribution and its extreme value version: Estimation, goodness of fit and application to earthquake data. *Journal of Applied Statistics*, 45, 187-209.
- Marchant, C., Leiva, V., Cysneiros, F.J.A., and Liu, S., 2018. Robust multivariate control charts based on Birnbaum-Saunders distributions. *Journal of Statistical Computation and Simulation*, 88, 182-202.
- Marchant, C., Leiva, V., Christakos, G., and Cavieres, M.A., 2019. Monitoring urban environmental pollution by bivariate control charts: new methodology and case study in Santiago, Chile. *Environmetrics*, pages in press, available at <https://doi.org/10.1002/env.2551>.
- Mazucheli, J., Menezes, A.F.B., and Dey, S., 2018. The unit-Birnbaum-Saunders distribution with applications. *Chilean Journal of Statistics*, 9, 47-57.
- Park, S., 2005. Testing exponentiality based on the Kullback-Leibler information with the type II censored data. *IEEE Transactions on Reliability*, 54, 22-26.
- Rad, A.H., Yousefzadeh F., and Balakrishnan N., 2011. Goodness-of-fit test based on Kullback-Leibler information for progressively type-II censored data. *IEEE Transactions on Reliability*, 60, 570-579.
- Shannon, C.E., 1948. A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423.
- Vasicek, O., 1976. A test for normality based on sample entropy. *Journal of the Royal Statistical Society B*, 38, 54-59.
- Vilca, F., Sanhueza, A., Leiva, V., and Christakos, G., 2010. An extended Birnbaum-Saunders model and its application in the study of environmental quality in Santiago, Chile. *Stochastic Environmental Research and Risk Assessment*, 24, 771-782.
- Vilca, F., Santana, L., Leiva, V., and Balakrishnan, N., 2011. Estimation of extreme percentiles in Birnbaum-Saunders distributions. *Computational Statistics and Data Analysis*, 55, 1665-1678.

DISTRIBUTION THEORY
RESEARCH PAPER

A new Birnbaum-Saunders type distribution based on the skew-normal model under a centered parameterization

NATHALIA L. CHAVES¹, CAIO L. N. AZEVEDO^{1,*}, FILIDOR VILCA-LABRA¹,
and JUVÊNCIO S. NOBRE²

¹Department of Statistics, Universidade Estadual de Campinas, Brazil

²Department of Statistics and Applied Mathematics, Universidade Federal de Ceará, Brazil

(Received: 22 February 2019 · Accepted in final form: 23 April 2019)

Abstract

In this paper, we introduce a new distribution for positively skewed data by combining the Birnbaum-Saunders and centered skew-normal distributions. Several of its properties are developed. Our model accommodates both positively and negatively skewed data. Also, we show that our proposal circumvents some problems related to another Birnbaum-Saunders distribution based on the usual skew-normal model, previously presented in the literature. We derive both maximum likelihood and Bayesian inference, comparing them through a suitable simulation study. The convergence of the expectation conditional maximization (for maximum likelihood inference) and MCMC algorithms (for Bayesian inference) are verified and several factors of interest are compared. In general, as the sample size increases, the results indicate that the Bayesian approach provided the most accurate estimates. Our model accommodates the asymmetry of the data more properly than the usual Birnbaum-Saunders distribution, which is illustrated through real data analysis.

Keywords: Bayesian inference · Birnbaum-Saunders distribution · ECM algorithm · Frequentist inference · MCMC algorithms · R software

Mathematics Subject Classification: Primary 60E05 · Secondary 62F15.

1. INTRODUCTION

The Birnbaum-Saunders (BS) distribution is characterized by two parameters and defined in terms of the standard normal distribution. The BS distribution has been received considerable attention over the past few years, since it has been used quite effectively to model positively skewed data, especially lifetime and crack growth data. Since the pioneering work of [Birnbaum and Saunders \(1969a\)](#) was published, several extensions of the BS distribution have been proposed in the literature and its parameters estimated under both frequentist and Bayesian paradigms.

*Corresponding author. Email: cnaber@ime.unicamp.br

From a frequentist viewpoint, [Birnbaum and Saunders \(1969b\)](#) presented a discussion on the maximum likelihood (ML) parameter estimation. [Mann et al. \(1974\)](#) showed that the BS distribution is unimodal. [Engelhardt et al. \(1981\)](#) developed confidence intervals and hypothesis tests for both parameters. [Desmond \(1985\)](#) developed a BS-type distribution based on a biological model. [Desmond \(1986\)](#) investigated the relationship between the BS distribution and the inverse Gaussian distribution. [Lu and Chang \(1997\)](#) used bootstrap methods to construct prediction intervals for future observations. In the linear regression context, [Rieck and Nedelman \(1991\)](#) developed a related log-linear model and showed that it can be used for modeling accelerated life tests and to compare average lifetime of different populations.

From a Bayesian perspective, there are few works on the BS distribution. The first one is due to [Achcar \(1993\)](#) who developed Bayesian estimation using numerical approximations for the marginal posterior distributions of interest based on the Laplace approximation. Also, [Xu and Tang \(2011\)](#) presented a Bayesian study with partial information, while [Wang et al. \(2016\)](#) assumed that the two parameters follow mutually independently inverse gamma distributions. All these results were studied under a normal distribution for generating the BS distribution.

In terms of modeling, most of the generalizations of the BS distribution are based on elliptical and skew-elliptical laws, in order to obtain more robust and flexible models. Some works developed extensions based on symmetric distributions as [Diaz-Garcia and Leiva \(2005\)](#) who generalized the BS model using elliptical distributions that includes the Cauchy, Laplace, logistic, normal and Student- t distributions as particular cases. Other works are: the generalized BS distribution ([Leiva et al., 2007](#)), the Student- t BS distribution ([Barros et al., 2008](#)), and the scale-mixture of normal BS distribution ([Balakrishnan et al., 2009](#)), among others. More information can be found in [Leiva \(2016\)](#), who presented a review on the BS distribution. Other generalizations have been obtained in different ways to those aforementioned, as [Owen and Padgett \(1999\)](#), who developed a three-parameter BS distribution and the β -BS distribution presented in [Cordeiro and Lemonte \(2011\)](#). Also, [Ferreira \(2013\)](#) proposed a based BS distribution useful for modeling tail events and [Mazucheli et al. \(2018\)](#) presented a distribution on the unit interval based on the BS model. In addition, [Balakrishnan et al. \(2017\)](#) and [Maehara \(2018\)](#) provided new families of BS distribution based on the skew scale mixture of normal models. Also, extensions of the BS distribution based on the skew-elliptical models can be found in [Vilca and Leiva \(2006\)](#), [Leiva et al. \(2007, 2008\)](#) and [Vilca et al. \(2011\)](#). In these works, theoretical results were obtained, extending the properties of the BS and log-BS distributions.

A Bayesian perspective for the BS distributions based on skew-normal (SN) distribution did not receive much attention in the literature. Indeed, [Vilca et al. \(2011\)](#) considered, under a frequentist perspective, the BS distribution based on the SN model. However, even though the SN distribution has been applied with success in several fields, when the related asymmetry parameter is equals to zero, the associated Fisher information matrix is singular. Recently, to overcome this problem, [Arellano and Azzalini \(2008\)](#) and [Azzalini \(2013\)](#) explored a SN distribution under a convenient parameterization (proposed by [Azzalini \(1985\)](#) and deeper explored by [Pewsey \(2000\)](#)), the so-called centered parametrization (CP), which leads to a non-singular Fisher information matrix. Moreover, the relative profile log-likelihood function (RPLL) for the Pearson index of skewness exhibits a more regular behavior, closer to a quadratic function, and without a stationary point under null asymmetry. The resulting empirical distributions of the estimators under the CP, named CP estimators, are much closer to the normality than those obtained under the usual SN distribution, which is named direct parametrization estimators. All these desirable properties, related to the CP, may be transferred to the respective BS distribution based on the centered SN (CSN) model. It is worthwhile to mention that all the aforementioned

BS models (that consider the SN model) used the direct parametrization that is, likely, they inherit the above problems.

The main objective of this work is to propose an alternative to the skew-normal BS (SNBS) distribution proposed by [Vilca et al. \(2011\)](#), considering the CSN distribution, as the generator variable. The resulting BS-type distribution has advantages, in inference terms, over the SNBS distributions (including those obtained as particular cases of the more general families as those of [Balakrishnan et al. \(2009\)](#) and [Maehara \(2018\)](#)), similarly to those related to the CSN distribution, compared with the SN distribution. The specific objectives of this work are: to develop a BS distribution based on the CSN model, named centered skew-normal BS (CSNBS) distribution, highlighting its advantages over the SNBS distribution proposed by [Vilca et al. \(2011\)](#), and its main properties. Also, estimation procedures under both frequentist and Bayesian approaches are developed and compared, considering different scenarios. In addition, some model comparison statistics are studied. Finally, two real data sets are analyzed showing some advantages of the CSNBS model compared to the usual BS distribution.

The paper is outlined as follows. In [Section 2](#), we present our distribution and some motivation for its development. In [Section 3](#), the estimation methods are proposed and some statistics of model comparison are presented. In [Section 4](#), some simulation studies are presented and two real data sets are analyzed. Finally, in [Section 5](#), some additional comments and conclusions are provided.

2. THE CENTERED SKEW-NORMAL BS DISTRIBUTION

2.1 THE CENTERED SKEW-NORMAL DISTRIBUTION

A random variable Y is said to have a CSN distribution, denoted by $Y \sim \text{CSN}(\mu, \sigma, \gamma)$, where μ , σ and γ are the mean, the standard deviation and the Pearson coefficient of skewness, respectively, if its density is given by

$$f_Y(y) = 2 \frac{\sigma_z}{\sigma} \phi\left(\mu_z + \frac{\sigma_z}{\sigma}(y - \mu)\right) \Phi\left[\lambda\left(\mu_z + \frac{\sigma_z}{\sigma}(y - \mu)\right)\right] \frac{2}{\omega} \phi\left(\frac{y - \xi}{\omega}\right) \Phi\left[\lambda\left(\frac{y - \xi}{\omega}\right)\right], y \in \mathbb{R}, \quad (1)$$

where $\mu_z = r\delta$, $\sigma_z^2 = 1 - \mu_z^2$, $\lambda = \gamma^{1/3}s/\sqrt{r^2 + s^2\gamma^{2/3}(r^2 - 1)}$, $r = \sqrt{2/\pi}$, $\gamma = r\delta^3(4/\pi - 1)(1 - \mu_z^2)^{-3/2}$, $\delta = \lambda/\sqrt{1 + \lambda^2}$, $\xi = \mu - \sigma\gamma^{1/3}s$, $\omega = \sigma\sqrt{1 + \gamma^{2/3}s^2}$, and $s = [2/(4 - \pi)]^{1/3}$. The quantity λ is the asymmetry parameter, see [Azzalini \(1985\)](#). For $\mu = 0$ and $\sigma = 1$, we have the standard CSN distribution, denoted by $Y \sim \text{CSN}(0, 1, \gamma)$, whose density is given by

$$f_Y(y) = \frac{2}{\omega} \phi\left(\frac{y - \xi}{\omega}\right) \Phi\left[\lambda\left(\frac{y - \xi}{\omega}\right)\right], y \in \mathbb{R},$$

where $\xi = -\gamma^{1/3}s$ and $\omega = \sqrt{1 + \gamma^{2/3}s^2}$. For inferential purposes, a useful stochastic representation of Y is given by

$$Y = \frac{1}{\sigma_z} \{ \delta |X_0| + (1 - \delta^2)X_1 - \mu_z \}, \quad (2)$$

where $X_i \sim N(0, 1)$, for $i = 0, 1$, are independent and so $H = |X_0|$ follows a half-normal (HN) distribution, denoted by $\text{HN}(0, 1)$.

2.2 THE PROPOSED DISTRIBUTION

Here, we present the CSNBS distribution, which is defined similarly to the usual BS and the SNBS distributions by

$$T = \beta \left[\frac{\alpha Y}{2} + \sqrt{\left(\frac{\alpha Y}{2}\right)^2 + 1} \right]^2, \quad (3)$$

where $Y \sim \text{CSN}(0, 1, \gamma)$, α is the shape parameter, β is the location parameter, and γ is the asymmetry parameter. We use the following notation $T \sim \text{CSNBS}(\alpha, \beta, \gamma)$. The vector $(\alpha, \beta, \gamma)^\top$ is called centered parameter and based on the SN distribution, that is, $(\alpha, \beta, \lambda)^\top$ is named direct parameters. Following the same steps as in the usual BS distribution, we have that its density is given by

$$f_T(t) = 2\phi[a_{t;\mu,\sigma}(\alpha, \beta)] \Phi[\lambda a_{t;\mu,\sigma}(\alpha, \beta)] A_{t;\sigma}(\alpha, \beta), t > 0, \quad (4)$$

where $a_{t;\mu,\sigma}(\alpha, \beta) = \mu_z + \sigma_z a_t(\alpha, \beta)$, $A_{t;\sigma}(\alpha, \beta) = \sigma_z A_t(\alpha, \beta)$, $a_t(\alpha, \beta) = (\sqrt{t/\beta} - \sqrt{\beta/t})/\alpha$, $A_t(\alpha, \beta) = da_t(\alpha, \beta)/dt = t^{-3/2}(t + \beta)/(2\alpha\beta^{1/2})$, and the other quantities are previously defined. Note that for $\gamma = 0$, we have the usual BS distribution. The mean and variance of T (see Appendix A for more details) are given, respectively, by

$$\mathbb{E}(T) = \beta \left(1 + \frac{\alpha^2}{2} \right) \quad \text{and} \quad \text{Var}(T) = (\alpha\beta)^2 \left\{ 1 + \frac{\alpha^2}{4} [2\Delta - 1] \right\},$$

where $\Delta = 2(\pi - 3)(4/\pi^2)\delta^4[1 - (2\delta^2/\pi)]^{-2} + 3$.

The following theorem is very useful to develop both classical and Bayesian approaches since they lead to conditional distributions that allow us to implement, more easily, the EM algorithm, and simplify the Bayesian developments. For the use of standard MCMC software, such as WinBUGS, OpenBUGS, JAGS or Stan, see [Lunn et al. \(2000\)](#), [Lunn et al. \(2009\)](#), [Depaoli et al. \(2016\)](#) and [Carpenter et al. \(2016\)](#).

THEOREM 2.1 Let $T \sim \text{CSNBS}(\alpha, \beta, \gamma)$ as in Equation (3), and Y and H as defined in Equation (2). Then,

- (i) The conditional density of T , given $H = h$, can be expressed as

$$f_{T|H}(t|h) = \phi(\nu_h + a_t(\alpha_\delta, \beta)) A_t(\alpha_\delta, \beta),$$

where $\alpha_\delta = \alpha\sqrt{(1 - \delta^2)/(1 - r^2\delta^2)}$ and $\nu_h = -(\delta(h - r))/\sqrt{1 - \delta^2}$.

- (ii) $f_{H|T}(h|t) = \frac{\phi\left(h|\delta\sqrt{1-r^2\delta^2}\left(a_t(\alpha,\beta) + \frac{r\delta}{\sqrt{1-r^2\delta^2}}\right), 1-\delta^2\right)}{\Phi\left(\lambda\sigma_z\left(a_t(\alpha,\beta) + \frac{r\delta}{\sqrt{1-r^2\delta^2}}\right)\right)}$, $h > 0$. Moreover,

$$\mathbb{E}(H|T = t) = \eta_t + W_\Phi\left(\frac{\eta_t}{\tau}\right)\tau \quad \text{and} \quad \mathbb{E}(H^2|T = t) = \eta_t^2 + \tau^2 + W_\Phi\left(\frac{\eta_t}{\tau}\right)\eta_t\tau,$$

where $\eta_t = \delta\sqrt{1 - r^2\delta^2}(a_{t_i}(\alpha, \beta) + (r\delta)/\sqrt{1 - r^2\delta^2})$.

The density in Theorem 2.1 corresponds to the extended Birnbaum-Saunders (EBS) discussed in [Leiva et al. \(2008\)](#) and denoted by $\text{EBS}(\alpha_\delta, \beta, \sigma = 2, \nu_h)$. The proof of Theorem 2.1 is in the Appendix B.

Figures 1-3 present the density of the CSNBS distribution for different values of α , β and γ . From Figure 1, we have that for $\alpha = 0.2$ the density is concentrated around β ($\beta = 1$), and for $\alpha = 0.8$ the density is more asymmetric, with a higher variability. As α increases, the density becomes more flat, positively skewed and more dispersed, as it can be seen in Figure 2, for different values of α , fixing the other parameters. In addition, Figure 3 shows densities more concentrated around β for different values of α and β , with $\gamma = 0.9$. It is also possible to see that for large values of β , the density is more negatively skewed. Note that the distribution tends to be symmetric around β , for $\gamma = 0$ (the usual BS distribution) and/or for small values of α . Positive asymmetry is observed as α increases, β decreases and/or γ assumes positive values. In addition, negative asymmetry is observed as α decreases, β increases and/or γ assumes negative values. Another interesting point is that the CSNBS distribution may be negatively skewed, which is an unusual behavior for positive random variables. This feature makes our distribution a very useful alternative for modeling positive skewed data.

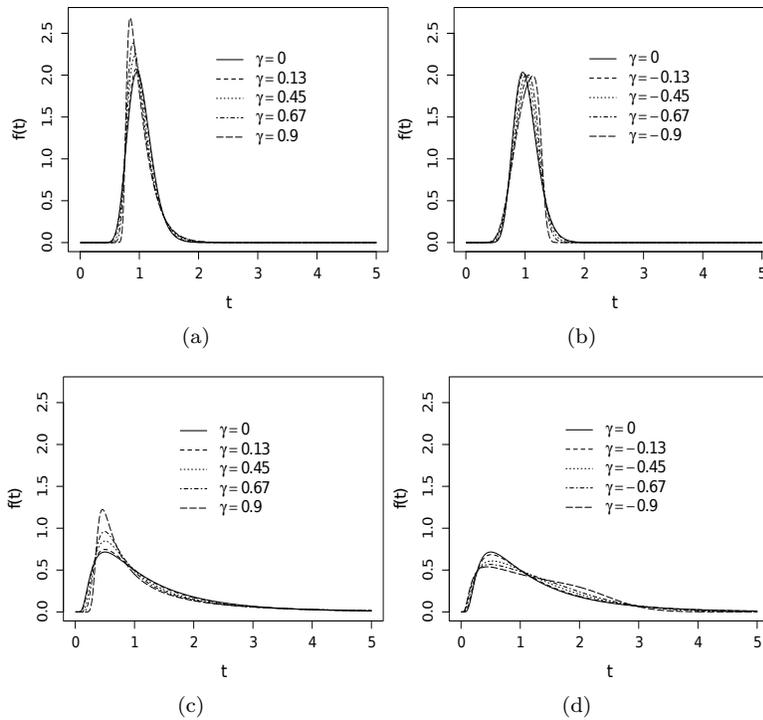


Figure 1. CSNBS density for different values of γ , with $\beta = 1$, $\alpha = 0.2$ (a)-(b) and $\alpha = 0.8$ (c)-(d).

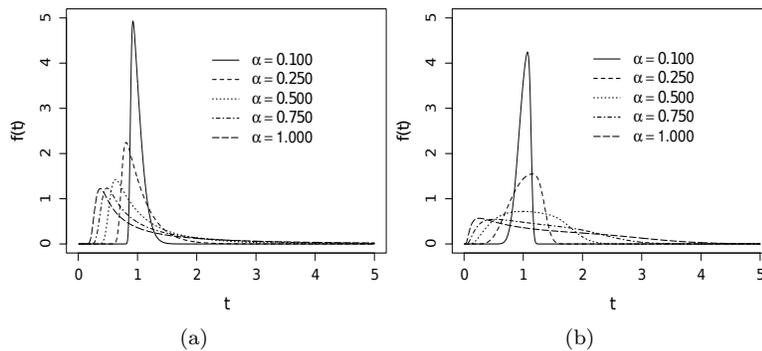


Figure 2. CSNBS density for different values of α , with $\beta = 1$, $\gamma = 0.9$ (a) and $\gamma = -0.9$ (b).

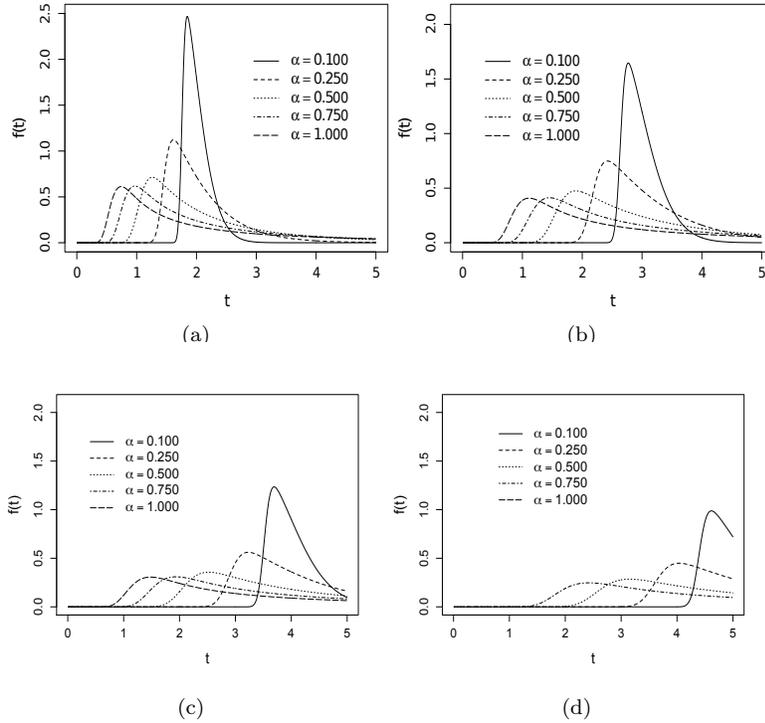


Figure 3. CSNBS density with $\beta = 2$ (a), $\beta = 3$ (b), $\beta = 4$ (c), and $\beta = 5$ (d) for indicated α and $\gamma = 0.9$.

2.3 SOME MOTIVATIONAL REMARKS ON THE PROPOSAL

- (i) It is well known that there is some difficulty in estimating the parameters of the SN distribution by the ML approach, when the asymmetry parameter is close to zero. Some problems seem to persist even if one switched to the Bayesian inference, unless a strongly informative prior is considered, as pointed out by [Arellano and Azzalini \(2008\)](#). The SNBS distribution seems to inherit such problems. Thus, the proposed CSNBS distribution can circumvent these problems, since it is based on the CSN model.
- (ii) When the asymmetry parameter is equals to zero, the Fisher information matrix is singular, even if all parameters are identifiable. This affects the behavior of the empirical distributions of the ML estimators and the Bayesian estimators. To get a direct perception of the problem, we generated 100 samples of size $n = 200$, from the SNBS distribution and for each sample, the ML and Bayesian estimates $(\hat{\alpha}, \hat{\beta}, \hat{\lambda})$ have been computed. In this case, we fix $\alpha = 0.5$, $\beta = 1$ and $\lambda = 1$, which induces a strong positively skewed behavior of the SNBS distribution. Figures 4 and 5 display the corresponding empirical distribution of $\hat{\alpha}$ and $(\hat{\alpha}, \hat{\beta})$, through a histogram and scatter plot, respectively. Moreover, 100 samples of size $n = 200$ are generated from the CSNBS distribution, and the respective ML and Bayesian estimates $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$ have been computed. In this case, we fix $\alpha = 0.5$, $\beta = 1$ and $\gamma = 0.137$, which induces a strong positively skewed behavior of the CSNBS model. The empirical distributions of $\hat{\alpha}$ and $(\hat{\alpha}, \hat{\beta})$ are shown in Figures 6 and 7, respectively. Clearly these empirical distributions are much closer to normality than those in Figures 4 and 5. In fact, it can be shown that the singularity of the expected Fisher information matrix, when the asymmetry parameter is null, no longer occurs.
- (iii) The CP circumvents the problem concerning the existence of an inflection point in the RPLL of this parameter. This can be seen in Figure 8, which refers to the plots of twice the RPLL function for λ , the asymmetry parameter of the SNBS distribution (left panel), and the for γ , the asymmetry parameter of the CSNBS distribution (right panel). The RPLL corresponds to $\ell(\hat{\alpha}(\gamma), \hat{\beta}(\gamma), \gamma) - \ell(\hat{\alpha}(\gamma), \hat{\beta}(\gamma), \hat{\gamma})$, where $\ell(\cdot)$ represents the

log-likelihood function. The respective plots are constructed by considering a random sample of both SNBS and CSBNS distributions, under suitable values of α , β and γ . We can notice a non-quadratic form of the log-likelihood function under the SNBS model, induced by the existence of an inflection point when the asymmetry parameter is very close to zero, making it difficult the obtaining of the ML estimates. However, the log-likelihood function of the CSNBS distribution presents a concave shape. Also, there is no inflection point when the asymmetry parameter is equals zero.

- (iv) The posterior distribution of λ for the SNBS distribution has a non-quadratic form, as it can be seen in Figure 9 (a), and this occurs even if we consider an informative prior. However, the posterior distribution of γ for the CSNBS distribution is well-behaved, presenting a concave shape, as it can be seen in Figure 9 (b).

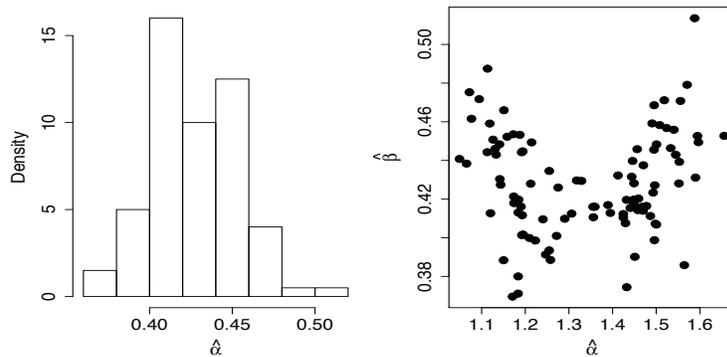


Figure 4. Estimated distributions of the ML estimates when samples of size $n = 200$ are drawn from SNBS; the left panel displays the histogram of $\hat{\alpha}$, the right panel displays the scatter plot of $(\hat{\alpha}, \hat{\beta})$.

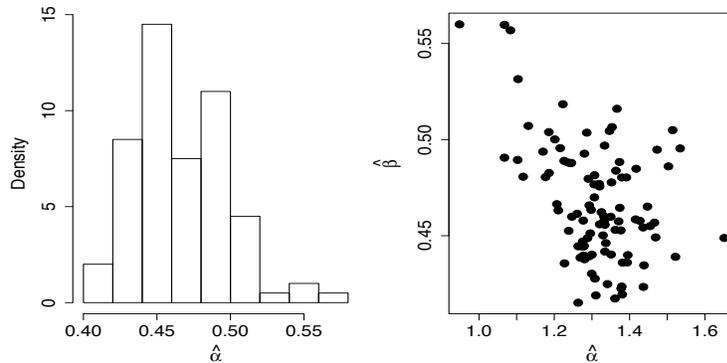


Figure 5. Estimated distributions of the Bayesian estimates when samples of size $n = 200$ are drawn from SNBS distribution; the left panel displays the histogram of $\hat{\alpha}$, the right panel displays the scatter plot of $(\hat{\alpha}, \hat{\beta})$.

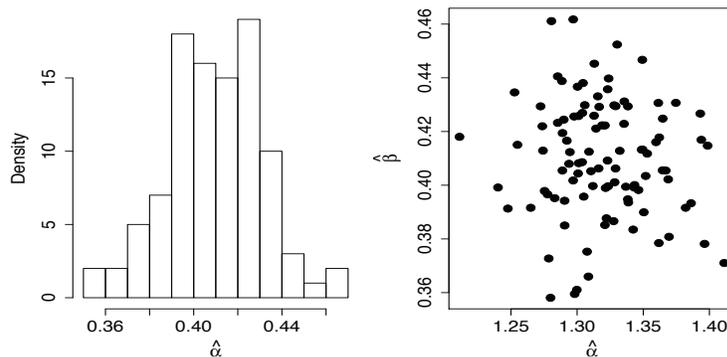


Figure 6. Estimated distributions of the ML estimates when samples of size $n = 200$ are drawn from CSNBS distribution; the left panel displays the histogram of $\hat{\alpha}$, the right panel displays the scatter plot of $(\hat{\alpha}, \hat{\beta})$.

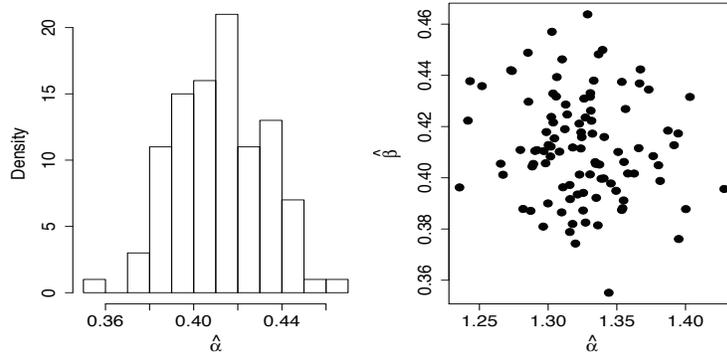


Figure 7. Estimated distributions of the Bayesian estimates when samples of size $n = 200$ are drawn from CSNBS distribution; histogram of $\hat{\alpha}$ (left) and scatter plot of $(\hat{\alpha}, \hat{\beta})$ (right).

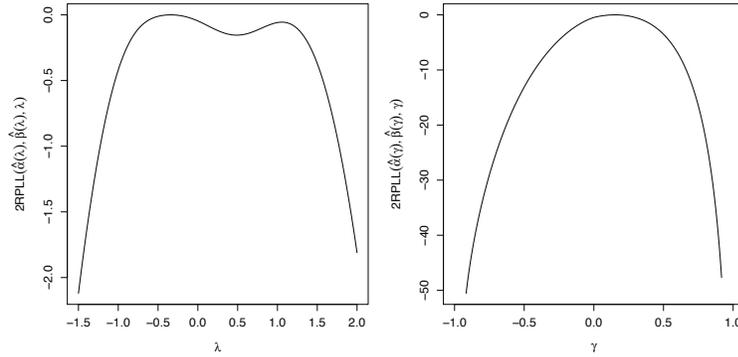


Figure 8. Twice the relative profiled log-likelihood function for the asymmetry parameter of the SNBS (left) and CSNBS (right) distributions.

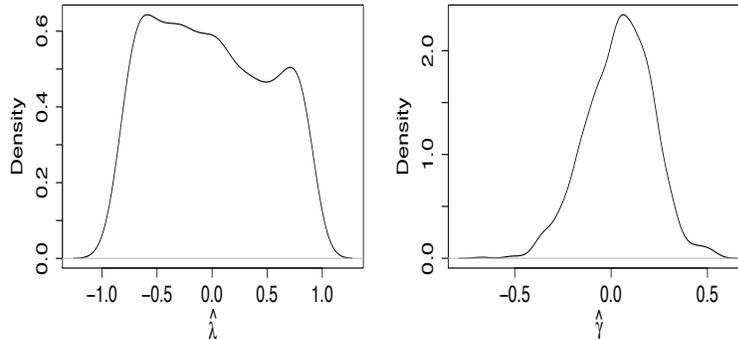


Figure 9. Posterior distribution of λ for the SNBS distribution (left) and of γ for the CSNBS distribution (right).

3. ESTIMATION AND INFERENCE

3.1 GENERAL CONTEXT

We present the ML estimation, based on the expectation conditional maximization (ECM) algorithm as in [Meng and Rubin \(1993\)](#), and the Bayesian approach, through MCMC algorithms. Let $T \sim \text{CSNBS}(\alpha, \beta, \gamma)$ and then, recall that, from [Theorem 2.1](#), we have $T|(H = h) \sim \text{EBS}(\alpha_\delta, \beta, \sigma = 2, \nu_h)$, where $H = |X_0| \sim \text{HN}(0, 1)$, $\alpha_\delta = \alpha\sqrt{(1 - \delta^2)/(1 - r^2\delta^2)}$ and $\nu_h = -\delta(h - r)/\sqrt{1 - \delta^2}$. In [Appendix B](#), we present some results that are useful for obtaining the ML estimators. For both methods, we consider a random sample T_1, \dots, T_n from $T \sim \text{SNBS}(\alpha, \beta, \gamma)$, where $\boldsymbol{\theta} = (\alpha, \beta, \gamma)^\top$.

3.2 THE ECM ALGORITHM AND ML ESTIMATION

Here, we discuss the ML estimation through the ECM algorithm. The log-likelihood function for $\boldsymbol{\theta}$ is given by $\ell(\boldsymbol{\theta}|\mathbf{t}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}|t_i)$, where

$$\ell_i(\boldsymbol{\theta}|t_i) = \log(2) + \log \{ \phi [a_{t_i;\mu,\sigma}(\alpha, \beta)] \} + \log \{ \Phi [\lambda a_{t_i;\mu,\sigma}(\alpha, \beta)] \} + \log [A_{t_i;\sigma}(\alpha, \beta)], \quad (5)$$

and $a_{t_i;\mu,\sigma}(\alpha, \beta)$ and $A_{t_i;\sigma}(\alpha, \beta)$ are given in Equation (4). Instead of considering the direct maximization of Equation (5), we obtain the ML estimates through the ECM algorithm, since it allows for a more tractable optimization process. In this case, we need to work with the so-called augmented likelihood function. Also, instead of working with $\boldsymbol{\theta}^* = (\alpha, \beta, \gamma)^\top$, we estimate $\boldsymbol{\theta} = (\alpha, \beta, \delta)^\top$, where δ is defined in Equation (1). Then, we recover γ through the invariance principle related to the ML estimators. This is performed since the related expressions (both analytically and computationally) are more tractable for $\boldsymbol{\theta}$.

Recall that, From Theorem 2.1, we have $T_i|H_i = h_i \stackrel{\text{IND}}{\sim} \text{EBS}(\alpha_\delta, \beta, \sigma = 2, \nu_{h_i})$ and $H_i \stackrel{\text{IND}}{\sim} \text{HN}(0, 1); i = 1, \dots, n$, where ‘‘IND’’ denotes ‘‘independent’’, $\alpha_\delta = \alpha\sqrt{(1 - \delta^2)/(1 - r^2\delta^2)}$ and $\nu_{h_i} = -(\delta(h_i - r))\sqrt{1 - \delta^2}$. Then, defining $\mathbf{t}_c = (\mathbf{t}^\top, \mathbf{h}^\top)^\top$, with $\mathbf{t} = (t_1, \dots, t_n)^\top$ and $\mathbf{h} = (h_1, \dots, h_n)^\top$, the augmented log-likelihood function can be written as

$$\begin{aligned} \ell(\boldsymbol{\theta}|\mathbf{t}_c) &= \sum_{i=1}^n \log[f_{T|H}(t_i|h_i)] + \sum_{i=1}^n f_H(h_i) \\ &= c - \frac{\delta^2}{2(1 - \delta^2)} \sum_{i=1}^n h_i^2 + \frac{r\delta^2}{(1 - \delta^2)} \sum_{i=1}^n h_i - \frac{nr^2\delta^2}{2(1 - \delta^2)} \\ &\quad + \frac{\delta\sqrt{1 - r^2\delta^2}}{1 - \delta^2} \sum_{i=1}^n h_i a_{t_i}(\alpha, \beta) - \frac{r\delta\sqrt{1 - r^2\delta^2}}{1 - \delta^2} \sum_{i=1}^n a_{t_i}(\alpha, \beta) - \frac{1 - r^2\delta^2}{2(1 - \delta^2)} \sum_{i=1}^n a_{t_i}^2(\alpha, \beta) \\ &\quad + \frac{n}{2} \log(1 - r^2\delta^2) + \sum_{i=1}^n \log(t_i + \beta) - \frac{n}{2} \log(1 - \delta^2) - n \log(\alpha) - \frac{n}{2} \log(\beta). \end{aligned}$$

For a current value of $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}$, the E-step requires the evaluation of $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \text{E}[\ell(\boldsymbol{\theta}|\mathbf{t}_c)|\mathbf{t}, \hat{\boldsymbol{\theta}}]$, where the expectation is taken with respect to the conditional distribution $H|(T = t)$, evaluated at $\hat{\boldsymbol{\theta}}$. For a estimate of $\boldsymbol{\theta}$ at r -th iteration, say $\boldsymbol{\theta}^{(r)} = (\alpha^{(r)}, \beta^{(r)}, \delta^{(r)})^\top$, consider $\hat{h}_i = \text{E}[H_i|\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}, t_i]$ and $\hat{h}_i^2 = \text{E}[H_i^2|\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}, t_i]$, given in Theorem 2.1, that is,

$$\hat{h}_i = \hat{\eta}_{t_i} + W_\Phi \left(\frac{\hat{\eta}_{t_i}}{\hat{\tau}} \right) \hat{\tau} \quad \text{and} \quad \hat{h}_i^2 = \hat{\eta}_{t_i}^2 + \hat{\tau}^2 + W_\Phi \left(\frac{\hat{\eta}_{t_i}}{\hat{\tau}} \right) (\hat{\eta}_{t_i} \hat{\tau}), \quad (6)$$

respectively, where $\hat{\eta}_{t_i} = \hat{\delta}\sqrt{1 - r^2\hat{\delta}^2} (a_{t_i}(\hat{\alpha}, \hat{\beta}) + r\hat{\delta}/\sqrt{1 - r^2\hat{\delta}^2})$, $\hat{\tau} = \sqrt{1 - \hat{\delta}^2}$ and $W_\Phi(z) = \phi(z)/\Phi(z)$, $z \in \mathbb{R}$. Then, let $\boldsymbol{\theta}^{(r)} = (\alpha^{(r)}, \beta^{(r)}, \delta^{(r)})^\top$ be the estimate of $\boldsymbol{\theta}$ at the k -th iteration. By considering Equation (6), we have that the augmented log-likelihood function becomes $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) = \text{E}[\ell(\boldsymbol{\theta}|\mathbf{t}_c)|\mathbf{t}, \hat{\boldsymbol{\theta}}^{(r)}]$, where

$$\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) &= c - \frac{\delta^{2(r)}}{2(1-\delta^{2(r)})} \sum_{i=1}^n \widehat{h}_i^{2(r)} + \frac{r\delta^{2(r)}}{(1-\delta^{2(r)})} \sum_{i=1}^n \widehat{h}_i^{(r)} - \frac{nr^2\delta^{2(r)}}{2\delta^{2(r)}} \\
&+ \frac{\delta^{(r)}\sqrt{1-r^2\delta^{2(r)}}}{\alpha^{(r)}(\delta^{2(r)})} \sum_{i=1}^n \widehat{h}_i^{(r)} a_{t_i}(1, \beta^{(r)}) - \frac{r\delta^{(r)}\sqrt{1-r^2\delta^{2(r)}}}{\alpha^{(r)}(1-\delta^{2(r)})} \\
&\times \sum_{i=1}^n a_{t_i}(1, \beta^{(r)}) - \frac{1-r^2\delta^{2(r)}}{2\alpha^{2(r)}(1-\delta^{2(r)})} \sum_{i=1}^n [a_{t_i}(1, \beta^{(r)})]^2 + \frac{n}{2} \log(1-\delta^{2(r)}) \\
&+ \sum_{i=1}^n \log(t_i + \beta^{(r)}) - \frac{n}{2} \log(1-\delta^{2(r)}) - n \log(\alpha^{(r)}) - \frac{n}{2} \log(\beta^{(r)}).
\end{aligned}$$

Hence, the ECM algorithm corresponds to iterate the following steps:

E-step: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(r)}$, compute \widehat{h}_i and \widehat{h}_i^2 , for $i = 1, \dots, n$ by using Equation (6);

CM-step 1: Fix $\beta = \widehat{\beta}^{(r)}$ and $\delta = \widehat{\delta}^{(r)}$ and update $\widehat{\alpha}^{(r)}$ through the positive root of $\widehat{\alpha}^2 + \widehat{b}^{(r)}\widehat{\alpha} + \widehat{c}^{(r)} = 0$, where

$$\begin{aligned}
\widehat{b}^{(r)} &= \frac{1}{(1-\widehat{\delta}^{2(r)})} \left[\widehat{\delta}^{(r)}\sqrt{1-r^2\widehat{\delta}^{2(r)}} \frac{1}{n} \sum_{i=1}^n \widehat{h}_i a_{t_i}(1, \widehat{\beta}^{(r)}) - r\widehat{\delta}^{(r)}\sqrt{1-r^2\widehat{\delta}^{2(r)}} \frac{1}{n} \sum_{i=1}^n a_{t_i}(1, \widehat{\beta}^{(r)}) \right], \\
\widehat{c}^{(r)} &= -\frac{(1-r^2\widehat{\delta}^{2(r)})}{(1-\widehat{\delta}^{2(r)})} \frac{1}{n} \sum_{i=1}^n \widehat{h}_i [a_{t_i}(1, \widehat{\beta}^{(r)})]^2,
\end{aligned}$$

that is, $\widehat{\alpha}^{(r+1)} = (-b(r+1) + \sqrt{b^2(r+1) - 4c(r+1)})/2$;

CM-step 2: Fix $\alpha = \widehat{\alpha}^{(r+1)}$ and update $\widehat{\beta}^{(r)}$ and $\widehat{\delta}^{(r)}$ using

$$\widehat{\beta}^{(r+1)} = \underset{\beta}{\operatorname{argmax}} Q\left(\widehat{\alpha}^{(r+1)}, \beta, \widehat{\delta}^{(r)}\right) \quad \text{and} \quad \widehat{\delta}^{(r+1)} = \underset{\delta}{\operatorname{argmax}} Q\left(\widehat{\alpha}^{(r+1)}, \widehat{\beta}^{(r+1)}, \delta\right).$$

The updating of $\widehat{\beta}^{(r+1)}$ and $\widehat{\delta}^{(r+1)}$ needs to be done through some numerical optimization method. In this work we use the function `optim`, available on the R software (see R Development Core Team, 2017), considering the L-BFGS-B optimization algorithm (see Byrd et al., 1995). Also, we start the ECM algorithm with initial values, say, $\widehat{\alpha}^{(0)}$, $\widehat{\beta}^{(0)}$ and $\widehat{\delta}^{(0)}$, using: $\widehat{\alpha}^{(0)} = [2(s/v) - 1]^{1/2}$ and $\widehat{\beta}^{(0)} = (sv)^{1/2}$, where $s = (1/n) \sum_{i=1}^n t_i$ and $v = [(1/n) \sum_{i=1}^n 1/t_i]^{-1}$, as in Vilca et al. (2011). After obtaining $\widehat{\alpha}^{(0)}$ and $\widehat{\beta}^{(0)}$, we can define $z_i = (1/\widehat{\alpha}^{(0)})[(t_i/\widehat{\beta}^{(0)})^{1/2} - (\widehat{\beta}^{(0)}/t_i)^{1/2}]$, for $i = 1, \dots, n$, which are observations related to a CSN distribution. Thus, $\widehat{\delta}^{(0)}$ can be obtained by maximizing (numerically) the log-likelihood function of a SN distribution with respect to δ , which is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \left[\log(2) + \log(\sigma_z) + \log[\phi(\mu_z + \sigma_z y_i)] + \log\{\Phi[\lambda(\mu_z + \sigma_z y_i)]\} \right].$$

According to Vilca et al. (2011), for ensuring that the true ML estimates are obtained, it is recommended to run the ECM algorithm using a range of different starting values and checking whether all of them result in similar estimates. The steps of the

ECM algorithm are repeated until a suitable convergence is attained, for example, using $\|\boldsymbol{\theta}^{(r)} - \boldsymbol{\theta}^{(r-1)}\| < \varepsilon$, with $\varepsilon > 0$. It is worthwhile to mention, under certain regularity conditions, that $\widehat{\boldsymbol{\theta}}$ converges in distribution to $N_3(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}})$, as $n \rightarrow \infty$. We approximate $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}$ by $I^{-1}(\boldsymbol{\theta})$, where $I(\boldsymbol{\theta}) = -\ddot{\ell}$, $\ddot{\ell} = [\ddot{\ell}_{\theta_1\theta_2}]$, $\theta_1, \theta_2 = \alpha, \beta$ or γ is the Hessian matrix, and $\ddot{\ell}_{\theta_1\theta_2} = \ddot{\ell}_{\theta_2\theta_1} = \partial^2 \ell(\boldsymbol{\theta}) / \partial \theta_1 \partial \theta_2 = \sum_{i=1}^n \partial^2 \ell_i(\boldsymbol{\theta}) / \partial \theta_1 \partial \theta_2$. The second derivatives of $\ell_i(\boldsymbol{\theta})$ are provided in Appendix C. The approximate standard errors (SE) of $\widehat{\boldsymbol{\theta}}$ can be estimated with the square roots of the diagonal elements of $I^{-1}(\boldsymbol{\theta})$, replacing $\boldsymbol{\theta}$ by $\widehat{\boldsymbol{\theta}}$.

3.3 BAYESIAN INFERENCE

Next, we present the developments related to the Bayesian inference through MCMC algorithms. We present the prior and the respective posterior distributions, along with suitable MCMC algorithms to sample from the respective marginal posterior distributions of interest. Consider both original and augmented likelihood functions (in order to compare them). The first of them is given by

$$L(\boldsymbol{\theta}|\mathbf{t}) = \prod_{i=1}^n 2\phi [a_{t_i;\mu,\sigma}(\alpha, \beta)] \Phi [\lambda a_{t_i;\mu,\sigma}(\alpha, \beta)] A_{t_i;\sigma}(\alpha, \beta).$$

We assume the following prior distributions: $\alpha \sim \text{gamma}(r_\alpha; \lambda_\alpha)$, $\beta \sim \text{gamma}(r_\beta; \lambda_\beta)$ and $\gamma \sim U(a; b)$, mutually independent, where $\text{gamma}(r, \lambda)$ stands for a gamma distribution such that $E(\alpha) = r/\lambda$ and $\text{Var}(\alpha) = r/\lambda^2$ and $U(a; b)$ stands for a continuous uniform distribution over the interval $[a, b]$. Combining the likelihood function with the prior distribution, we have that the joint posterior distribution is given by

$$\pi(\boldsymbol{\theta}|\mathbf{t}) \propto \alpha^{r_\alpha-1} \beta^{r_\beta-1} \exp[-(\alpha\lambda_\alpha + \beta\lambda_\beta)] \prod_{i=1}^n \phi [a_{t_i;\mu,\sigma}(\alpha, \beta)] \Phi [\lambda a_{t_i;\mu,\sigma}(\alpha, \beta)] A_{t_i;\sigma}(\alpha, \beta),$$

and the respective full conditional distributions, given by

$$\pi(\alpha|\beta, \gamma, \mathbf{t}) \propto \alpha^{r_\alpha-1} \exp(-\alpha\lambda_\alpha) \prod_{i=1}^n \phi [a_{t_i;\mu,\sigma}(\alpha, \beta)] \Phi [\lambda a_{t_i;\mu,\sigma}(\alpha, \beta)] A_{t_i;\sigma}(\alpha, \beta),$$

$$\pi(\beta|\alpha, \gamma, \mathbf{t}) \propto \beta^{r_\beta-1} \exp(-\beta\lambda_\beta) \prod_{i=1}^n \phi [a_{t_i;\mu,\sigma}(\alpha, \beta)] \Phi [\lambda a_{t_i;\mu,\sigma}(\alpha, \beta)] A_{t_i;\sigma}(\alpha, \beta),$$

$$\pi(\gamma|\alpha, \beta, \mathbf{t}) \propto \prod_{i=1}^n \phi [a_{t_i;\mu,\sigma}(\alpha, \beta)] \Phi [\lambda a_{t_i;\mu,\sigma}(\alpha, \beta)] A_{t_i;\sigma}(\alpha, \beta).$$

In addition, the augmented likelihood function is given by

$$L(\boldsymbol{\theta}|\mathbf{t}_c) = \prod_{i=1}^n \sqrt{2/\pi} \phi [\nu_{h_i} + a_{t_i}(\alpha, \beta)] A_{t_i}(\alpha, \beta) \exp\left(-\frac{h_i^2}{2}\right).$$

Similarly, combining the augmented likelihood function with the above prior distribution, we have that the posterior distribution is expressed as

$$\pi(\boldsymbol{\theta}, \mathbf{h}|\mathbf{t}) \propto \alpha^{r_\alpha-1} \beta^{r_\beta-1} \prod_{i=1}^n \phi [a_{t_i,h_i}(\alpha, \beta)] A_{t_i}(\alpha, \beta) \exp\left[-\frac{1}{2} (h_i^2 + 2\alpha\lambda_\alpha + 2\beta\lambda_\beta)\right]$$

and the respective full conditional distributions are given by

$$\begin{aligned}\pi(\mathbf{h}|\alpha, \beta, \gamma, \mathbf{t}_c) &\propto \prod_{i=1}^n \phi[a_{t_i, h_i}(\alpha, \beta)] A_{t_i}(\alpha, \beta) \exp\left(-\frac{h_i^2}{2}\right), \\ \pi(\alpha|\beta, \gamma, \mathbf{t}_c) &\propto \alpha^{r_\alpha-1} \prod_{i=1}^n \phi[a_{t_i, h_i}(\alpha, \beta)] A_{t_i}(\alpha, \beta) \exp\left[-\frac{1}{2}(h_i^2 + 2\alpha\lambda_\alpha)\right], \\ \pi(\beta|\alpha, \gamma, \mathbf{t}_c) &\propto \beta^{r_\beta-1} \prod_{i=1}^n \phi[a_{t_i, h_i}(\alpha, \beta)] A_{t_i}(\alpha, \beta) \exp\left[-\frac{1}{2}(h_i^2 + 2\beta\lambda_\beta)\right], \\ \pi(\gamma|\alpha, \beta, \mathbf{t}_c) &\propto \prod_{i=1}^n \phi[a_{t_i, h_i}(\alpha, \beta)] A_{t_i}(\alpha, \beta) \exp\left(-\frac{h_i^2}{2}\right),\end{aligned}$$

where $a_{t_i, h_i}(\alpha, \beta) = \nu_{h_i} + a_{t_i}(\alpha, \beta)$. We can see that both posterior distributions are not analytically tractable. Therefore, some numerical method must be employed to obtain suitable numerical approximations for the respective marginal posterior distributions. The above full conditional distributions do not correspond to known distributions, but they can be simulated through some auxiliary algorithm such as the Metropolis-Hastings, slice sampling or adaptive rejection. All these algorithms can be easily implemented in the R program. In addition, which is the approach pursued here, we can use a general MCMC computational framework, such `OpenBUGS`, see [Lunn et al. \(2009\)](#). In this case, it is necessary to provide the original or the augmented likelihood function, along with the prior distributions, such that the full conditional distributions are simulated through suitable algorithms, following a pre-defined hierarchy available on the `OpenBUGS`. We made all simulations using the R package `R2OpenBUGS`.

4. NUMERICAL ASPECTS

4.1 SIMULATION STUDY I

A simulation study is conducted to assess the behavior of the ECM algorithm, in terms of parameter recovery, and the accuracy of the corresponding SEs, calculated through the observed Fisher information matrix. For that, $N = 1,000$ replications are generated considering $n = 500$ and $\boldsymbol{\theta}^\top = (\alpha, \beta, \gamma) = (0.5, 1.0, 0.67)$, which induces a strong positively skewed behavior of the SNBS distribution. In [Table 1](#) we can see the mean of the estimates ($\widehat{\boldsymbol{\theta}}$), the mean of the theoretical (asymptotic) SE ($\text{SE}(\widehat{\boldsymbol{\theta}})$) and the empirical SE (SE_{emp}). We can notice that the parameters are well recovered and that the empirical SE are close to the theoretical ones, which indicates that the use of the observed Fisher information matrix, to obtain the corresponding SE, is appropriate.

Table 1. Results of the simulation study I.

	$\widehat{\boldsymbol{\theta}}$	$\text{SE}(\widehat{\boldsymbol{\theta}})$	SE_{emp}
$\widehat{\alpha}$	0.495	0.019	0.021
$\widehat{\beta}$	1.003	0.032	0.028
$\widehat{\gamma}$	0.667	0.015	0.012

4.2 SIMULATION STUDY II

We consider a total of 30 scenarios, resulting from the combination of the levels of three different sample sizes (n) (10, 50, 200), under $\alpha \in (0.5; 1.5)$, $\beta = 1$ and $\gamma \in (-0.67; -0.45; 0; 0.45; 0.67)$. The sample sizes are chosen in order to verify the properties of the estimators, as consistency, and to compare their behavior, in terms of accuracy. The values of α and β are chosen in order to induce different shapes and small variability, whereas the values of γ induce from null to high positive/negative asymmetry. We calculated the usual statistics to measure the accuracy of the estimates: bias, variance (Var), root mean squared error (RMSE) and absolute value of the relative bias (AVRB). Let θ be the parameter of interest, $\hat{\theta}_r$ be some estimate related to the replica r and $\bar{\hat{\theta}} = (1/R) \sum_{r=1}^R \hat{\theta}_r$. The adopted statistics are: Bias = $\bar{\hat{\theta}} - \theta$, Variance = $(1/R) \sum_{r=1}^R (\hat{\theta}_r - \bar{\hat{\theta}})^2$, RMSE = $((1/R) \sum_{r=1}^R (\hat{\theta}_r - \bar{\hat{\theta}})^2)^{1/2}$, AVR B = $|\bar{\hat{\theta}} - \theta|/|\theta|$.

The usual tools for monitoring the convergence of the MCMC algorithms, see [Gaman and Lopes \(2006\)](#), indicate that a burn-in of 4,000, a thin of 100, simulating a total of 100,000 values, are enough to produce valid MCMC samples of size 1,000 for each parameter. Since the other results are similar (they are omitted here but they are available under request from the authors), we present only those related to the scenario where $\alpha = 0.5$, $\beta = 1$, $\gamma = -0.67$, varying the sample size. We used (< 0.001) to represent positive values (statistics and/or estimates) and (> -0.001) to denote negative values, when they are close to zero. In addition, we refer the Bayesian estimates as “augmented”, when the augmented likelihood function is used, and “original”, whenever the original likelihood function is considered. The selected results can be seen in [Table 2](#). In general, we can see that, as the sample size increases, the estimates obtained by the three approaches tend to the correspondent the respective true values. When $\alpha = 0.5$, the ML estimates are more accurate than the Bayesian ones, especially considering the bias and AVR B metrics. In other scenarios (not shown), when $\alpha = 1.5$, the opposite occurs for all sample sizes. Concerning β and γ , it is possible to notice that, under the smallest sample size ($n = 10$), the ML approach presents more accurate estimates than the Bayesian ones. In addition, for $n = 50$ and $n = 200$, Bayesian estimates, for both parameters, are closer to the respective true values. In conclusion, we can say that all estimators, mainly the Bayesian ones, are consistent, since both bias and RMSE tend to decrease, as the sample size increases. Furthermore, the results indicate (including those not shown here) that the Bayesian approach provided the most accurate estimates. Moreover, we can notice that the original and augmented approaches, performed quite similarly. Therefore, we decide to use the original likelihood function) approach, since it is easier to implement and faster.

4.3 REAL DATA ANALYSIS I

We analyze a data set corresponding to self-efficacy, which is available in the R software and can be accessed from the `EstCRM` package through the command `data(SelfEff)`. A group of 307 pre-service teachers, graduated from various departments in the college of education, are asked to check on a 11 cm line segment with two end points (can not do at all, highly certain can do) using their own judgment for the 10 items that measure teacher self-efficacy on different activities. We take, as response variable, the teacher self-efficacy in the creation of learning environments in which students can effectively express themselves. [Table 3](#) presents some descriptive statistics, including location measures, standard deviation (SD), coefficient of skewness (CS), and kurtosis (CK). We can notice that the distribution is strongly negatively skewed. We fit the CSNBS and BS distributions, using the Bayesian augmented and the ML method, to the data. The results obtained considering the frequentist approach are omitted here but they are available under request from

Table 2. Results of simulation study II with $\gamma = -0.67$.

Parameter	n	Method	Mean	Variance	Bias	RMSE	AVRB
α	10	Augmented	0.577	< 0.001	0.077	0.081	0.154
		Original	0.578	0.001	0.078	0.082	0.156
		ML	0.520	0.071	0.020	0.267	0.040
	50	Augmented	0.511	< 0.001	0.011	0.016	0.022
		Original	0.511	< 0.001	0.011	0.015	0.021
		ML	0.498	0.001	-0.002	0.033	0.004
	200	Augmented	0.502	< 0.001	0.002	0.005	0.004
		Original	0.502	< 0.001	0.002	0.005	0.004
		ML	0.490	< 0.001	-0.010	0.012	0.019
β	10	Augmented	1.006	< 0.001	0.006	0.023	0.006
		Original	1.004	< 0.001	0.004	0.021	0.004
		ML	1.105	0.214	0.105	0.474	0.105
	50	Augmented	0.996	< 0.001	-0.004	0.009	0.004
		Original	0.997	< 0.001	-0.003	0.009	0.003
		ML	1.039	0.018	0.039	0.140	0.039
	200	Augmented	0.999	< 0.001	-0.001	0.005	0.001
		Original	0.999	< 0.001	-0.001	0.005	0.001
		ML	0.997	< 0.001	-0.003	0.004	0.003
γ	10	Augmented	-0.157	0.067	0.513	0.575	0.766
		Original	-0.182	0.054	0.488	0.540	0.728
		ML	-0.603	0.028	0.067	0.179	0.100
	50	Augmented	-0.493	0.059	0.177	0.301	0.264
		Original	-0.505	0.049	0.165	0.276	0.247
		ML	-0.569	0.012	0.101	0.148	0.150
	200	Augmented	-0.614	0.017	0.056	0.142	0.083
		Original	-0.601	0.015	0.069	0.141	0.103
		ML	-0.523	0.002	0.147	0.153	0.220

the authors. The prior distributions are the same used in Section 3. In Table 4, in addition to the posterior expectations (PE), the posterior standard deviations (PSD) and the 95% equi-tailed credibility intervals (CI), we also present the model selection criteria. We consider the usual statistics of model comparison for both frequentist (AIC, BIC) and Bayesian (DIC, EAIC, EBIC and LPLM) see, respectively (Akaike, 1974; Schwarz, 1978; Spiegelhalter et al., 2014). The smaller values of AIC and BIC indicates the model that fits the data better. In addition, the smaller the values of DIC, EAIC, EBIC, the better the model fit, occurring the opposite with the LPML. We can notice that the estimates of α and γ (under the CSNBS model) indicate that the distribution is strongly negatively skewed. Notice also that we have indications that the asymmetry parameter is different from zero, since this value does not belong to the CI. Moreover, the criteria indicated the CSNBS model is the best. Figure 10 (left) presents the histogram of the observations and estimated densities. We can notice that the CSNBS distribution presents an advantage over the BS model. From Figure 10, we can notice that the CSNBS distribution predicts better the observations than the BS distribution. In conclusion, we can say that the CSNBS model is preferable to the BS model.

Table 3. Descriptive statistics for the teacher self-efficacy data.

Mean	Median	Minimum	Maximum	SD	Asymmetry	Kurtosis
9.205	9.700	1.650	10.900	1.365	-1.752	7.781

4.4 REAL DATA ANALYSIS II

We analyze now a data set corresponding to prices of bottles of Barolo wine and discussed in Azzalini (2013). It concerns the price (in euros) of bottles (75 cl) of Barolo wine. The data have been obtained in July 2010 from the websites of four Italian wine resellers, selecting only quotations of Barolo wine, which is produced in the Piedmont region of

Table 4. Posterior expectations (PE), posterior standard deviations (PSD), equi-tailed 95% CI and model selection criteria.

Parameter	PE	PSD	CI _{95%}
CSNBS			
α	0.154	0.002	[0.151; 0.157]
β	8.871	0.016	[8.836; 8.903]
γ	-0.971	0.003	[-0.978; -0.966]
EAIC		1,021.912	
EBIC		1,033.093	
DIC		3,047.154	
LPML		-508.531	
BS			
α	0.205	0.008	[0.190; 0.222]
β	9.016	0.105	[8.815; 9.229]
EAIC		1,252.772	
EBIC		1,260.226	
DIC		3,744.335	
LPML		-632.9564	

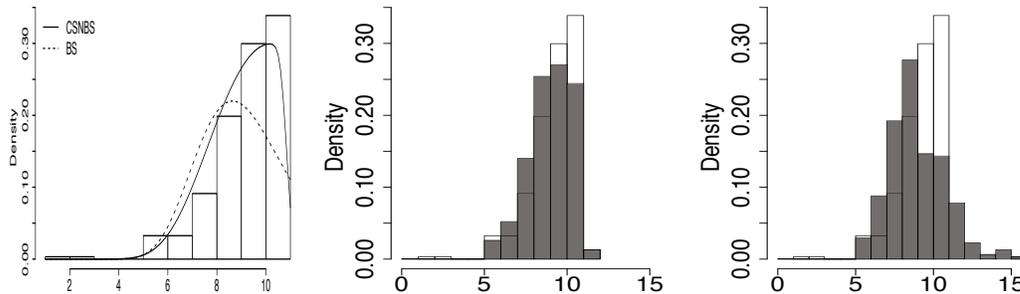


Figure 10. Histogram of the observations and estimated densities (left), histogram of the predicted and observed distributions for the CSNBS (center) and BS (right) models.

Italy. The price does not include the delivery charge. In Table 5 and Figure 11 (left), we present a descriptive analysis. It is possible to see that the distribution is positively skewed and more concentrated in the first class [0,100]. We fit the CSNBS and BS distributions, using the Bayesian augmented and the ML method, to the data. The results obtained considering the frequentist approach are omitted here but they are available under request from the authors. The prior distributions are the same used in Section 3. In Table 6, in addition to the posterior expectations (PE), the posterior standard deviations (PSD) and the 95% equi-tailed CI, we also present the Bayesian criteria. Table 6 shows that the estimates of α and γ (under the CSNBS model) indicate that the distribution of the prices is strongly positively skewed. Notice also that we have indications that the asymmetry parameter is different from zero, since this value does not belong to the CI. Moreover, the criteria indicated the CSNBS model is the best. Also, we construct QQ plots with simulated envelopes. Similar to Vilca et al. (2011), we considered the Bayesian estimates of α and β in $d(\alpha, \beta) = (1/\alpha^2)(T/\beta + \beta/T - 2)$. When $T \sim BS(\alpha, \beta)$, it is known that $d(\alpha, \beta) \sim N(0, 1)$. Since the observations $d(\hat{\alpha}, \hat{\beta})$ are expected to follow a standard normal distribution, under the well fit of the model, the envelopes are simulated from the standard

normal distribution, as described in Atkinson (1985). Similarly, if $T \sim \text{CSNBS}(\alpha, \beta, \gamma)$, thus $d(\alpha, \beta) \sim \text{CSN}(0, 1, \gamma)$. Since the observations $d(\hat{\alpha}, \hat{\beta})$ are expected to follow a CSN distribution, under the well fit of the model, the envelopes are simulated from the CSN distribution. These plots are presented in Figure 11 (lines represent the 5th percentile, the mean, and the 95th percentile of 100 simulated points). From those figures, we conclude that the CSNBS distribution provides a better fit than the BS model. Specifically, from the QQ plot shown in Figure 11 (a), we notice that the observations appear to form a slight upward-facing concave. However, the QQ plot shown in Figures 11 (b) indicate that the CSNBS distribution offers an excellent fit, provided that the majority of observations are inside of the envelope.

Table 5. Descriptive statistics for the prices of bottles of Barolo wine.

Mean	Median	Minimum	Maximum	SD	Asymmetry	Kurtosis
124.617	72	14	1000	37.041	2.903	12.982

Table 6. Posterior expectations (PE), posterior standard deviations (PSD), equi-tailed 95% CI and model selection criteria.

Parameter	PE	PSD	CI _{95%}
CSNBS			
α	0.844	0.037	[0.775; 0.917]
β	89.576	3.911	[82.260; 97.871]
γ	0.690	0.070	[0.541; 0.809]
EAIC		3,437.879	
EBIC		3,449.060	
DIC		10,292.690	
LPML		-1,718.110	
BS			
α	0.858	0.035	[0.794; 0.929]
β	92.444	4.264	[84.778; 101.302]
EAIC		3,474.893	
EBIC		3,482.346	
DIC		10,410.620	
LPML		-1,736.669	

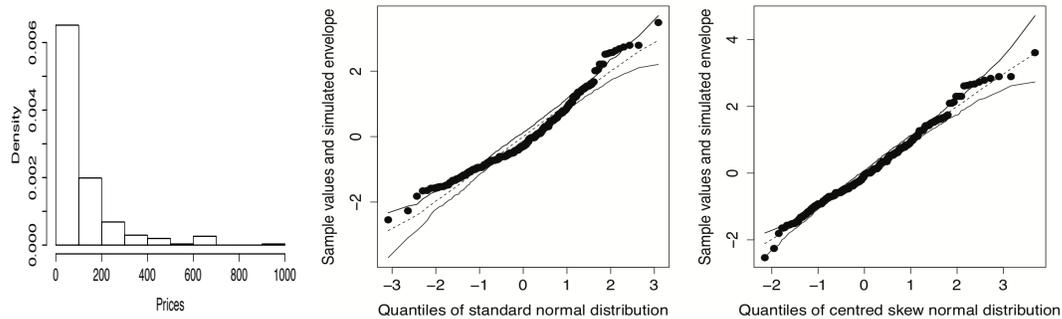


Figure 11. Histogram of the prices of bottles of Barolo wine (left), QQ plots with envelopes for BS (center) and CSNBS (right) distributions for the data of Barolo wine bottle prices.

5. CONCLUDING REMARKS

In this paper, we introduced a new distribution for modeling positive data which can present both positive and negative asymmetry, by combining the Birnbaum-Saunders and the centered skew normal distributions. We developed both maximum likelihood and Bayesian estimation procedures, comparing them through a suitable simulation study. The convergence of the conditional expectation maximization and MCMC algorithms were verified and several factors of interest were compared in the parameter recovery study. In general, as the sample size increases, the results indicated that the Bayesian approach provided the most accurate estimates. In future works we can consider the development of predictive posterior checking to detect the goodness of fit. Furthermore, we suggest the use of Jeffreys-rule prior and independence Jeffreys prior. Other auxiliary algorithms as the Hamiltonian Monte Carlo (see [Homand and Gelman, 2014](#); [Carpenter et al., 2016](#)), adaptive reject sampling and slice sampling (see [Gamerman and Lopes, 2006](#)) can be used and compared. Other family of distributions could be used instead of the centered skew normal distribution, as the scale mixture of the SN distributions, to generate new family of Birnbaum-Saunders-type distributions. Finally, other numerical methods to obtain approximation for the marginal posterior distributions, such as the INLA algorithm, can be considered (see [Rue and Martino, 2009](#)).

6. APPENDIX

APPENDIX A. MOMENTS OF THE CSNBS DISTRIBUTION

THEOREM A.1 Let $T \sim \text{CSNBS}(\alpha, \beta, \gamma)$ and $Y \sim \text{CSN}(0, 1, \gamma)$. If $E[Y^{2(r-j+i)}] < \infty$, then the moments of T are given by

$$E(T^r) = \beta^r \sum_{j=0}^r \binom{2r}{2j} \sum_{i=0}^j E[Y^{2(r-j+i)}] (\alpha/2)^{2(r-j+i)}.$$

Proof of Theorem A.1 From Equation (3), we have that

$$E\left[\left(\frac{T}{\beta}\right)^r\right] = E\left\{\left[\left(\frac{\alpha Y}{2} + \sqrt{\left(\frac{\alpha Y}{2}\right)^2 + 1}\right)^2\right]^r\right\}.$$

From the binomial theorem, that is, $(a + b)^m = \sum_{k=0}^m \binom{m}{k} a^{m-k} b^k$, we have that

$$E\left[\left(\frac{T}{\beta}\right)^r\right] = \sum_{k=0}^{2r} \binom{2r}{k} E\left\{\left[\left(\frac{\alpha Y}{2}\right)^2 + 1\right]^{k/2} \left(\frac{\alpha Y}{2}\right)^{2r-k}\right\}.$$

Considering $k = 2j$, that is, $j = k/2$, it comes that

$$E\left[\left(\frac{T}{\beta}\right)^r\right] = \sum_{j=0}^r \binom{2r}{2j} E\left\{\left[\left(\frac{\alpha Y}{2}\right)^2 + 1\right]^j \left(\frac{\alpha Y}{2}\right)^{2(r-j)}\right\}.$$

From the binomial theorem again, we have

$$\begin{aligned} \mathbb{E} \left[\left(\frac{T}{\beta} \right)^r \right] &= \sum_{j=0}^r \binom{2r}{2j} \mathbb{E} \left\{ \sum_{i=0}^j \binom{j}{i} \left(\frac{\alpha Y}{2} \right)^{2i} \left(\frac{\alpha Y}{2} \right)^{2(r-j)} \right\} \\ &= \sum_{j=0}^r \binom{2r}{2j} \sum_{i=0}^j \binom{j}{i} \mathbb{E} \left[\left(\frac{\alpha Y}{2} \right)^{2(r-j+i)} \right] \\ &= \sum_{j=0}^r \binom{2r}{2j} \sum_{i=0}^j \binom{j}{i} \mathbb{E} \left[Y^{2(r-j+i)} \right] \left(\frac{\alpha}{2} \right)^{2(r-j+i)}. \end{aligned}$$

Therefore,

$$\mathbb{E}(T^r) = \beta^r \sum_{j=0}^r \binom{2r}{2j} \sum_{i=0}^j \binom{j}{i} \mathbb{E} \left[Y^{2(r-j+i)} \right] (\alpha/2)^{2(r-j+i)}. \quad (\text{A1})$$

From Equation (A1), we get

$$\mathbb{E}(T) = \beta \sum_{j=0}^1 \binom{2}{2j} \sum_{i=0}^j \binom{j}{i} \mathbb{E} \left[Y^{2(1-j+i)} \right] (\alpha/2)^{2(1-j+i)}.$$

For $j = 0$, the first term of the sum in Equation (A1) is equal to $\beta \mathbb{E}(Y^2)(\alpha/2)^2$. For $j = 1$, the second term of the sum in Equation (A1) is equal to $\beta [1 + \mathbb{E}(Y^2)(\alpha/2)^2]$. Hence, by adding these two terms, we have

$$\mathbb{E}(T) = \beta [1 + (\alpha^2/2)].$$

Furthermore, from Equation (A1), we have

$$\mathbb{E}(T^2) = \beta^2 \sum_{j=0}^2 \binom{4}{2j} \sum_{i=0}^j \binom{j}{i} \mathbb{E} \left[Y^{2(2-j+i)} \right] (\alpha/2)^{2(2-j+i)}.$$

Developing the above sum in j , we obtain

$$\mathbb{E}(T^2) = \beta^2 \left[1 + \frac{\alpha^4}{2} \Delta + 2\alpha^2 \right].$$

Thus,

$$\begin{aligned} \text{Var}(T) &= \mathbb{E}(T^2) - [\mathbb{E}(T)]^2 \\ &= (\alpha\beta)^2 \left\{ 1 + \frac{\alpha^2}{4} [2\Delta - 1] \right\}. \end{aligned}$$

where $\Delta = \mathbb{E}(Y^4) = 2(\pi - 3)(4/\pi^2)\delta^4[1 - (2\delta^2/\pi)]^{-2} + 3$.

APPENDIX B. THE ECM ALGORITHM

The following result is used in the proof of Theorem 2.1.

Lemma 1. Let $X \sim N(\eta, \tau^2)$, thus $\forall a \in \mathbb{R}$

$$\mathbb{E}(X|X > a) = \eta + \frac{\phi\left(\frac{a-\eta}{\tau}\right)}{1 - \Phi\left(\frac{a-\eta}{\tau}\right)}\tau; \quad \mathbb{E}(X^2|X > a) = \eta^2 + \tau^2 + \frac{\phi\left(\frac{a-\eta}{\tau}\right)}{1 - \Phi\left(\frac{a-\eta}{\tau}\right)}(\eta + a)\tau.$$

Proof of Theorem 2.1

- (i) Since $Y \sim \text{CSN}(0, 1, \gamma)$, using the stochastic representation given by Equation (2), we can define

$$Y = \frac{1}{\sigma_z} \left[\delta H + \sqrt{1 - \delta^2} X_1 - \mu_z \right] = \frac{1}{\alpha} \left[\sqrt{T/\beta} - \sqrt{\beta/T} \right].$$

Therefore,

$$Y|(H = h) = \frac{1}{\alpha} \left(\sqrt{T/\beta} - \sqrt{\beta/T} \right) \Big| (H = h) \sim N(\mu_h, \sigma^2),$$

where $\mu_h = \delta(h - r)/(1 - r^2\delta^2)^{1/2}$ and $\sigma^2 = (1 - \delta^2)/(1 - r^2\delta^2)$. Then,

$$W|(H = h) = -\frac{\mu_h}{\sigma} + \frac{1}{\sigma\alpha} \left(\sqrt{T/\beta} - \sqrt{\beta/T} \right) \Big| (H = h) \sim N(0, 1)$$

and

$$T = \beta \left[\frac{\alpha}{2} (\sigma W + \mu_h) + \sqrt{\left[\frac{\alpha}{2} (\sigma W + \mu_h) \right]^2 + 1} \right].$$

From the above result, the proof is completed.

- (ii) As $f_H(h) = 2\phi(h|0, 1)$, $h > 0$ and

$$\phi(\nu_h + a_t(\alpha, \beta)) = \frac{\sqrt{1 - \delta^2}}{\sqrt{1 - r^2\delta^2}} \phi \left(a_t(\alpha, \beta) \Big| \frac{\delta(h - r)}{\sqrt{1 - r^2\delta^2}}, \frac{1 - \delta^2}{1 - r^2\delta^2} \right).$$

Then, we have

$$\begin{aligned} \phi \left(a_t(\alpha, \beta) \Big| \frac{\delta(h - r)}{\sqrt{1 - r^2\delta^2}}, \frac{1 - \delta^2}{1 - r^2\delta^2} \right) \phi(h|0, 1) &= \phi \left(a_t(\alpha, \beta) \Big| -\frac{r\delta}{\sqrt{1 - r^2\delta^2}}, \frac{1}{1 - r^2\delta^2} \right) \\ &\times \phi \left(h \Big| \delta\sqrt{1 - r^2\delta^2} (a_t(\alpha, \beta) + \frac{r\delta}{\sqrt{1 - r^2\delta^2}}), 1 - \delta^2 \right). \end{aligned}$$

Therefore, the proof of (i) follows directly from that $f_{H|T}(h|t) = f_{T|H}(t|h)f_H(h)/f_T(t)$. To demonstrate (ii)-(iii), notice, for $k = 1, 2$, we have that

$$\mathbb{E} \left[H^k | T \right] = \frac{1}{\Phi \left(\lambda \sigma_z \left(a_t(\alpha, \beta) + \frac{r\delta}{\sqrt{1 - r^2\delta^2}} \right) \right)} \int_0^\infty h^k \phi \{ h | \eta_t, 1 - \delta^2 \} dh = \mathbb{E}(X^k | X > 0).$$

Then, using some properties of the HN distribution from Lemma 1, the proof is completed.

APPENDIX C. THE OBSERVED FISHER INFORMATION MATRIX

The necessary expressions are given below. For the sake of simplicity, we consider the following notation to obtain the necessary expressions, $a_{t_i;\mu,\sigma} = a_{t_i;\mu,\sigma}(\alpha, \beta)$ and $A_{t_i;\sigma} = A_{t_i;\sigma}(\alpha, \beta)$.

$$\begin{aligned} \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_2} &= -\frac{1}{A_{t_i;\sigma}^2} \frac{\partial A_{t_i;\sigma}}{\partial \theta_1} \frac{\partial A_{t_i;\sigma}}{\partial \theta_2} + \frac{1}{A_{t_i;\sigma}} \frac{\partial^2 A_{t_i;\sigma}}{\partial \theta_1 \partial \theta_2} - \frac{1}{2} \frac{\partial^2 a_{t_i;\mu,\sigma}^2}{\partial \theta_1 \partial \theta_2} + \lambda^2 W_{\Phi}'[\lambda a_{t_i;\mu,\sigma}] \frac{\partial a_{t_i;\mu,\sigma}}{\partial \theta_1} \frac{\partial a_{t_i;\mu,\sigma}}{\partial \theta_2} \\ &\quad + \lambda W_{\Phi}[\lambda a_{t_i;\mu,\sigma}] \frac{\partial^2 a_{t_i;\mu,\sigma}}{\partial \theta_1 \partial \theta_2}, \quad \theta_1, \theta_2 = \alpha, \beta \\ \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \theta_3 \partial \gamma} &= -\frac{1}{A_{t_i;\sigma}^2} \frac{\partial A_{t_i;\sigma}}{\partial \theta_3} \frac{\partial A_{t_i;\sigma}}{\partial \gamma} + \frac{1}{A_{t_i;\sigma}} \frac{\partial^2 A_{t_i;\sigma}}{\partial \theta_3 \partial \gamma} - \frac{1}{2} \frac{\partial^2 a_{t_i;\mu,\sigma}^2}{\partial \theta_3 \partial \gamma} + \lambda W_{\Phi}[\lambda a_{t_i;\mu,\sigma}] \frac{\partial^2 a_{t_i;\mu,\sigma}}{\partial \theta_3 \partial \gamma} \\ &\quad + \left\{ \lambda W_{\Phi}'[\lambda a_{t_i;\mu,\sigma}] \frac{\partial \lambda a_{t_i;\mu,\sigma}}{\partial \gamma} + W_{\Phi}[\lambda a_{t_i;\mu,\sigma}] \frac{\partial \lambda}{\partial \gamma} \right\} \frac{\partial a_{t_i;\mu,\sigma}}{\partial \theta_3}, \quad \theta_3 = \alpha, \beta \\ \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \gamma^2} &= -\frac{1}{A_{t_i;\sigma}^2(\alpha, \beta)} \frac{\partial^2 A_{t_i;\sigma}}{\partial \gamma^2} + \frac{1}{A_{t_i;\sigma}(\alpha, \beta)} \frac{\partial^2 A_{t_i;\sigma}}{\partial \gamma^2} - \frac{1}{2} \frac{\partial^2 a_{t_i;\mu,\sigma}^2}{\partial \gamma^2} + W_{\Phi}'[\lambda a_{t_i;\mu,\sigma}] \left(\frac{\partial \lambda a_{t_i;\mu,\sigma}}{\partial \gamma} \right)^2 \\ &\quad + W_{\Phi}[\lambda a_{t_i;\mu,\sigma}] \frac{\partial^2 \lambda a_{t_i;\mu,\sigma}}{\partial \gamma^2}, \end{aligned}$$

where $W_{\Phi}'(x) = -W_{\Phi}(x)[x + W_{\Phi}(x)]$ is the derivative of $W_{\Phi}(x)$ with respect to x , see [Vilca et al. \(2011\)](#), and the other quantities are as before defined.

ACKNOWLEDGEMENTS

The authors would like to thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for the financial support through a Master Scholarship granted to the first author under the guidance of the second and also for Conselho Nacional de Desenvolvimento Científico e Tecnológico, grants numbers 308339/2015-0 and 305336/2017-7, for a research scholarship granted to the second and fourth authors, respectively.

REFERENCES

- Achcar, J.A., 1993. Inferences for the Birnbaum-Saunders fatigue life model using Bayesian methods. *Computational Statistics and Data Analysis*, 15, 367-380.
- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716-723.
- Atkinson, A.C., 1985. *Plots, Transformations and Regressions*. Oxford University Press, Oxford, UK.
- Arellano-Valle, R. and Azzalini, A., 2008. The centered parametrization for the multivariate skew-normal distribution. *Journal of Multivariate Analysis*, 99, 1362-1382.
- Azzalini, A., 1985. A class of distribution which includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171-178.
- Azzalini, A., 2013. *The skew-normal and related families*. Cambridge University Press, Cambridge, UK.
- Balakrishnan, N., Leiva, V., Sanhueza, A., and Vilca-Labra, F., 2009. Estimation in the Birnbaum-Saunders distribution based on scale-mixture of normals and the EM-algorithm. *SORT*, 33, 171-192.

- Balakrishnan, N., Saulo, H., and Leao, J., 2017. On a new class of skewed Birnbaum-Saunders models. *Journal of Statistical Theory and Practice*, 11, 573-593.
- Barros, M., Paula, G.A., and Leiva, V., 2008. A New Class of Survival Regression Models with Heavy-Tailed Errors: Robustness and Diagnostics. *Lifetime Data Analysis*, 14, 1-17.
- Birnbaum, Z.W. and Saunders, S.C., 1969a. A new family of life distributions. *Journal of Applied Probability*, 6, 637-652.
- Birnbaum, Z.W. and Saunders, S.C., 1969b. Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability*, 6, 328-347.
- Byrd, R.H., Lu, P., Nocedal, J., and Zhu, C., 1995. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 15, 1190-1208.
- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M.A., Guo, J., Li, P., and Riddell, A., 2016. Stan: A probabilistic programming language. *Journal of Statistical Software*, 20, 1-37.
- Cordeiro, G.M. and Lemonte, A.J., 2011. The β -Birnbaum-Saunders distribution: An improved distribution for fatigue life modeling. *Computational Statistics and Data Analysis*, 55, 116-124.
- Depaoli, S., Clifton, J.P., and Cobb, P. R., 2016. Just another Gibbs sampler (JAGS) flexible software for MCMC implementation. *Journal of Educational and Behavioral Statistics*, 41, 628-649.
- Desmond, A.F., 1985. Stochastic models of failure in random environments. *Canadian Journal of Statistics*, 13, 171-183.
- Desmond, A.F., 1986. On the relationship between two fatigue-life models. *IEEE Transactions on Reliability*, 35, 167-169.
- Diaz-Garcia, J.A. and Leiva, V., 2005. New family of life distributions based on the elliptically contoured distributions. *Journal of Statistical Planning and Inference*, 128, 445-457.
- Engelhardt, M., Bain, L.J., and Wright, F. ., 1981. Inferences on the parameters of the Birnbaum-Saunders fatigue life distribution based on maximum likelihood estimation. *Technometrics*, 23, 251-256.
- Ferreira, M.S., 2013. A study of exponential-type tails applied to Birnbaum-Saunders models. *Chilean Journal of Statistics*, 4, 87-97.
- Gamerman, D. and Lopes, H., 2006. *MCMC-Stochastic Simulation for Bayesian Inference*. Chapman and Hall/CRC, Boca Raton, FL, US.
- Homan, M.D. and Gelman, A., 2014. The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *The Journal of Machine Learning Research*, 15, 1593-1623.
- Leiva, V., Barros, M., Paula, G.A., and Sanhueza, A., 2007. Generalized Birnbaum-Saunders distributions applied to air pollutant concentration. *Environmetrics*, 19, 235-249.
- Leiva, V., Vilca, F., Balakrishnan, N., and Sanhueza, A., 2008. A skewed sinh-normal distribution and its properties and application to air pollution. *Communications in Statistics Theory and Methods*, 39, 426-443.
- Leiva, V., 2016. *The Birnbaum-Saunders Distribution*. Academic Press, New York.
- Lu, M.C. and Chang, D.S., 1997. Bootstrap prediction intervals for the Birnbaum-Saunders distribution. *Microelectronics Reliability*, 37, 1213-1216.
- Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D., 2000. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10, 325-337.
- Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N., 2009. The BUGS project: Evolution, critique and future directions (with discussion). *Statistics in Medicine*, 28, 3049-3082.

- Maehara, R.P. (2018). An extension of Birnbaum-Saunders distributions based on scale mixtures of skew-normal distributions with applications to regression models. PhD thesis, University of Sao Paulo, available at <http://www.teses.usp.br/teses/disponiveis/45/45133/tde-14052018-202935/es.php>.
- Mann, N.R., Singpurwalla, N.D., and Schafer, R.E., 1974. *Methods for Statistical Analysis of Reliability and Life Data*. Wiley, New York.
- Mazucheli, J.M., Menezes, A.F.B., and Dey, S., 2018. The unit-Birnbaum-Saunders distribution with applications. *Chilean Journal of Statistics*, 9, 47-57.
- Meng, X.L. and Rubin, D.B., 1993. Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, 80, 267-278.
- Owen, W.J. and Padgett, W.J., 1999. Accelerated test models for system strength based on Birnbaum-Saunders distributions. *Lifetime Data Analysis*, 5, 133-147.
- Pewsey, A., 2000. Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics*, 27, 859-870.
- R Development Core Team, 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rieck, J.R. and Nedelman, J.R., 1991. A log-linear model for the Birnbaum-Saunders distribution. *Technometrics*, 33, 51-60.
- Rue, H. and Martino, S., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society B*, 71, 319-392.
- Schwarz, G., 1978. Estimating the dimension of a model. *The Annals of Statistics*, 6, 461-464.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and Linde, A., 2014. The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society B*, 76, 485-493.
- Vilca, F. and Leiva, V., 2006. A new fatigue life model based on the family of skew-elliptical distributions. *Communications in Statistics: Theory and Methods*, 35, 229-244.
- Vilca, F., Santana, L., Leiva, V., and Balakrishnan, N., 2011. Estimation of extreme percentiles in Birnbaum-Saunders distributions. *Computational Statistics and Data Analysis*, 55, 1665-1678.
- Wang, M., Sun, X., and Park, C., 2016. Bayesian analysis of BirnbaumSaunders distribution via the generalized ratio-of-uniforms method. *Computational Statistics*, 31, 207-225.
- Xu, A. and Tang, Y., 2011. Bayesian analysis of Birnbaum-Saunders distribution with partial information. *Computational Statistics and Data Analysis*, 55, 2324-2333.

DISTRIBUTION THEORY
RESEARCH PAPER

The Harris extended Lindley distribution for modeling hydrological data

GAUSS M. CORDEIRO¹, M. MANSOOR^{2,*} and SERGE B. PROVOST³

¹Department of Statistics, Federal University of Pernambuco, Recife, PE, Brazil

²Department of Statistics, Government Sadiq Egerton College Bahawalpur, Bahawalpur, Pakistan

³Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Canada

(Received: 17 April 2018 · Accepted in final form: 14 February 2019)

Abstract

We introduce a three-parameter extension of the Lindley distribution, which has as sub-models the Lindley and Marshall-Olkin Lindley distributions. The proposed model turns out to be quite flexible: its probability density function can be decreasing or unimodal and its associated hazard rate may be increasing, decreasing, unimodal or bathtub-shaped. Since this new distribution has a survival function and a hazard rate that can be expressed in closed form, it can readily be simulated and used to analyze censored data. Computable expressions are obtained for certain statistical functions such as its quantile function, ordinary and incomplete moments, moment generating function, order statistics and reliability function. The maximum likelihood method is utilized to obtain estimates of the model parameters and a simulation study is carried out to assess the performance of the corresponding maximum likelihood estimators. Two illustrative examples involving hydrological data sets are presented.

Keywords: Data modeling · Extended distributions · Hazard rate · Maximum likelihood estimation · Monte Carlo simulations · Precipitation data.

Mathematics Subject Classification: Primary 60E05 · Secondary 62E10 · 62N05

1. INTRODUCTION

Lindley (1958) introduced a one-parameter distribution in the context of fiducial and Bayesian statistics, which is obtained as a mixture of exponential(λ) and gamma($2, \lambda$) probability density functions (PDFs), as defined in Equation (2). Aly and Benkherouf (2011) recently proposed a convenient method for adding two parameters to a baseline distribution, which gives rise to what is referred to as the Harris extended (HE) family of distributions. This family includes the baseline distribution itself as a basic exemplar and provides more flexibility for modeling various types of data. This novel approach is based on the probability generating function of a discrete distribution introduced by Harris (1948). In this paper, we define a three-parameter generalization of the Lindley distribution by applying to it the HE generator, the resulting model being named the Harris extended Lindley (HEL) distribution. This distribution is in fact an extension of the Marshall-Olkin

*Corresponding author. Email: mansoor.abbasi143@gmail.com

extended Lindley (MOL) distribution that was proposed by [Ghitany et al. \(2012\)](#), and its additional shape parameter α ought to provide an improved fit related to the MOL distribution. This extra parameter helps in controlling the shape of the HE PDF and enables us to model heavy-tailed distributions which are fairly common in hydrology; see, e.g., [Li et al. \(2013\)](#) and [Ashkar and El Adlouni \(2014\)](#). Moreover, the new distribution has an interesting physical interpretation when α is a positive integer and $0 < \theta < 1$: it is indeed the distribution of the time until failure of a device composed of N serial components having constant failure rate, where N is a random variable which arises from a branching process such as that described in [Harris \(1948\)](#). This distribution can be utilized for modeling purposes in research fields such as hydrology, engineering, insurance, biology and epidemiology wherein skewed positive data are frequently encountered.

One of the most crucial aspects of hydrological data analysis consists in achieving a close fit to the experimental data by employing proper statistical models. The Gumbel, Weibull, gamma, generalized logistic as well as other well-known distributions have been extensively utilized for modeling hydrological observations such as rainfall, flood, precipitation and stream flow data; see, e.g., [Zelenhasic \(1970\)](#), [Chadwick et al. \(2004\)](#), [Heo and Boes \(2011\)](#), [Bhunya et al. \(2012\)](#) and [Kang et al. \(2015\)](#). Yet, there exists a need for developing more flexible statistical models that would be applicable to data sets related to hydrological structures and phenomena or water resource planning and management, and the proposed three-parameter generalization of the Lindley distribution fits the purpose.

Although little attention has been paid to the Lindley distribution, there has recently been a surge of interest in this model, generalizations thereof and related applications. [Nadarajah et al. \(2007\)](#) introduced the exponentiated Lindley distribution as an alternative to the gamma, log-normal, Weibull and exponentiated exponential distributions; see also [Cordeiro et al. \(2016\)](#). Several properties of the Lindley distribution have been studied by [Ghitany et al. \(2008\)](#) who have shown that, for instance, it can provide a better fit than the exponential distribution. [Ghitany et al. \(2011\)](#) studied another two parameter extension of Lindley distribution and called it the weighted Lindley distribution. By making use of the Marshall-Olkin method, [Ghitany et al. \(2012\)](#) introduced and studied another extension of the Lindley model called the Marshall-Olkin extended Lindley (MOL) distribution. [Ghitany et al. \(2013\)](#) introduced a two-parameter power Lindley distribution and discussed its properties. A three-parameter generalization of the Lindley model was introduced by [Mervoci and Sharma \(2014\)](#). This extension, referred to as the beta Lindley (BL) distribution, is generated from the logit of a beta random variable. [Ghitany et al. \(2015\)](#) considered the problem of estimating the stress-strength parameter of the power Lindley distribution. [Mazucheli et al. \(2016\)](#) developed some statistical for testing hypotheses on the parameters of the weighted Lindley distribution. [Alizadeh et al. \(2017\)](#) introduced another extension of the power Lindley distribution.

The objective of this work is to derive the HEL distribution focusing on its probabilistic and statistics aspects, as well as applications in hydrology.

The remainder of the paper is organized as follows. We define the new distribution in Section 2. In Section 3, we provide computable expressions for some of its statistical functions such as its quantile function (QF), ordinary and incomplete moments, mean deviations, moment generating function (MGF) and order statistics. In Section 4, the model parameters are estimated by making use of the maximum likelihood (ML) method and a simulation study is carried out. In Section 5, we illustrate the usefulness of the proposed distribution by modeling two hydrological data sets. Finally, Section 6 offers some concluding remarks.

2. THE HEL DISTRIBUTION

In this section, we provide probabilistic aspects of the HEL distribution. The survival function (SF) and PDF of the distribution introduced by Lindley (1958) are respectively given by

$$\bar{G}_L(x) = \left(\frac{1 + \lambda + \lambda x}{1 + \lambda} \right) e^{-\lambda x}, \quad x > 0, \quad (1)$$

and

$$g_L(x) = \frac{\lambda^2}{\lambda + 1} (1 + x) e^{-\lambda x}, \quad x > 0, \quad (2)$$

where the parameter λ is assumed to be positive. We now describe a technique whereby the so-called Harris extended family of distributions can be generated and apply it to the Lindley distribution. The resulting distribution is referred to as the Harris extended Lindley (HEL) distribution. Let $G(x) = G(x; \xi)$ be a baseline cumulative distribution function (CDF) and

$$\bar{G}(x) = \bar{G}(x; \xi) = 1 - G(x; \xi)$$

be the corresponding SF of a lifetime random variable W , where $\xi = (\xi_1, \dots, \xi_q)$ is a parameter vector of dimension q . Furthermore, let $g(x) = g(x; \xi)$ be the PDF of W . The SF of the HE family is then defined by

$$\bar{F}_{\text{HE}}(x) = \frac{\theta^{1/\alpha} \bar{G}(x)}{[1 - \bar{\theta} \bar{G}(x)^\alpha]^{1/\alpha}}, \quad x > 0, \quad (3)$$

where $\bar{\theta} = 1 - \theta$, the parameters $\theta > 0$ and $\alpha > 0$ being additional shape parameters that allow for greater flexibility. Thereupon, the HE PDF has the form

$$f_{\text{HE}}(x) = \frac{\theta^{1/\alpha} g(x)}{[1 - \bar{\theta} \bar{G}(x)^\alpha]^{1+1/\alpha}}, \quad x > 0.$$

Aly and Benkherouf (2011) pointed out that when $\alpha > 0$ is a positive integer, the HE family can be looked upon as resulting from examining a simple discrete branching process where a particle either splits into $(\alpha + 1)$ identical branches or remains the same during a short interval. Clearly, Equation (3) constitutes a flexible generator for obtaining new parametric distributions from existing ones. For $\theta = 1$, $\bar{F}(x) = \bar{G}(x)$ and $\bar{G}(x)$ is thus a basic exemplar of the distribution. Additionally, the Marshall and Olkin (1997) extended (MOE) family arises from Equation (3) by letting $\alpha = 1$. Accordingly, the HE family can be viewed as a generalization of the MOE family.

The SF of the HEL distribution is defined as

$$\bar{F}(x) = \frac{\theta^{1/\alpha} \bar{G}_L(x)}{[1 - \bar{\theta} \bar{G}_L(x)^\alpha]^{1/\alpha}}, \quad x > 0, \quad (4)$$

for $\alpha > 0$, $\theta > 0$, $\lambda > 0$, where $\bar{G}_L(x)$ is given in Equation (1), with its PDF corresponding

to Equation (4) being

$$f(x) = \frac{\theta^{1/\alpha} \lambda^2 (1+x) e^{-\lambda x}}{(1+\lambda) [1 - \bar{\theta} \bar{G}_L(x)^\alpha]^{1+1/\alpha}}, \quad x > 0. \quad (5)$$

Henceforth, a random variable X having the PDF specified in Equation (5) is denoted by $X \sim \text{HEL}(\theta, \alpha, \lambda)$. This three-parameter PDF has two shape parameters and one scale parameter, and it can be either decreasing or unimodal. The two main special cases of the HEL model are: (i) the MOL distribution in which case $\alpha = 1$; (ii) the Lindley distribution which is obtained by letting $\alpha = \theta = 1$. The hazard rate (HR) associated with HEL model is given by

$$h(x) = \frac{\lambda^2 (1+x)}{(\lambda + 1 + \lambda x)} [1 - \bar{\theta} \bar{G}_L(x)^\alpha]^{-1}, \quad x > 0.$$

This HR can assume the four principal shapes associated with increasing, decreasing, bathtub-shaped or upside-down bathtub-shaped HRs. The HEL model is thus most appropriate to analyze a variety of hydrological and lifetime data sets. We note that there appears to be very few three-parameter distributions in the literature whose HR can take on the four main shapes of an HR. Moreover, the SF and HR of the HEL distribution have closed-form representations. Accordingly, this model can readily be utilized to analyze censored data sets. As well, simulating it is straightforward.

Figures 1 and 2 display some plots of the PDF and HR of the HEL distribution for certain parameter values. Figure 1 indicates that the HEL PDF can be right-skewed and reversed-J shaped. Figure 2 reveals that the HEL HR can be increasing (IFR), decreasing (DFR), upside-down bathtub (UBT) or bathtub-shaped (BT).

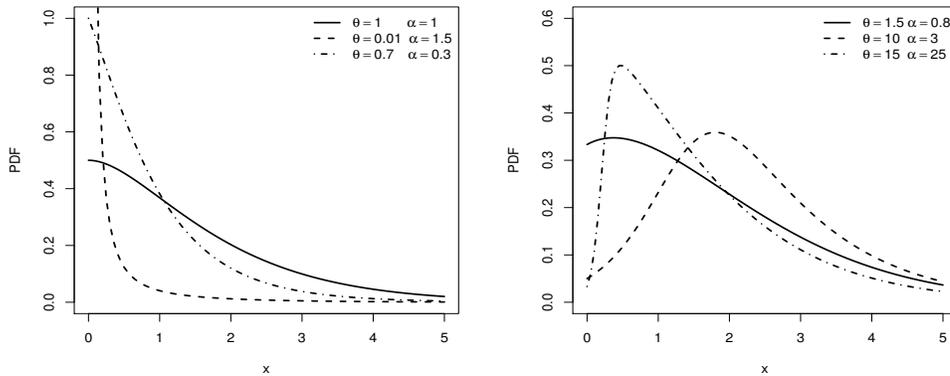


Figure 1. Plots of the HEL PDF for certain parameter values.

Given the functional form of the HEL PDF denoted by $f(x)$, a general representation of the mode that would be expressible in terms of the parameters of the distribution does not appear to be tractable. However, for a specific set of parameters, the command `NSolve[f'[x]==0,x,Reals]` in *Mathematica* can readily be utilized to determine the mode. If the solution happens to be greater than zero, then the PDF has a mode at that point; otherwise, it is strictly decreasing on the positive half-line. The extremum of the HR can be similarly obtained whenever it exists.

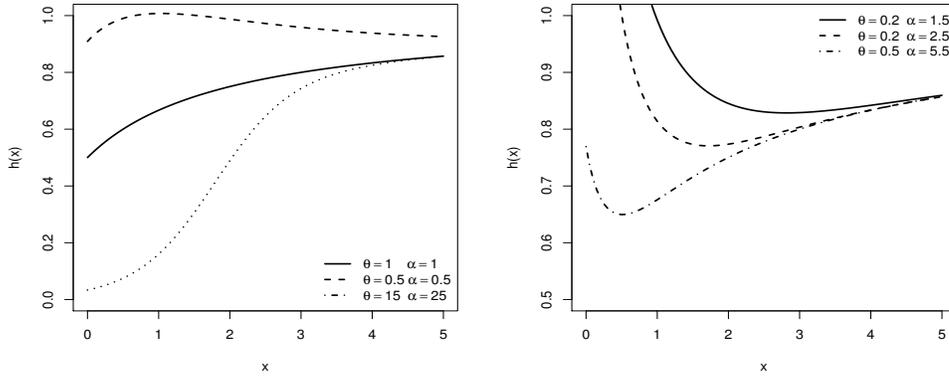


Figure 2. Plots of the HEL HR for certain parameter values.

3. STATISTICAL FUNCTIONS OF THE HEL DISTRIBUTION

In this section, we provide computable representations of certain statistical functions of the HEL distribution. More specifically, we focus, in order, on the quantile function, some useful expansions, the moments, including the incomplete ones, the moment generating function and the order statistics. The derived expressions can be easily evaluated by most symbolic computation software packages such as **Maple**, **Mathematica** and **Matlab**. These platforms can process analytic expressions of great complexity. Whenever available, an explicit representation of a statistical function is preferable to its determination by numerical integration.

The QF of a distribution has numerous uses in both statistical theory and applications. In the case of the HEL distribution, its QF is obtained by inverting the HEL CDF and is given by

$$Q(u) = -1 - \frac{1}{\lambda} - \frac{1}{\lambda} W \left[-(1 + \lambda) \frac{1 - \tau}{e^{1+\lambda}} \right], \quad 0 < u < 1, \tag{6}$$

where $\tau = 1 - (1 - u) [\theta + \bar{\theta}(1 - u)^\alpha]^{-1/\alpha}$ and $W(x)$ is the negative branch of the Lambert W function, see [Corless et al. \(1996\)](#) and [Jodrá \(2010\)](#) for details on its properties. The Lambert function cannot be expressed in terms of elementary functions. However, it is analytically differentiable and integrable and its principal branch satisfies $x = W(x e^x)$, $x \geq -1$. Furthermore, whenever $|x| \leq e^{-1}$, $W(x) = \sum_{n=1}^{\infty} (-n)^{n-1} x^n / n$. Clearly, if U has a uniform distribution in the interval $(0, 1)$, then $X = Q(U)$ has the PDF specified in Equation (5). The Lambert W function is implemented within various scientific libraries, as for example, in the R software (by the `lamW` package), **Mathematica** (by the `ProductLog` function), **Matlab** (by the `lambertw` function) and **Maple** (by the `LambertW` function), thus allowing for efficient evaluation of the QF of the HEL distribution.

Some useful expansions are now provided. Let $g_a(x) = a g(x) \bar{G}(x)^{a-1}$ be the Lehmann type-II-G (LII-G) PDF with power parameter $a > 0$. We demonstrate that the HEL PDF can be expressed as a linear combination of LII-Lindley (LIIL) PDFs. First, for $0 < \theta < 1$, we consider the negative binomial series

$$(1 - z)^{-p} = \sum_{i=0}^{\infty} \frac{\Gamma(p + i)}{\Gamma(p) i!} z^i,$$

which holds for $|z| < 1$ and any real number $p > 0$, where $\Gamma(a) = \int_0^{\infty} z^{a-1} e^{-z} dz$ is the

complete gamma function. Using this power series in Equation (5), we have

$$f(x) = \theta^{1/\alpha} g_L(x) \sum_{j=0}^{\infty} \bar{\theta}^j \frac{\Gamma(\alpha^{-1} + 1 + j)}{\Gamma(\alpha^{-1} + 1)j!} \bar{G}_L(x)^{j\alpha},$$

where $\bar{G}_L(x)$ and $g_L(x)$ are the SF and PDF of the Lindley distribution as provided by Equations (1) and (2). Note that for $\theta > 1$, we can write

$$f(x) = \theta^{-1} g_L(x) \sum_{j=0}^{\infty} \sum_{\ell=j}^{\infty} (-1)^j \left(\frac{\theta-1}{\theta}\right)^\ell \binom{\ell}{j} \frac{\Gamma(\alpha^{-1} + 1 + \ell)}{\Gamma(\alpha^{-1} + 1)\ell!} \bar{G}_L(x)^{j\alpha}.$$

On combining the last two expressions for $f(x)$ in a single one, we have

$$f(x) = \sum_{j=0}^{\infty} w_j h_{j\alpha+1}(x), \quad (7)$$

where $h_{j\alpha+1}(x) = (j\alpha + 1) g_L(x) \bar{G}_L(x)^{j\alpha}$ is the LIIL PDF with power parameter $j\alpha + 1$ and

$$w_j = w_j(\alpha, \theta) = \begin{cases} \frac{\theta^{1/\alpha} \bar{\theta}^j \Gamma(\alpha^{-1} + 1 + j)}{(j\alpha + 1) \Gamma(\alpha^{-1} + 1) j!}, & 0 < \theta < 1 \\ \frac{(-1)^j \theta^{-1}}{(j\alpha + 1)} \sum_{\ell=j}^{\infty} \left(\frac{\theta-1}{\theta}\right)^\ell \binom{\ell}{j} \frac{\Gamma(\alpha^{-1} + 1 + \ell)}{\Gamma(\alpha^{-1} + 1) \ell!}, & \theta > 1. \end{cases}$$

Equation (7) reveals that the HEL PDF (for any $\theta > 0$) can indeed be expressed as a linear combination of LIIL PDFs. It can also be shown that the HEL PDF can be expressed as a linear combination of gamma PDFs. Given Equations (1) and (2), it follows from the representation of Equation (7) that

$$f(x) = \sum_{j=0}^{\infty} w_j (j\alpha + 1) \left(\frac{\lambda^2}{\lambda + 1}\right) (1 + x) \left(1 + \frac{\lambda x}{1 + \lambda}\right)^{j\alpha} e^{-(j\alpha+1)\lambda x}.$$

On expanding $[1 + \lambda x / (1 + \lambda)]^{j\alpha}$ and using the Taylor series $z^\beta = \sum_{k=0}^{\infty} (\beta)_k (z-1)^k / k!$, where $(\beta)_k = \beta(\beta-1)\cdots(\beta-k+1)$ is the falling factorial, after some algebra, we obtain

$$f(x) = \sum_{i,j=0}^{\infty} v_{i,j} x^i (1+x) e^{-(j\alpha+1)\lambda x}, \quad (8)$$

where $v_{i,j} = (j\alpha + 1) w_j [\lambda^{2+i} / (\lambda + 1)^{i+1}] (j\alpha)_i / i!$ for $i, j = 0, 1, 2, \dots$

Letting $\pi(x; \alpha, \beta) = \beta^\alpha x^{\alpha-1} e^{-\beta x} / \Gamma(\alpha)$ be the gamma PDF with shape parameter $\alpha > 0$ and rate parameter $\beta > 0$, we can then rewrite Equation (8) as

$$f(x) = \sum_{i,j=0}^{\infty} \left[v_{i,j}^{(1)} \pi(x; i+1, (j\alpha + 1)\lambda) + v_{i,j}^{(2)} \pi(x; i+2, (j\alpha + 1)\lambda) \right], \quad (9)$$

where $v_{i,j}^{(1)} = i! v_{i,j} / [(j\alpha + 1)\lambda]^{i+1}$ and $v_{i,j}^{(2)} = (i+1)! v_{i,j} / [(j\alpha + 1)\lambda]^{i+2}$.

Equation (9) indicates that the HEL PDF can also be expressed as a linear combination of gamma PDFs. Thus, this representation can be used to obtain explicit expressions for

the ordinary and incomplete moments and the MGF of the HEL distribution from the corresponding quantities associated with the gamma distribution. Equations (7) and (9) constitute the main results of this section.

Certain of the main characteristics of a distribution such as tendency, dispersion, skewness and kurtosis can be investigated via its moments. We now establish that the ordinary moments of the HEL distribution can be obtained as infinite power series. It follows from Equation (7) that

$$\mu'_r = E(X^r) = \frac{\lambda^2}{1 + \lambda} \sum_{j=0}^{\infty} w_j \int_0^{\infty} x^r (1 + x) \left(1 + \frac{\lambda x}{1 + \lambda}\right)^{j\alpha} e^{-\lambda(j\alpha+1)x} dx,$$

or equivalently

$$\mu'_r = \frac{\lambda^2}{1 + \lambda} \sum_{j=0}^{\infty} w_j \int_0^{\infty} x^r (1 + x) \sum_{i=0}^{\infty} \left(\frac{\lambda}{1 + \lambda}\right)^i x^i \frac{(j\alpha)_i}{i!} e^{-\lambda(j\alpha+1)x} dx.$$

After some algebra, we obtain

$$\mu'_r = \frac{\lambda^2}{1 + \lambda} \sum_{i,j=0}^{\infty} p_{i,j} \frac{\Gamma(r + i + 1)}{[\lambda(j\alpha + 1)]^{r+i+1}} \left(1 + \frac{r + i + 1}{\lambda(j\alpha + 1)}\right), \tag{10}$$

where $p_{i,j} = w_j [(j\alpha)_i / i!] (\lambda / (1 + \lambda))^i$.

Table 1 includes numerical values for the first four ordinary moments of the HEL distribution as evaluated from Equation (10) by truncating the series to 100 terms and computed by numerical integration for some parameter values. We note that the numerical values obtained from both approaches are consistently in close agreement.

Table 1. Ordinary moments of the HEL distribution for certain parameter values with $\lambda = 10$.

μ'_r	$\alpha = 0.5$		$\alpha = 1.5$	
	Numerical	Equation (10)	Numerical	Equation (10)
$\theta = 0.5$				
μ'_1	0.0670906	0.0670905	0.0833919	0.08156687
μ'_2	0.0105268	0.01052653	0.0158697	0.01586975
μ'_3	0.00276376	0.002763106	0.00492889	0.004928885
μ'_4	0.0010382	0.001036676	0.0020813	0.002081299
$\theta = 1.5$				
μ'_1	0.141446	0.1414455	0.127601	0.1276013
μ'_2	0.0364545	0.0364543	0.0295221	0.02952214
μ'_3	0.0132554	0.01325516	0.0098071	0.009807097
μ'_4	0.00616152	0.006160951	0.00425269	0.004252694

The r th incomplete moment of X is given by $m_r(y) = \int_0^y x^r f(x) dx$. On making use of Equation (7) and proceeding as in the case of ordinary moments, we obtain

$$m_r(y) = \frac{\lambda^2}{1 + \lambda} \sum_{j,i=0}^{\infty} w_j \left(\frac{\lambda}{1 + \lambda}\right)^i \frac{(j\alpha)_i}{i!} \int_0^y x^{r+i} (1 + x) e^{-\lambda(j\alpha+1)x} dx. \tag{11}$$

On expressing the integral in Equation (11) in terms of the incomplete gamma function

$\gamma(a, y) = \int_0^y z^{a-1} e^{-z} dz$, we have

$$m_r(y) = \frac{\lambda^2}{1+\lambda} \sum_{i,j=0}^{\infty} K_{i,j} \left\{ \frac{\gamma(r+i+1, (j\alpha+1)\lambda y)}{[(j\alpha+1)\lambda]^{r+i+1}} + \frac{\gamma(r+i+2, (j\alpha+1)\lambda y)}{[(j\alpha+1)\lambda]^{r+i+2}} \right\}, \quad (12)$$

where $K_{i,j} = w_j [\lambda/(1+\lambda)]^i (j\alpha)_i / i!$ for $i, j = 0, 1, \dots$

Bonferroni and Lorenz curves as well as mean deviations can be determined by letting $r = 1$ in Equation (12). The Bonferroni and Lorenz curves are defined (for a given probability π) as $B(\pi) = m_1(q)/(\pi\mu'_1)$ and $L(\pi) = m_1(q)/\mu'_1$, respectively, where $q = Q(\pi)$ may be established from Equation (6). The mean deviations about the mean and about the median are given by $\delta_1 = E(|X - \mu'_1|) = 2\mu'_1 F(\mu'_1) - 2m_1(\mu'_1)$ and $\delta_2 = E(|X - M|) = \mu'_1 - 2m_1(M)$, where the median M and the mean μ'_1 can be evaluated from Equations (6) and (10), respectively. We now provide a general formula for $M(t) = E(e^{tX})$, the MGF of X . The MGF of the gamma PDF with parameters α and β is $(1 - t/\beta)^{-\alpha}$ ($t < \beta$). Then, it follows from Equation (9) that, for $t < \lambda$,

$$M(t) = \sum_{i,j=0}^{\infty} \left[v_{i,j}^{(1)} \left(1 - \frac{t}{(j\alpha+1)\lambda}\right)^{-i-1} + v_{i,j}^{(2)} \left(1 - \frac{t}{(j\alpha+1)\lambda}\right)^{-i-2} \right].$$

The last aspect being discussed in this section is the distribution of order statistics. Order statistics appear in many areas of statistical theory and practice. Suppose X_1, \dots, X_n is a random sample from the HEL distribution and let $X_{i:n}$ denote the i th order statistic. The PDF of $X_{i:n}$ can be expressed as

$$f_{i:n}(x) = K \sum_{k=0}^{n-i} (-1)^k \binom{n-i}{k} f(x) F(x)^{k+i-1}, \quad (13)$$

where $K = 1/B(i, n-i+1)$ and $B(p, q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$ is the beta function.

Consider the following representation available from [Gradshteyn and Ryzhik \(2000\)](#) for a power series raised to a positive integer n :

$$\left(\sum_{j=0}^{\infty} a_j u^j \right)^n = \sum_{j=0}^{\infty} b_{n,j} u^j, \quad (14)$$

where the coefficients $b_{n,j}$, for $n = 1, 2, \dots$ and $j = 1, 2, \dots$, are obtained from the recursive equation

$$b_{n,j} = (j a_0)^{-1} \sum_{m=1}^j [m(n+1) - j] a_m b_{n,j-m},$$

with $b_{n,0} = a_0^n$. On integrating the right-hand side of Equation (7), we can write

$$F(x) = \bar{G}_L(x) \sum_{j=0}^{\infty} w_j \bar{G}_L(x)^{j\alpha},$$

and then making use of Equation (14), we have

$$F(x)^{k+i-1} = \sum_{j=0}^{\infty} t_{k+i-1,j} \bar{G}_L(x)^{j \alpha + k + i - 1},$$

where $t_{k+i-1,j} = (j w_0)^{-1} \sum_{m=1}^j [m(k+i) - j] w_m t_{k+i-1,i-m}$ for $j \geq 1$ and $t_{k+i-1,0} = w_0^{k+i-1}$. Inserting the previous expression for $F(x)^{k+i-1}$ and the representation of Equation (7) of the PDF appearing in Equation (13) gives

$$f_{i:n}(x) = K \sum_{r,j=0}^{\infty} \sum_{k=0}^{n-i} v_{r,j,k} h_{(r+j) \alpha + k + i}(x), \quad (15)$$

where

$$v_{r,j,k} = \frac{(-1)^k (r \alpha + 1) w_r t_{k+i-1,j}}{(r+j) \alpha + k + i} \binom{n-i}{k}.$$

Equation (15) reveals that the PDF of the HEL order statistics can be expressed as a triple linear combination of LIII PDFs. Accordingly, certain mathematical properties of the HEL order statistics could be determined from those of the LIII distribution.

4. PARAMETER ESTIMATION

We now discuss the estimation of the model parameters using the ML method. There exist several approaches for estimating parameters; however, the ML method is the most commonly employed. The ML estimators enjoy several desirable properties and can be utilized in the construction of confidence intervals for the model parameters. They also appear in some test statistics. The normal approximation to the distribution of these estimators follows from large sample distribution theory.

Let X_1, \dots, X_n be a sample of size n from the HEL distribution whose associated PDF is given in Equation (5). The log-likelihood function $\ell = \ell(\Theta)$ of the vector of parameters $\Theta = (\theta, \alpha, \lambda)^\top$ is given by

$$\ell = \frac{n}{\alpha} \log \theta + n \log \left(\frac{\lambda^2}{1 + \lambda} \right) + \sum_{i=1}^n \log(1 + x_i) - \lambda x_i - \left(1 + \frac{1}{\alpha}\right) \sum_{i=1}^n \log[1 - \theta \bar{G}_L(x)^\alpha]. \quad (16)$$

The ML estimates $\hat{\theta}$, $\hat{\alpha}$ and $\hat{\lambda}$ are determined by maximizing the log-likelihood function of Equation (16) with respect to the parameters θ , α and λ . In general, there is no closed-form representation for these estimates, which are determined in practice the by making use of numerical methods. Equation (16) can be maximized either directly by using the R (`optim` function), SAS (`NLMixed` procedure) or Ox (`MaxBFGS` function), or by solving the nonlinear likelihood equations obtained by equating the partial derivatives of ℓ with respect to each parameter to zero.

The components of the score vector $U(\Theta)$ are expressed as

$$U_\theta = \frac{n}{\alpha\theta} - \left(1 + \frac{1}{\alpha}\right) \sum_{i=1}^n \frac{\bar{G}_L(x)^\alpha}{1 - \theta\bar{G}_L(x)^\alpha},$$

$$U_\alpha = -\frac{n}{\alpha^2} \log \theta - \frac{1}{\alpha^2} \sum_{i=1}^n \frac{\bar{G}_L(x)^\alpha \log \bar{G}_L(x)}{1 - \theta\bar{G}_L(x)^\alpha},$$

$$U_\lambda = \frac{n(2 + \lambda)}{\lambda + \lambda^2} + x_i + \left(1 + \frac{1}{\alpha}\right) \sum_{i=1}^n \left[\frac{\alpha \theta \bar{G}_L(x)^{\alpha-1}}{(1 + \lambda)^2 [1 - \theta\bar{G}_L(x)^\alpha]} \lambda x_i [2 + \lambda + (1 + \lambda)x_i] e^{-\lambda x_i} \right].$$

Setting these equations to zero and solving them simultaneously yields the ML estimates of the model parameters.

We now assess the performance of the ML estimators of the model parameters by means of Monte Carlo simulations. The simulations are replicated 1,000 times with samples of sizes $n = 50, 100, 200$ and the following parameter values: I: $\theta = 0.5, \alpha = 0.5$ and $\lambda = 1$; II: $\theta = 0.1, \alpha = 1.5$ and $\lambda = 1$; III: $\theta = 1.5, \alpha = 0.5$ and $\lambda = 1$; IV: $\theta = 1.5, \alpha = 1.5$ and $\lambda = 1$. Table 2 lists the average bias (Bias) of the ML estimators, mean squared errors (MSE), coverage probabilities (CP) and average widths (AW) of the confidence intervals for the parameters θ, α and λ and the three sample sizes. From these results, we conclude that the ML estimators perform well when it comes to estimating the parameters of the HEL distribution. In general, the biases, MSEs and AWs decrease when the sample size increases. Moreover, the CPs of the confidence intervals are quite close to the 95% nominal level. Thus, the ML estimators and their asymptotic distributional properties can be adopted for constructing approximate confidence intervals for the parameters of the HEL distribution.

5. EMPIRICAL ILLUSTRATIONS WITH HYDROLOGICAL DATA

In this section, we fit the HEL model and some other competing models to two hydrological data sets. We assess how well the HEL distribution performs as compared to the beta-Lindley (BL) studied by [Mervoci and Sharma \(2014\)](#), exponentiated power Lindley (EPL) due to [Ashour and Eltehiwy \(2015\)](#), beta-exponential (BE) proposed by [Nadarajah and Kotz \(2006\)](#), exponentiated Nadarajah and Haghghi (ENH) defined by [Lemonte \(2013\)](#), Harris extended exponential (HEE) discussed by [Pinho et al. \(2015\)](#), exponentiated Weibull (EW) studied by [Mudholkar and Sharivastava \(1993\)](#), power Lindley (PL) introduced by [Ghitany et al. \(2013\)](#), exponentiated Lindley defined by [Nadarajah et al. \(2007\)](#) and Lindley distributions. For each model, we estimated the parameters by the ML method and assessed the goodness-of-fit by means of the Akaike information criterion (AIC), Cramér-von Mises (W), Anderson-Darling (AD), Kolmogrov-Smirnov (KS) and average scaled absolute error (ASAE) statistics. The ASAE is defined as (see [Castilo and Hadi, 2005](#)) $ASAE = (1/n) \sum_{i=1}^n (|x_{(i)} - \hat{x}_{(i)}|) / (x_{(n)} - x_{(1)})$, where $x_{(i)}$ is the observed value of i th order statistic, and $\hat{x}_{(i)}$ is obtained from the QF, $Q(u_i)$, wherein the ML estimates are substituted to the parameters, with $u_i = i/(n+1)$. The ASAE statistic is useful for measuring the accuracy of the fitted model. In general, the smaller values of the above statistics indicate a better fit to the data.

Table 2. Monte Carlo simulation results for the listed statistical indicator.

Parameter	n	Bias	MSE	CP	AW
I					
θ	50	-0.044	0.112	0.92	1.483
	100	-0.037	0.045	0.95	0.979
	200	-0.037	0.033	0.98	0.749
α	50	0.626	1.690	0.96	2.079
	100	0.419	0.429	0.95	0.983
	200	0.314	0.110	0.95	0.799
λ	50	-0.028	0.193	0.93	1.459
	100	-0.042	0.111	0.96	1.154
	200	-0.046	0.079	0.95	0.123
II					
θ	50	0.022	0.007	0.93	0.368
	100	0.012	0.003	0.96	0.232
	200	0.004	0.001	0.95	0.153
α	50	0.621	1.340	0.95	4.809
	100	0.199	0.537	0.95	2.475
	200	0.078	0.167	0.95	1.588
λ	50	0.162	0.293	0.91	2.117
	100	0.080	0.133	0.94	1.436
	200	0.026	0.063	0.95	0.994
III					
θ	50	1.317	0.589	0.98	1.508
	100	0.609	0.371	0.98	1.192
	200	0.288	0.148	0.96	0.506
α	50	1.375	0.473	0.90	1.624
	100	0.563	0.171	0.98	1.270
	200	0.157	0.049	0.95	0.014
λ	50	0.264	0.479	0.91	1.006
	100	0.204	0.278	0.95	0.214
	200	0.199	0.130	0.96	0.102
IV					
θ	50	0.638	3.602	0.90	2.835
	100	0.237	1.276	0.91	1.401
	200	0.141	0.629	0.94	0.038
α	50	-0.003	0.083	0.96	1.156
	100	0.015	0.042	0.96	0.818
	200	-0.001	0.021	0.95	0.571
λ	50	0.117	0.255	0.96	1.977
	100	0.035	0.104	0.96	1.323
	200	0.024	0.055	0.96	0.923

The CDFs of the BL, EPL, BE, ENH, HEE, EW, MOL, PL and EL distributions are given by

$$F_{\text{BL}}(x, a, b, \theta) = I_{1 - (1 + \frac{\theta x}{1 + \theta})e^{-\theta x}}(a, b), \quad x, \theta > 0,$$

$$F_{\text{EPL}}(x, \alpha, \beta, \theta) = \left(1 - \left(1 + \frac{\theta x^\beta}{1 + \theta}\right)e^{-\theta x^\beta}\right)^\alpha, \quad x, \alpha, \beta, \theta > 0,$$

$$F_{\text{BE}}(x, a, b, \lambda) = I_{1 - e^{-\lambda x}}(a, b), \quad x, a, b, \lambda > 0.$$

$$F_{\text{ENH}}(x, \beta, \alpha, \lambda) = \left(1 - e^{1 - (1 + \lambda x)^\alpha}\right)^\beta, \quad x, \beta, \alpha, \lambda > 0,$$

$$F_{\text{HEE}}(x, \beta, k, \lambda) = \frac{\beta^{1/k} e^{-\lambda x}}{[1 - (1 - \beta)e^{-\lambda k x}]^{1/k}}, \quad x, \beta, k, \lambda > 0,$$

$$F_{\text{EW}}(x; c, \alpha, \lambda) = \left(1 - e^{-(x/\lambda)^c}\right)^\alpha, \quad x, c, \alpha, \lambda > 0,$$

$$F_{\text{MOL}}(x, \alpha, \lambda) = \frac{1 - (1 + \lambda)^{-1}[1 + \lambda + \lambda x]e^{-\lambda x}}{1 - (1 - \alpha)(1 + \lambda)^{-1}[1 + \lambda + \lambda x]e^{-\lambda x}}, \quad x, \alpha, \lambda > 0,$$

$$F_{\text{PL}}(x, \beta, \theta) = 1 - \left(1 + \frac{\theta x^\beta}{1 + \theta}\right)e^{-\theta x^\beta}, \quad x, \beta, \theta > 0,$$

$$F_{\text{EL}} = \left[1 - \left(\frac{1 + \theta + \theta x}{1 + \theta}\right)e^{-\theta x}\right]^\alpha, \quad x, \theta > 0,$$

respectively, where $I_z(p, q)$ denotes the incomplete beta function.

First, we consider a data set consisting of s exceedances (rounded to one decimal place) of flood peaks (in m^3/s) of the Wheaton river, which is located in the Yukon Territory, Canada, for the years 1958-1984. The data set is the following: 1.7, 2.2, 14.4, 1.1, 0.4, 20.6, 5.3, 0.7, 1.9, 13.0, 12.0, 9.3, 1.4, 18.7, 8.5, 25.5, 11.6, 14.1, 22.1, 1.1, 2.5, 14.4, 1.7, 37.6, 0.6, 2.2, 39.0, 0.3, 15.0, 11.0, 7.3, 22.9, 1.7, 0.1, 1.1, 0.6, 9.0, 1.7, 7.0, 20.1, 0.4, 2.8, 14.1, 9.9, 10.4, 10.7, 30.0, 3.6, 5.6, 30.8, 13.3, 4.2, 25.5, 3.4, 11.9, 21.5, 27.6, 36.4, 2.7, 64.0, 1.5, 2.5, 27.4, 1.0, 27.1, 20.2, 16.8, 5.3, 9.7, 27.5, 2.5, 27.0. Some summary statistics of these data are: $n = 72$, $\bar{x} = 12.20417$, $s = 12.29722$, coefficient of skewness = 1.47251 and coefficient of kurtosis = 2.88955. The boxplot of these observations displayed in Figure 3(a) indicates that the distribution is right-skewed. The TTT (total time on test) plot (see, e.g., Gill, 1986; Aarset, 1987) of these data is shown in Figure 3(b). It is first convex and then concave, which suggests a bathtub-shaped failure rate. Accordingly, the HEL distribution could, in principle, be appropriate for modeling these data. The ML estimates (with the corresponding standard errors -SEs- in parentheses) as well as the ASAE, AIC, KS, CM and AD statistics are given in Table 3. All five goodness-of-fit statistics indicate that the HEL model provides the best fit. For a visual comparison, the empirical SF (ESF) and estimated SF associated with the HEL model as well as a theoretical versus empirical probability (PP) plot, which compares the empirical CDF of the data with the fitted CDF, are respectively included in Figures 4(a) and 4(b). Clearly, the HEL model closely fits the data distribution.

In this second illustration, the data set, which is freely available on the Korea Meteorological Administration (KMA) website (<http://www.kma.go.kr>), represents the annual maximum daily rainfall amounts in millimeters in Seoul (Korea) during the period 1961-2002. Some summary statistics of these precipitation data are: $n = 128$, $\bar{x} = 144.5991$, $s = 66.17812$, coefficient of skewness = 0.94067 and coefficient of kurtosis = 0.80435. The boxplot of these observations that is displayed in Figure 5(a) indicates that the distribution is right-skewed. The TTT plot appearing in Figure 5(b) suggests an increasing failure

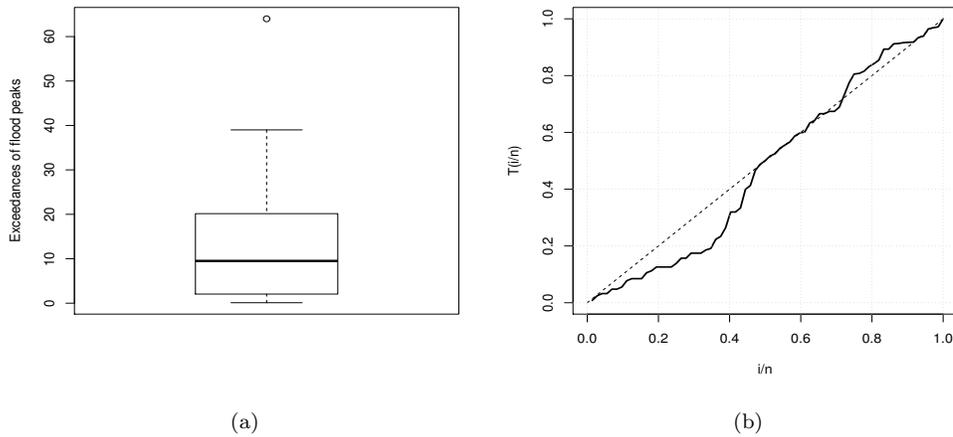


Figure 3. Boxplot (a) and TTT plot (b) for the flood data.

Table 3. ML estimates, SEs (in parentheses) and goodness-of-fit measures for the flood data.

Distribution	Estimates			ASAE	AIC	KS	CM	AD
HEL(θ, α, λ)	0.077 (0.038)	6.135 (2.031)	0.110 (0.014)	0.017	503.194	0.073	0.054	0.338
BL(a, b, θ)	0.556 (0.098)	0.275 (0.241)	0.334 (0.273)	0.020	510.206	0.115	0.126	0.775
EPL(α, β, θ)	0.916 (0.595)	0.730 (0.235)	0.300 (0.279)	0.025	510.425	0.106	0.149	0.857
BE(a, b, λ)	0.812 (0.137)	0.412 (0.290)	0.179 (0.131)	0.023	508.465	0.098	0.122	0.705
ENH(β, α, λ)	0.732 (0.137)	1.675 (0.143)	0.032 (0.032)	0.019	507.850	0.106	0.104	0.632
HEE(β, k, λ)	0.433 (0.193)	5.086 (0.147)	0.071 (0.011)	0.023	506.460	0.078	0.094	0.550
EW(c, α, λ)	1.387 (0.587)	0.519 (0.308)	0.016 (0.036)	0.403	508.050	0.107	0.105	0.642
MOL(α, λ)	0.216 (0.128)	0.090 (0.023)		0.044	522.571	0.175	0.582	4.148
PL(β, θ)	0.700 (0.057)	0.339 (0.056)		0.026	508.444	0.105	0.154	0.877
EL(α, θ)	0.509 (0.077)	0.104 (0.015)		0.021	509.349	0.117	0.135	0.833
L(θ)	0.153 (0.013)			0.044	530.424	0.241	0.819	7.424

rate. The estimates of the parameters of the fitted distributions are listed in Table 4. We note that the HEL model has the lowest ASAE, AIC, KS, CM and AD values, which indicate that it provides the most accurate fit to the data. Furthermore, the ESF and estimated SF and PP plots shown in Figures 6(a) and 6(b) also suggest a close fit to the data distribution.

A likelihood ratio test can be utilized to compare a distribution having additional parameters with some of its sub-models. Accordingly, we made use of the likelihood ratio test to assess the improvement in fit that the HEL distribution produces with respect to

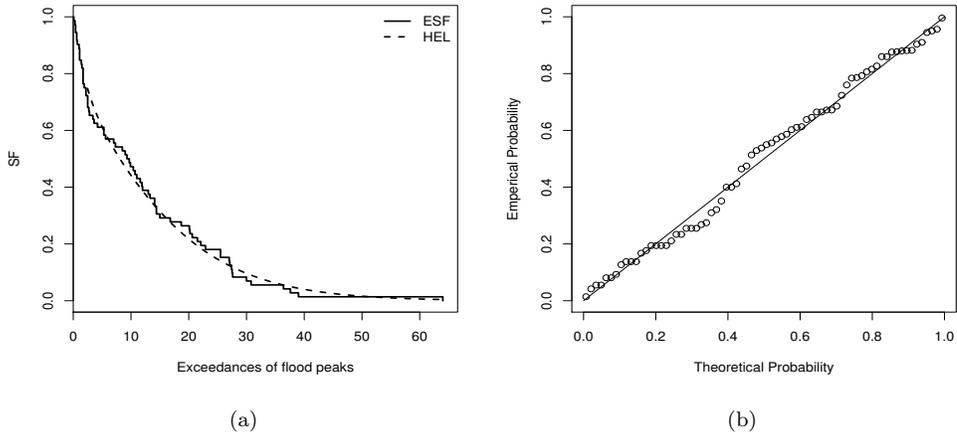


Figure 4. Empirical SF and estimated HEL SF (a) and PP plot (b) for the flood data.

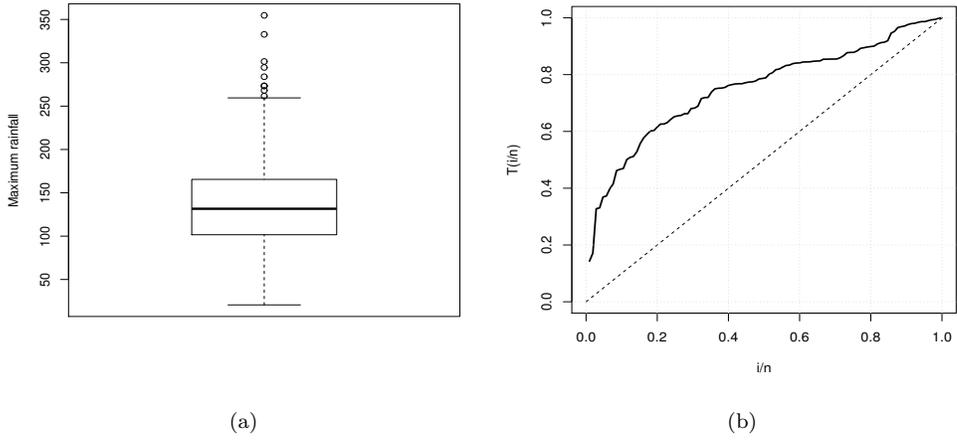


Figure 5. Boxplot (a) and TTT plot (b) for the precipitation data.

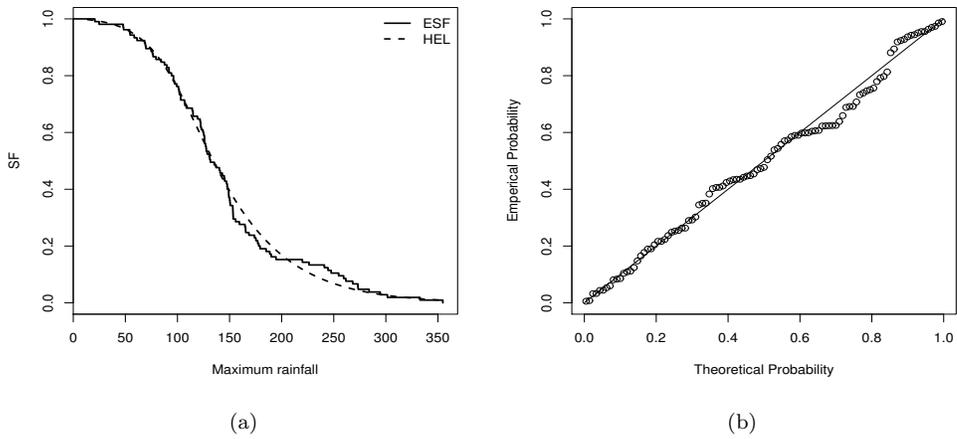


Figure 6. Empirical SF and estimated HEL SF (a) and PP plot (b) for the precipitation data.

Table 4. ML estimates, SEs (in parentheses) and goodness-of-fit measures for the precipitation data.

Distribution	Estimates			ASAE	AIC	KS	CM	AD
HEL(θ, α, λ)	17.443 (9.276)	3.081 (1.069)	0.022 (0.003)	0.019	1165.064	0.077	0.077	0.490
BL(a, b, θ)	2.776 (0.622)	1.117 (0.577)	0.020 (0.007)	0.022	1169.396	0.085	0.144	0.809
EPL(α, β, θ)	1.530 (0.225)	1.318 (0.025)	0.003 (0.004)	0.024	1168.717	0.097	0.150	0.862
BE(a, b, λ)	4.433 (0.685)	1.448 (0.535)	0.012 (0.003)	0.029	1172.022	0.092	0.263	1.412
ENH(β, α, λ)	4.183 (0.687)	1.694 (0.217)	0.006 (0.001)	0.024	1168.620	0.095	0.146	0.837
HEE(β, k, λ)	1.535 (0.299)	1.860 (0.847)	0.008 (0.001)	0.137	1241.535	0.276	2.569	13.078
EW(c, α, λ)	1.411 (0.334)	2.907 (1.519)	98.866 (29.851)	0.433	1168.586	0.093	0.142	0.821
MOL(α, λ)	10.455 (4.118)	0.029 (0.003)		0.032	1171.003	0.103	0.184	1.330
PL(β, θ)	0.014 (0.007)	16.182 (2.037)		1.433	4820.512	0.999	34.999	1631.130
EL(α, θ)	2.871 (0.501)	0.022 (0.002)		0.022	1167.600	0.084	0.146	0.818
L(θ)	0.014 (0.001)			0.584	1199.216	1.187	0.519	6.508

the Lindley and MOL distributions. It is known that, under the null hypothesis,

$$-2 \log \left(\frac{\text{likelihood under the null hypothesis}}{\text{likelihood in the whole parameter space}} \right) \sim \chi^2(d),$$

where, asymptotically, $\chi^2(d)$ follows a chi-square distribution having d degrees of freedom, d being equal to the number of additional parameters in the extended model. Using this result and standard statistical tables, we can obtain critical values for the test statistic. Table 5 includes the likelihood ratio statistics and corresponding p-values for the two data sets. Given the values of these statistics and their associated p-values, we reject the null hypotheses for both data sets and conclude that the HEL model provides a significantly better representation of the distribution of these data than the Lindley or MOL distributions. The 95% bootstrap confidence intervals obtained for the parameters θ, α and λ are given in Table 6.

Table 5. Likelihood ratio statistics and their p-values.

Hypothesis	Flood data	Precipitation data
H ₀ : $\alpha=1$ (MOL) H ₁ : $\alpha \neq 1$ (HEL)	21.377 (< 0.000)	7.939 (0.005)
H ₀ : $\alpha=\theta=1$ (L) H ₁ : $\alpha \neq 1, \theta \neq 1$ (HEL)	31.229 (< 0.000)	38.151 (<0.000)

Table 6. 95% bootstrap confidence intervals for the parameters θ , α and λ .

Data set	θ	α	λ
Flood data	(0.039, 0.225)	(3.036, 10.429)	(0.087, 0.146)
Precipitation data	(8.243, 20.463)	(1.378, 5.027)	(0.018, 0.031)

Next, we present the concepts of return period, mean deviation about a return level and the r th moment of the order statistics. For a given a data set, the return period can be estimated by $\hat{T} = 1/\bar{F}(x)$, where $\bar{F}(x) = 1 - F(x)$ and $F(x)$ denote the CDF of the distribution. The estimated return periods (\hat{T}) correspond to the return levels (x_T) for each of these two data sets. They are reported in Table 7 and have been computed as $T = 1/\bar{F}(x_T)$, where $\bar{F}(\cdot)$ is as given in Equation (4). The mean deviation about a return level which is the mean of the distances of the values from their return level is given by $\eta = 2x_T F(x_T) - x_T - \mu + 2 \int_{x_T}^{\infty} x f(x) dx$, where $f(\cdot)$ and $F(\cdot)$ denote the HEL PDF and CDF. Table 7 provides the mean deviations about certain values of the return levels (\bar{x}_T) for both the flood and precipitation data sets.

Table 7. Estimated return periods (\hat{T}) and mean deviations about the return levels (η).

Flood data			Precipitation data		
x_T	\hat{T}	η	x_T	\hat{T}	η
140	499147.836	127.800	410	315.215	265.623
100	8350.571	87.802	375.5	160.422	435.000
50	62.48360	38.135	315.5	50.389	172.849
30	10.375	19.949	260	17.693	121.247
10	2.265	9.337	210	7.093	80.513

In order to be able to plan for future emergencies in connection with various hydrological events, it is useful to ascertain some distributional results on certain of the order statistics. To that end, we determine the r th moment, for $r = 1, 2, 3, 4$, of some order statistics for each data sets under the HEL model wherein the parameters are replaced by their ML estimates. Those moments are included in Table 8 for each data set.

Table 8. Some numerical values of $E(X_{i:n}^r)$ for the indicated data set.

Flood data			Precipitation data		
i	r	$E(X_{i:72}^r)$	i	r	$E(X_{i:128}^r)$
1	1	0.097	1	1	21.409
	2	0.019		2	585.869
	3	0.006		3	18628.800
	4	0.002		4	658641.210
20	1	2.868	15	1	77.111
	2	8.962		2	5989.380
	3	30.433		3	468486.450
	4	111.999		4	3.689×10^4
60	1	22.898	30	1	98.427
	2	543.677		2	9719.320
	3	12726.600		3	962824.794
	4	308653.083		4	9.568×10^7

6. CONCLUDING REMARKS

We introduced a three-parameter extension of the Lindley distribution referred to as the Harris extended Lindley (HEL) distribution, which is obtained by applying the Harris extended method to the Lindley distribution. The proposed model has two shape parameters and one scale parameter. It includes as sub-models the Marshall-Olkin Lindley and Lindley distributions. The HEL PDF can be decreasing or unimodal. Moreover, the HEL HR can be increasing, decreasing, unimodal (upside-down bathtub) or bathtub-shaped. We gave explicit expressions for the ordinary and incomplete moments, mean deviations, Bonferroni and Lorenz curves and order statistics associated with the proposed distribution. The estimation of the model parameters was successfully carried out by making use of the maximum likelihood method. In conclusion, the HEL distribution provides a very flexible model for fitting the wide spectrum of positive data sets arising in engineering, survival analysis, hydrology, economics, biology as well as numerous other fields of scientific investigation. All the calculations were performed with the symbolic computing software Mathematica, the code being available from the authors upon request.

ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to the referees, an Associate Editor and the Editor-in-Chief for their valuable comments and suggestions which significantly improved the presentation of the results. The financial support of the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged by the the third author.

REFERENCES

- Aarset, M.V., 1987. How to identify bathtub hazard rate. *IEEE Transactions on Reliability*, 36, 106-108.
- Alizadeh, M., MirMostafaei, S.M.T.K., and Ghosh., I., 2017. A new extension of power Lindley distribution for analyzing bimodal data. *Chilean Journal of Statistics*, 8, 67-86.
- Aly, E.A.A. and Benkherouf, L., 2011. A new family of distributions based on probability generating functions. *Sankhyā B*, 73, 70-80.
- Ashkar, F. and El Adlouni, S., 2014. Correcting confidence intervals for quantiles of a heavy-tailed distribution: Case study of two-parameter kappa distribution. *Journal of Hydrology*, 512, 498-505.
- Ashour, S.K. and Eltehiwy, M.A., 2015. Exponentiated power Lindley distribution. *Journal of Advance Research*, 6, 895-905
- Bhunya, P.K., Singh, R.D., Berndtsson, R., and Panda, S.N., 2012. Flood analysis using generalized logistic models in partial duration series, *Journal of Hydrology*, 420, 59-71.
- Castillo, E., Hadi, A.S., Balakrishnan, N., and Sarabia, J.M., 2005. *Extreme Value and Related Models with Applications in Engineering and Science*, Wiley, Hoboken.
- Chadwick, A., Morfett, J., and Borthwick, M., 2004. *Hydraulics in Civil and Environmental Engineering*, Spon Press, London.
- Cordeiro, G.M., Alizadeh, M., Nascimento, A.D., and Rasekhi, M., 2016. The exponentiated Gompertz generated family of distributions: Properties and applications. *Chilean Journal of Statistics*, 7, 29-50
- Corless, R.M., Gonnet, G.H., Hare, D.E.G., Jeffrey, D.J., and Knuth, D.E., 1996. On the Lambert W function. *Advances in Computational Mathematics*, 5, 329-359.

- Ghitany, M.E., Al-Mutairi, D.K., and Aboukhamseen, S.M., 2015. Estimation of the reliability of a stress-strength system from power Lindley distributions. *Communications in Statistics - Theory and Methods*, 44, 118-136.
- Ghitany M.E., Al-Mutairi D.K., Al-Awadhi, F.A., and Al-Burais M.M., 2012. Marshall-Olkin extended Lindley distribution and its application. *International Journal of Applied Mathematics*, 25, 709-721.
- Ghitany, M.E., Al-Mutairi, D.K., Balakrishnan, N., and Al-Enezi, L.J., 2013. Power Lindley distribution and associated inference. *Computational Statistics and Data Analysis*, 64, 20-33.
- Ghitany, M.E., Alqallaf, F., Al-Mutairi, D.K. and Husain, H.A., 2011. A two-parameter weighted Lindley distribution and its applications to survival data. *Mathematics and Computers in Simulation*, 81, 1190-1201.
- Ghitany, M.E., Atieh, B., and Nadarajah, S., 2008. Lindley distribution and its application. *Mathematics in Computer and Simulation*, 78, 493-506.
- Gill, D.R. (1986). The total time on test plot and the cumulative total time on test statistic for a counting process. *The Annals of Statistics*, 14, 1234-1239.
- Gradshteyn, I.S. and Ryzhik, I.M., 2000. *Tables of Integrals, Series, and Products*, Academic Press, New York.
- Harris, T.E., 1948. Branching processes. *Annals of Mathematical Statistics*, 19, 474-494.
- Heo, J. and Boes, D.C., 2011. Regional flood frequency analysis based on a Weibull model: Part 2. Simulations and applications. *Journal of Hydrology*, 242, 171-182.
- Jodrá, J., 2010. Computer generation of random variables with Lindley or Poisson-Lindley distribution via the Lambert W function. *Mathematics and Computers in Simulation*, 81, 851-859.
- Kang, D., Ko, K., and Huh, J., 2015. Determination of extreme wind values using the Gumbel distribution. *Energy*, 86, 51-58.
- Lemonte, A.J., 2013. A new exponential-type distribution with constant, decreasing, increasing, upside-down bathtub and bathtub-shaped failure rate function. *Computational Statistics and Data Analysis*, 62, 149-170.
- Li, C., Singh, V.P., and Mishra, A.K. 2013. A bivariate mixed distribution with heavy-tailed component and its application to single-site daily rainfall simulation. *Water Resource Research*. 49, 767-789.
- Lindley, D.V., 1958. Fiducial distributions and Bayes theorem. *Journal of Royal Statistical Society B*, 20, 102-107.
- Marshall, A.N. and Olkin, I., 1997. A new method for adding a parameter to a family of distributions with applications to the exponential and Weibull families. *Biometrika*, 84, 641-652.
- Mazucheli, J., Coelho-Barros, E.A., and Louzada, F., 2016. On the hypothesis testing for the weighted Lindley distribution. *Chilean Journal of Statistics*, 7, 17-27.
- Mervoci, F. and Sharma, V.K., 2014. The beta Lindley distribution: Properties and Applications. *Journal of Applied Mathematics*, Art.ID 198951, 10 p.
- Mudholkar, G.S. and Srivastava, D.K., 1993. Exponentiated Weibull family for analyzing bathtub failure data. *IEEE Transactions on Reliability*, 42, 299-302.
- Nadarajah, S., Bakouch, H., and Tahmasbi, R., 2007. A generalized Lindley distribution. *Sankhya B*, 73, 331-359.
- Nadarajah, S. and Kotz, S., 2006. The beta exponential distribution. *Reliability Engineering and System Safety*, 91, 689-697.
- Pinho, L.G.B., Cordeiro, G.M., and Nobre, J.S., 2015. The Harris extended exponential distribution. *Communications in Statistics - Theory and Methods*, 44, 3486-3502.
- Zelenhasic, E.F., 1970. Theoretical probability distributions for flood peaks. *Hydrology papers (Colorado State University)*, ID 42.

INFORMATION FOR AUTHORS

The editorial board of the Chilean Journal of Statistics (ChJS) is seeking papers, which will be refereed. We encourage the authors to submit a PDF electronic version of the manuscript in a free format to Víctor Leiva, Editor-in-Chief of the ChJS (E-mail: chilean.journal.of.statistics@gmail.com). Submitted manuscripts must be written in English and contain the name and affiliation of each author followed by a leading abstract and keywords. The authors must include a "cover letter" presenting their manuscript and mentioning: "We confirm that this manuscript has been read and approved by all named authors. In addition, we declare that the manuscript is original and it is not being published or submitted for publication elsewhere".

PREPARATION OF ACCEPTED MANUSCRIPTS

Manuscripts accepted in the ChJS must be prepared in Latex using the ChJS format. The Latex template and ChJS class files for preparation of accepted manuscripts are available at <http://chjs.mat.utfsm.cl/files/ChJS.zip>. Such as its submitted version, manuscripts accepted in the ChJS must be written in English and contain the name and affiliation of each author, followed by a leading abstract and keywords, but now mathematics subject classification (primary and secondary) are required. AMS classification is available at <http://www.ams.org/mathscinet/msc/>. Sections must be numbered 1, 2, etc., where Section 1 is the introduction part. References must be collected at the end of the manuscript in alphabetical order as in the following examples:

Arellano-Valle, R., (1994). Elliptical Distributions: Properties, Inference and Applications in Regression Models. Unpublished Ph.D. Thesis. Department of Statistics, University of São Paulo, Brazil.

Cook, R.D., (1997). Local influence. In Kotz, S., Read, C.B. and Banks, D.L. (Eds.), Encyclopedia of Statistical Sciences, Update, Vol. 1, Wiley, New York, pp. 380-385.

Rukhin, A.L., (2009). Identities for negative moments of quadratic forms in normal variables. Statistics and Probability Letters 79, 1004-1007.

Stein, M.L., (1999). Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.

Tsay, R.S., Pena, D., and Pankratz, A.E., (2000). Outliers in multivariate time series. Biometrika 87, 789-804.

References in the text must be given by the author's name and year of publication, e.g., Gelfand and Smith (1990). In the case of more than two authors, the citation must be written as Tsay et al. (2000).

COPYRIGHT

Authors who publish their articles in the ChJS automatically transfer their copyright to the Chilean Statistical Society. This enables full copyright protection and wide dissemination of the articles and the journal in any format. The ChJS grants permission to use figures, tables and brief extracts from its collection of articles in scientific and educational works, in which case the source that provides these issues (Chilean Journal of Statistics) must be clearly acknowledged.

AIMS

The Chilean Journal of Statistics (ChJS) is an official publication of the Chilean Statistical Society (www.soche.cl). The ChJS takes the place of *Revista de la Sociedad Chilena de Estadística*, which was published from 1984 to 2000.

The ChJS covers a broad range of topics in statistics, including research, survey and teaching articles, reviews, and material for statistical discussion. In particular, the ChJS considers timely articles organized into the following sections: Theory and methods, computation, simulation, applications and case studies, education and teaching, development, evaluation, review, and validation of statistical software and algorithms, review articles, letters to the editor.

The ChJS editorial board plans to publish one volume per year, with two issues in each volume. On some occasions, certain events or topics may be published in one or more special issues prepared by a guest editor.

EDITOR-IN-CHIEF

Víctor Leiva *Pontificia Universidad Católica de Valparaíso, Chile.*

EDITORS

Héctor Allende Cid	<i>Pontificia Universidad Católica de Valparaíso, Chile</i>
José M. Angulo	<i>Universidad de Granada, Spain</i>
Roberto G. Aykkroyd	<i>University of Leeds, UK</i>
Narayanaswamy Balakrishnan	<i>McMaster University, Canada</i>
Carmen Batanero	<i>Universidad de Granada, Spain</i>
Ionut Bebu	<i>The George Washington University, US</i>
Marcelo Bourguignon	<i>Universidade Federal do Rio Grande do Norte, Brazil</i>
Márcia Branco	<i>Universidade de São Paulo, Brazil</i>
Oscar Bustos	<i>Universidad Nacional de Córdoba, Argentina</i>
Luis M. Castro	<i>Pontificia Universidad Católica de Chile</i>
George Christakos	<i>San Diego State University, US</i>
Enrico Colosimo	<i>Universidade Federal de Minas Gerais, Brazil</i>
Gauss Cordeiro	<i>Universidade Federal de Pernambuco, Brazil</i>
Francisco Cribari-Neto	<i>Universidade Federal de Pernambuco, Brazil</i>
Francisco Cysneiros	<i>Universidade Federal de Pernambuco, Brazil</i>
Mario de Castro	<i>Universidade de São Paulo, São Carlos, Brazil</i>
José A. Díaz-García	<i>Universidad Autónoma Agraria Antonio Narro, Mexico</i>
Raul Fierro	<i>Universidad de Valparaíso, Chile</i>
Jorge Figueroa	<i>Universidad de Concepción, Chile</i>
Isabel Fraga	<i>Universidade de Lisboa, Portugal</i>
Manuel Galea	<i>Pontificia Universidad Católica de Chile</i>
Christian Genest	<i>McGill University, Canada</i>
Marc G. Genton	<i>King Abdullah University of Science and Technology, Saudi Arabia</i>
Patricia Giménez	<i>Universidad Nacional de Mar del Plata, Argentina</i>
Hector Gómez	<i>Universidad de Antofagasta, Chile</i>
Daniel Griffith	<i>University of Texas at Dallas, US</i>
Eduardo Gutiérrez-Peña	<i>Universidad Nacional Autónoma de Mexico</i>
Nikolai Kolev	<i>Universidade de São Paulo, Brazil</i>
Eduardo Lalla	<i>University of Twente, Netherlands</i>
Shuangzhe Liu	<i>University of Canberra, Australia</i>
Jesús López-Fidalgo	<i>Universidad de Navarra, Spain</i>
Liliana López-Kleine	<i>Universidad Nacional de Colombia</i>
Rosangela H. Loschi	<i>Universidade Federal de Minas Gerais, Brazil</i>
Carolina Marchant	<i>Universidad Católica del Maule, Chile</i>
Manuel Mendoza	<i>Instituto Tecnológico Autónomo de Mexico</i>
Orietta Nicolis	<i>Universidad Andrés Bello, Chile</i>
Felipe Osorio	<i>Universidad Técnica Federico Santa María, Chile</i>
Carlos D. Paulino	<i>Instituto Superior Técnico, Portugal</i>
Fernando Quintana	<i>Pontificia Universidad Católica de Chile</i>
Nalini Ravishanker	<i>University of Connecticut, US</i>
Fabrizio Ruggeri	<i>Consiglio Nazionale delle Ricerche, Italy</i>
José M. Sarabia	<i>Universidad de Cantabria, Spain</i>
Helton Saulo	<i>Universidade de Brasília, Brazil</i>
Pranab K. Sen	<i>University of North Carolina at Chapel Hill, US</i>
Julio Singer	<i>Universidade de São Paulo, Brazil</i>
Milan Stehlik	<i>Johannes Kepler University, Austria</i>
M. Dolores Ugarte	<i>Universidad Pública de Navarra, Spain</i>
Andrei Volodin	<i>University of Regina, Canada</i>

MANAGING EDITOR

Marcelo Rodríguez *Universidad Católica del Maule, Chile*

FOUNDING EDITOR

Guido del Pino *Pontificia Universidad Católica de Chile.*

Chilean Journal of Statistics

VOLUME 10, NUMBER 1

APRIL 2019

CONTENTS

Víctor Leiva <i>“Chilean Journal of Statistics” Ten years after its launch: A message from the new Editor-in-Chief</i>	1
Jhonnata B. de Carvalho, Murilo C. Silva, George F. von Borries, André L.S. de Pinho and Ricardo F. von Borries <i>A combined Fourier analysis and support vector machine for EEG classification</i>	3
Luis Benites, Rocío Maehara, Victor H. Lachos and Heleno Bolfarine <i>Linear regression models using finite mixtures of skew heavy-tailed distributions</i>	21
Ednário Mendonça, Michelli Barros and Joelson Campos <i>Goodness-of-fit test for the Birnbaum-Saunders distribution based on the Kullback-Leibler information</i>	41
Nathalia L. Chaves, Caio L. N. Azevedo, Filidor Vilca-Labra and Juvêncio S. Nobre <i>A new Birnbaum-Saunders type distribution based on the skew-normal model under a centered parameterization</i>	55
Gauss M. Cordeiro, M. Mansoor and Serge B. Provost <i>The Harris extended Lindley distribution for modeling hydrological data</i>	77

TENTH VOLUME – FIRST NUMBER
EDITORIAL PAPER

**“Chilean Journal of Statistics”
Ten years after its launch:
A message from the new Editor-in-Chief**

Welcome to the first issue of the tenth volume of the Chilean Journal of Statistics (ChJS). Today, April 29, 2019, the ChJS celebrates ten years of life and begins leaving its childhood, walking quickly to become a teenager. I remember perfectly well when the baby ChJS was born. I was present at that birth and I accompanied the baby during its first three years of life as its Executive Editor. The first volume of the ChJS had two issues, published in April and September 2010, which paid tribute to Dr. Pilar Iglesias, a beloved Chilean statistician. Pilar was the main motivation for the Chilean editorial board to launch this journal, which has as its ancestor the *Revista de la Sociedad Chilena de Estadística*, published in Spanish from 1984 to 2000. I would like to name its dearest uncles who helped the ChJS to survive. Among them are Marcia Branco and Rosangela Loschi from Brazil, Eduardo Gutiérrez-Peña and Manuel Mendoza from Mexico, Marc Genton from Switzerland, as well as Guido del Pino, Manuel Galea, Ronny Vallejos, and Reinaldo Arellano from Chile. I would like to take this opportunity to congratulate Reinaldo, who has honored the Chilean statistical community as the recent winner of the “Mahalanobis Prize 2019” awarded by the International Statistical Institute. Obviously, the ChJS would be nothing without the valuable contributions of renowned international researchers who have honored us by publishing their interesting works in our journal; all of these papers are available for free at <http://chjs.mat.utfsm.cl/issues.html>. We also thank all the anonymous reviewers who have contributed to keeping the top quality standards of the ChJS.

Although the ChJS is published by the Chilean Statistical Society (www.soche.cl) and belongs to the Chilean statistical community, our journal can be recognized as an international publication since its editorial board is composed of colleagues from practically the five continents. Our Editors are from Argentina, Australia, Austria, Bulgaria, Brazil, Canada, Chile, China, Colombia, Greece, India, Italy, Mexico, Netherlands, Peru, Portugal, Romania, Saudi Arabia, Spain, Switzerland, UK, and US. Our current Editorial Board, presented at <http://chjs.mat.utfsm.cl/board.html>, is a mixture of experienced editors and talented young researchers, the latter mainly from Chile and Brazil, who with great interest and enthusiasm have honored us by accepting to be part of the ChJS. They are having their first editorial experiences, although they all have extensive experience as researchers as well as reviewers for prestigious international journals.

I would also like to thank the members of the Directory of the Chilean Statistical Society (<https://soche.cl/quienes-somos>) headed by its President, Dr. Mauricio Castro, for the trust placed in me to be the new Editor-in-Chief of the ChJS. They can rest assured that, just as I did in the past as its Executive Editor, I will make my best effort to bring the ChJS to the highest standards of professionalism, impartiality and quality that all scientific journals must strive for.

In addition to this presentation note, the first issue of the tenth volume of the ChJS comprises five papers. Jhonnata B. de Carvalho, Murilo C. Silva, George F. von Borries, André L.S. de Pinho, and Ricardo F. von Borries, from Brazil and US, combined Fourier analysis and support vector machines to conduct an interesting work for classification of electroencephalograms, a relevant current theme related to data science. Luis Benites, Rocío Maehara, Víctor H. Lachos, and Heleno Bolfarine, from Peru, US and Brazil, proposed a regression model based on a finite mixture of skew heavy-tailed distributions, a widely studied topic by Brazilian and Chilean researchers within the context of statistical modeling. Ednário Mendonça, Michelli Barros, and Joelson Campos, from Brazil, derived goodness-of-fit tests based on the Kullback-Leibler information for the Birnbaum-Saunders model, a distribution which has had some of its more important developments in Chile and Brazil. Nathalia L. Chaves, Caio L.N. Azevedo, Filidor Vilca, and Juvêncio S. Nobre, from Brazil and Chile, introduced a new distribution to describe data with positive support and asymmetry by combining the Birnbaum-Saunders and centered skew-normal models, providing different statistical and mathematical features for this new model. Finally, our fifth paper is presented by Gauss M. Cordeiro, M. Mansoor, and Serge B. Provost, from Brazil, Pakistan and Canada, who derived their work in the setting of distribution theory, an area of wide development around the world, connecting the Harris and Lindley distributions to perform an interesting study which was applied to the modeling of hydrological data.

As a final comment, I would like the Chilean statistical community, as well as the international statistical community, our prestigious Editorial Board and past authors to champion ChJS as an emerging international journal and to encourage others to submit new works to the ChJS. Currently, we are indexed by several international systems, including the Institute for Scientific Information (ISI) Web of Science in the Emerging Sources Citation Index. The ChJS faces important challenges for the near future, such as reaching the Science Citation Index and looking for partnerships with prestigious publishers, societies and associations. However, just as with statistics itself, our success will depend on a team effort. Each one of us is important in meeting these challenges. We need you all.

Víctor Leiva
Editor-in-Chief
Chilean Journal of Statistics
<http://www.victorleiva.cl>

DATA SCIENCE
RESEARCH PAPER

A combined Fourier analysis and support vector machine for EEG classification

JHONNATA B. DE CARVALHO^{1,*}, MURILO C. SILVA², GEORGE F. VON BORRIES²,
ANDRÉ L.S. DE PINHO¹ and RICARDO F. VON BORRIES³

¹Department of Statistics, Universidade Federal do Rio Grande do Norte, Natal, Brazil,

²Department of Statistics, Universidade de Brasília, Brasília, Brazil,

³Department of Electrical and Computer Engineering, University of Texas at El Paso, El Paso, United States of America

(Received: 30 September 2018 · Accepted in final form: 10 March 2019)

Abstract

This paper introduces a method for the classification of electroencephalogram (EEG) data combining Fourier analysis, support vector machine (SVM) and a weighting system, called WFF-SVM, that provides high correct classification rates (accuracy) using a small training data set. Basically, an SVM classifier is calculated for each frequency in the periodogram and a proposed weighting system, based on the error rate of each SVM classifier, is used to obtain a final decision value. Also, it is shown that principal component analysis can be used to identify the best group of EEG channels to apply to the classification method, improving the correct classification rate. Two applications with real data are presented. The first application uses a public data set of epileptic patients and compares the proposed method with other methods presented in the literature. In this case, the correct classification rate obtained was 100%. The second application consists of EEG data collected from a subject submitted to 10 visual stimuli and the correct classification rate obtained was 95.31%. The classifier WFF-SVM combines multiple existing techniques, each one of them widely used in time series and dimensionality reduction problems. Our paper combines standard signal processing techniques to obtain high classification rates of EEG data.

Keywords: Epilepsy data · Periodogram · Principal components analysis · Simple moving averages · Supervised learning.

Mathematics Subject Classification: Primary 62H25 · Secondary 68Q32.

1. INTRODUCTION

Machine learning (ML) techniques have been gaining prominence due to real-world problems as well as large databases. Basically, one can divide ML methods into two classes, supervised learning and unsupervised learning. In unsupervised learning, the method has to recognize the groups by existing standards with a certain criterion. This type of learning tries to gain some understanding of the process that generated the data, e.g., the K-means method applied in DNA gene expression and Internet newsgroups (Ding and He, 2004),

*Corresponding author. Email: jhon_dbz@yahoo.com.br

clustering with hill-climbing optimization method applied to bee species (Friedman and Rubin, 1967), botanical data (Rubin, 1967) and in clustering of plants, wines and heart diseases (Souza et al., 2017). In supervised learning, groups (or classes) are known a priori and it is necessary to provide examples for method training. These methods are often used in classification and regression problems, e.g., logistic regression in the prediction of a financial crisis in Latin American companies (Giampaoli et al., 2016), in the fault diagnosis in chemical processes using Fisher discriminant analysis (Chiang et al., 2000), SVM classification in validation of cancer tissue samples (Furey et al., 2000). However, our interest is in the classification of electroencephalography signals.

An EEG are recordings of the electrical potentials produced by the brain (Bronzino, 1999; Buzsaki, 2006). Basically, the digital EEG is a time series containing information of the electrical activity generated by the brain. EEG has vast application in areas such as epilepsy detection (Andrzejak et al., 2001), emotion regulation using neurofeedback (Ruiz et al., 2014), affective neuroscience (Sitaram et al., 2011), and brain computer interface (Kübler et al., 2001; Wolpaw et al., 2002). For an efficient classification of EEG, an algorithm should address two main problems: feature extraction and classification method. Several methods have been used to extract features of EEG data, such as discrete wavelet transforms (DWT) (Jahankhani et al., 2006; Subasi, 2007; Subasi and Gursoy, 2010), amplitude values (Kaper et al., 2004), clustering techniques (Li and Wen, 2011), autoregressive and adaptive autoregressive parameters (Penny et al., 2000; Pfurtscheller et al., 1998), wavelet packet decomposition and extracted eigenvalues from the resultant wavelet coefficients using principal component analysis (PCA) (Acharya et al., 2012), continuous wavelet transform (CWT), higher order spectra (Acharya et al., 2013), approximate entropy and DWT (Ocak, 2009), analytic time-frequency flexible wavelet transform and fractal dimension (Sharma et al., 2017).

In order to classify a set of extracted features, several pattern recognition methods have been used, such as artificial neural network (Guo et al., 2009; Jahankhani et al., 2006; Nigam and Graupe, 2004; Subasi, 2007), mixture of expert model (Subasi, 2007), linear discriminant analysis (Subasi and Gursoy, 2010), SVM (Chandaka et al., 2009; Subasi and Gursoy, 2010), decision trees (Polat and Günes, 2007), least squares SVM (Li and Wen, 2011; Übeyli, 2010) and hidden markov models (Chiappa and Bengio, 2004). For a more complete review refer to Lotte et al. (2007).

Recently several algorithms have been developed to classify EEG in a variety of applications, such as in Zhang et al. (2016), which proposed a linear Bayesian discriminant with a Laplace prior, named sparse Bayesian method by exploiting a Laplace prior. A major advantage of this method is that it estimates automatically all the parameters of the classifier, without the need to use cross-validation. However, we point out that any Bayesian procedure needs a suitable prior distribution and although the Laplace distribution has been suggested it is conceivable that for a particular application a better prior distribution can be found. Wang et al. (2016) introduces a new approach that utilizes spatiotemporal feature extraction with multivariate linear regression (MLR) to learn discriminative of steady-state visual evoked potentials (SSVEP) features, for improving the detection accuracy. SSVEP are signals that are natural responses to visual stimulation at specific frequencies. MLR is implemented on dimensionality reduced EEG training data and a constructed label matrix to find optimally discriminative subspaces. Jiao et al. (2017) proposed a method that is an extension of multiset canonical correlation analysis (MsetCCA), called multilayer correlation maximization (MCM) model for further improving SSVEP recognition accuracy. MCM combines advantages of both Canonical Correlation Analysis and MsetCCA by carrying out three layers of correlation maximization processes. Zhang et al. (2018) introduced a new method, called multi-kernel extreme learning machine (MKELM) to EEG classification. Basically, this method transforms the EEG through the common

spatial pattern (CSP) and inserts a kernel function in the extreme learning machine (ELM). The MKELM provides a way to circumvent calculation of the hidden layer outputs and inherently encode it in a kernel matrix.

The proposed WFF-SVM is a classifier based on the SVM and the Fourier transform, providing the periodogram as feature extraction. In addition, it uses a weighting system based on the error rate. Thus, we call this classifier weighted Fourier frequencies and SVM, WFF-SVM for short. The WFF-SVM classifier differs from the other methods because it requires just one data transformation (Fourier), which leads to a good capacity to discriminate among groups. The PCA is used to identify the most active regions of the brain, providing the use of fewer electrodes and reducing the complexity of the data, since some electrodes pick up only noises, whereas the other methods ended up losing information by reducing the dimension based on the application of CSP or PCA. In relation to the Fourier transform, we observed that analyzing the signals in the frequency domain (periodogram), as shown in Figure 1, allows us to discriminate the signals for some frequencies. Our classifier takes into account the most distinct frequencies for classification through the weighting system. However, we point out that the choice of the kernel function is not unique, but for our applications the results are virtually the same by considering different kernels, suggesting a robust procedure.

Visual stimuli are commonly used to understand different components, such as color, texture, motion, objects, readability (text versus nontext), and others (Thomas and Vinod, 2017). Moreover, visual stimuli are also used in biometric authentication (Zuquete et al., 2010), emotion classification (Wang et al., 2014), person identification (Das et al., 2009), and others. We tested our classification method using real-world EEG data of two main applications: epilepsy and vision. The first application (described in Subsection 4.1) uses a publicly available data set described in Andrzejak et al. (2001), already used in previous works on EEG classification, and it allows a direct comparison of our classification method to other methods presented in the literature. In this application, the proposed method achieved a correct classification rate of 100.00% under a relatively simple model, showing that the proposed method performs well compared to other methods in the literature. The second application (described in Subsection 4.2) uses a data set collected in an experiment conducted at the University of Texas at El Paso in which the EEG data are acquired while the subject is submitted to visual stimuli. The proposed method showed a high correct classification rate of 95.31% using only three signals from each class in the training phase.

This paper is organized as follows. Section 2 provides a brief review of the SVM classifier relevant for our work and presents the periodogram, which is used for feature extraction. Section 3 presents our classification method integrating Fourier data analysis, SVM and a weighting system. Section 4 reports the performance of our method using real-world data of two applications and compares it with concurrent methods found in the literature. Section 5 provides some discussions, conclusions and recommendations for future work.

2. BACKGROUND

In this section, the methods used in the WFF-SVM classifier are described. The first method is the SVM and it includes three main blocks: the basic classifier, parameters estimation and SVM with nonlinear functions. The other methods are the Fourier analysis, periodogram, and the technique of simple moving averages.

2.1 SUPPORT VECTOR MACHINE

The SVM is a pattern recognition technique that has been widely used in problems like regression and classification (Hastie et al., 2008; Hornik et al., 2006; Theodoridis

and Koutroumbas, 2008; Vapnik, 1996). In classification problems the SVM technique separates two classes (say W_1 and W_{-1}) by a hyperplane $\langle \boldsymbol{\beta}, \mathbf{x} \rangle + \beta_0 = 0$, where $\langle \cdot, \cdot \rangle$ is the inner product, $\mathbf{x}, \boldsymbol{\beta} \in \mathbb{R}^D$ and $\beta_0 \in \mathbb{R}$, corresponding to the decision function

$$f(\mathbf{x}) = \text{sign}(\langle \boldsymbol{\beta}, \mathbf{x} \rangle + \beta_0). \quad (1)$$

The optimal hyperplane is defined as the one maximizing the margin of separation between classes. Note that the optimal hyperplane does not necessarily guarantee a complete separation of points from the two classes. This hyperplane can be constructed using Lagrange multipliers and then solving a constrained convex optimization problem.

Consider a set of training samples \mathbf{x}_i with $i = 1, 2, \dots, N$, then the primal optimization problem along with the soft margin method (Cortes and Vapnik, 1995) is given by

$$\begin{aligned} \min_{\boldsymbol{\beta}, \beta_0, \xi_i} \quad & \frac{1}{2} \|\boldsymbol{\beta}\|^2 + c \sum_{i=1}^N \xi_i, \\ \text{subject to} \quad & \begin{cases} y_i (\langle \boldsymbol{\beta}, \mathbf{x}_i \rangle + \beta_0) \geq 1 - \xi_i, \\ \xi_i \geq 0, \text{ for } i = 1, \dots, N, \end{cases} \end{aligned} \quad (2)$$

where the constant c is previously chosen and determines the influence of the two terms in the minimization problem. The variables ξ_i are known as slack variables measuring the proportional amount of predictions that fall on the wrong side of the margin, and y_i is an indicator variable defined by

$$y_i = \begin{cases} +1, & \text{if } \mathbf{x}_i \in W_1, \\ -1, & \text{if } \mathbf{x}_i \in W_{-1}. \end{cases}$$

Using Lagrange multipliers (Hastie et al., 2008), one can obtain the Wolfe dual function given by

$$L_D = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^N \alpha_i \alpha_k y_i y_k \langle \mathbf{x}_i, \mathbf{x}_k \rangle. \quad (3)$$

The solution is obtained by maximizing L_D , a simple convex optimization problem which must satisfy the conditions $0 \leq \alpha_i \leq c$ and $\sum_{i=1}^N \alpha_i y_i = 0$.

One can also generalize the SVM technique using a non-linear discriminant (unlike the hyperplane). In this case, a mapping is used in a larger number of dimensions. It can be shown (Theodoridis and Koutroumbas, 2008) that this mapping in a larger number of dimensions can be implemented without increasing the computational demand by replacing the inner product $\langle \mathbf{x}_i, \mathbf{x}_k \rangle$ in Equation (3) by a kernel $K(\mathbf{x}_i, \mathbf{x}_k)$ to compute the inner product in a higher dimensional space. In this study, we consider two popularly used kernels:

- Gaussian kernel: $K_1(\mathbf{x}_i, \mathbf{x}_j) = \exp \{-\sigma \|\mathbf{x}_i - \mathbf{x}_j\|^2\}$;
- Polynomial kernel: $K_2(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^d$;

where σ and d are kernel width and polynomial degree, respectively. Note when $d = 1$, the polynomial kernel is called linear kernel.

2.2 FOURIER ANALYSIS

Fourier frequency analysis is a very important tool in signal processing and the periodogram is one of its subproducts (Fuller, 1996). The periodogram shows how the covariance of a time series is distributed in frequency. Any stationary time series can be represented as a sum of sines and cosines (Fuller, 1996), that is, a discrete stationary time series $\{X_t\}$, where $t = 1, \dots, n$, (n being odd) can be represented by

$$X_t = \frac{a_0}{2} + \sum_{j=1}^{\lfloor n/2 \rfloor} a_j \cos(\omega_j t) + b_j \sin(\omega_j t),$$

where $\lfloor n/2 \rfloor$ is the largest integer less than or equal to $n/2$, a_k and b_k are parameters to be estimated. Also, the Fourier frequencies are defined by

$$\omega_k = \frac{2\pi k}{n}, \quad k = 0, \dots, \left\lfloor \frac{n}{2} \right\rfloor.$$

The periodogram can be defined as the sequence $\{J_k\}$, where

$$J_k = \frac{n}{2} (a_k^2 + b_k^2), \quad (4)$$

and the sum of squares removed by $\cos(\omega_k t)$ and $\sin(\omega_k t)$ is

$$J_k = \frac{2}{n} \left[\left(\sum_{t=1}^n X_t \cos(\omega_k t) \right)^2 + \left(\sum_{t=1}^n X_t \sin(\omega_k t) \right)^2 \right].$$

Thus, the value of the periodogram at frequency ω_k is the contribution from this frequency to the sum of squares of $\{X_t\}$ or, equivalently, its energy.

Some periodograms shown in this paper are smoothed using a moving average technique (Brockwell and Davis, 2002). Considering $\{J_k\}$ a sequence of points in the periodogram, for some $\alpha \in \mathbb{N}$, we define the smoothing by

$$J_k^\alpha = \frac{1}{\alpha} \sum_{j=1}^{\alpha} J_{k+j-1}, \quad k = 0, 1, \dots, \left\lfloor \frac{n}{2} \right\rfloor + 1 - \alpha, \quad (5)$$

where J_k^α is the average of α terms in sequence starting at the point J_k , meaning that each point J_k^α is the average contribution of α frequencies for the total energy of the series.

Let $\mathbf{X}_{i,1}$ and $\mathbf{X}_{i,2} \in \mathbb{R}^{P \times C}$ EEG samples of two classes from the i -trial with C and P being the number of channels and samples, respectively. The application of the Fourier transform will be in each column (channel) of $\mathbf{X}_{i,1}$ and $\mathbf{X}_{i,2}$ from the i -trial, building a vector

$$\mathbf{J}_{\ell,k}^\alpha = (J_{\ell,k_{i,g}}^\alpha)^\top, \quad (6)$$

with $i = 1, \dots, N_g$, N_g being the number of trials belonging to class g ($g = 1, 2$) and $\ell = 1, \dots, C$. These vectors together with the vector of labels $\mathbf{y} = (y_1, y_2, \dots, y_{N_1+N_2})^\top$ are the inputs of the classifier WFF-SVM.

3. NEW METHOD FOR EEG CLASSIFICATION

The classification of EEG data is a difficult task, with the analysis disturbed because most of the EEG channels may not be relevant to the classification at hand. Usually, traditional classification techniques alone do not provide good results when applied to EEG data. Therefore, it is important to construct a new method able to distinguish important brain regions and to capture the essential information contained in the data.

3.1 MOTIVATION

The Fourier analysis, especially the periodogram, can reveal hidden patterns in signals. Figure 1 has a set of 4 plots, all of which represent the signals generated by two different stimuli and captured by a channel of the EEG data (red for W_1 class and black for W_{-1} class) for a visual stimuli study (see Section 4.2). The top-left graph represents the superimposed plots of the original EEG signals. Note that it is difficult to visually distinguish two different classes in the time-domain plots presented in this graph. The top-right and bottom-left graphs represent the periodogram and the smoothed periodogram (J_k^4 in Equation (5)) of the signals, respectively. Now, it is easier to notice hidden patterns revealed by the periodograms of the data.

The plots indicate that the periodograms of W_1 have higher values at central frequencies than the periodograms of W_{-1} . In fact, the bottom-right graph in Figure 1 shows a possible discriminant (the dashed line) for these periodograms. Note that the periodograms of W_1 always have values above this hypothetical discrimination line for the central frequencies of the periodogram. However, it should be noted that this type of pattern does not occur for all the channels nor in all regions of the brain. It is necessary to use methods that identify both the relevant channels and the relevant frequencies in a set of periodograms, so that in an application, such as epilepsy detection of signals can be automatically classified into one of the expected classes.

3.2 CALCULATING THE DISCRIMINANT

The graphs in Figure 1 are revealing. It is easy to discriminate the periodograms for certain frequencies, but this separation is not so clear for other frequencies. It is noticeable that each frequency has its own importance and, therefore, could be evaluated individually and not as a whole. Thus, this paper describes a method in which a different discriminant is calculated for each frequency using the SVM classifier.

Considering the set of training $\mathbf{J}_{\ell,k}^\alpha$ of Equation (6) and the label vector \mathbf{y} with C channels, $\ell = 1, 2, \dots, C$ and a set of F frequencies, $k = 0, 1, \dots, F$ (k -th point of the smoothed periodogram and $F = \lfloor n/2 \rfloor$), define $\text{SVM}_{\ell,k}[j_{\ell,k}^\alpha]$ as the discriminant function generated by SVM, given by Equation (1), that classifies a new value $j_{\ell,k}^\alpha$ of the periodogram for a test signal into one of two classes, W_1 or W_{-1} , according to

$$\text{SVM}_{\ell,k}[j_{\ell,k}^\alpha] = \begin{cases} +1, & \text{if } j_{\ell,k}^\alpha \text{ is classified in } W_1, \\ -1, & \text{if } j_{\ell,k}^\alpha \text{ is classified in } W_{-1}. \end{cases} \quad (7)$$

Then, each discriminant will classify a new signal between two classes depending on whether the periodogram has higher or lower value at a particular frequency. Figure 2 shows an example of these discriminants. Note that each discriminant function $\text{SVM}_{\ell,k}[\cdot]$ could present a different decision. Thus, in order to unify these decisions, the next two sections present a weighting system that generates a single answer to the decision problem.

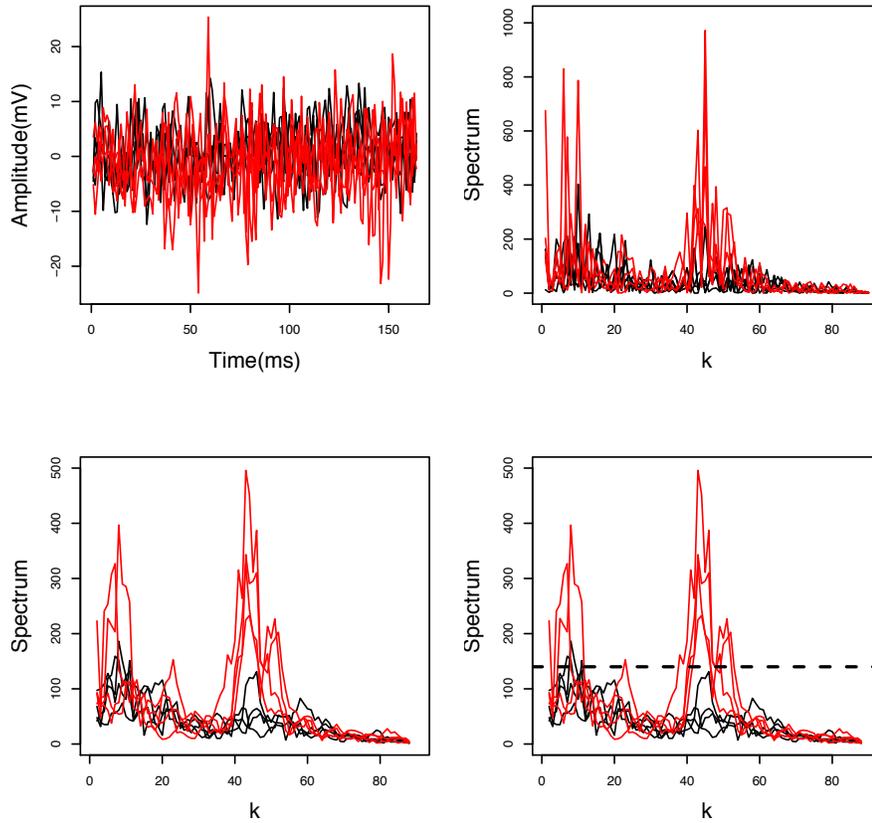


Figure 1. Representations of a set of signals generated by two stimuli. Each line is a signal from the W_1 class (red/lighter lines) or W_{-1} class (black/darker lines). Top-left: original signals. Top-right: periodogram of the signals. Bottom-left: smoothed periodogram of the signals. Bottom-right: smoothed periodogram of the signals with a possible naive discriminant (dashed line). These data are obtained at the Multi-Sensing-Processing and Learning Laboratory (MSPL) at the University of Texas at El Paso (UTEP).

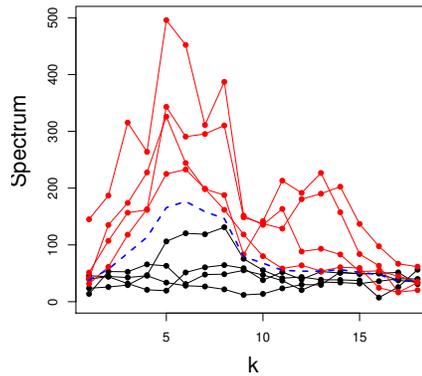


Figure 2. Some discriminating points (dashed line) for some Fourier frequencies ω_k for classes W_1 and W_{-1} . Red (lighter lines) represents class W_1 and black (darker lines) represents class W_{-1} .

3.3 WEIGHTING SYSTEM

Now, we have several discriminant functions, one for each EEG channel and each point in frequency, with discriminant functions producing different decisions. However, it is clear

that there are some discriminants more reliable than others and this reliability is determined by the incorrect classification rate (or error rate) on the training phase of the classification problem. For example, if for some channel ℓ and frequency k the discriminant function $\text{SVM}_{\ell,k}[\cdot]$ provides a low error rate on the training phase, then it is considered more reliable than another discriminant function with a higher error rate. Having this in mind, we introduce a weighting system based on the error rate for each discriminant.

The weight for channel ℓ and frequency k is defined as

$$\Psi_{\ell,k} = [1 - 2 \cdot \min(\text{Error Rate}, 0.5)]^{\hat{\rho}_{\ell,k}}, \quad (8)$$

where $\text{Error Rate} \in [0, 1]$ and $\hat{\rho}_{\ell,k} \geq 1$ is a constant given by

$$\hat{\rho}_{\ell,k} = \frac{\text{SS}_{\text{Total}}}{\text{SS}_{\text{Treatment}}}, \quad (9)$$

where $\text{SS}_{\text{Total}} = \sum_{i=1}^{n_c} \sum_{j=1}^{N_i} (J_{i,j}^\alpha - \bar{J})^2$ and $\text{SS}_{\text{Treatment}} = \sum_{i=1}^{n_c} \sum_{j=1}^{N_i} (\bar{J}_i - \bar{J})^2$, with n_c representing the number of classes (in this case we have $n_c = 2$), N_i is the number of frequencies of the smoothed periodogram of the i -th class, $J_{i,j}^\alpha$ is the j -th smoothed periodogram of the i -th class, \bar{J}_i is the arithmetic mean of the i -th class and \bar{J} is the mean of all smoothed periodograms. The basic concept of our truncated weighting system is to allocate 0 to the ones that have at least a 50% error rate, since $\min\{0, 0.5\} = 0$ implies zero weight. This is so because, based on our experience, it does not make sense to consider classifiers that provide over 50% error rate. On the other hand, the weighting system is an increasing function as the error rate tends to zero, achieving its maximum value when the error rate is zero. Finally, the power $\hat{\rho}_{\ell,k}$ is used to penalize the classifiers that have an error rate between 0 and 50%.

There are several advantages in the use of the exponent $\hat{\rho}_{\ell,k}$ in Equation (9) for the weighting system. It only involves sums, is easy to implement, does not involve optimization, has computational cost almost zero, it uses the data for calculation, it measures the distance between the groups taking into account the variability between and within the groups, and each frequency will have its own weight for SVM.

It is very important to use this kind of information to classify EEG data because much of the data contain non-relevant information of non-activated brain regions such as artifacts in EEG or noise. The next section will show how to use these weights to produce a single decision between one of the two classes W_1 or W_{-1} for new signals.

The implementation of the WFF-SVM method is presented in Algorithm 1. In Figure 3 we display a flowchart of the SVM framework that summarizes all the steps proposed. This classifier is denominated weighted Fourier and support vector machine (WFF-SVM).

Algorithm 1 Training WFF-SVM algorithm.

- 1:** Let $\mathbf{X}_{1,i} \in \mathbb{R}^{P \times C}$ and $\mathbf{X}_{2,i} \in \mathbb{R}^{P \times C}$ denote EEG samples of two classes recorded from the i -th trial. Choose the SVM kernel, the value of c and α smoothing parameter of Equation (5);
 - 2:** Apply the Fourier transform of Equation (4) in each column (channel) of $\mathbf{X}_{i,1}$ and $\mathbf{X}_{i,2}$ from the i -trial and use the moving average technique of Equation (5);
 - 3:** Use the SVM in the smoothed periodograms in step 2, totalizing $C \times F$ models;
 - 4:** Calculate the training error rate to each model in step 3 and the respective weight of Equation (8).
-

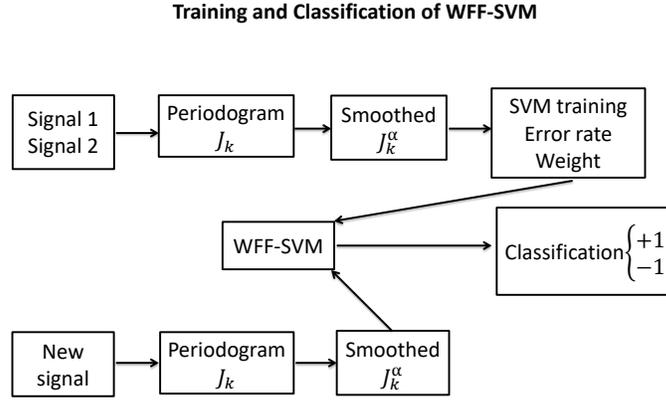


Figure 3. Flowchart for the training and classification phase of a new signal.

3.4 TEST PHASE FOR PRACTICAL APPLICATION

On the test phase for a practical application, we have a new set of signals (one signal per channel) to be classified as class W_1 or W_{-1} . This is done in two different ways (which will be compared later in this paper) using the discriminant function of Equation (7) associated with the weight of Equation (8).

The proposed classification method comprises the following main steps: first, consider a new stimulus $\mathbf{X} \in \mathbb{R}^{P \times C}$ and for each channel $\ell (\ell = 1, \dots, C)$ calculate the periodogram $\{J_{\ell,k}^\alpha\}$. Then, for each channel ℓ and frequency k of the periodogram use the discriminant function $\text{SVM}_{\ell,k}[J_{\ell,k}^\alpha]$ given by Equation (7) to obtain a particular decision (+1 or -1). Finally, using the weights, two decision methods are devised to classify the EEG signals.

In the first decision method, which we label as D_1 , each decision $\text{SVM}_{\ell,k}[J_{\ell,k}^\alpha]$ is weighted by $\Psi_{\ell,k}$ and each channel has its own decision weighting as in

$$D_1 = \text{sign} \left\{ \sum_{\ell=1}^C \text{sign} \left\{ \sum_{k=0}^F \Psi_{\ell,k} \times \text{SVM}_{\ell,k}[J_{\ell,k}^\alpha] \right\} \right\}. \quad (10)$$

In the second decision method, which we label as D_2 , each channel has its own decision weighting $\text{SVM}_{\ell,k}[J_{\ell,k}^\alpha]$ by $\Psi_{\ell,k}$, and the final decision is a pool between channels. Thus, we define

$$D_2 = \text{sign} \left\{ \frac{\sum_{\ell=1}^C \sum_{k=0}^F \text{SVM}_{\ell,k}[J_{\ell,k}^\alpha] \times \Psi_{\ell,k}}{\sum_{k=0}^F \sum_{\ell=1}^C \Psi_{\ell,k}} \right\}. \quad (11)$$

Basically, this decision system takes into account the performance of the channel in the training phase, because if there is a considerable disagreement regarding the classifiers in a given channel, the contribution of this channel to the final classification will not have a

great influence. Then, for both decision methods, we apply the criteria

$$\text{Decision} = \begin{cases} W_1, & \text{if } D_j = +1, \\ W_{-1}, & \text{if } D_j = -1, \\ \text{None}, & \text{if } D_j = 0, \end{cases} \quad (12)$$

for $j = 1, 2$. The implementation of the classification of a new signal is presented in Algorithm 2.

Algorithm 2 Classification of a new signal in WFF-SVM algorithm.

- 1:** Let $\mathbf{X}_{\text{new}} \in \mathbb{R}^{P \times C}$ denote EEG sample of a new recorded;
 - 2:** Apply the Fourier transform of Equation (4) in each column (channel) of \mathbf{X}_{new} and use the moving average technique of Equation (5);
 - 3:** Apply the $C \times F$ SVM models calculated by Algorithm 1 in the smoothed periodograms of step 2, totalizing $C \times F$ of values of Equation (7);
 - 4:** Use the $C \times F$ values calculated in the step 3 and use the decision weighting of Equations (10) or (11).
-

The following sections present two applications with real EEG data. First the proposed method is compared to other methods proposed in the literature, then we use it with a new data set.

4. APPLICATIONS AND RESULTS

This section presents two applications of our classification method. The first application uses a publicly available data set described in Andrzejak et al. (2001) which is used in several papers and is very useful to compare the proposed classification method with other methods. The second application uses a data set collected in an experiment conducted by the MSPL at UTEP. The classifier is implemented in the R software and to have access to the respective code, visit <https://carvalhomyssearches.weebly.com>; see R (2018).

4.1 EPILEPSY DATA CLASSIFICATION

The epilepsy data consists of five distinct sets each containing 100 single-channel EEG segments (Andrzejak et al., 2001). Two of these sets, denoted A and B, are obtained from EEG recordings from five healthy volunteers in an awake state with eyes open and eyes closed, respectively. Sets C, D, and E originated from an EEG archive of pre-surgical diagnosis. Segments in set D are recorded from within the epileptogenic zone, and those in set C from the hippocampal formation of the opposite hemisphere of the brain. While sets C and D contained only activity measured during seizure free intervals, set E only contained seizure activity (for more details about these data sets see Andrzejak et al. (2001)). As in previous studies (Nigam and Graupe, 2004; Subasi, 2007; Subasi and Gursoy, 2010), we used only two datasets (A and E) to test the classifier.

Both sets A and E have 100 signals each, one signal for each channel and each signal corresponding to 4097 samples. To perform the classification it is cut out the beginning and the end of the signals and subsampled them into 20 signals (components) of 200 samples each. Then, for each set A and E, we randomly selected 10 of the corresponding 20 signals to use in the training phase. In the test phase we repeated this same subsampling process to all the signals in both sets A and E. Thus, it is generated 2000 signals to use in the test phase.

Many authors also proposed methods for the classification of EEG data using data sets A and E to test their classifiers. Table 1 has a summary of the overall results and also the result with the application of the proposed method, named WFF-SVM. In WFF-SVM is used the linear kernel, $c = 1$, $\alpha = 5$ and D_2 as described in Equations (2), (5) and (11), respectively.

According to Zhang et al. (2018), the MKELM is more efficient than the following methods: multilayer perceptron with a single hidden layer; the conventional SVM; SVM with Gaussian and polynomial kernel; multi-kernel SVM using both Gaussian and polynomial kernels; the conventional ELM; ELM with Gaussian kernel; ELM with polynomial kernel, and finally, the multi-kernel ELM using both Gaussian and polynomial kernels. Therefore, we also considered in the comparison the new classifier proposed by Zhang et al. (2018), called MKELM, in both applications.

Table 1. Comparison of results for epilepsy data.

Reference	% Accuracy	Method
Subasi (2007)	94.50	ME
	93.20	MLPNN
	98.75	DWT, PCA and SVM
Subasi and Gursoy (2010)	99.50	DWT, ICA and SVM
	100.00	DWT, LDA and SVM
Jahankhani et al. (2006)	98.00	NN
Guo et al. (2009)	95.00	RWE and NN
Nigam and Graupe (2004)	97.20	NN
Polat and Günes (2007)	98.72	TRF
Li and Wen (2011)	99.90	LS-SVM
Chandaka et al. (2009)	95.96	SVM
Übeyli (2010)	99.56	LS-SVM
Zhang et al. (2018)	100.00	MKELM
Proposed method	100.00	WFF-SVM

where ME is mixed of experts; MLPNN is multi-layer perceptron neural network; DWT is discrete wavelet transform; LDA is linear discriminant analysis; ICA is independent component analysis; NN is neural networks; RWE is relative wavelet energy; LS-SVM is least square support vector machine; MKELM is multi-kernel extreme learning machine using both Gaussian and polynomial kernels with CSP feature.

Note that the proposed method is as efficient as (or more efficient than) the other methods. A possible reason for this improvement is the weighting system capturing the most important regions for classification, strengthening the process.

Despite the greater efficiency of the proposed method, it can be noted that all methods are very efficient for this problem. The main reason for this result is that it is relatively easy to classify the epilepsy data; in fact, neurologists can visually distinguish the EEG patterns of epileptic patients and non-epileptics patients. For this reason, the following example presents a more complex application that uses EEG data collected in an experiment based on visual stimuli with a set of tasks to classify.

4.2 CLASSIFICATION OF VISUAL STIMULI

In the visual stimuli application, the objective is to calculate the discriminant function so that, given a new visual stimulus event, our classification method is capable of identifying the slide presented to the subject from the EEG data recordings only. To do this, the proposed method is used after a selection of activated channels using PCA.

Experimental Design The data set used in this application is acquired at the MSPL at UTEP. The EEG data are recorded from a volunteer test subject using a Biosemi EEG acquisition system with 128 channels. The acquisition system recorded EEG signals

corresponding to 10 different visual stimuli, each one presented multiple times in random order and during a regular interval of time. The visual stimuli used correspond to the slides shown in Figure 4. Each stimulus is shown on a computer monitor screen 4 times (in random order) with a five seconds break between each slide, corresponding to a blank screen. An audible tone alerted the subject each time a new slide is about to be displayed. Thus, the EEG data set of the second experiment comprised of 4 EEG signals for each one of the 10 visual stimuli, acquired by 128 channels.

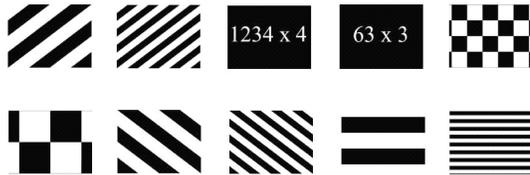


Figure 4. Ten visual stimuli shown to the test subject during EEG signal acquisition.

Using PCA for source localization The PCA is used to explain the variance-covariance structure of a set of variables by a smaller set of variables formed by linear combinations of the original ones (Johnson and Wichern, 2007). Generally, in databases that contain strongly correlated variables (as in EEG data) the PCA is very useful to reduce the dimensionality of the problem. In PCA, the first principal component is the linear combination with the highest possible variance. This means, in the case of EEG data, that the most important channels for the composition of the first principal component are the channels that capture signals with higher variance (the channels corresponding to the activated brain regions) as described in von Borries et al. (2013). Figure 5 shows contours obtained for the first principal component when PCA is applied to EEG signals from 128 channels of the visual stimuli experiment. One can observe that most of the variability in this experiment is present in the channels located on the brain’s frontal lobe. The next sections show that, in fact, this region is the most important for classification and the other regions basically do not bring relevant information to the classification problem at hand. Actually, our results show that the correct classification rate increases when the signals from those regions are not included in classification.

Data analysis First, we train the classifier. Since the proposed method is a binary classifier and we have 10 apparently different visual stimuli, the classification process is implemented sequentially by pairs of visual stimuli. Moreover, as many images are very similar, the classification is performed only with abstract images against images with arithmetic operations, making a total of 16 discriminants (or 16 pairs). Cross-validation is used to approximate the correct classification rate of this method, as follows: for each pair of images analyzed, the first repetition of each image (independent of the others) is excluded in the training phase to be used in the testing phase. Then, the second repetition of each image (independent of the others) is excluded in the training phase to be used in the testing phase, and so on. Thus, $4 \times 16 \times 2 = 128$ signals are used in the test phase. Note that the signals used in the test phase are not used to build the discriminant, resulting in a reliable analysis. The first test is done using the periodogram with the configurations $\alpha = 1$ and 4, linear kernel and using $c = 1$. Note in Table 2, the classification rates for each configuration. There is an increase of around 10% for all configurations when the smoothed periodogram ($\alpha = 4$) is used, indicating that smoothing is a good option to improve the classification rate. Furthermore, D_1 method is better than the D_2 , but not having a very large difference between the rates. Figure 6 shows a contour plot of the accuracy of each brain region. It should be noted that the EEG signals located at the brain’s frontal lobe

had the best correct classification rates. The similarity between Figures 5 and 6 is remarkable, indicating that the regions identified using PCA actually correspond to the regions of higher correct classification rates. Therefore, one might think that the non-activated regions contain non-relevant information that actually disturbs the classification. Thus, the cross-validation process is repeated using 53 channels with the highest hit rates, where most are from the front of the brain, with parameters $c = 1, 10, 100$. The results presented in Table 3 indicate that the correct classification rates increase when using the smoothed periodograms and specially when selecting only the most relevant channels. Therefore, it appears to be extremely important, in a classification analysis of EEG data, to remove from the analysis the channels that appear basically to capture non-relevant information. However, the cost value does not seem to influence much on the results and the classification rates are very similar for all values of c , so, for the analyzes that will be done from now on, will be used $c = 1$.

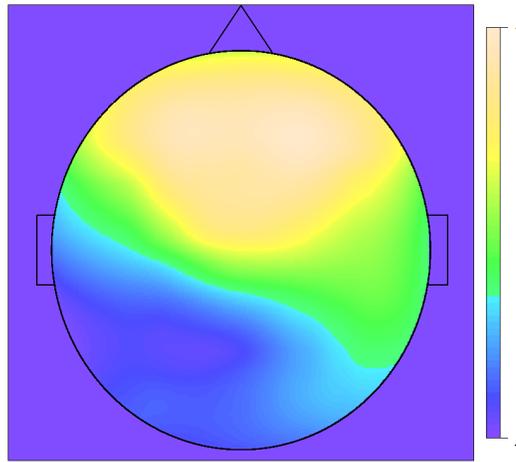


Figure 5. Variability of signals through the Brain. Contours for the first principal component when PCA is applied to EEG signals from 128 channels of the visual stimuli experiment. The front of the brain presents most of the signal variability.

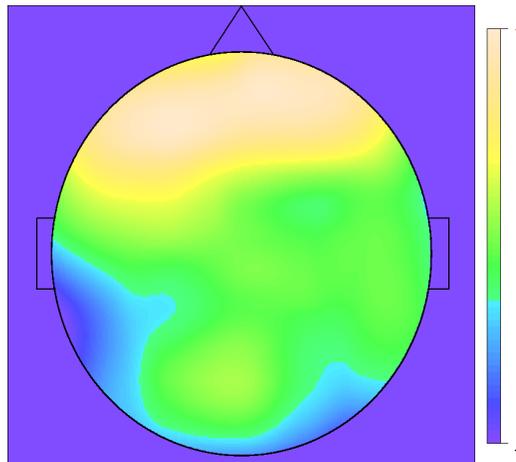


Figure 6. Contour lines for the correct classification rates by channel: new method with the smoothed periodogram, $\alpha = 4$ and $c = 1$.

Table 2. Accuracy for some settings of smoothing parameter in the WFF-SVM algorithm.

Classifier	α	Method	% accuracy
WFF-SVM	1	D_1	75.00
		D_2	73.44
	4	D_1	87.50
		D_2	84.37

Table 3. Results using $\alpha = 4$ for the accuracy using some values of cost (c), number of channels and type of decision.

Cost	Channel	Method	% accuracy
$c = 1$	128	D_1	87.50
		D_2	84.37
	53	D_1	92.97
		D_2	92.97
$c = 10$	128	D_1	85.94
		D_2	85.16
	53	D_1	92.97
		D_2	92.97
$c = 100$	128	D_1	85.95
		D_2	85.16
	53	D_1	92.97
		D_2	92.97

After some α variations, we obtained a classification rate of 95.31% with $c = 1$, $\alpha = 5$, using D_2 with 53 channels, and 73.44% to MKELM using all the channels with CSP feature. These are the best results found in this study. The non-requirement of an extensive training data set constitutes an important characteristic of the proposed classification method since in real-world applications the collection of signals available to train the classifier can be limited to only a few cases.

5. DISCUSSION AND CONCLUSIONS

EEG technique is employed to help in a variety of diagnosis, such as posttraumatic stress, human emotions and epilepsy. Regarding the latter one, there is a special interest to detect as early as possible epilepsy in order to initiate the proper treatment and mitigate this neurological disorder effects. Several studies were conducted with this objective, such as [Fergus et al. \(2015\)](#) who uses machine learning, whereas [Thodoroff et al. \(2016\)](#) and [Acharya et al. \(2018\)](#) have used the deep learning (DL) approach. The DL method has been used in several problems as in image recognition ([Krizhevsky et al., 2012](#)), diagnosis of Alzheimer’s disease ([Ortiz et al., 2016](#)), prediction of sale prices of real estate units ([Rafiei and Adeli, 2015](#)) and in the estimation of concrete compressive strength [Rafiei et al. \(2017\)](#). There are examples in the literature that use SVM and DL, such as in [Tang \(2013\)](#), who developed an approach in DL replacing the softmax layer by a linear SVM. [Erfani et al. \(2016\)](#) used a hybrid model where an unsupervised deep belief networks is trained to extract generic underlying features, and one class SVM is trained from the features learned by the deep belief networks. Therefore, these works show that the use of SVM in DL is not new and suggests that in future works WFF-SVM in DL can also be contemplated in order to search for more efficient methods. The WFF-SVM can be used in any type of signal, EEG, electrocardiogram, electromyogram, etc. In order to accomplish

that, it is sufficient to represent the data as a time series or in a certain proper order. This proposed paper is based on a broader study found in [Carvalho \(2016\)](#), in which electromyogram data were also considered. Furthermore, this classifier can be used in clinical application or any other application. Regarding the computational intensive aspect, with the rapidly increasing performance of new computers, including parallel programming and the promising quantum programming the tendency is to be feasible. The application using epilepsy data showed that the proposed method has no better competitor among other methods presented in the literature. This paper presents a second and more complete application. This application using EEG data captured during an experiment involving visual stimuli showed a number of specific features for the classification of EEG data. In particular, this application showed that the brain region identified using PCA was similar to the region where the channels had the best individual correct classification rates. In fact, the correct classification rate increased significantly by discarding the EEG channels that had non-relevant information. The proposed method of using smoothed periodograms and assigning weights to the channels based on their individual error rates resulted in higher correct classification rates than other methods reported in the literature. It should be noted that the proposed method showed a high correct classification rate of 95.31% using only three signals from each class in the training phase. Thus, a topic for future research is to extend the WFF-SVM to accept more than two groups for training and classification. In addition, it would be important to propose some sort of threshold for decision-making, in guiding the decision Equation (12) on how far it must be from zero to have a more objective classification.

This paper presented a new method for classification of EEG data that uses Fourier analysis and SVM. The proposed method employs a specific SVM decision value for each frequency of the periodogram. In addition, a simple weighting system based on the performance of the classifier, obtained in the training phase, is applied to the classification phase. We used two data sets to test the performance of the proposed classifier. The first data set referred to EEG of an epilepsy study and the second to EEG of a visual stimulation study. Finally, one point for improvement include the extension of our classification method to more than two classes and to expand the performance comparison with other methods.

ACKNOWLEDGMENT

The authors thank and acknowledge financial support from, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Universidade Federal do Rio Grande do Norte (UFRN), Universidade de Brasília (UnB) and University of Texas at El Paso (UTEP). We also thank the Department of Statistics at UFRN for inviting the fourth author to a visit that generate a partnership that led to this research. Furthermore, we thank the anonymous reviewers for their careful reading of our manuscript and their many insightful comments and suggestions.

REFERENCES

- Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., and Adeli, H., 2018. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Computers in Biology and Medicine*, 100, 270-278.
- Acharya, U.R., Sree, S.V., Alvin, A.P.C., and Suri, J.S., 2012. Use of principal component analysis for automatic classification of epileptic EEG activities in wavelet framework. *Expert Systems with Applications*, 39, 9072-9078.
- Acharya, U.R., Yanti, R., Zheng, J.W., Krishnan, M.M.R., Tan, J.H., Martis, R.J., and

- Lim, C.M., 2013. Automated diagnosis of epilepsy using CWT, HOS and texture parameters. *International Journal of Neural Systems*, 23, 1350009.
- Andrzejak, R.G., Lehnertz, K., Mormann, F., Rieke, C., David, P., and Elger, C.E., 2001. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, 64, 061907.
- Brockwell, P.J. and Davis, R.A., 2002. *Introduction to Time Series and Forecasting*. Springer, New York.
- Bronzino, J.D., 1999. *Biomedical engineering handbook*. CRC press, Boca Raton, US.
- Buzsaki, G., 2006. *Rhythms of the Brain*. Oxford University Press, Oxford, UK.
- Carvalho, J.B., 2016. *Classificador Máquina de Suporte Vetorial com Análise de Fourier Aplicada em Dados de EEG e EMG*. Master's thesis. Federal University of Rio Grande do Norte, available at <https://repositorio.ufrn.br/jspui/handle/123456789/20964>.
- Chandaka, S., Chatterjee, A., and Munshi, S., 2009. Cross-correlation aided support vector machine classifier for classification of EEG signals. *Expert Systems with Applications*, 36, 1329-1336.
- Chiang, L.H., Russell, E.L., and Braatz, R.D., 2000. Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 50, 243-252.
- Chiappa, S., and Bengio, S., 2004. HMM and IOHMM modeling of EEG rhythms for asynchronous BCI systems. *IDIAP*.
- Cortes, C., and Vapnik, V., 1995. Support-vector networks. *Machine Learning*, 20, 273-297.
- Das, K., Zhang, S., Giesbrecht, B., and Eckstein, M.P., 2009. Using rapid visually evoked EEG activity for person identification. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 2490-2493). IEEE.
- Ding, C. and He, X., 2004. K-means clustering via principal component analysis. In *Proceedings of the Twenty-first International Conference on Machine Learning* (p. 29). ACM.
- Erfani, S.M., Rajasegarar, S., Karunasekera, S., and Leckie, C., 2016. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. *Pattern Recognition*, 58, 121-134.
- Fergus, P., Hussain, A., Hignett, D., Al-Jumeily, D., Abdel-Aziz, K., and Hamdan, H., 2015. A machine learning system for automated whole-brain seizure detection. *Applied Computing and Informatics*, 12, 70-89.
- Friedman, H.P., and Rubin, J., 1967. On some invariant criteria for grouping data. *Journal of the American Statistical Association*, 62, 1159-1178.
- Fuller, W.A., 1996. *Introduction to Statistical Time Series*. Wiley, New York
- Furey, T.S., Cristianini, N., Duffy, N., Bednarski, D.W., Schummer, M., and Haussler, D., 2000. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 16, 906-914.
- Giampaoli, V., Tamura, K.A., Caro, N.P., and de Araujo, L.J.S., 2016. Prediction of a financial crisis in Latin American companies using the mixed logistic regression model. *Chilean Journal of Statistics*, 7, 31-41.
- Guo, L., Rivero, D., Seoane, J.A., and Pazos, A., 2009. Classification of EEG signals using relative wavelet energy and artificial neural networks. In *Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation* (pp. 177-184). ACM.
- Hastie, T., Tibshirani, R., and Friedman, J., 2008. *The Elements of Statistical Learning*. Springer, New York.
- Hornik, K., Meyer, D., and Karatzoglou, A., 2006. Support vector machines in R. *Journal of statistical software*, 15, 1-28.

- Jahankhani, P., Kodogiannis, V., and Revett, K., 2006. EEG signal classification using wavelet feature extraction and neural networks. In *Modern Computing, 2006. JVA'06. IEEE John Vincent Atanasoff 2006 International Symposium on* (pp. 120-124). IEEE.
- Jiao, Y., Zhang, Y., Wang, Y., Wang, B., Jin, J., and Wang, X., 2018. A novel multi-layer correlation maximization model for improving CCA-based frequency recognition in SSVEP braincomputer interface. *International Journal of Neural Systems*, 28, 1750039.
- Johnson, R.A. and Wichern D.W., 2007. *Applied Multivariate Statistical Analysis*. Prentice-Hall, New Jersey.
- Kaper, M., Meinicke, P., Grossekhoefer, U., Lingner, T., and Ritter, H., 2004. BCI competition 2003-data set IIb: support vector machines for the P300 speller paradigm. *IEEE Transactions on Biomedical Engineering*, 51, 1073-1076.
- Krizhevsky, A., Sutskever, I., and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (pp. 1097-1105).
- Kübler, A., Kotchoubey, B., Kaiser, J., Wolpaw, J.R., and Birbaumer, N., 2001. Brain-computer communication: Unlocking the locked in. *Psychological Bulletin*, 127, 358.
- Li, Y. and Wen, P.P., 2011. Clustering technique-based least square support vector machine for EEG signal classification. *Computer Methods and Programs in Biomedicine*, 104, 358-372.
- Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., and Arnaldi, B., 2007. A review of classification algorithms for EEG-based braincomputer interfaces. *Journal of Neural Engineering*, 4, R1.
- Nigam, V.P., and Graupe, D., 2004. A neural-network-based detection of epilepsy. *Neurological Research*, 26, 55-60.
- Ocak, H., 2009. Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy. *Expert Systems with Applications*, 36, 2027-2036.
- Ortiz, A., Munilla, J., Gorriz, J.M., and Ramirez, J., 2016. Ensembles of deep learning architectures for the early diagnosis of the Alzheimers disease. *International Journal of Neural Systems*, 26, 1650025.
- Penny, W.D., Roberts, S.J., Curran, E.A., and Stokes, M.J., 2000. EEG-based communication: a pattern recognition approach. *IEEE Transactions on Rehabilitation Engineering*, 8, 214-215.
- Pfurtscheller G., Neuper C., Schlogl A., and Lugger K., 1998. Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters. *IEEE Transactions on Rehabilitation Engineering*, 6, 316-325.
- Polat, K. and Günes, S., 2007. Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. *Applied Mathematics and Computation*, 187, 1017-1026.
- R Core Team, 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rafiei, M.H. and Adeli, H., 2015. A novel machine learning model for estimation of sale prices of real estate units. *Journal of Construction Engineering and Management*, 142, 04015066.
- Rafiei, M.H., Khushefati, W.H., Demirboga, R., and Adeli, H., 2017. Supervised deep restricted Boltzmann machine for estimation of concrete. *ACI Materials Journal*, 114, 237-244.
- Rubin, J., 1967. Optimal classification into groups: an approach for solving the taxonomy problem. *Journal of Theoretical Biology*, 15, 103-144.
- Ruiz, S., Buyukturkoglu, K., Rana, M., Birbaumer, N., and Sitaram, R., 2014. Real-time fMRI brain computer interfaces: self-regulation of single brain regions to networks. *Biological Psychology*, 95, 4-20.

- Sharma, M., Pachori, R.B., and Acharya, U.R., 2017. A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension. *Pattern Recognition Letters*, 94, 172-179.
- Sitaram, R., Lee, S., Ruiz, S., Rana, M., Veit, R., and Birbaumer, N., 2011. Real-time support vector classification and feedback of multiple emotional brain states. *Neuroimage*, 56, 753-765.
- Souza, M.P.S., Filho, T.M.S., Amaral, G.J.A., and Souza, R.M.C.R., 2019. Investigating different fitness criteria for swarm-based clustering. *International Journal of Business Intelligence and Data Mining*, pages in press, available at <http://dx.doi.org/10.1504/IJBIDM.2017.10007517>.
- Subasi, A., 2007. EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications*, 32, 1084-1093.
- Subasi, A. and Gursoy, M.I., 2010. EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Systems with Applications*, 37, 8659-8666.
- Tang, Y., 2013. Deep learning using linear support vector machines. Working paper, available at <https://arxiv.org/abs/1306.0239>.
- Theodoridis, S. and Koutroumbas, K., 2008. *Pattern Recognition*. Academic Press, New York.
- Thodoroff, P., Pineau, J., and Lim, A., 2016. Learning robust features using deep learning for automatic seizure detection. In *Machine Learning for Healthcare Conference* (pp. 178-190).
- Thomas, K.P. and Vinod, A.P., 2017. Toward EEG-based biometric systems: the great potential of brain-wave-based biometrics. *IEEE Systems, Man, and Cybernetics Magazine*, 3, 6-15.
- Übeyli, E.D., 2010. Least squares support vector machine employing model-based methods coefficients for analysis of EEG signals. *Expert Systems with Applications*, 37, 233-239.
- Vapnik, V., 1996. *Statistical Learning Theory*. Wiley, New York.
- von Borries, G.F., Rocha, L.C.S., Coutinho, M.S., and von Borries, R.F., 2013. Identification of brain-generated activity with EEG using visual stimuli. *Revista Brasileira de Biometria*, 31, 28-42.
- Wang, H., Zhang, Y., Waytowich, N.R., Krusienski, D.J., Zhou, G., Jin, J., Wang, X., and Cichocki, A., 2016. Discriminative feature extraction via multivariate linear regression for SSVEP-based BCI. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24, 532-541.
- Wang, X.W., Nie, D., and Lu, B.L., 2014. Emotional state classification from EEG data using machine learning approach. *Neurocomputing*, 129, 94-106.
- Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., and Vaughan, T.M., 2002. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113, 767-791.
- Zhang, Y., Wang, Y., Zhou, G., Jin, J., Wang, B., Wang, X., and Cichocki, A., 2018. Multi-kernel extreme learning machine for EEG classification in brain-computer interfaces. *Expert Systems with Applications*, 96, 302-310.
- Zhang, Y., Zhou, G., Jin, J., Zhao, Q., Wang, X., and Cichocki, A., 2016. Sparse Bayesian classification of EEG for brain-computer interface. *IEEE Transactions on Neural Networks and Learning Systems*, 27, 2256-2267.
- Zuquete, A., Quintela, B., and Cunha, J.P.S., 2010. Biometric authentication using electroencephalograms: a practical study using visual evoked potentials. *Electrónica e Telecomunicações*, 5, 185-194.

STATISTICAL MODELING
RESEARCH PAPER

Linear regression models using finite mixtures of skew heavy-tailed distributions

LUIS BENITES^{1,*}, ROCÍO MAEHARA², VÍCTOR H. LACHOS³, and HELENO BOLFARINE⁴

¹Departamento de Ciencias, Pontificia Universidad Católica del Perú, Perú,

²Departamento de Ingeniería, Universidad del Pacífico, Lima, Perú,

³Department of Statistics, University of Connecticut, Storrs, Connecticut, USA,

⁴Departamento de Estatística, Universidade de São Paulo, Brazil

(Received: 15 March 2019 · Accepted in final form: 15 April 2019)

Abstract

In this paper, we propose a regression model based on the assumption that the error term follows a mixture of normal distributions. Specifically, we consider a finite scale mixture of skew-normal distributions, a rich family that contains the skew-normal, skew-t, skew-slash and skew-contaminated normal distributions as members. This model allows us to describe data with high flexibility, simultaneously accommodating multimodality, skewness and heavy tails. We develop a simple EM-type algorithm to perform maximum likelihood inference of the parameters of the proposed model with closed-form expressions for both E- and M-steps. Furthermore, the observed information matrix is derived analytically to account for the corresponding standard errors and a bootstrap procedure is implemented to test the number of components in the mixture. The practical utility of the new model is illustrated with a real dataset and several simulation studies. The proposed algorithm and methods are implemented in an R package named `FMsmnReg`.

Keywords: ECME algorithm · Mixture model · Non-normal error distribution · Scale mixtures of skew-normal distributions

Mathematics Subject Classification: Primary 62J05 · Secondary 62J99

1. BIBLIOGRAPHICAL REVIEW AND MOTIVATING EXAMPLE

1.1 INTRODUCTION

A basic assumption of the linear regression (LR) model is that the error term follows a normal distribution. However, it is well known that data from some phenomena do not always satisfy this assumption, instead having a distribution with heavy tails, skewness or multimodality. Many extensions of this classic model have been proposed to broaden the applicability of Gaussian linear regression (N-LR) analysis to situations where the Gaussian error term assumption may be inadequate, such as, the use of the Student-t distribution (Lange et al., 1989), which is appropriate for datasets involving errors with

*Corresponding author. Email: lbenitess@pucp.edu.pe

longer than normal tails. Other extensions include the use of the symmetrical class of scale mixtures of normal (SMN) distributions (Andrews and Mallows, 1974; Lange and Sinsheimer, 1993), as discussed in Galea et al. (1997), the asymmetrical class of skew-normal (SMSN) distributions proposed by Branco and Dey (2001) or the unified skew-elliptical distributions proposed by Arellano and Genton (2010). However, in practice when nothing is known about the true distribution of the error terms, a risk exists that linear regression analysis based on any of the above models will be performed using an incorrectly specified model. There can also be situations where a single parametric family is unable to provide a satisfactory model for local variations in the observed data.

To overcome these problems, solutions that use finite mixture (FM-LR) models have been recently proposed. For instance, Bartolucci and Scaccia (2005), Soffritti and Galimberti (2011) and Galimberti and Soffritti (2014) developed methods for linear regression analysis by assuming a finite mixture of Gaussian (FM-N-LR) and Student-t (FM-T-LR) components for the error terms.

The classic approach to finite mixture modeling has several challenging aspects. There are nontrivial issues, like non-identifiability and an unbounded likelihood. In this context, Holzmann and Munk (2006) established the identifiability of finite mixtures of elliptical distributions under conditions of the characteristic or probability density function (PDF) generators. More recently, Otianiano et al. (2015) established the identifiability of finite mixture of skew-normal (Azzalini, 1985) and skew-t (Azzalini and Genton, 2008) distributions.

The class of SMSN distributions, proposed by Branco and Dey (2001), is attractive since it simultaneously models skewness with heavy tails (Prates et al., 2012) and contains as proper elements distributions such as the skew-normal, skew-t, skew-slash, skew-contaminated normal and all the symmetric class of scale mixtures of normal (SMN) distributions defined by Andrews and Mallows (1974). Besides this, it has a stochastic representation for easy implementation of the Expectation-Maximization (EM) algorithm (Dempster et al., 1977) and it also facilitates the study of many useful properties. Thus, this extension results in a flexible class of models for robust estimation and inference in FM-LR models.

The objective of this paper is to propose a mixture regression model (and associated likelihood inference) based on the mixtures of the class of scale mixtures of skew-normal (SMSN) distributions, by extending the mixture model based on symmetrical distributions. An advantage of this model is the possibility of fitting multimodality, heavy tails and skewness simultaneously. We derive a mixture model for the random errors based on the class of SMSN distributions (FM-SMSN-LR model) and evaluate the performance of the FM-SMSN-LR model by simulations. In order to motivate our research, we describe the following example with a dataset from the Australian Institute of Sport data (AIS).

1.2 MOTIVATING EXAMPLE

Before discussing the goal of this work, we present a motivating example. More specifically, we extend the linear regression model proposed by Bartolucci and Scaccia (2005), which is defined as

$$Y_i = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad f(\varepsilon_i) = \sum_{j=1}^g p_j \phi(\varepsilon_i | \mu_j, \sigma_j^2), \quad i = 1, \dots, n,$$

where Y_i is the response of case i , $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ is a vector of explanatory variable values, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ is a vector of unknown linear parameters, p_j are positive weights summing to 1, the μ_j terms satisfy the constraint $\sum_{j=1}^g p_j \mu_j = 0$, $\phi(\cdot; \mu_j, \sigma_j^2)$ denotes

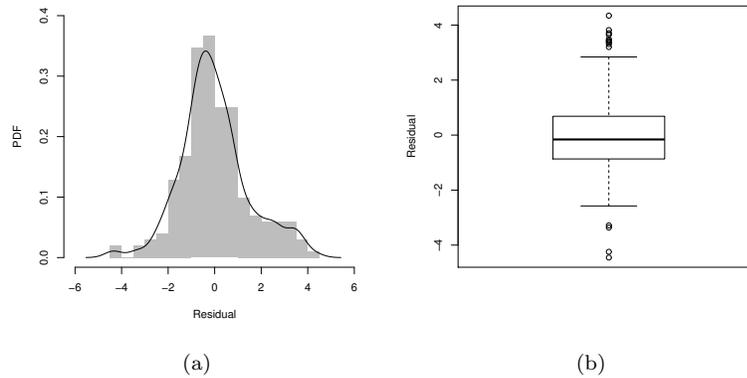


Figure 1. Histogram with a kernel PDF estimate superimposed (a) and the boxplot of ordinary residuals (b) with AIS data.

the PDF of the normal distribution, by assuming that the distribution of the error terms follows a finite mixture of SMSN distribution, so that the FM-SMSN-LR is defined. It is important to stress that our proposal is different from that of Zeller et al. (2016), where the linear regression is modeled with different regression functions, the so-called mixture of regressions or switching regression (Quandt and Ramsey, 1978). An important question that is addressed in this paper is whether a mixture model ($g \geq 2$) is needed instead of a one-component model. Thus, we use the parametric bootstrap log-likelihood ratio statistic, which was proposed by Turner (2000).

To test our proposed model, we use the AIS data available in an R package named `FMsmnReg`. Figure 1 (panels a and b) displays the histogram with a kernel PDF estimate superimposed and the boxplot of ordinary residuals, respectively, obtained by fitting a N-LR model to the AIS data. The plots reveal the existence of multimodal residuals, with evident presence of outliers. In summary, it is necessary to consider a more robust structure in the error. Therefore, this example serves as a motivation for the FM-SMSN-LR model.

1.3 ORGANIZATION OF THE PAPER

The remainder of the paper is organized as follows. In Section 2, we briefly discuss some properties of the univariate SMSN family. In Section 3, we present the FM-SMSN-LR model, including the EM-type algorithm for maximum likelihood (ML) estimation, and derive the empirical information matrix analytically to obtain the standard errors. In Section 4, numerical samples using both simulated and real data are given to illustrate the performance of the proposed model. Finally, some concluding remarks are presented in Section 5.

2. BACKGROUND

2.1 SCALE MIXTURES OF SKEW-NORMAL DISTRIBUTIONS

Next, we start by defining the skew-normal (SN) distribution and then we introduce some useful properties. As defined by Azzalini (1985), a random variable Z has a skew-normal distribution with location parameter μ , scale parameter σ^2 and skewness parameter $\lambda \in \mathbb{R}$, denoted by $Z \sim \text{SN}(\mu, \sigma^2, \lambda)$, if its PDF is given by

$$\phi_{\text{SN}}(z|\mu, \sigma^2, \lambda) = 2\phi(z|\mu, \sigma^2)\Phi(\lambda(z - \mu)/\sigma).$$

The relation between the SMSN class and the SN distribution is provided in the next definition.

DEFINITION 2.1 A random variable Y has an SMSN distribution with location parameter μ , scale parameter σ^2 and skewness parameter λ , denoted by $\text{SMSN}(\mu, \sigma^2, \lambda; H)$, if it has the stochastic representation

$$Y = \mu + \kappa^{1/2}(U)Z, \quad U \perp Z,$$

where $Z \sim \text{SN}(0, \sigma^2, \lambda)$, U is a positive random variable with cumulative distribution function $H(\cdot | \boldsymbol{\nu})$ indexed by a scalar or vector parameter $\boldsymbol{\nu}$ and $\kappa(u)$ is a positive function of u .

The mean and variance of Y are given respectively by

$$\text{E}[Y] = \mu + \sqrt{\frac{2}{\pi}} K_1 \Delta, \quad \text{Var}[Y] = \sigma^2 \left(K_2 - \frac{2}{\pi} K_1^2 \delta^2 \right), \quad (1)$$

where $\Delta = \sigma\delta$, with $\delta = \lambda/\sqrt{1 + \lambda^2}$ and $K_r = \text{E}[U^{-r/2}]$, $r = 1, 2, \dots$. Although we can deal with any $\kappa(\cdot)$ function, in this paper we restrict our attention to the case where $\kappa(u) = 1/u$, since it leads to good mathematical properties. Given $U = u$, we have that $Y|U = u \sim \text{SN}(\mu, u^{-1}\sigma^2, \lambda)$. Thus, the PDF of Y is expressed as

$$f(y) = \phi_{\text{SMSN}}(y|\mu, \sigma^2, \lambda, \boldsymbol{\nu}) = 2 \int_0^\infty \phi(y|\mu, u^{-1}\sigma^2) \Phi\left(u^{1/2}\lambda(y - \mu)/\sigma\right) dH(u|\boldsymbol{\nu}). \quad (2)$$

When H is degenerate, with $u = 1$, we obtain the $\text{SN}(\mu, \sigma^2, \lambda)$ distribution, and when $\lambda = 0$, the SMSN distribution reduces to the class of scale-mixtures of the normal (SMN) distribution represented by the PDF $f_0(y) = \phi_{\text{SMN}}(y|\mu, \sigma^2, \boldsymbol{\nu}) = \int_0^\infty \phi(y|\mu, u^{-1}\sigma^2) dH(u|\boldsymbol{\nu})$.

2.2 SPECIAL CASES OF THE SMSN DISTRIBUTIONS

Some special families of SMSN distributions are the following:

- The skew-t distribution with ν degrees of freedom. In this case, the PDF of Y takes the form

$$\phi_{\text{T}}(y|\mu, \sigma^2, \lambda, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu\sigma}} \left(1 + \frac{d}{\nu}\right)^{-\frac{\nu+1}{2}} T\left(\sqrt{\frac{\nu+1}{d+\nu}} A|\nu+1\right), \quad y \in \mathbb{R},$$

where $d = (y - \mu)^2/\sigma^2$, $A = \lambda(y - \mu)/\sigma$ and $T(\cdot|\nu)$ denotes the distribution function of the standard Student-t distribution, with location zero, scale one and ν degrees of freedom, namely $t(0, 1, \nu)$. We use the notation $Y \sim \text{ST}(\mu, \sigma^2, \lambda, \nu)$.

- The skew-slash distribution. It is denoted by $Y \sim \text{SSL}(\mu, \sigma^2, \lambda, \nu)$ and the associated PDF is given by

$$\phi_{\text{SL}}(y|\mu, \sigma^2, \lambda, \nu) = 2\nu \int_0^1 u^{\nu-1} \phi(y|\mu, u^{-1}\sigma^2) \Phi(u^{1/2}A) du, \quad y \in \mathbb{R}.$$

The skew-slash is a heavy-tailed distribution having as limiting distribution the skew-normal one (when $\nu \rightarrow \infty$).

- The skew contaminated normal distribution. We denote it by $Y \sim \text{SCN}(\mu, \sigma^2, \lambda, \nu, \gamma)$. Its PDF is given by

$$\phi_{\text{SCN}}(y|\mu, \sigma^2, \lambda, \nu) = 2\{\nu\phi(y|\mu, \gamma^{-1}\sigma^2)\Phi(\gamma^{1/2}A) + (1-\nu)\phi(y|\mu, \sigma^2)\Phi(A)\}, \quad \nu, \gamma \in (0, 1].$$

The parameters ν and γ can be interpreted as the proportion of outliers and a scale factor, respectively. The skew contaminated normal distribution reduces to the skew-normal distribution when $\gamma = 1$.

2.3 COMPUTATIONAL FRAMEWORK

The R software (R Core Team, 2016) produces statistical analyses, with its open source codes. This non-commercial computational program may be downloaded from <http://www.r-project.org>. Our method was implemented in R and its codes are available through the `FMsmnReg` package (Benites et al., 2016). We use the `mixmsmn` package, which allows the simulation of mixture the class of scale mixture of skew-normal distributions, see Prates et al. (2013). This computational framework is useful for conducting the simulation studies and the empirical illustration carried out in Section 4.

3. THE LINEAR REGRESSION MODEL WITH FM-SMSN ERRORS

3.1 GENERAL CONTEXT

Next, we introduce the linear regression model using finite mixture of skew heavy tailed distributions where the distribution of the error terms follows a finite mixture of scale mixture of skew-normal distributions (FM-SMSN-LR), following a similar setup as that developed by Bartolucci and Scaccia (2005). Consider the linear regression model expressed as

$$Y_i = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, n, \quad (3)$$

where Y_i is the response of case i , $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ is a vector of explanatory variables of dimension $(p+1) \times 1$, and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ is the regression parameter vector. Furthermore, we assume that

$$f(\varepsilon_i) = \sum_{j=1}^g p_j \phi_{\text{SMSN}}(\varepsilon_i | \mu_j + b\Delta_j, \sigma_j^2, \lambda_j, \boldsymbol{\nu}_j), \quad i = 1, \dots, n, \quad (4)$$

where p_j are positive weights summing to 1, the μ_j s satisfy the identifiability constraint $\sum_{j=1}^g p_j \mu_j = 0$, $b = -\sqrt{2/\pi}K_1$, $K_1 = \text{E}[U^{-1/2}]$, $\Delta_j = \sigma_j \delta_j$ with $\delta_j = \lambda_j / \sqrt{1 + \lambda_j^2}$. Then from Equation (1), we have that $\text{E}(\varepsilon_i) = 0$. Thus, for linearity of SMSN distributions, the PDF of Y_i is expressed as

$$f(y_i | \boldsymbol{\theta}) = \sum_{j=1}^g p_j \phi_{\text{SMSN}}(y_i | \mu_{ij} + b\Delta_j, \sigma_j^2, \lambda_j, \boldsymbol{\nu}_j), \quad \mu_{ij} = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + \mu_j = \vartheta_j + \mathbf{x}_i^\top \boldsymbol{\beta}, \quad (5)$$

where $\mu_{ij} = \mathbf{x}_i^\top \boldsymbol{\beta} + \vartheta_j$, $\vartheta_j = \beta_0 + \mu_j$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, (p_1, \dots, p_{g-1})^\top, \vartheta_1, \dots, \vartheta_g, \sigma_1^2, \dots, \sigma_g^2, \lambda_1, \dots, \lambda_g, \nu_1, \dots, \nu_g)^\top$ is the vector with all parameters. Concerning the parameter $\boldsymbol{\nu}_j$ of the mixing distribution $H(\cdot | \boldsymbol{\nu}_j)$, for $j = 1, \dots, g$, it can be a vector of parameters, e.g.,

the contaminated normal distribution. Thus, for computational convenience we assume that $\boldsymbol{\nu}_1 = \dots = \boldsymbol{\nu}_g = \boldsymbol{\nu}$. This strategy works very well in the empirical studies that we have conducted and greatly simplifies the optimization problem. For $U = 1$, Equations (3) and (4) lead to the FM-N-LR defined by [Bartolucci and Scaccia \(2005\)](#). Moreover, when $g = 1$ and a nonlinear function is used instead of $\mathbf{x}_i^\top \boldsymbol{\beta}$, the FM-SMSN-LR framework reduces to the model discussed by [Garay et al. \(2011\)](#). For each i and j , consider the latent indicator variable Z_{ij} , such that

$$Z_{ij} = \begin{cases} 1, & \text{if the } i\text{th subject is from the } j\text{th component;} \\ 0, & \text{otherwise.} \end{cases}$$

Observe that $Z_{ij} = 1$ if and only if $Z_i = j$. Then

$$P(Z_{ij} = 1) = 1 - P(Z_{ij} = 0) = p_j \quad \text{and} \quad y_i | Z_{ij} = 1 \sim \text{SMSN}(\mu_{ij} + b\Delta_j, \sigma_j^2, \lambda_j; H(\boldsymbol{\nu})). \quad (6)$$

Note that by integrating out $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{ig})^\top$, we obtain the marginal PDF presented in Equation (2) and $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ are independent random vectors, each one having a multinomial distribution with PDF defined as $f(\mathbf{z}_i) = p_1^{z_{i1}} p_2^{z_{i2}} \dots (1 - p_1 - \dots - p_{g-1})^{z_{ig}}$, which we denote by $\mathbf{Z}_i \sim M(1; p_1, \dots, p_g)$. These latent vectors appear in the hierarchical representation given next, which is used to build the Expectation Conditional Maximization Either (ECME) algorithm as proposed by [Liu and Rubin \(1994\)](#), which is a variant of the EM algorithm [Dempster et al. \(1977\)](#). From Equation (6) along with Definition 2.1, the FM-SMSN-LR model can be represented as

$$Y_i | u_i, t_i, Z_{ij} = 1 \stackrel{\text{IND}}{\sim} N(\mu_{ij} + \Delta_j t_i, u_i^{-1} \Gamma_j), \quad (7)$$

$$T_i | u_i, Z_{ij} = 1 \stackrel{\text{IND}}{\sim} \text{TN}(b, u_i^{-1}, (b, \infty)),$$

$$U_i | Z_{ij} = 1 \stackrel{\text{IND}}{\sim} H(u_i; \boldsymbol{\nu}),$$

$$\mathbf{Z}_i \stackrel{\text{IID}}{\sim} M(1; p_1, \dots, p_g), \quad i = 1, \dots, n, \quad j = 1, \dots, g, \quad (8)$$

where IND denotes independent, whereas IID stands for independent and identically distributed, with $\Gamma_j = (1 - \delta_j^2)\sigma_j^2$, $\Delta_j = \sigma_j \delta_j$ and $\delta_j = \lambda_j / \sqrt{1 + \lambda_j^2}$.

3.2 PARAMETER ESTIMATION VIA THE ECME ALGORITHM

Next, we show how to implement the ECME algorithm for ML estimation of the parameters of the FM-SMSN-LR model. By using Equations (7) to (8), we have that the complete-data log-likelihood function is given by

$$\begin{aligned} \ell_c(\boldsymbol{\theta} | \mathbf{y}, \mathbf{t}, \mathbf{u}, \mathbf{z}) = & c + \sum_{i=1}^n \sum_{j=1}^g Z_{ij} \left\{ \log(p_j) - \frac{1}{2} \log(\Gamma_j) - \frac{u_i}{2\Gamma_j} (y_i - \mu_{ij} - \Delta_j t_i)^2 \right. \\ & \left. + \log(h(u_i | \boldsymbol{\nu})) + \log[\phi_{\text{TN}}(t_i | b, u_i^{-1}, (b, \infty))] \right\}, \end{aligned}$$

where c is a constant that is independent of the parameter vector $\boldsymbol{\theta}$. By defining the quantities $\hat{z}_{ij} = E[Z_{ij} | \hat{\boldsymbol{\theta}}, y_i]$, $\hat{s}_{1ij} = E[Z_{ij} U_i | \hat{\boldsymbol{\theta}}, y_i]$, $\hat{s}_{2ij} = E[Z_{ij} U_i T_i | \hat{\boldsymbol{\theta}}, y_i]$ and $\hat{s}_{3ij} =$

$E[Z_{ij}U_iT_i^2|\hat{\boldsymbol{\theta}}, y_i]$, as having known properties of conditional expectation, we obtain

$$\hat{z}_{ij} = \frac{\hat{p}_j \phi_{\text{SMSN}}(y_i|\mu_{ij} + b\Delta_j, \sigma_j^2, \lambda_j, \boldsymbol{\nu})}{\sum_{j=1}^g \hat{p}_j \phi_{\text{SMSN}}(y_i|\mu_{ij} + b\Delta_j, \sigma_j^2, \lambda_j, \boldsymbol{\nu})},$$

$\hat{s}_{1ij} = \hat{z}_{ij}\hat{u}_{ij}$, $\hat{s}_{2ij} = \hat{z}_{ij}(\hat{u}_{ij}\hat{\mu}_{T_{ij}} + \widehat{M}_{T_j}\hat{\tau}_{1_{ij}})$ and $\hat{s}_{3ij} = \hat{z}_{ij}(\hat{u}_{ij}\hat{\mu}_{T_{ij}}^2 + \widehat{M}_{T_j}^2 + \widehat{M}_{T_j}(\hat{\mu}_{T_{ij}} + b)\hat{\tau}_{1_{ij}})$, where

$$\hat{\tau}_{1_{ij}} = E \left[U_i^{1/2} W_{\Phi_1} \left(\frac{U_i^{1/2} \hat{\mu}_{T_{ij}}}{\widehat{M}_{T_j}} \right) \mid \hat{\boldsymbol{\theta}}, y_i, Z_{ij} = 1 \right], \quad i = 1, \dots, n, \quad j = 1, \dots, g,$$

$$\widehat{M}_{T_j}^2 = \frac{\Gamma_j}{\Gamma_j + \Delta_j^2}, \quad \hat{\mu}_{T_{ij}} = b + \frac{\Delta_j}{\Gamma_j + \Delta_j^2} (y_i - \mu_{ij} - \Delta b) \quad \text{and} \quad \hat{u}_{ij} = E[U_j|\hat{\boldsymbol{\theta}}, y_i, Z_{ij} = 1].$$

Once again, at each step the conditional expectations \hat{u}_{ij} and $\hat{\tau}_{1_{ij}}$ can be easily derived from the results given in [Basso et al. \(2010\)](#). Thus, the Q -function is given by

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = c + \sum_{i=1}^n \sum_{j=1}^g \left(\hat{z}_{ij}^{(k)} (\log(p_j) - \frac{1}{2} \log(\Gamma_j) - \frac{1}{2\Gamma_j} (\hat{s}_{1ij}^{(k)} (y_i - \mu_{ij})^2 - 2(y_i - \mu_{ij}) \Delta_j \hat{s}_{2ij}^{(k)} + \Delta_j^2 \hat{s}_{3ij}^{(k)})) + E[Z_{ij} \log(h(U_i|\boldsymbol{\nu}))|\hat{\boldsymbol{\theta}}^{(k)}, y_i] + E[Z_{ij} \log(\phi_{\text{TN}}(T_i|b, u_i^{-1}, (b, \infty)))|\hat{\boldsymbol{\theta}}^{(k)}, y_i] \right).$$

In the CML-step we update the estimate of $\boldsymbol{\nu}$ by direct maximization of the marginal log-likelihood, circumventing the computation of the conditional expectations $\hat{s}_{4ij} = E[Z_{ij} \log(h(U_i|\boldsymbol{\nu}))|\hat{\boldsymbol{\theta}}, y_i]$ and $\hat{s}_{5ij} = E[Z_{ij} \log(\phi_{\text{TN}}(T_i|b, u_i^{-1}, (b, \infty)))|\hat{\boldsymbol{\theta}}^{(k)}, y_i]$. Thus, the ECME algorithm for ML estimation of $\boldsymbol{\theta}$ is defined as follows:

E-step: Given a current estimate $\hat{\boldsymbol{\theta}}^{(k)}$, compute \hat{z}_{ij} , \hat{s}_{1ij} , \hat{s}_{2ij} , \hat{s}_{3ij} , for $i = 1, \dots, n$ and $j = 1, \dots, g$.

CM-steps: Update $\hat{\boldsymbol{\theta}}^{(k)}$ by maximizing $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = E[\ell_c(\boldsymbol{\theta})|\mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)}]$ over $\boldsymbol{\theta}$, which leads to the closed-form expressions given by

$$\begin{aligned} \hat{p}_j^{(k+1)} &= n^{-1} \sum_{i=1}^n \hat{z}_{ij}^{(k)}, \\ \hat{\vartheta}_j^{(k+1)} &= \left(\sum_{i=1}^n (\hat{s}_{1ij}^{(k)} (y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}) - \hat{\Delta}_j^{(k)} \hat{s}_{2ij}^{(k)}) \right) / \sum_{i=1}^n \hat{s}_{1ij}^{(k)}, \\ \hat{\boldsymbol{\beta}}^{(k+1)} &= \left(\sum_{i=1}^n \sum_{j=1}^g \frac{\hat{s}_{1ij}^{(k)} \mathbf{x}_i \mathbf{x}_i^\top}{\hat{\Gamma}_j^{(k)}} \right)^{-1} \sum_{i=1}^n \sum_{j=1}^g \frac{1}{\hat{\Gamma}_j^{(k)}} [\hat{s}_{1ij}^{(k)} (y_i - \hat{\vartheta}_j^{(k+1)}) - \hat{\Delta}_j^{(k)} \hat{s}_{2ij}^{(k)}] \mathbf{x}_i, \\ \hat{\Delta}_j^{(k+1)} &= \left(\sum_{i=1}^n (y_i - \hat{\mu}_{ij}^{(k+1)}) \hat{s}_{2ij}^{(k)} \right) / \sum_{i=1}^n \hat{s}_{3ij}^{(k)} \\ \hat{\Gamma}_j^{(k+1)} &= \sum_{i=1}^n \left(\hat{s}_{1ij}^{(k)} (y_i - \hat{\mu}_{ij}^{(k+1)})^2 - 2(y_i - \hat{\mu}_{ij}^{(k+1)}) \hat{\Delta}_j^{(k+1)} \hat{s}_{2ij}^{(k)} + \hat{\Delta}_j^{2(k+1)} \hat{s}_{3ij}^{(k)} \right) / \sum_{i=1}^n \hat{z}_{ij}^{(k)}. \end{aligned}$$

CML-step: Update $\widehat{\boldsymbol{\nu}}^{(k)}$ by maximizing the current marginal log-likelihood function, obtaining

$$\boldsymbol{\nu}^{(k+1)} = \operatorname{argmax}_{\boldsymbol{\nu}} \sum_{i=1}^n \log \left(\sum_{j=1}^g p_j^{(k+1)} \phi_{\text{SMSN}} \left(y_i | \mu_{ij}^{(k+1)} + b(\boldsymbol{\nu}) \Delta_j^{(k+1)}, \sigma_j^{2(k+1)}, \lambda_j^{(k+1)}, \boldsymbol{\nu} \right) \right).$$

Through constraint $\sum_{j=1}^g p_j \mu_j = 0$ (Bartolucci and Scaccia, 2005), we obtain the estimates of β_0 and μ_j as

$$\widehat{\beta}_0^{(k+1)} = \sum_{j=1}^g \widehat{p}_j^{(k+1)} \widehat{\vartheta}_j^{(k+1)} \quad \text{and} \quad \widehat{\mu}_j^{(k+1)} = \widehat{\vartheta}_j^{(k+1)} - \widehat{\beta}_0^{(k+1)},$$

respectively, for $j = 1, \dots, g$. This process is iterated until a suitable stopping criterion is satisfied. To avoid an indication of lack of progress of the algorithm (McNicholas et al., 2010), we adopted the Aitken acceleration method as the stopping criterion. At iteration k , we first compute the Aitken acceleration factor $c^{(k)} = (\ell^{(k+1)} - \ell^{(k)}) / (\ell^{(k)} - \ell^{(k-1)})$, where following Böhning et al. (1994), the asymptotic estimate of the log-likelihood at iteration $k + 1$ is given by

$$\ell_{\infty}^{(k+1)} = \ell^{(k)} + \frac{1}{1 - c^{(k)}} \left[\ell^{(k+1)} - \ell^{(k)} \right]. \quad (9)$$

As pointed out by Lindsay (1995), the algorithm is considered to reach convergence when $\ell_{\infty}^{(k+1)} - \ell^{(k+1)} < \varepsilon$, where ε is the desired tolerance (we use $\varepsilon = 10^{-6}$). A usual criticism is that EM-type procedures tend to get stuck in local modes. A convenient way to avoid this limitation is to try several EM iterations with a variety of starting values. If there are several modes, one can find the global mode by comparing their relative masses and log-likelihood values. We suggest the following strategy: For β_0 and $\boldsymbol{\beta}$ use the ordinary least-squares (OLS) estimate. Initial values for $p_j, \mu_j, \sigma_j^2, \lambda_j$ and $\boldsymbol{\nu}$, $j = 1, \dots, g$, are obtained by fitting the mixture model given in Equation (3) to the OLS residuals (Bartolucci and Scaccia, 2005), which can be done through the FMsmnReg package (Benites et al., 2016).

3.3 MODEL SELECTION AND APPROXIMATE STANDARD ERRORS

Consider the problem of comparing several FM-SMSN-LR models, with different numbers of component PDFs. Here, we use two model selection criteria, the Akaike information criterion plus a bias correction term (Hurvich and Tsai, 1989), denoted by (AIC_c) , and the adjusted Bayesian information criterion (Schlove, 1987), denoted by (BIC_a) . These criteria are defined as

$$\text{AIC}_c = -2\ell(\widehat{\boldsymbol{\theta}}) + \frac{2n\rho}{n - \rho - 1} \quad \text{and} \quad \text{BIC}_a = -2\ell(\widehat{\boldsymbol{\theta}}) + \rho \log \left(\frac{n+2}{2} \right),$$

where $\ell(\boldsymbol{\theta})$ is the actual log-likelihood, ρ is the number of free parameters that have to be estimated in the model, and n is the sample size.

A simple way of obtaining the standard errors of ML estimators of mixture model parameters is to approximate the asymptotic covariance matrix of $\widehat{\boldsymbol{\theta}}$ by the inverse of the observed information matrix. Let $\mathbf{I}_o(\boldsymbol{\theta}) = -\partial^2 \ell(\boldsymbol{\theta} | \mathbf{y}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$ be the observed information matrix, where $\ell(\boldsymbol{\theta} | \mathbf{y})$ is the observed log-likelihood function, which is obtained using Equation (5). In this work we use the alternative method suggested by Basford et al. (1997),

which consists of approximating the inverse of the covariance matrix by

$$\mathbf{I}_o(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \hat{\mathbf{s}}_i \hat{\mathbf{s}}_i^\top, \quad \text{where} \quad \hat{\mathbf{s}}_i = \left. \frac{\partial}{\partial \boldsymbol{\theta}} \log [f(y_i|\boldsymbol{\theta})] \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad (10)$$

where $\hat{\mathbf{s}}_i = (\hat{s}_{i,\boldsymbol{\beta}}^\top, \hat{s}_{i,p_1}, \dots, \hat{s}_{i,p_{g-1}}, \hat{s}_{i,\vartheta_1}, \dots, \hat{s}_{i,\vartheta_g}, \hat{s}_{i,\sigma_1^2}, \dots, \hat{s}_{i,\sigma_g^2}, \hat{s}_{i,\lambda_1}, \dots, \hat{s}_{i,\lambda_g}, \hat{s}_{i,\boldsymbol{\nu}})^\top$. It is important to stress that the standard error of $\boldsymbol{\nu}$, obtained from $\hat{s}_{i,\boldsymbol{\nu}}$, depends heavily on the calculation of conditional expectation $E[\log(U_i)|y_{\text{obs}_i}, \hat{\boldsymbol{\theta}}]$, which relies on computationally intensive Monte Carlo integrations, since no analytical expression for this expected value exists. Therefore, the expressions for the elements $\hat{s}_{i,\boldsymbol{\beta}}, \hat{s}_{i,p_j}, \hat{s}_{i,\vartheta_j}, \hat{s}_{i,\sigma_j^2}, \hat{s}_{i,\lambda_j}$, for $j = 1, \dots, g$, are given as

$$\begin{aligned} \hat{s}_{i,\boldsymbol{\beta}}^\top &= \frac{\sum_{j=1}^G p_j D_{\boldsymbol{\beta}}(y_i; \boldsymbol{\theta}_j)}{f(y_i; \boldsymbol{\theta})}, \quad \hat{s}_{i,\vartheta_j} = \frac{p_j D_{\vartheta_j}(y_i; \boldsymbol{\theta}_j)}{f(y_i; \boldsymbol{\theta})}, \quad \hat{s}_{i,\sigma_j^2} = \frac{p_j D_{\sigma_j^2}(y_i; \boldsymbol{\theta}_j)}{f(y_i; \boldsymbol{\theta})}, \quad \hat{s}_{i,\lambda_j} = \frac{p_j D_{\lambda_j}(y_i; \boldsymbol{\theta}_j)}{f(y_i; \boldsymbol{\theta})}, \\ \hat{s}_{i,p_j} &= \frac{1}{f(y_i; \boldsymbol{\theta})} [\phi_{\text{SMSN}}(y_i|\mu_{ij} + b\Delta_j, \sigma_j^2, \lambda_j, \boldsymbol{\nu}) - \phi_{\text{SMSN}}(y_i|\mu_{ig} + b\Delta_g, \sigma_g^2, \lambda_g, \boldsymbol{\nu})], \end{aligned}$$

with

$$D_{\vartheta_j}(y_i; \boldsymbol{\theta}_j) = \frac{\partial}{\partial \vartheta_j} \left(\phi_{\text{SMSN}}(y_i|\mu_{ij} + b\Delta_j, \sigma_j^2, \lambda_j, \boldsymbol{\nu}) \right).$$

After some algebraic manipulation, we obtain

$$\begin{aligned} D_{\boldsymbol{\beta}}(y_i; \boldsymbol{\theta}_j) &= \frac{2}{\sqrt{2\pi\sigma_j^2}} \left[\sigma^{-2}(y_i - \mu_{ij} - b\Delta_j) I_{ij}^\Phi(3/2) - \sigma_j^{-1} \lambda_j I_{ij}^\phi(1) \right] \mathbf{x}_i, \\ D_{\vartheta_j}(y_i; \boldsymbol{\theta}_j) &= \frac{2}{\sqrt{2\pi\sigma_j^2}} \left[\sigma_j^{-2}(y_i - \mu_{ij} - b\Delta_j) I_{ij}^\Phi(3/2) - \sigma_j^{-1} \lambda_j I_{ij}^\phi(1) \right], \\ D_{\lambda_j}(y_i; \boldsymbol{\theta}_j) &= \frac{2}{\sqrt{2\pi\sigma_j^2}} \left[\frac{(y_i - \mu_{ij} - b\Delta_j)b}{(1 + \lambda_j^2)^{(3/2)}} I_{ij}^\Phi(3/2) + \left((y_i - \mu_{ij} - b\Delta_j) - \frac{b\Delta_j}{1 + \lambda_j^2} I_{ij}^\phi(1) \right) \right], \\ D_{\sigma_j^2}(y_i; \boldsymbol{\theta}_j) &= \frac{1}{\sqrt{2\pi\sigma_j^2}} \left[-\sigma_j^{-2} I_{ij}^\Phi(1/2) + \sigma_j^{-4}(y_i - \mu_{ij} - b\Delta_j)^2 I_{ij}^\Phi(3/2) \right. \\ &\quad \left. + \sigma_j^{-4}(y_i - \mu_{ij} - b\Delta_j)b\Delta_j I_{ij}^\Phi(3/2) - \lambda_j \sigma_j^{-3}(y_i - \mu_{ij}) I_{ij}^\phi(1) \right] \end{aligned}$$

where the expressions $I_{ij}^\Phi(w)$ and $I_{ij}^\phi(w)$ are given in [Basso et al. \(2010\)](#). The information-based approximation defined in Equation (10) is asymptotically applicable. However, it is less reliable unless the sample size is sufficiently large. Observe that the asymptotic covariance matrix of the ML estimates, that is, the inverse of Equation (10), was obtained using the parametrization $\varphi_j = \beta_0 + \mu_j$, $j = 1, \dots, g$. We can use the traditional delta method (see [Rao, 1973](#), Sec. 6a.2), to obtain standard errors using the original parameterization.

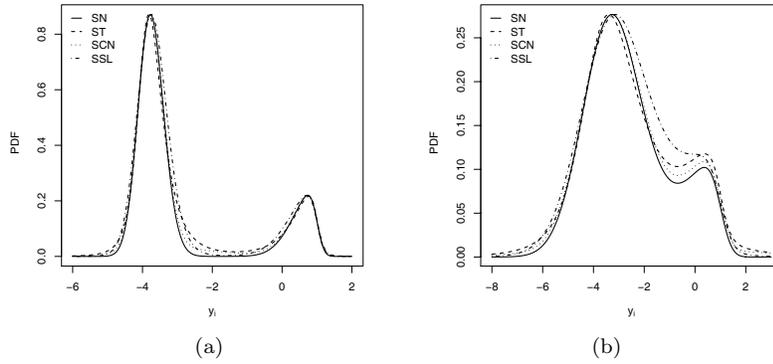


Figure 2. Target mixture PDFs from simulated data in Scenario 1 (a) and Scenario 2 (b).

4. NUMERICAL STUDIES

4.1 PARAMETER RECOVERY (SIMULATION STUDY I)

We conduct three simulation studies to illustrate the performance of our proposed model. The first simulation presented below reports the consistency of the approximate standard errors for the ML estimators of parameters through the EM algorithm with each sample under the stopping criterion in Equation (9), whereas the contents of the second and third simulations are described in the corresponding subsections. In addition, we finish this section of numerical studies with an empirical illustration based on real data.

Here, we consider two scenarios for simulation in order to verify if we can estimate the true parameter values accurately by using the proposed ECME algorithm. This is the first step to ensure that the estimation procedure works satisfactorily. We fit data that were artificially generated from the following model with two components

$$f(y_i|\boldsymbol{\theta}) = \sum_{j=1}^2 p_j \phi_{\text{SMSN}}(y_i|\mu_{ij} + b\Delta_j, \sigma_j^2, \lambda_j, \boldsymbol{\nu}), \quad i = 1, \dots, n,$$

where Z_{ij} is a component indicator of Y_i with $P(Z_{ij} = 1) = p_j$, $j = 1, 2$, $x_i^\top = (x_{i1}, x_{i2})$, such that $x_{i1} \sim U(0, 1)$ and $x_{i2} \sim U(0, 1)$, for $i = 1, \dots, n$, and ε_1 and ε_2 follow a distribution as in the assumption given in Equation (3). We consider the following parameter values: $\beta_0 = -1$, $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top = (-4, -3)^\top$, $\mu_1 = -4$, $\mu_2 = 1$, $\lambda_1 = 1$, $\lambda_2 = -4$ and $p_1 = 0.2$. In addition, we consider the following scenarios (depicted in Figure 2): scenario 1 (well separated components) with $\sigma_1^2 = 0.2$ and $\sigma_2^2 = 0.4$, and scenario 2 (poorly separated components) with $\sigma_1^2 = 2$ and $\sigma_2^2 = 2$. For each combination of parameters, we generated 1000 Monte Carlo samples of size $n = 1000$ from the FM-SMSN-LR models, under four different situations: FM-SN-LR, FM-ST-LR ($\nu = 3$), FM-SSL-LR ($\nu = 3$) and FM-SCN-LR ($\boldsymbol{\nu}^\top = (0.1, 0.1)$). The average values and standard deviations (MC SD) of the estimators across the 1000 Monte Carlo samples were computed, along with the average (IM SE) values of the approximate standard deviations of the estimates obtained through the method described in the Subsection 3.3. Moreover, we compute coverage probability of each parameter (COV), which is defined by $\text{COV}(\hat{\boldsymbol{\theta}}) = (1/m) \sum_{j=1}^m I(\boldsymbol{\theta} \in [\hat{\boldsymbol{\theta}}_L, \hat{\boldsymbol{\theta}}_U])$, where I is the indicator function such that $\boldsymbol{\theta}$ lies in the interval $[\hat{\boldsymbol{\theta}}_L, \hat{\boldsymbol{\theta}}_U]$, with $\hat{\boldsymbol{\theta}}_L$ and $\hat{\boldsymbol{\theta}}_U$ being estimated lower and upper bounds of the 95% CI, respectively. The results are presented in Table 1. Note that under both scenarios (well and poorly separated components), the results suggest that the proposed FM-SMSN-LR model produces satisfactory estimates.

It can be seen from this table that the estimation method of the standard errors provides relatively close results (IM SE and MC SD), indicating that the proposed asymptotic

Table 1. Simulation study I: mean and MC SD are the respective estimated means and standard deviations from fitting a FM-SMSN-LR model based on 1000 samples. IM SE is the average value of the approximate standard error obtained through the information-based method. COV is the coverage probability. True values of parameters are in parentheses.

Parameter		Scenario 1: $\sigma_1^2 = 0.2, \sigma_2^2 = 0.4$				Scenario 2: $\sigma_1^2 = \sigma_2^2 = 2$			
		SN	ST($\nu = 3$)	SCN ($\nu = 0.1$)	SSL($\nu = 3$)	SN	ST($\nu = 3$)	SCN ($\nu = 0.1$)	SSL($\nu = 3$)
$\beta_0(-1)$	Mean	-0.9971	-1.0038	-0.9953	-0.9989	-1.0119	-1.0070	-0.9965	-1.0413
	IM SE	0.0602	0.0859	0.0777	0.0883	0.1928	0.3345	0.2369	0.3238
	MC SD	0.0698	0.0755	0.0713	0.0770	0.0925	0.1214	0.1324	0.1284
	COV	90.6%	96.7%	96.6%	96.0%	99.4%	95.7%	91.8%	95.8%
$\beta_1(-4)$	Mean	-4.0002	-3.9985	-3.9996	-3.9947	-3.9949	-3.9958	-3.9963	-4.0005
	IM SE	0.0368	0.0418	0.0402	0.0423	0.0889	0.1021	0.0974	0.0985
	MC SD	0.0365	0.0426	0.0403	0.0449	0.0899	0.1076	0.0950	0.1031
	COV	94.7%	94.2%	95.5%	95.0%	95.0%	92.9%	95.4%	93.3%
$\beta_2(-3)$	Mean	-3.0012	-2.9998	-3.0014	-2.9938	-2.9994	-2.9989	-2.9967	-3.0013
	IM SE	0.0374	0.0424	0.0410	0.0432	0.0859	0.1005	0.0975	0.1020
	MC SD	0.0370	0.0442	0.0413	0.0430	0.0836	0.1046	0.0977	0.1109
	COV	95.6%	93.7%	94.0%	96.0%	96.2%	94.4%	94.2%	92.0%
$\mu_1(-4)$	Mean	-4.0026	-3.9945	-4.0040	-4.0166	-4.0295	-3.9806	-4.0899	-3.9924
	IM SE	0.0853	0.0800	0.0894	0.0854	0.1396	0.2782	0.1896	0.2531
	MC SD	0.0691	0.0876	0.0744	0.0859	0.1111	0.3161	0.2483	0.2202
	COV	98.2%	99.8%	98.6%	98.6%	97.3%	92.3%	84.5%	94.8%
$\mu_2(1)$	Mean	0.9992	1.0012	1.0007	0.9945	0.9990	1.0103	1.0391	0.9955
	IM SE	0.0837	0.0878	0.0862	0.0873	0.0744	0.1098	0.0861	0.0983
	MC SD	0.0630	0.0625	0.0656	0.0625	0.0692	0.1000	0.1060	0.0813
	COV	98.3%	99.7%	98.4%	99.0%	96.7%	96.7%	86.4%	97.7%
σ_1^2	Mean	0.2097	0.2089	0.2084	0.1946	2.0069	2.2009	1.9385	1.9221
	IM SE	0.0680	0.0575	0.0643	0.0543	1.4238	0.9880	0.7385	1.5234
	MC SD	0.0427	0.0639	0.0644	0.0539	0.5626	1.0118	0.8238	0.9698
	COV	88.7%	89.8%	88.9%	89.0%	99.6%	87.3%	83.3%	89.1%
σ_2^2	Mean	0.3991	0.4026	0.3940	0.3988	2.0452	1.9839	1.8290	2.1521
	IM SE	0.0274	0.0385	0.0343	0.0381	0.1978	0.3796	0.1898	0.2758
	MC SD	0.0283	0.0501	0.0423	0.0463	0.1816	0.2642	0.3309	0.3109
	COV	94.0%	85.9%	85.5%	88.0%	95.9%	93.7%	72.5%	89.2%
$\lambda_1(1)$	Mean	1.0916	1.0534	1.0894	0.9679	1.1614	1.0068	0.6175	0.8514
	IM SE	0.7420	0.4956	0.6466	0.4814	1.4279	1.0923	1.2206	2.7316
	MC SD	0.8216	0.4983	0.6441	0.4385	0.4974	0.7792	1.3124	1.1426
	COV	94.3%	96.3%	95.9%	98.0%	99.6%	96.9%	88.4%	92.4%
$\lambda_2(-4)$	Mean	-4.0874	-4.1108	-4.0739	-4.1418	-4.2153	-4.0168	-3.7773	-4.0682
	IM SE	0.5446	0.5969	0.5971	0.6086	0.6299	0.8950	0.6262	0.6219
	MC SD	0.5406	0.6141	0.6007	0.5477	0.5967	0.6555	0.8671	0.6494
	COV	96.8%	95.5%	94.3%	96.0%	96.8%	94.5%	86.8%	93.6%
$p_1(0.2)$	Mean	0.1998	0.2004	0.1999	0.1985	0.1987	0.2033	0.2028	0.2000
	IM SE	0.0126	0.0131	0.0130	0.0131	0.0146	0.2218	0.0159	0.0204
	MC SD	0.0126	0.0125	0.0129	0.0127	0.0138	0.0235	0.0213	0.0191
	COV	95.3%	95.8%	95.0%	94.0%	96.3%	92.9%	87.3%	94.6%
ν	Mean	-	3.0735	0.1070	2.9791	-	3.2216	0.1342	4.4543
$\gamma(0.1)$	Mean	-	-	0.1098	-	-	-	0.1415	-

approximation for the variances of the ML estimates of Equation (10) is reliable. Note also that the coverage probability (COV) for the regression parameters is quite stable for two scenarios, indicating that the proposed asymptotic approximation for the variance estimates of the ML estimates is reliable.

4.2 ASYMPTOTIC PROPERTIES OF THE EM ESTIMATES (SIMULATION STUDY II)

The main focus in this simulation study is to show the asymptotic properties of the EM estimates. Our strategy is to generate artificial samples from the FM-SMSN-LR model with $x_i^\top = (x_{i1}, x_{i1})$, such that $x_{i1} \sim U(0, 1)$ and $x_{i2} \sim U(0, 1)$, for $i = 1, \dots, n$. We choose sample sizes $n = 100, 250, 500, 1000, 2500$ and 5000 . The true values of the parameters were taken as $\beta_0 = -1, \beta = (\beta_1, \beta_2)^\top = (-4, -3)^\top, \mu_1 = -4, \mu_2 = 1, \sigma_1^2 = 0.2, \sigma_2^2 = 0.4$ and $p_1 = 0.2$. For each combination of parameters and sample sizes, we generated 1000 random samples from the FM-SMSN-LR models, under three different situations: FM-SN-LR, FM-ST-LR ($\nu = 3$), FM-SSL-LR ($\nu = 3$) and FM-SCN-LR ($\nu^\top = (0.1, 0.1)$). In order to analyze asymptotic properties of the EM estimates, we computed the bias and the relative root mean square error (RMSE) for each combination of sample size and

parameter values. For θ_i , they are given by

$$\text{Bias}(\theta_i) = \frac{1}{1000} \sum_{i=1}^{1000} (\theta_i^{(j)} - \theta_i) \text{ and } \text{RMSE}(\theta_i) = \sqrt{\frac{1}{1000} \sum_{i=1}^{1000} (\theta_i^{(j)} - \theta_i)^2},$$

where $\hat{\theta}_i^{(j)}$ is the estimate of θ_i for the j th sample. The results for β_0 , β_1 and β_2 are shown in Figure 3; the results for μ_1 , σ_1 and λ_1 are shown in Figure 4; the results for μ_2 , σ_2 , λ_2 are shown in Figure 5; and the results for p_1 are shown in Figure 6. One can see a pattern of convergence to zero of the bias and RMSE when n increases for all the parameters. As a general rule, we can say that Bias and RMSE tend to approach zero when the sample size increases, indicating that the estimates based on the proposed EM-type algorithm under the FM-SMSN-LR model do provide good asymptotic properties.

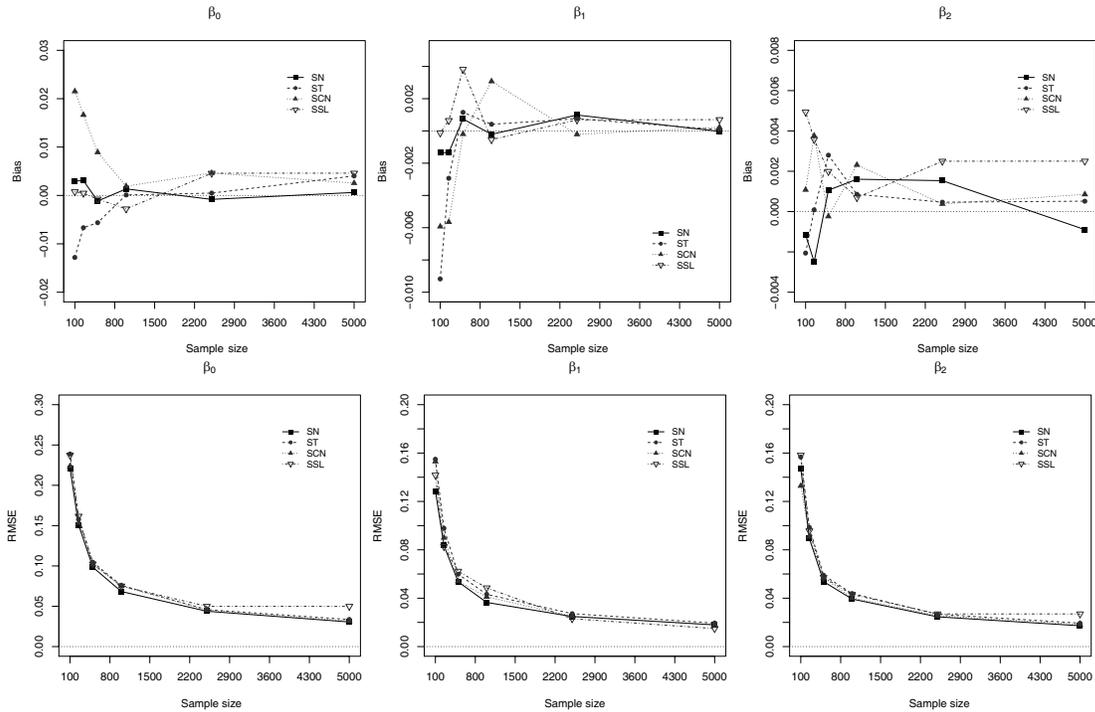


Figure 3. Average bias (1st row) and average RMSE (2nd row) of the estimators of $\beta_0, \beta_1, \beta_2$ for simulation II.

4.3 ROBUSTNESS OF THE EM ESTIMATES (SIMULATION STUDY III)

The purpose of this simulation study is to compare the effect of the robustness of the estimates of the FM-SMSN-LR models in the presence of outliers on the response variable. We compare the FM-SN-LR, FM-ST-LR ($\nu = 3$), FM-SSL-LR ($\nu = 3$) and the FM-CN-LR ($(\nu, \gamma) = (0.1, 0.1)$) models. In this scenario, we generated 500 samples of size $n = 500$ of the FM-SMSN-LR model with $f(\varepsilon_i) = \sum_{j=1}^2 p_j \phi_{\text{SMSN}}(\varepsilon_i | \mu_j + b\Delta_j, \sigma_j^2, \lambda_j, \nu)$. The true values of the parameters were taken as $\beta_0 = -1$, $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top = (-4, -3)^\top$, $\mu_1 = -4$, $\mu_2 = 1$, $\sigma_1^2 = 0.2$, $\sigma_2^2 = 0.4$ and $p_1 = 0.2$. To assess how much the EM estimates are influenced by the presence of outliers, we replaced observation y_{150} by $y_{150}(v) = y_{150} + v$, with $v = 1, 2, \dots, 10$. For each replication, we obtained the parameter estimates with and without outliers, with the three FM-SMSN-LR models.

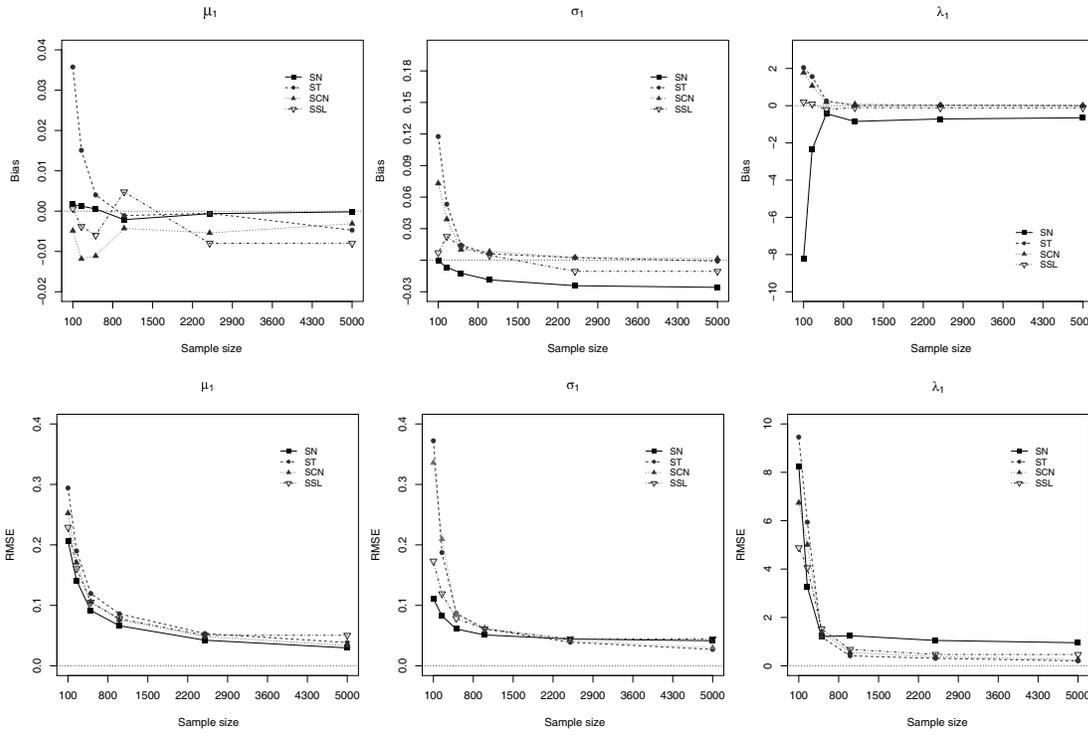


Figure 4. Average bias (1st row) and average RMSE (2nd row) of the estimators of $\mu_1, \sigma_1, \lambda_1$ for simulation II.

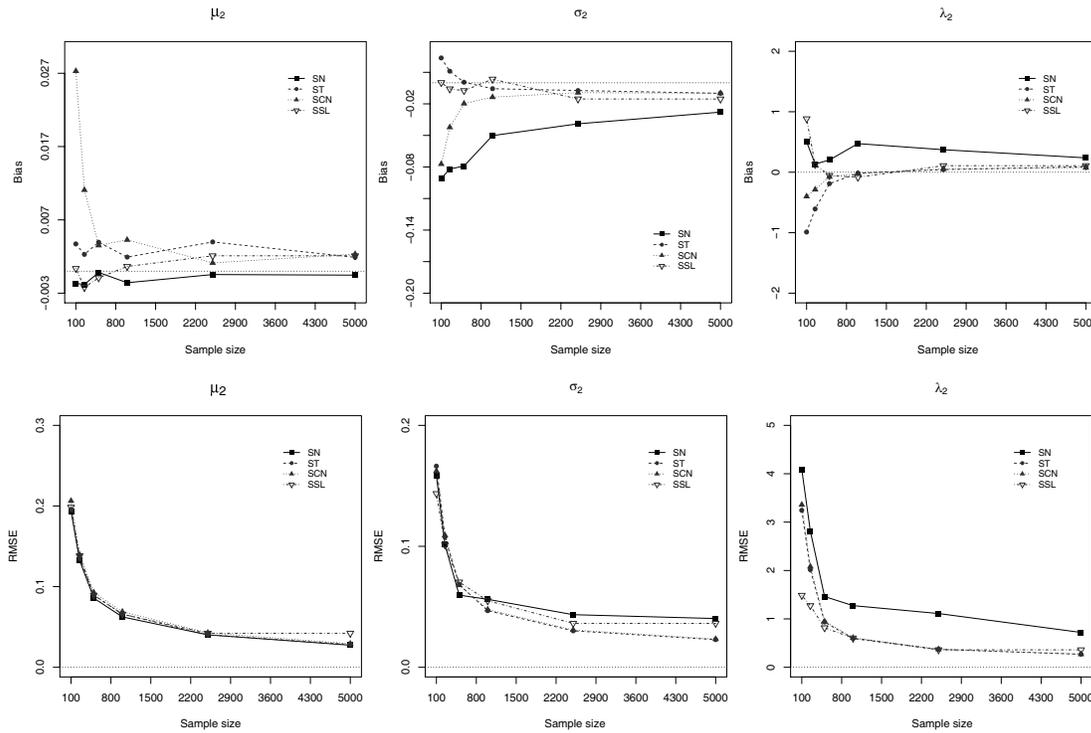


Figure 5. Average bias (1st row) and average RMSE (2nd row) of the estimators of $\mu_2, \sigma_2, \lambda_2$ for simulation II.

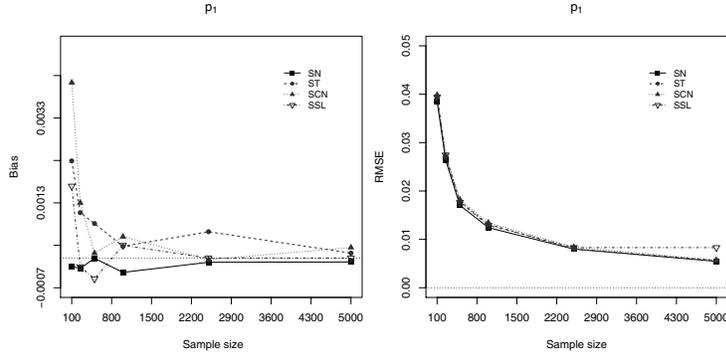


Figure 6. Average bias (1st row) and average RMSE (2nd row) of the estimators of p_1 for simulation II.

We are interested in evaluating the relative change (RC) in the estimates as a function of ν . Given $\Theta = (\beta_1, \beta_2, p_1, p_2, \theta_1, \theta_2)$, with $\theta_j = (\beta_0, \mu_j, \sigma_j^2, \lambda_j)$, $j = 1, 2$, the RC is defined by

$$\text{RC} \left(\hat{\Theta}_i(\nu) \right) = \left| \frac{\hat{\Theta}_i(\nu) - \hat{\Theta}_i}{\hat{\Theta}_i} \right|,$$

where $\hat{\Theta}_i(\nu)$ and $\hat{\Theta}_i$ denote the EM estimates of Θ_i with and without perturbation, respectively.

Figure 7 shows the average values of the relative changes undergone by all the parameters. We note that for all parameters, the average RCs suddenly increase under FM-SN-LR model as the ν value grows. In contrast, for the FM-SMSN-LR models with heavy tails, namely the FM-ST-LR ($\nu = 3$) and FM-SCN-LR ($\nu = (0.1, 0.1)$), the measures vary little, indicating they are more robust than the FM-SN-LR model in the ability to accommodate discrepant observations.

4.4 EMPIRICAL ILLUSTRATION

Next, the proposed techniques are illustrated with the analysis a real dataset, the one previously analyzed by [Cook and Weisberg \(1982\)](#) in a normal regression setting. The dataset comes from the Australian Institute of Sport (AIS) and consists of measurements of 202 athletes. Here, we focus on percent body fat (Bfat), which is assumed to be explained by the sum of skin folds (ssf) and height in cm (Ht). Thus, we consider the FM-SMSN-LR model given by

$$\text{Bfat}_i = \beta_0 + \beta_1 \text{ssf}_i + \beta_2 \text{Ht}_i + \varepsilon_i, \quad i = 1, \dots, 202,$$

where ε_i belongs to the FM-SMSN family.

By using the `FMsmnsnReg` package (see the appendix), we fit the FM-SMSN-LR models as was described in Section 3. Table 2 compares the fit of various mixture models for $g = 1$ to 5 components, using the model selection criteria discussed in Subsection 3.3. Note from this table that, as expected, the heavy-tailed models perform significantly better than the SN model (and the symmetric counterparts such as the normal and Student-t models), with mixtures of two ($g = 2$) components being significantly better in all cases, except for the normal case (FM-N), where a mixture of $g = 3$ is needed.

Moreover, the 2-component FM-ST-LR model fits the data substantially better. This conclusion also is verified through a hypotheses procedure for testing the number of components in the FM-ST-LR model. As reported by [Turner \(2000\)](#), we can use parametric

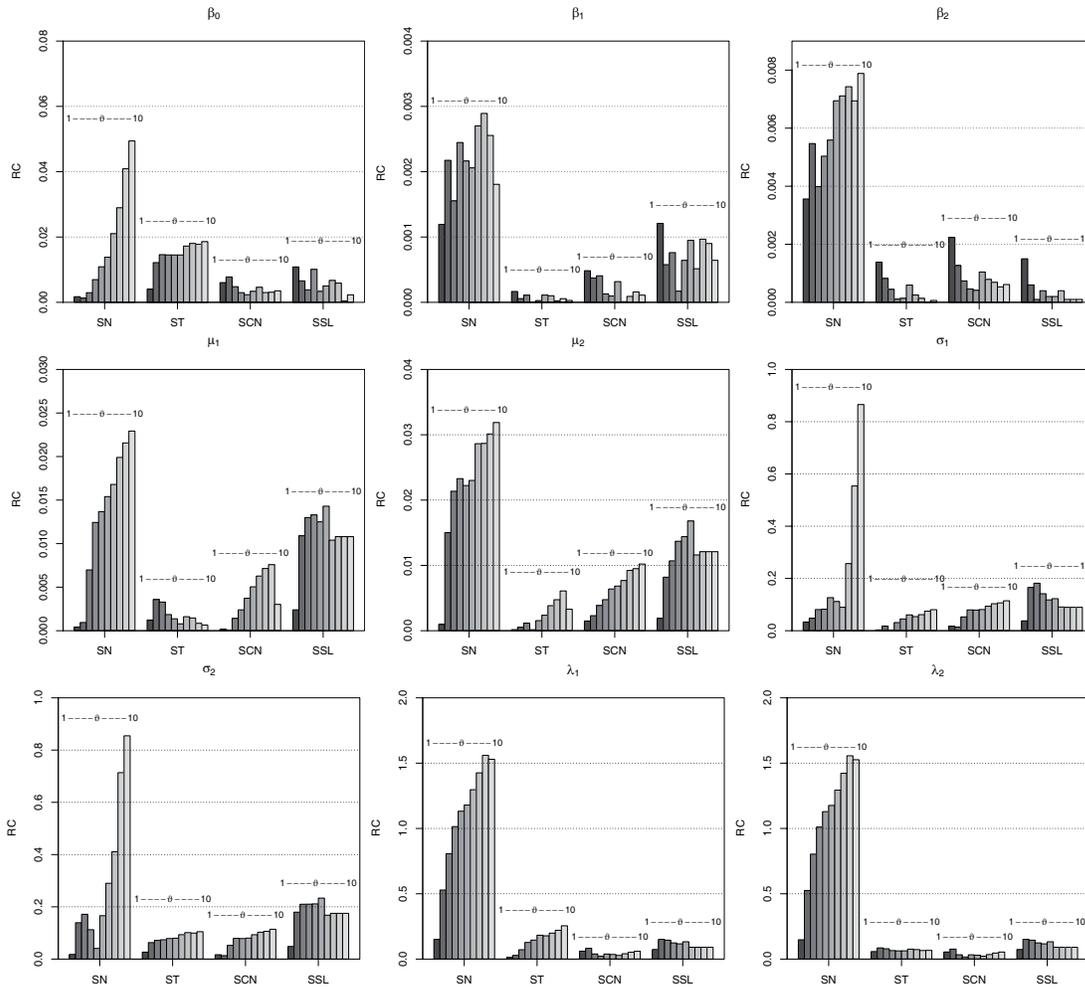


Figure 7. Average RCs of estimates with different perturbations v for simulation study III.

Table 2. Comparison of maximum log-likelihood, AIC_c and BIC_A for fitted FM-SMSN-LR models using the AIS data. The number of parameters is denoted by m .

Model	g	m	log-lik	AIC_c	BIC_a
FM-N	1	5	-367.2395	744.7850	745.1792
FM-N	2	8	-359.2902	735.3265	735.7009
FM-N	3	11	-355.2892	733.9679	734.1192
FM-T	1	6	-363.9525	738.2111	738.6053
FM-T	2	9	-358.2494	733.2449	733.6194
FM-T	3	12	-356.3237	736.0369	736.1881
FM-SN	1	6	-363.0346	738.5001	738.9097
FM-SN	2	10	-356.3079	733.7675	734.0164
FM-SN	3	14	-354.1438	738.5336	738.2486
FM-SN	4	18	-353.1388	746.0152	744.7987
FM-SN	5	22	-352.2579	754.1695	751.5973
FM-ST	1	7	-360.7632	736.1038	736.5070
FM-ST	2	11	-353.9696	731.3286	731.4799
FM-ST	3	15	-353.8492	740.2790	739.7994
FM-ST	4	19	-352.3138	746.8034	745.2888
FM-ST	5	23	-351.7865	755.7752	752.7944
FM-SCN	1	8	-357.0375	738.5001	738.9097
FM-SCN	2	12	-353.7235	733.0978	733.1278
FM-SCN	3	16	-354.1656	743.2717	742.5722
FM-SCN	4	20	-352.0380	748.7169	746.8773
FM-SCN	5	24	-352.8184	760.4164	756.9983
FM-SSL	1	7	-362.3246	739.2264	739.6296
FM-SSL	2	11	-354.1580	731.7054	731.8566
FM-SSL	3	15	-354.1941	740.9689	740.4892
FM-SSL	4	19	-352.2586	746.6930	745.1785
FM-SSL	5	23	-352.3504	756.9031	753.9224

Table 3. AIS data. Parameter estimates of the FM-SMSN- LR models with $g = 2$. SE denotes the corresponding standard errors obtained via the information-based matrix.

Parameter	FM-SN		FM-ST		FM-SCN		FM-SSL	
	ML	SE	ML	SE	ML	SE	ML	SE
β_0	14.7241	0.0001	14.51593	0.00253	14.6622	0.0025	14.7475	0.0025
β_1	0.1799	0.0012	0.17972	0.00850	0.1805	0.0089	0.1796	0.0091
β_2	-0.0757	0.1302	-0.07536	0.19264	-0.0757	0.1458	-0.0754	0.1513
p_1	0.1543	0.9295	0.15418	1.04192	0.1483	1.0841	0.1514	1.0393
μ_1	2.5504	2.2932	1.93244	4.00942	2.3654	3.8355	2.3891	3.9553
μ_2	-0.4652	1.8546	-0.35226	2.94875	-0.4120	2.5091	-0.4263	2.6266
σ_1^2	0.8483	0.5074	3.80681	1.57056	2.2957	1.6255	2.3158	1.6615
σ_2^2	2.2793	0.4021	1.06550	11.56693	1.1240	7.1021	0.9740	7.0029
λ_1	0.1624	0.8467	-5.70438	0.52991	-3.5415	0.4408	-4.8612	0.3724
λ_2	-2.2318	1.7509	-0.62860	9.52263	-1.0111	7.9389	-1.0144	11.9961
ν	-	-	7.45874	-	0.2270	-	2.3036	-
γ	-	-	-	-	0.3075	-	-	-

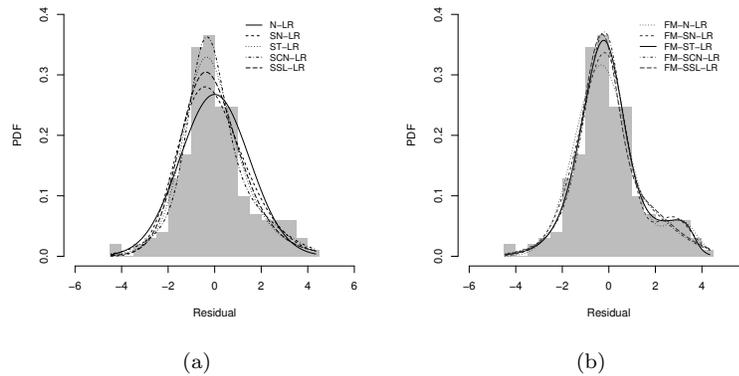


Figure 8. Panels (a) and (b) display the histogram ordinary residuals superimposed on the FM-SMSN-LR residual PDF for $g = 1$ and $g = 2$ components, respectively with AIS dataset.

or semiparametric bootstrap to test hypotheses concerning the number of components in the mixture. Following the method proposed by Turner (2000), we considered 1000 bootstrap statistics for testing $g = 1$ versus $g = 2$, in which case the p -value was 0.027 for the parametric bootstrap. Accordingly, there is strong evidence that at least two components are needed. For testing $g = 2$ versus $g = 3$, the bootstrap p -value was 0.984, so there is no evidence that more than two components are required to model the AIS dataset.

Table 3 presents the ML estimates of the parameters considering the four models with $g = 2$, say FM-SN-LR, FM-ST-LR, FM-SCN-LR and the FM-SSL-LR, along with the corresponding standard errors (SE), obtained via the information-based procedure presented in Subsection 3.3. Notice from Table 3 that the small value of the estimate of ν for the FM-ST-LR and FM-SSL-LR models indicates a lack of adequacy of the SN assumption.

In Figure 8, we plot the histogram of OLS residuals and then display the residual PDFs for the four FM-SMSN-LR models superimposed on a single set of coordinate axes, with $g = 1$ and $g = 2$ components respectively. Additional results related to $g = 3$ and $g = 4$ components are given in Figure 10. Based on this graphical representation, it appears once again that the FM-ST-LR, FT-SCN-LR and the FT-SSL-LR models have quite reasonable and better fit than the FM-SN-LR model with $g = 2$ components.

In order to detect incorrect specification of the error distribution for our best model (FM-ST-LR), we present quantile versus quantile (QQ) plots and simulated envelopes for the residuals $(y - \hat{y})$ in Figure 9. The QQ plots for the other models are given in Figure 11. This figure provides strong evidence that the FM-ST-LR (with $g = 2$ components) yields a better fit to the current data than the ST-LR model (with $g = 1$ component), since there are no observations falling outside the envelope.

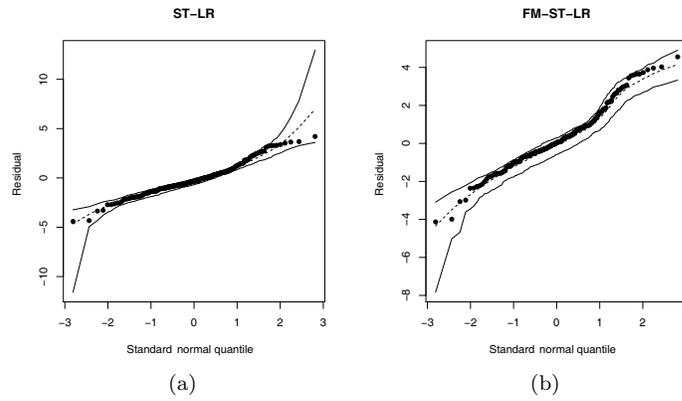


Figure 9. Panels (a) and (b) display the QQ plots and simulated envelopes for the residual $(y - \hat{y})$ with for $g = 1$ and $g = 2$ components, respectively with AIS dataset.

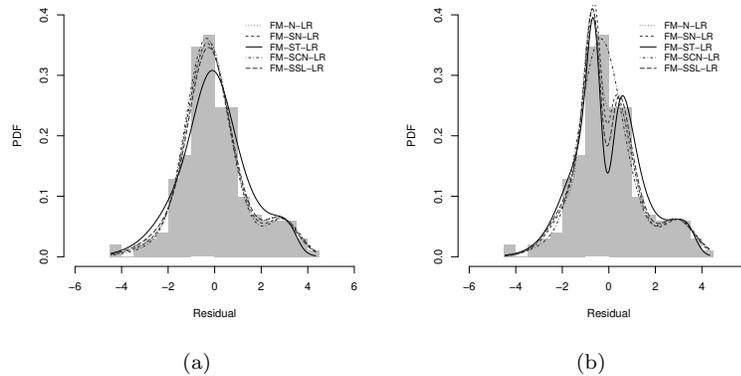


Figure 10. Panels (a) and (b) display the histogram of ordinary residuals with FM-SMSN-LR residual with for $g = 3$ and $g = 4$ components, respectively with AIS dataset.

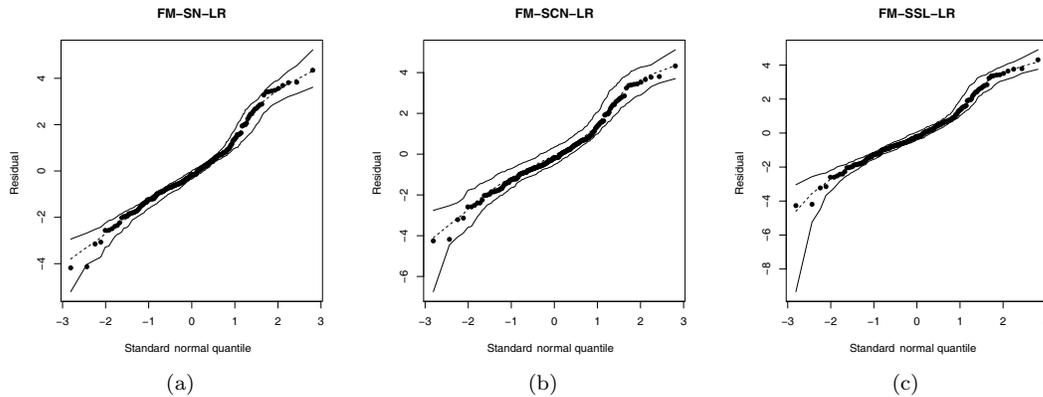


Figure 11. Panels (a), (b) and (c) display the QQ plots and simulated envelopes for the residual $(y - \hat{y})$ for $g = 2$ components based on FM-SN, FM-SCN and FM-SSL distributions, respectively with AIS dataset.

5. CONCLUSIONS

In this paper we consider a regression model whose error term follows a finite mixture of SMSN distributions, which is a rich class of distributions that contains the skew-normal, skew-t, skew-slash and skew-contaminated normal distributions as proper elements. This

approach allows us to model data with great flexibility, simultaneously accommodating multimodality, skewness and heavy tails for the random error in linear regression models. It is important to stress that our proposal is different from that of [Zeller et al. \(2016\)](#), where they use a finite mixture of linear regression models, the so-called switching regression. In this paper, instead of mixtures of regressions, mixtures are exploited as a convenient semiparametric method, which lies between parametric models and kernel PDF estimators, to model the unknown distributional shape of the errors. For this structure we developed a simple EM-type algorithm to perform ML inference of the parameters with closed-form expression at the E-step. The proposed methods are implemented using the `FMsmnReg` package, providing practitioners with a convenient tool for further applications in their domain. The practical utility of the new method is illustrated with the analysis of a real dataset and several simulation studies.

The proposed methods can be extended to multivariate settings using the multivariate SMSN class of distributions ([Cabral et al., 2012](#)), such as the recent proposals of [Soffritti and Galimberti \(2011\)](#) and [Galimberti and Soffritti \(2014\)](#). Due to the popularity of Markov chain Monte Carlo techniques, another potential work is to pursue a fully Bayesian treatment in this context for producing posterior inference. The method can also be extended to mixtures of regressions with skewed and heavy-tailed censored responses, based on recent approaches by [Caudill \(2012\)](#) and [Karlsson and Laitila \(2014\)](#).

APPENDIX: SAMPLE OUTPUT FROM THE `FMsmnReg` PACKAGE

```

-----
Finite Mixture of Scale Mixture Skew Normal Regression Model
-----
Observations = 202
Family = Skew.t

-----
Estimates
-----
      Estimate      SE
beta0  14.51593  0.00253
beta1   0.17972  0.00850
beta2  -0.07536  0.19264
mu1     1.93244  4.00942
mu2    -0.35226  2.94875
sigma1   3.80681  1.57056
sigma2   1.06550 11.5669
shape1  -5.70438  0.52991
shape2  -0.62860  9.52263
pi11     0.15418  1.04192
nu       7.45874    NA

-----
Model selection criteria
-----
Loglik      AIC      BIC      EDC      ICL
Value -357.030 730.235 766.626 739.502 2916.687

-----
Details
-----
Convergence reached? = TRUE
EM iterations = 147 / 500
Criteria = 1e-07
Processing time = 27.11465 secs

```

ACKNOWLEDGMENT

The authors thank the Editors and Reviewers for their constructive comments on an earlier version of this manuscript. Luis Benites and Rocío Maehara acknowledges support from CNPq-Brazil. Victor H. Lachos was supported from CNPq-Brazil (Grant 306334/2015-1). Partial support from CAPES is also acknowledged.

REFERENCES

- Andrews, D.F. and Mallows, C.L., 1974. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society B*, 36, 99-102.
- Arellano-Valle, R.B. and Genton, M.G., 2010. Multivariate unified skew-elliptical distributions. *Chilean Journal of Statistics*, 1, 17-33.
- Azzalini, A., 1985. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171-178.
- Azzalini, A. and Genton, M., 2008. Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review*, 76, 1490-1507.
- Bartolucci, F. and Scaccia, L., 2005. The use of mixtures for dealing with non-normal regression errors. *Computational Statistics and Data Analysis*, 48, 821-834.
- Basford, K., Greenway, D., McLachlan, G., and Peel, D., 1997. Standard errors of fitted component means of normal mixtures. *Computational Statistics*, 12, 1-18.
- Basso, R.M., Lachos, V.H., Cabral, C.R.B., Ghosh, P., 2010. Robust mixture modeling based on scale mixtures of skew-normal distributions. *Computational Statistics and Data Analysis*, 54, 2926-2941.
- Benites, L., Maehara, R., and Lachos, V.H., 2016. FMsmnReg: Regression Models with Finite Mixtures of Skew Heavy-Tailed Errors. R package version 1.0. <http://CRAN.R-project.org/package=FMsmnReg>
- Böhning, D., Dietz, E., Schaub, R., Schlattmann, P., and Lindsay, B.G., 1994. The distribution of the likelihood ratio for mixtures of densities from the one-parameter exponential family. *Annals of the Institute of Statistical Mathematics*, 46, 373-388.
- Branco, M.D. and Dey, D.K., 2001. A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, 79, 99-113.
- Cabral, C.R.B., Lachos, V.H., and Prates, M.O., 2012. Multivariate mixture modeling using skew-normal independent distributions. *Computational Statistics and Data Analysis*, 56, 126-142.
- Caudill, S.B., 2012. A partially adaptive estimator for the censored regression model based on a mixture of normal distributions. *Statistical Methods and Applications*, 21, 121-137.
- Cook, R.D. and Weisberg, S., 1982. *Residuals and Influence in Regression*. Chapman and Hall/CRC, Boca Raton, FL, US.
- Dempster, A., Laird, N., and Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39, 1-38.
- Galea, M., Paula, G.A., and Bolfarine, H., 1997. Local influence in elliptical linear regression models. *Journal of the Royal Statistical Society D*, 46, 71-79.
- Galimberti, G. and Soffritti, G., 2014. A multivariate linear regression analysis using finite mixtures of t distributions. *Computational Statistics and Data Analysis*, 71, 138-150.
- Garay, A.M., Lachos, V.H., and Abanto-Valle, C.A., 2011. Nonlinear regression models based on scale mixtures of skew-normal distributions. *Journal of the Korean Statistical Society*, 40, 115-124.
- Holzmann, H. and Munk, A., 2006. Identifiability of finite mixture of elliptical distributions. *Scandinavian Journal of Statistics*, 33, 753-763.

- Hurvich, C. and Tsai, C.L., 1989. Regression and time series model selection in small samples. *Biometrika*, 76, 297-307.
- Karlsson, M. and Laitila, T., 2014. Finite mixture modeling of censored regression models. *Statistical Papers*, 55, 627-642.
- Lange, K.L., Little, R., and Taylor, J., 1989. Robust statistical modeling using t distribution. *Journal of the American Statistical Association*, 84, 881-896.
- Lange, K.L. and Sinsheimer, J.S., 1993. Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics*, 2, 175-198.
- Lindsay, B., 1995. Mixture models: Theory, geometry and applications. Institute of Mathematical Statistics, Hayward, CA, US.
- Liu, C. and Rubin, D.B., 1994. The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika*, 80, 267-278.
- McNicholas, P., Murphy, T., and McDaid, A.D.F., 2010. Serial and parallel implementations of model-based clustering via parsimonious gaussian mixture models. *Computational Statistics and Data Analysis*, 54, 711-723.
- Otianiano, C., Rathie, P., and Ozelim, L., 2015. On the identifiability of finite mixture of skew-normal and skew-t distributions. *Statistics and Probability Letters*, 106, 103-108.
- Prates, M.O., Lachos, V.H., and Cabral, C.B., 2013. mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions. *Journal of Statistical Software*, 12, 1-20.
- Prates, M.O., Dey, D.K., and Lachos, V.H., 2012. A dengue fever study in the state of Rio de Janeiro with the use of generalized skew-normal/independent spatial fields. *Chilean Journal of Statistics*, 3, 143-155.
- Quandt, R.E. and Ramsey, J.B., 1978. Estimating mixtures of normal distributions and switching regressions. *Journal of the American Statistical Association*, 73, 730-738.
- R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rao, C.R., 1973. *Linear Statistical Inference and its Applications*. Wiley, New York.
- Sclove, S., 1987. Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52, 333-343.
- Soffritti, G. and Galimberti, G., 2011. Multivariate linear regression with non-normal errors: a solution based on mixture models. *Statistics and Computing*, 21, 523-536.
- Turner, T.R., 2000. Estimating the propagation rate of a viral infection of potato plants via mixtures of regressions. *Journal of the Royal Statistical Society C*, 49, 371-384.
- Zeller, C.B., Cabral, C.R., Lachos, V.H., 2016. Robust mixture regression modeling based on scale mixtures of skew-normal distributions. *TEST*, 25, 375-396.

GOODNESS-OF-FIT METHODS
RESEARCH PAPER

Goodness-of-fit test for the Birnbaum-Saunders distribution based on the Kullback-Leibler information

EDNÁRIO MENDONÇA¹, MICHELLI BARROS^{2,*}, and JOELSON CAMPOS²

¹Departamento de Estatística, Universidade Estadual da Paraíba, Campina Grande, Brazil,

²Departamento de Estatística, Universidade Federal de Campina Grande, Campina Grande, Brazil

(Received: 08 March 2019 · Accepted in final form: 18 April 2019)

Abstract

In this work, we propose a goodness-of-fit test based on the Kullback-Leibler information for the Birnbaum-Saunders distribution. We use Monte Carlo simulations to evaluate the size and power of the proposed test for several alternative hypotheses under different sample sizes. We compare the powers with standard goodness-of-fit tests based as the Anderson-Darling and Cramér-von Mises tests. Finally, we illustrate the proposed test with a real data set to show its potential applications.

Keywords: Anderson-Darling and Cramér-von Mises tests · Information measures · Maximum likelihood estimation · Monte Carlo method · Power test · R software

Mathematics Subject Classification: Primary 62J20 · Secondary 62J99.

1. INTRODUCTION

The Birnbaum-Saunders (BS) model, proposed by [Birnbaum and Saunders \(1969\)](#), is a life distribution originating from a material fatigue problem, which relates the time to the occurrence of failure with some cumulative damage that is assumed to be Gaussian distributed. The BS model has received much attention in the last decades due to its wide applicability. Based on its genesis from material fatigue, different cumulative damage processes can be modeled by this distribution, including natural engineering applications, but the BS model can also be applied to other areas as: medicine ([Leiva et al., 2007](#); [Barros et al., 2008](#); [Azevedo et al., 2012](#); [Gomes et al., 2012](#); [Desousa et al., 2018](#); [Leao et al., 2018](#)), atmospheric contamination ([Leiva et al., 2008, 2010, 2015a](#); [Vilca et al., 2011](#); [Ferreira, 2013](#); [Marchant et al., 2018, 2019](#)), water quality ([Leiva et al., 2009](#); [Vilca et al., 2010](#)), neuronal sciences ([Leiva et al., 2015b](#)), human aging ([Leiva and Saunders, 2015](#)), and earthquakes ([Lillo et al., 2018](#)), among others. However, because the BS model is a statistical distribution, we can apply it to several other fields, for example, business, finance, industry, science management, and quality control. For more details about various

*Corresponding author. Email: michelli.karinne@gmail.com

developments on the BS distribution, see [Leiva \(2016\)](#) and references cited therein. The BS model has also been used to construct new more flexible models having heavier and lighter tails than the standard BS distribution, as well as in the construction of models in the unit interval; see [Barros et al. \(2008\)](#), [Azevedo et al. \(2012\)](#), [Mazucheli et al. \(2018\)](#) and [Athayde et al. \(2019\)](#).

In statistics, it is of great interest to determine whether a probabilistic model fits a data set well or not, which could indicate whether these data may have been generated from this model or not. In this sense, several goodness-of-fit tests have been proposed for different probability distributions. Since goodness-of-fit tests measure the discrepancy between a theoretical model and a data set, they can be done in a variety of ways, such as, for example, formulated by chi-squared type tests, by statistics based on the empirical cumulative distribution function or empirical characteristic function. Further details on goodness-of-fit tests can be found in [D'Agostino and Stephens \(1986\)](#), [Castro-Kuriss \(2011\)](#) and [Barros et al. \(2014\)](#).

The Anderson-Darling (AD) and Cramér-von Mises (CM) statistics are often used to test normality. These statistics are based on the distance between the empirical distribution function and the theoretical distribution function. [Chen and Balakrishnam \(1995\)](#) proposed a general purpose approximate goodness-of-fit test based on these statistics which may be used to test the validity of different families of skew distributions. Note that the Kullback-Leibler (KL) criterion is an information measure, which can be used to evaluate the discrepancy between two distribution functions. Such a measure of information has shown good results in testing fitting of models to data sets, in the sense of obtaining more powerful tests than the standard tests; see [Park \(2005\)](#) and [Rad et al. \(2011\)](#). Then, due to the wide applicability of the BS distribution, the objective of this paper is to propose a goodness-of-fit test for the BS distribution based on the KL information and investigate if the proposed test is most powerful than in the case of standard AD and CM tests.

The rest of this paper is organized as follows. In [Section 2](#), we present the methodology with the definitions of entropy, KL information, and a brief review of the BS distribution, as well as an estimation method of its parameters. In addition, in this section, goodness-of-fit test for the BS distribution based on KL information are derived. In [Section 3](#), a simulation study based on the Monte Carlo method is conducted to evaluate the size and power of the proposed test. Also in this section, we illustrate the proposed methodology with a real data set. Finally, [Section 4](#) provides the conclusions of this work and some comments on future research related to this topic.

2. METHODOLOGY

2.1 ENTROPY AND KULLBACK-LEIBLER INFORMATION

In order to quantify the degree of disorder in a physical system the German Rudfold Clausius introduced in [Clausius \(1867\)](#) a new quantity in thermodynamics which he called entropy. Since this concept was introduced in studies of information theory by [Shannon \(1948\)](#). Shannon's idea was to measure the degree of disorder of the occurrence of the values of a random variable (RV) in the sense that the more distinct rare events occur.

Let X be an RV with cumulative distribution function (CDF) F and probability density function (PDF) f . The differential entropy $H(f)$ of X is defined in [Shannon \(1948\)](#) by

$$H(f) = - \int_{-\infty}^{\infty} f(x) \log(f(x)) dx.$$

Let X_1, \dots, X_n , with $n \geq 3$, be a sample from the distribution F , and let $X_{(1)} \leq \dots \leq X_{(n)}$

be their corresponding order statistics. A nonparametric estimator of $H(f)$, proposed by Vasicek (1976), is given by

$$H_{mn} = \frac{1}{n} \sum_{i=1}^n \log \left\{ \frac{n}{2m} (x_{(i+m)} - x_{(i-m)}) \right\}, \quad (1)$$

where the window m is a positive integer less than $n/2$ and $x_{(i-m)} = x_{(1)}$, for $i - m < 1$ and $x_{(i+m)} = x_{(n)}$, for $i + m > n$, such that $x_{(i)}$ is i -th observed value of the corresponding order statistic.

Let $f(x)$ and $g(x)$ be PDFs. The KL information is defined in Kullback and Leibler (1951) as

$$I(f: g) = \int_{-\infty}^{\infty} f(x) \log \left[\frac{f(x)}{g(x)} \right] dx, \quad (2)$$

so that $I(f: g)$ measures the divergence between the PDFs f and g . By using the Gibbs inequality, we can show that $I(f: g) \geq 0$ and $I(f: g) = 0$ if and only if $f(x) = g(x)$. Thus, the sample estimate of the KL information can also be considered for goodness of fit.

2.2 THE BIRNBAUM-SAUNDERS DISTRIBUTION

Let X be a nonnegative RV. Then, X follows a BS distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$, if the CDF of X is given by

$$F(x) = \Phi \left[\frac{1}{\alpha} \left(\sqrt{\frac{x}{\beta}} - \sqrt{\frac{\beta}{x}} \right) \right], \quad x > 0.$$

We use the notation $X \sim \text{BS}(\alpha, \beta)$ for indicating an RV X with BS distribution of shape and scale parameters α and β , respectively. Consequently, the PDF of X is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2\alpha^2} \left(\frac{x}{\beta} + \frac{\beta}{x} - 2 \right) \right] \frac{x^{-3/2}(x + \beta)}{2\alpha\sqrt{\beta}}, \quad x > 0. \quad (3)$$

If $X \sim \text{BS}(\alpha, \beta)$, then the following properties are satisfied:

- (i) The parameter β is also the median of the distribution.
- (ii) If $Z \sim \text{N}(0, 1)$, then X and Z are related by $X = \beta(\alpha Z + (\alpha^2 Z^2 + 4)^{1/2})^2/4$. Thus, $Z = (1/\alpha)[(X/\beta)^{1/2} - (\beta/X)^{1/2}] \sim \text{N}(0, 1)$.
- (iii) $cX \sim \text{BS}(\alpha, c\beta)$, if $c > 0$ and $1/X \sim \text{BS}(\alpha, 1/\beta)$.
- (iv) $E(X) = \beta(1 + \alpha^2/2)$ and $\text{Var}(X) = \beta^2\alpha^2(1 + 5\alpha^2/4)$.
- (v) The q th quantile of X is given by $x_q = \beta(\alpha z_q + (\alpha^2 z_q^2 + 4)^{1/2})^2/4$, where $z_q = \Phi^{-1}(q)$, $\text{N}(0, 1)$ q th quantile.
- (vi) The survival function is expressed as $S(x; \alpha, \beta) = \Phi\{(1/\alpha)[(\beta/x)^{1/2} - (x/\beta)^{1/2}]\}$.

For estimation of the model parameters, we consider the maximum likelihood (ML) method. Let X_1, \dots, X_n be a random sample of size n from $X \sim \text{BS}(\alpha, \beta)$ with PDF given by Equation PDF), so that x_1, \dots, x_n are their respective observed values. Then,

the log-likelihood function for $\boldsymbol{\theta} = (\alpha, \beta)^\top$ is given by

$$\ell(\boldsymbol{\theta}) = K - \frac{1}{2\alpha^2} \sum_{i=1}^n \left(\frac{x_i}{\beta} + \frac{\beta}{x_i} - 2 \right) + \sum_{i=1}^n \log(x_i + \beta) - n \log(\alpha) - \frac{n}{2} \log(\beta),$$

where $K = n(\log(1/\sqrt{2\pi}) - \log(2)) - 3/2 \sum_{i=1}^n \log(x_i)$. The ML estimate of α is defined as

$$\hat{\alpha} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{x_i}{\hat{\beta}} + \frac{\hat{\beta}}{x_i} - 2 \right)}.$$

In the case of the parameter β , the ML estimate do not have closed form requiring the use of a numerical method. Under regularity conditions (see [Cox and Hinkley, 1974](#)), the estimators $\hat{\alpha}$ and $\hat{\beta}$ are consistent and have a bivariate normal joint asymptotic distribution with asymptotic means α and β , respectively, and an asymptotic covariance matrix $\boldsymbol{\Sigma}_{\hat{\theta}}$ that can be obtained from the inverse of the Fisher information matrix given by

$$\mathcal{I}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{2n}{\alpha^2} & 0 \\ 0 & \frac{n}{\beta^2} \left(\frac{1}{4} + \frac{1}{\alpha^2} + I(\alpha) \right) \end{pmatrix},$$

where

$$I(\alpha) = 2 \int_0^\infty \left(\frac{1}{1 + \frac{1}{\xi(az)}} - \frac{1}{2} \right)^2 \phi(z) dz,$$

with ϕ being the PDF of $Z \sim N(0, 1)$ and $\xi(u) = u^{1/2} - u^{-1/2}$. For more details, see [Leiva \(2016\)](#).

2.3 GOODNESS-OF-FIT TESTS FOR THE BS DISTRIBUTION

Given a random sample X_1, \dots, X_n of the RV X , we are interested in testing H_0 : the RV X follows the $BS(\alpha, \beta)$ distribution with PDF given in Equation (3) against H_1 : the RV X does not follow the BS distribution. Note that Equation (2) can be written as

$$\begin{aligned} I(f: g) &= \int_{-\infty}^{\infty} f(x) [\log(f(x)) - \log(g(x))] dx \\ &= -H(f) - \int_{-\infty}^{\infty} f(x) \log(g(x)) dx. \end{aligned} \quad (4)$$

Then, from Equation (4), an estimate of the KL information can be obtained. For doing this, we replace $H(f)$ by its estimate given in Equation (1) and we use the estimated values of the parameters in f . Thus, under the null hypothesis that $f(x) = g(x)$, we can estimate the information of KL using

$$I_{mn} = -H_{mn} - \int_{-\infty}^{\infty} f(x; \hat{\boldsymbol{\theta}}) \log(f(x; \hat{\boldsymbol{\theta}})) dx,$$

where $\hat{\boldsymbol{\theta}}$ is a consistent estimator for $\boldsymbol{\theta}$. Therefore, I_{mn} is a test statistic to verify the suitability of a continuous probabilistic model with PDF given by f to a data set.

For $X \sim \text{BS}(\alpha, \beta)$ and f given in Equation (3), we obtain

$$I_{mn} = -H_{mn} - \log \frac{1}{\sqrt{2\pi}} - \frac{1}{\hat{\alpha}^2} + \log \left(2\hat{\alpha}\sqrt{\hat{\beta}} \right) + \frac{1}{\hat{\alpha}^2} \left(1 + \frac{\hat{\alpha}^2}{2} \right) \\ + \frac{3}{2n} \sum_{i=1}^n \log(x_{(i)}) - \frac{1}{n} \sum_{i=1}^n \log(x_{(i)} + \hat{\beta}),$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the ML estimates of α and β , respectively. Thus, following [Arizono and Ohta \(1989\)](#), we introduce the statistic

$$\text{KL}_{mn} = \frac{1}{\exp(I_{mn})},$$

with $0 \leq \text{KL}_{mn} \leq 1$ since $I_{mn} \in [0, \infty)$. Note that KL_{mn} can be used as test statistic for testing the goodness-of-fit of the BS distribution to a data set. The decision rule is to reject the hypothesis H_0 if $\text{KL}_{mn} \leq \text{KL}_{mn}^*(\rho)$, where $\text{KL}_{mn}^*(\rho)$ is the critical value for a significance level ρ . As we do not have an exact distribution of KL_{mn} , then we obtain $\text{KL}_{mn}^*(\rho)$ through Monte Carlo simulations.

3. NUMERICAL STUDIES

3.1 CRITICAL VALUES FOR THE SIMULATIONS

To obtain the critical values of the proposed test, we conduct Monte Carlo simulation studies with $R = 10,000$ replications each. These studies are based on $n \in \{10, 30, 50, 100\}$, $\alpha \in \{0.5, 1.0, 1.5\}$, and significance level $\rho = 0.05$. In addition, we fix, without loss of generality, $\beta = 1$, since this is a scale parameter. The values considered for the window m are those returned the maximum critical value, according to [Arizono and Ohta \(1989\)](#). This procedure is described in Algorithm 1. All simulations are obtained from implementations in the R statistical software, which is freely distributed from www.R-project.org. For parameters estimation we use the `maxLik` package.

Algorithm 1: Obtaining the critical values of the proposed test.

- 1: Fix n , α and β ;
- 2: Generate 10,000 random samples of size n from $X \sim \text{BS}(\alpha, \beta)$;
- 3: For each sample, estimate the parameter vector $\boldsymbol{\theta} = (\alpha, \beta)^\top$ consistently, through the ML method;
- 4: For each sample, obtain the values of the test statistic KL_{mn} ;
- 5: Sort the test statistic values obtained in the previous step and determine the 5th quantile and then obtain the critical values for the respective significance level.

The critical values obtained, considering the $\text{BS}(0.5, 1)$, $\text{BS}(1, 1)$ and $\text{BS}(1.5, 1)$ distributions are presented in Tables 1-3.

3.2 EVALUATING THE EMPIRICAL SIZE AND POWER OF THE TEST

Next, the empirical size and power of the proposed test are evaluated for different sample sizes based on the Monte Carlo method. We make a comparison among the AD, CM and KL tests, whose statistics are denoted by A^2 , W^2 , KL, and verify in what situations the test based on the KL information is better, in the sense of being most powerful.

Table 1. Critical values for the statistic KL_{mn} considering the $BS(0.5,1)$ distribution and significance level 5%.

n	m									
	1	2	3	4	5	6	7	8	9	10
3	0.2462									
4	0.2577									
5	0.2925	0.4221								
6	0.3256	0.4404								
7	0.3544	0.4620	0.4835							
8	0.3866	0.4935	0.5083							
9	0.4054	0.5102	0.5319	0.5168						
10	0.4250	0.5340	0.5481	0.5401						
12	0.4614	0.5689	0.5840	0.5760	0.5625					
14	0.4911	0.5908	0.6114	0.6072	0.5973	0.5771				
16	0.5159	0.6207	0.6383	0.6354	0.6227	0.6069	0.5880			
18	0.5308	0.6396	0.6597	0.6605	0.6461	0.6331	0.6184	0.5980		
20	0.5499	0.6564	0.6820	0.6796	0.6674	0.6542	0.6428	0.6250	0.6082	
25	0.5754	0.6871	0.7176	0.7194	0.7124	0.7042	0.6905	0.6769	0.6617	0.6489
30	0.5976	0.7132	0.7421	0.7474	0.7481	0.7384	0.7280	0.7153	0.7036	0.6899
35	0.6122	0.7297	0.7593	0.7699	0.7707	0.7655	0.7577	0.7473	0.7352	0.7254
40	0.6243	0.7423	0.7766	0.7904	0.7900	0.7860	0.7789	0.7720	0.7620	0.7527
45	0.6343	0.7547	0.7887	0.8007	0.8053	0.8034	0.7975	0.7917	0.7832	0.7765
50	0.6426	0.7634	0.7982	0.8129	0.8165	0.8142	0.8146	0.8062	0.8027	0.7935
60	0.6568	0.7755	0.8135	0.8291	0.8350	0.8368	0.8355	0.8330	0.8274	0.8235
70	0.6646	0.7854	0.8251	0.8421	0.8498	0.8515	0.8522	0.8501	0.8476	0.8435
80	0.6751	0.7959	0.8349	0.8514	0.8596	0.8641	0.8644	0.8649	0.8628	0.8595
90	0.6804	0.8012	0.8408	0.8598	0.8687	0.8735	0.8758	0.8742	0.8733	0.8718
100	0.6858	0.8075	0.8471	0.8656	0.8760	0.8818	0.8833	0.8841	0.8826	0.8813

Under same the conditions of the obtained critical values, we calculate the empirical size of the test. Algorithm 2 displays this procedure. The results of our simulation study are presented in Table 4. Note that the empirical size is close to the nominal level for all situations considered, indicating that the test is controlled.

Algorithm 2: Obtaining the empirical size of the proposed test.

- 1: Fix n , α and β ;
- 2: Generate 10,000 random samples of size n from $X \sim BS(\alpha, \beta)$;
- 3: For each sample, estimate the parameter vector $\theta = (\alpha, \beta)^\top$ consistently, through the ML method;
- 4: For each sample, obtain the values of the test statistic KL_{mn} ;
- 5: Obtain the empirical size of the test by calculating the proportion of replications that present test statistic value less than the critical value for the corresponding values of n and m .

To determine the empirical power, we consider some probability distributions for the alternative hypothesis. These distributions are chosen and grouped into classes to be analyzed according to the shape of their hazard function: increasing, decreasing and non-monotonous. The probability distributions considered in the evaluation of the power test

Table 2. Critical values for the statistic KL_{mn} considering the BS(1,1) distribution and significance level 5%.

n	m									
	1	2	3	4	5	6	7	8	9	10
3	0.2618									
4	0.2724									
5	0.3066	0.4446								
6	0.3369	0.4698								
7	0.3653	0.4947	0.5095							
8	0.3974	0.5215	0.5398							
9	0.4132	0.5349	0.5650	0.5420						
10	0.4350	0.5575	0.5809	0.5691						
12	0.4681	0.5870	0.6143	0.6133	0.5941					
14	0.4960	0.6066	0.6390	0.6408	0.6338	0.6115				
16	0.5202	0.6338	0.6610	0.6654	0.6591	0.6464	0.6292			
18	0.5342	0.6508	0.6803	0.6871	0.6807	0.6709	0.6592	0.6416		
20	0.5539	0.6671	0.6974	0.7049	0.6989	0.6923	0.6854	0.6683	0.6558	
25	0.5778	0.6946	0.7299	0.7387	0.7371	0.7358	0.7279	0.7177	0.7056	0.6986
30	0.5987	0.7193	0.7525	0.7621	0.7686	0.7639	0.7587	0.7502	0.7436	0.7357
35	0.6140	0.7341	0.7680	0.7823	0.7863	0.7856	0.7827	0.7776	0.7709	0.7651
40	0.6258	0.7460	0.7830	0.7990	0.8038	0.8031	0.7997	0.7968	0.7911	0.7869
45	0.6355	0.7572	0.7942	0.8089	0.8156	0.8169	0.8157	0.8122	0.8085	0.8053
50	0.6436	0.7663	0.8028	0.8196	0.8261	0.8261	0.8289	0.8249	0.8231	0.8175
60	0.6575	0.7776	0.8169	0.8338	0.8422	0.8453	0.8471	0.8463	0.8431	0.8427
70	0.6651	0.7872	0.8279	0.8461	0.8555	0.8583	0.8603	0.8606	0.8596	0.8577
80	0.6753	0.7967	0.8375	0.8546	0.8637	0.8694	0.8717	0.8730	0.8727	0.8716
90	0.6806	0.8024	0.8429	0.8621	0.8722	0.8781	0.8815	0.8809	0.8816	0.8809
100	0.6859	0.8084	0.8484	0.8675	0.8789	0.8850	0.8877	0.8898	0.8895	0.8895

are: gamma, generalized exponential, beta, Pareto type I, Weibull, and half-normal, whose PDFs are the following:

- Gamma($\kappa; \theta$) with PDF

$$f_1(x; \kappa, \theta) = \frac{1}{\Gamma(\kappa)\theta^\kappa} x^{\kappa-1} \exp\left(-\frac{x}{\theta}\right), \quad x > 0, \kappa, \theta > 0$$

and CDF denoted by F_1 .

- GExp($\kappa; \theta$) with PDF

$$f_2(x; \kappa, \theta) = \kappa\theta x \exp\{-\theta x\}[1 - \exp(-\theta x)]^{\kappa-1}, \quad x > 0,$$

$\kappa, \theta > 0$, and CDF denoted by F_2 .

- Beta($\kappa; \theta$), with PDF

$$f_3(x; \kappa, \theta) = \frac{\Gamma(\kappa + \theta)}{\Gamma(\kappa)\Gamma(\theta)} x^{\kappa-1}(1 - x)^{\theta-1}, \quad 0 < x < 1,$$

$\kappa, \theta > 0$, and CDF denoted by F_3 .

Table 3. Critical values for the statistic KL_{mn} considering the BS(1.5,1) distribution and significance level 5%.

n	m									
	1	2	3	4	5	6	7	8	9	10
3	0.2819									
4	0.2911									
5	0.3237	0.4760								
6	0.3514	0.5053								
7	0.3796	0.5303	0.5440							
8	0.4102	0.5547	0.5791							
9	0.4256	0.5665	0.6065	0.5852						
10	0.4449	0.5865	0.6206	0.6114						
12	0.4763	0.6121	0.6529	0.6591	0.6446					
14	0.5036	0.6274	0.6734	0.6850	0.6848	0.6684				
16	0.5270	0.6509	0.6916	0.7056	0.7080	0.7029	0.6938			
18	0.5385	0.6651	0.7067	0.7234	0.7267	0.7247	0.7198	0.7136		
20	0.5564	0.6787	0.7191	0.7371	0.7430	0.7454	0.7462	0.7343	0.7317	
25	0.5806	0.7025	0.7460	0.7630	0.7704	0.7791	0.7791	0.7785	0.7741	0.7770
30	0.6010	0.7260	0.7643	0.7816	0.7940	0.7992	0.8004	0.8024	0.8021	0.8048
35	0.6149	0.7396	0.7777	0.7984	0.8080	0.8131	0.8167	0.8198	0.8209	0.8238
40	0.6273	0.7499	0.7900	0.8116	0.8210	0.8254	0.8293	0.8318	0.8340	0.8349
45	0.6365	0.7602	0.8008	0.8193	0.8298	0.8358	0.8404	0.8425	0.8447	0.8478
50	0.6443	0.7685	0.8080	0.8285	0.8385	0.8425	0.8496	0.8507	0.8541	0.8544
60	0.6581	0.7794	0.8206	0.8402	0.8507	0.8572	0.8614	0.8645	0.8661	0.8708
70	0.6657	0.7884	0.8309	0.8506	0.8618	0.8677	0.8720	0.8749	0.8778	0.8789
80	0.6758	0.7977	0.8397	0.8580	0.8694	0.8761	0.8808	0.8847	0.8867	0.8886
90	0.6807	0.8026	0.8449	0.8652	0.8760	0.8838	0.8891	0.8904	0.8934	0.8946
100	0.6856	0.8089	0.8502	0.8699	0.8824	0.8895	0.8936	0.8972	0.8988	0.9012

Table 4. Empirical size for different sample size and values of the parameter α indicated.

n	m	BS(0.5,1)	BS(1,1)	BS(1.5,1)
10	3	0.0473	0.0588	0.0494
30	5	0.0564	0.0563	0.0513
50	7	0.0520	0.0514	0.0454
100	8	0.0504	0.0515	0.0482

- Pareto($\kappa; \theta$), with PDF

$$f_4(x; \kappa, \theta) = \frac{\kappa \theta^\kappa}{x^{\kappa+1}}, \quad x \in [\theta, \infty), \quad \kappa, \theta > 0,$$

and CDF denoted by F_4 .

- Weibull($\kappa; \theta$), with PDF

$$f_5(x; \kappa, \theta) = \frac{\kappa}{\theta} \left(\frac{x}{\theta}\right)^{\kappa-1} \exp\left\{-\left(\frac{x}{\theta}\right)^\kappa\right\}, \quad x > 0,$$

$\kappa, \theta > 0$ and CDF denoted by F_5 .

- $HN(\theta)$, with PDF

$$f_6(x; \theta) = \frac{2\theta}{\pi} \exp\left(-\frac{x^2\theta^2}{\pi}\right), \quad x \geq 0, \theta > 0,$$

and CDF denoted by F_6 .

The power of the test is calculated based on testing the hypotheses

$$\begin{cases} H_0: X \sim BS(\alpha, \beta), \text{ for some } \alpha > 0 \text{ and } \beta > 0; \\ H_1: X \sim F_i(\boldsymbol{\theta}), \text{ with } \boldsymbol{\theta} > 0 \text{ and } i = 1, \dots, 6. \end{cases}$$

In the procedure, 10,000 Monte Carlo replications and sample sizes $n = 10, 30, 50, 100$ are considered. The powers of the tests are obtained at the significance level $\rho = 0.05$. For each value of n and each distribution in H_1 , with different parameters, the 10,000 samples are generated and the respective values of the test statistic are calculated. Based on the critical values presented in Tables 1-3, we obtain the rejection proportions based on the 10,000 simulated samples. In addition, the power of the test is evaluated based on the CM and AD statistics using the procedure proposed by [Chen and Balakrishnam \(1995\)](#). We make a comparison among the tests and verify in what situations the test based on the KL information is better, in the sense of being most powerful. Tables 5-8 present the powers for the test in question with sample sizes of $n = 10, n = 30, n = 50$ and $n = 100$, respectively.

Table 5. Empirical power for different forms of hazard functions and different distributions considering sample size $n = 10$.

Hazard function	Alternatives	KL_{mn}	W^2	A^2
Increasing	Gamma(3; 1)	0.1534	0.0873	0.0937
	GExp(3; 1)	0.1288	0.0805	0.0825
	Beta(2; 1)	0.6841	0.3970	0.4282
Decreasing	Gamma(0.5; 1)	0.0376	0.0890	0.0959
	GExp(0.5; 1)	0.0428	0.0938	0.1025
Nonmonotone	Pareto(2; 1)	0.4748	0.4070	0.4342
	Weibull(2; 1)	0.2405	0.1507	0.1617
	HN(3)	0.2434	0.2096	0.2298

Table 6. Empirical power for different forms of hazard functions and different distributions considering sample size $n = 30$.

Hazard function	Alternatives	KL_{mn}	W^2	A^2
Increasing	Gamma(3; 1)	0.2656	0.1856	0.2082
	GExp(3; 1)	0.2050	0.1544	0.1719
	Beta(2; 1)	0.9970	0.9053	0.9388
Decreasing	Gamma(0.5; 1)	0.3465	0.3959	0.5343
	GExp(0.5; 1)	0.3638	0.3937	0.5442
Nonmonotone	Pareto(2; 1)	0.9767	0.9039	0.9365
	Weibull(2; 1)	0.5458	0.4172	0.4559
	HN(3)	0.7164	0.6576	0.6987

Table 7. Empirical power for different forms of hazard functions and different distributions considering sample size $n = 50$.

Hazard function	Alternatives	KL_{mn}	W^2	A^2
Increasing	Gamma(3; 1)	0.3575	0.2774	0.3099
	GExp(3; 1)	0.2716	0.2187	0.2498
	Beta(2; 1)	1.0000	0.9905	0.9965
Decreasing	Gamma(0.5; 1)	0.6622	0.7369	0.8711
	GExp(0.5; 1)	0.6779	0.7394	0.8748
Nonmonotone	Pareto(2; 1)	0.9993	0.9922	0.9970
	Weibull(2; 1)	0.7317	0.6190	0.6671
	HN(3)	0.9026	0.8701	0.8999

Table 8. Empirical power for different forms of hazard functions and different distributions considering sample size $n = 100$.

Hazard function	Alternatives	KL_{mn}	W^2	A^2
Increasing	Gamma(3; 1)	0.4937	0.4829	0.5364
	GExp(3; 1)	0.3744	0.3807	0.4276
	Beta(2; 1)	1.0000	1.0000	1.0000
Decreasing	Gamma(0.5; 1)	0.9861	0.9786	0.9969
	GExp(0.5; 1)	0.9882	0.9772	0.9957
Nonmonotone	Pareto(2; 1)	1.0000	1.0000	1.0000
	Weibull(2; 1)	0.9134	0.8922	0.9256
	HN(3)	0.9948	0.9915	0.9958

According to our simulation study, we conclude that the goodness-of-fit test based on the KL information, in general, presents greater powers when compared to standard AD and CM tests, for small sample size. When the hazard function under alternative hypothesis is decreasing, the proposed test has difficulties in discriminating the models, leading to powers close to nominal levels. This is because the hazard functions considered under the alternative hypothesis closely approximate the hazard function of the BS distribution. In addition, as the sample size increases, the power of the test also increases, as expected.

3.3 EMPIRICAL ILLUSTRATION

Next, we consider a set of data related to fatigue life cycles of samples of 6061-T6 aluminum presented in [Birnbaum and Saunders \(1969\)](#). These specimens were cut at an angle parallel to the direction of rotation, oscillating at 18 cycles per second. They were exposed to a pressure with a maximum stress of 26000 psi (pounds per square inch). The data are presented in [Table 9](#).

We want to test the null hypothesis that the sample presented in [Table 9](#) follows the BS distribution. The model parameter estimates are $\hat{\alpha} = 0.1614$ and $\hat{\beta} = 392.7622$. The value observed for the test statistic is $kl_{mn} = 0.9270$, and the critical value for this case is $KL_{mn}^*(\rho) = 0.8834$, at the 5% significance level. Therefore, we do not reject the hypothesis that the data follow the BS distribution. [Figure 1](#) compares the empirical distribution function with the theoretical one. We can observe from this figure that the empirical and theoretical distribution functions are very close, which reinforces the conclusion reached by the test.

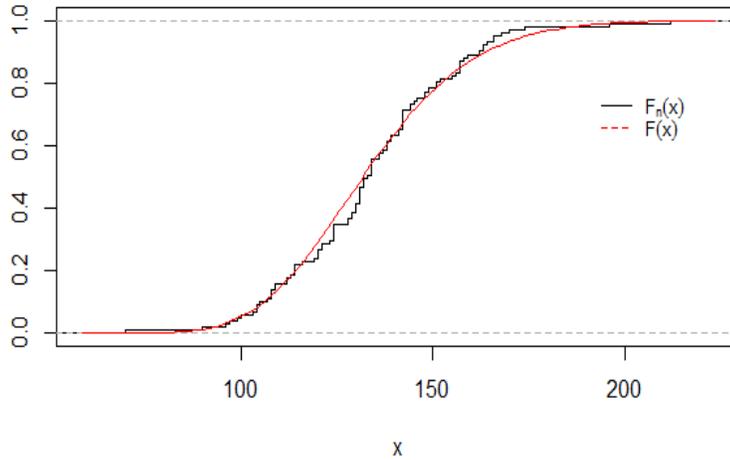


Figure 1. Empirical and theoretical distribution functions BS for aluminum data.

Table 9. Data set of aluminum lifetimes (26.000 psi).

233	258	268	276	290	310	312	315	318	321
321	329	335	336	338	338	342	342	342	344
349	350	350	351	351	352	352	356	358	358
360	362	363	366	367	370	370	372	372	374
375	376	379	379	380	382	389	389	395	396
400	400	400	403	404	406	408	408	410	412
414	416	416	416	420	422	423	426	428	432
432	433	433	437	438	439	439	443	445	445
452	456	456	460	464	466	468	470	470	473
474	476	476	486	488	489	490	491	503	517
540	560								

4. CONCLUSIONS AND FUTURE RESEARCH

In this paper, we proposed a goodness-of-fit test for the Birnbaum-Saunders distribution based on the Kullback-Leibler information. The proposed goodness-of-fit test performed better than the standard Anderson-Darling and Cramér-von Mises tests, in the sense that the proposed test had greater power for the alternatives considered with increasing and nonmonotone hazard functions. When the distribution of the alternative hypothesis had a decreasing hazard function, the test based in KL information presented less power than the Anderson-Darling and Cramér-von Mises tests. In general, the proposed test proved to be a good alternative to the standard Anderson-Darling and Cramér-von Mises tests. As future research, we hope to obtain new tests for the Birnbaum-Saunders distribution based on information measures for censored data, more specifically, for type II and progressively Type-II censored samples.

ACKNOWLEDGMENT

The authors wish to thank the Editors and anonymous referees for their constructive comments on an earlier version of this manuscript, which resulted in this improved version. The study was supported partially by CNPq and CAPES.

REFERENCES

- Arizono, I. and Ohta, H., 1989. A test for normality based on Kullback-Leibler information. *The American Statistician*, 43, 20-22.
- Athayde, E., Azevedo, A., Barros, M., and Leiva, V., 2019. Failure rate of Birnbaum-Saunders distributions: shape, change-point, estimation and robustness. *Brazilian Journal of Probability and Statistics*, 33, 301-328.
- Azevedo, C., Leiva, V., Athayde, E., and Balakrishnan, N., 2012. Shape and change point analyses of the Birnbaum-Saunders-t hazard rate and associated estimation. *Computational Statistics and Data Analysis*, 56, 3887-3897.
- Barros, M., Paula, G.A., and Leiva, V., 2008. A new class of survival regression models with heavy-tailed errors: robustness and diagnostics. *Lifetime Data Analysis*, 14, 316-332.
- Barros, M., Leiva, V., Ospina, R., and Tsuyuguchi, A., 2014. Goodness-of-fit tests for the Birnbaum-Saunders distribution with censored reliability data. *IEEE Transactions on Reliability*, 63, 543-554.
- Birnbaum, Z.W. and Saunders, S.C., 1969. Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability*, 6, 328-347.
- Castro-Kuriss, C., 2011. On a goodness-of-fit test for censored data from a location-scale distribution with application. *Chilean Journal of Statistics*, 37, 15-136.
- Chen, G. and Balakrishnan, N., 1995. A general purpose approximate goodness-of-fit test. *Journal of Quality Technology*, 27, 154-161.
- Clausius, R., 1867. *The mechanical theory of heat: with its applications to the steam-engine and to the physical properties of bodies*. J. van Voorst.
- Cox, D., Hinkley, D., 1974. *Theoretical Statistics*. Chapman and Hall, London, UK.
- D'Agostino, C. and Stephens, M.A., 1986. *Goodness of Fit Techniques*. Marcel Dekker, New York.
- Desousa, M., Saulo, H., Leiva, V., and Scalco, P., 2018. On a tobit-Birnbaum-Saunders model with an application to medical data. *Journal of Applied Statistics*, 45, 932-955.
- Ferreira, M.S., 2013. A study of exponential-type tails applied to Birnbaum-Saunders models. *Chilean Journal of Statistics*, 4, 87-97.
- Gomes, M.I., Ferreira, M., and Leiva, V., 2012. The extreme value Birnbaum-Saunders model and its moments and application in biometry. *Biometrical Letters*, 49, 81-94.
- Kullback, S. and Leibler, R.A., 1951. On information and sufficiency. *The annals of mathematical statistics*, 22, 79-86.
- Leao, J., Leiva, V., Saulo, H., and Tomazella, V., 2018. Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data. *Statistics in Medicine*, 37, 4421-4440.
- Leiva, V., Barros, M., Paula, G.A., and Galea, M., 2007. Influence diagnostics in log Birnbaum-Saunders regression models with censored data. *Computational Statistics and Data Analysis*, 51, 5694-5707.
- Leiva, V., Barros, M., Paula, G.A., and Sanhueza, A., 2008. Generalized Birnbaum-Saunders distributions applied to air pollutant concentration. *Environmetrics*, 19, 235-249.
- Leiva, V., Sanhueza, A., and Angulo, J.M., 2009. A length-biased version of the Birnbaum-Saunders distribution with application in water quality. *Stochastic Environmental Research and Risk Assessment*, 23, 299-307.
- Leiva, V., Vilca, F., Balakrishnan, N., and Sanhueza, A., 2010. A skewed sinhnormal distribution and its properties and application to air pollution. *Communications in Statistics*, 39, 426-443.
- Leiva, V. and Saunders, S.C., 2015. Cumulative damage models. *Wiley StatsRef: Statistics Reference Online* 1-10.

- Leiva, V., Marchant, C., Ruggeri, F., and Saulo, H., 2015a. A criterion for environmental assessment using Birnbaum-Saunders attribute control charts. *Environmetrics*, 26, 463-476.
- Leiva, V., Tejo, M., Guiraud, P., Schmachtenberg, O., Orio, P., and Marmolejo F., 2015b. Modeling neural activity with cumulative damage distributions. *Biological Cybernetics*, 109, 421-433.
- Leiva, V., 2016. *The Birnbaum-Saunders Distribution*. Academic Press, New York.
- Lillo, C., Leiva, V., Nicolis, O., and Aykroyd, R.G., 2018. L-moments of the Birnbaum-Saunders distribution and its extreme value version: Estimation, goodness of fit and application to earthquake data. *Journal of Applied Statistics*, 45, 187-209.
- Marchant, C., Leiva, V., Cysneiros, F.J.A., and Liu, S., 2018. Robust multivariate control charts based on Birnbaum-Saunders distributions. *Journal of Statistical Computation and Simulation*, 88, 182-202.
- Marchant, C., Leiva, V., Christakos, G., and Cavieres, M.A., 2019. Monitoring urban environmental pollution by bivariate control charts: new methodology and case study in Santiago, Chile. *Environmetrics*, pages in press, available at <https://doi.org/10.1002/env.2551>.
- Mazucheli, J., Menezes, A.F.B., and Dey, S., 2018. The unit-Birnbaum-Saunders distribution with applications. *Chilean Journal of Statistics*, 9, 47-57.
- Park, S., 2005. Testing exponentiality based on the Kullback-Leibler information with the type II censored data. *IEEE Transactions on Reliability*, 54, 22-26.
- Rad, A.H., Yousefzadeh F., and Balakrishnan N., 2011. Goodness-of-fit test based on Kullback-Leibler information for progressively type-II censored data. *IEEE Transactions on Reliability*, 60, 570-579.
- Shannon, C.E., 1948. A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423.
- Vasicek, O., 1976. A test for normality based on sample entropy. *Journal of the Royal Statistical Society B*, 38, 54-59.
- Vilca, F., Sanhueza, A., Leiva, V., and Christakos, G., 2010. An extended Birnbaum-Saunders model and its application in the study of environmental quality in Santiago, Chile. *Stochastic Environmental Research and Risk Assessment*, 24, 771-782.
- Vilca, F., Santana, L., Leiva, V., and Balakrishnan, N., 2011. Estimation of extreme percentiles in Birnbaum-Saunders distributions. *Computational Statistics and Data Analysis*, 55, 1665-1678.

DISTRIBUTION THEORY
RESEARCH PAPER

A new Birnbaum-Saunders type distribution based on the skew-normal model under a centered parameterization

NATHALIA L. CHAVES¹, CAIO L. N. AZEVEDO^{1,*}, FILIDOR VILCA-LABRA¹,
and JUVÊNCIO S. NOBRE²

¹Department of Statistics, Universidade Estadual de Campinas, Brazil

²Department of Statistics and Applied Mathematics, Universidade Federal de Ceará, Brazil

(Received: 22 February 2019 · Accepted in final form: 23 April 2019)

Abstract

In this paper, we introduce a new distribution for positively skewed data by combining the Birnbaum-Saunders and centered skew-normal distributions. Several of its properties are developed. Our model accommodates both positively and negatively skewed data. Also, we show that our proposal circumvents some problems related to another Birnbaum-Saunders distribution based on the usual skew-normal model, previously presented in the literature. We derive both maximum likelihood and Bayesian inference, comparing them through a suitable simulation study. The convergence of the expectation conditional maximization (for maximum likelihood inference) and MCMC algorithms (for Bayesian inference) are verified and several factors of interest are compared. In general, as the sample size increases, the results indicate that the Bayesian approach provided the most accurate estimates. Our model accommodates the asymmetry of the data more properly than the usual Birnbaum-Saunders distribution, which is illustrated through real data analysis.

Keywords: Bayesian inference · Birnbaum-Saunders distribution · ECM algorithm · Frequentist inference · MCMC algorithms · R software

Mathematics Subject Classification: Primary 60E05 · Secondary 62F15.

1. INTRODUCTION

The Birnbaum-Saunders (BS) distribution is characterized by two parameters and defined in terms of the standard normal distribution. The BS distribution has been received considerable attention over the past few years, since it has been used quite effectively to model positively skewed data, especially lifetime and crack growth data. Since the pioneering work of [Birnbaum and Saunders \(1969a\)](#) was published, several extensions of the BS distribution have been proposed in the literature and its parameters estimated under both frequentist and Bayesian paradigms.

*Corresponding author. Email: cnaber@ime.unicamp.br

From a frequentist viewpoint, [Birnbaum and Saunders \(1969b\)](#) presented a discussion on the maximum likelihood (ML) parameter estimation. [Mann et al. \(1974\)](#) showed that the BS distribution is unimodal. [Engelhardt et al. \(1981\)](#) developed confidence intervals and hypothesis tests for both parameters. [Desmond \(1985\)](#) developed a BS-type distribution based on a biological model. [Desmond \(1986\)](#) investigated the relationship between the BS distribution and the inverse Gaussian distribution. [Lu and Chang \(1997\)](#) used bootstrap methods to construct prediction intervals for future observations. In the linear regression context, [Rieck and Nedelman \(1991\)](#) developed a related log-linear model and showed that it can be used for modeling accelerated life tests and to compare average lifetime of different populations.

From a Bayesian perspective, there are few works on the BS distribution. The first one is due to [Achcar \(1993\)](#) who developed Bayesian estimation using numerical approximations for the marginal posterior distributions of interest based on the Laplace approximation. Also, [Xu and Tang \(2011\)](#) presented a Bayesian study with partial information, while [Wang et al. \(2016\)](#) assumed that the two parameters follow mutually independently inverse gamma distributions. All these results were studied under a normal distribution for generating the BS distribution.

In terms of modeling, most of the generalizations of the BS distribution are based on elliptical and skew-elliptical laws, in order to obtain more robust and flexible models. Some works developed extensions based on symmetric distributions as [Diaz-Garcia and Leiva \(2005\)](#) who generalized the BS model using elliptical distributions that includes the Cauchy, Laplace, logistic, normal and Student- t distributions as particular cases. Other works are: the generalized BS distribution ([Leiva et al., 2007](#)), the Student- t BS distribution ([Barros et al., 2008](#)), and the scale-mixture of normal BS distribution ([Balakrishnan et al., 2009](#)), among others. More information can be found in [Leiva \(2016\)](#), who presented a review on the BS distribution. Other generalizations have been obtained in different ways to those aforementioned, as [Owen and Padgett \(1999\)](#), who developed a three-parameter BS distribution and the β -BS distribution presented in [Cordeiro and Lemonte \(2011\)](#). Also, [Ferreira \(2013\)](#) proposed a based BS distribution useful for modeling tail events and [Mazucheli et al. \(2018\)](#) presented a distribution on the unit interval based on the BS model. In addition, [Balakrishnan et al. \(2017\)](#) and [Maehara \(2018\)](#) provided new families of BS distribution based on the skew scale mixture of normal models. Also, extensions of the BS distribution based on the skew-elliptical models can be found in [Vilca and Leiva \(2006\)](#), [Leiva et al. \(2007, 2008\)](#) and [Vilca et al. \(2011\)](#). In these works, theoretical results were obtained, extending the properties of the BS and log-BS distributions.

A Bayesian perspective for the BS distributions based on skew-normal (SN) distribution did not receive much attention in the literature. Indeed, [Vilca et al. \(2011\)](#) considered, under a frequentist perspective, the BS distribution based on the SN model. However, even though the SN distribution has been applied with success in several fields, when the related asymmetry parameter is equals to zero, the associated Fisher information matrix is singular. Recently, to overcome this problem, [Arellano and Azzalini \(2008\)](#) and [Azzalini \(2013\)](#) explored a SN distribution under a convenient parameterization (proposed by [Azzalini \(1985\)](#) and deeper explored by [Pewsey \(2000\)](#)), the so-called centered parametrization (CP), which leads to a non-singular Fisher information matrix. Moreover, the relative profile log-likelihood function (RPLL) for the Pearson index of skewness exhibits a more regular behavior, closer to a quadratic function, and without a stationary point under null asymmetry. The resulting empirical distributions of the estimators under the CP, named CP estimators, are much closer to the normality than those obtained under the usual SN distribution, which is named direct parametrization estimators. All these desirable properties, related to the CP, may be transferred to the respective BS distribution based on the centered SN (CSN) model. It is worthwhile to mention that all the aforementioned

BS models (that consider the SN model) used the direct parametrization that is, likely, they inherit the above problems.

The main objective of this work is to propose an alternative to the skew-normal BS (SNBS) distribution proposed by [Vilca et al. \(2011\)](#), considering the CSN distribution, as the generator variable. The resulting BS-type distribution has advantages, in inference terms, over the SNBS distributions (including those obtained as particular cases of the more general families as those of [Balakrishnan et al. \(2009\)](#) and [Maehara \(2018\)](#)), similarly to those related to the CSN distribution, compared with the SN distribution. The specific objectives of this work are: to develop a BS distribution based on the CSN model, named centered skew-normal BS (CSNBS) distribution, highlighting its advantages over the SNBS distribution proposed by [Vilca et al. \(2011\)](#), and its main properties. Also, estimation procedures under both frequentist and Bayesian approaches are developed and compared, considering different scenarios. In addition, some model comparison statistics are studied. Finally, two real data sets are analyzed showing some advantages of the CSNBS model compared to the usual BS distribution.

The paper is outlined as follows. In [Section 2](#), we present our distribution and some motivation for its development. In [Section 3](#), the estimation methods are proposed and some statistics of model comparison are presented. In [Section 4](#), some simulation studies are presented and two real data sets are analyzed. Finally, in [Section 5](#), some additional comments and conclusions are provided.

2. THE CENTERED SKEW-NORMAL BS DISTRIBUTION

2.1 THE CENTERED SKEW-NORMAL DISTRIBUTION

A random variable Y is said to have a CSN distribution, denoted by $Y \sim \text{CSN}(\mu, \sigma, \gamma)$, where μ , σ and γ are the mean, the standard deviation and the Pearson coefficient of skewness, respectively, if its density is given by

$$f_Y(y) = 2 \frac{\sigma_z}{\sigma} \phi\left(\mu_z + \frac{\sigma_z}{\sigma}(y - \mu)\right) \Phi\left[\lambda\left(\mu_z + \frac{\sigma_z}{\sigma}(y - \mu)\right)\right] \frac{2}{\omega} \phi\left(\frac{y - \xi}{\omega}\right) \Phi\left[\lambda\left(\frac{y - \xi}{\omega}\right)\right], y \in \mathbb{R}, \quad (1)$$

where $\mu_z = r\delta$, $\sigma_z^2 = 1 - \mu_z^2$, $\lambda = \gamma^{1/3}s/\sqrt{r^2 + s^2\gamma^{2/3}(r^2 - 1)}$, $r = \sqrt{2/\pi}$, $\gamma = r\delta^3(4/\pi - 1)(1 - \mu_z^2)^{-3/2}$, $\delta = \lambda/\sqrt{1 + \lambda^2}$, $\xi = \mu - \sigma\gamma^{1/3}s$, $\omega = \sigma\sqrt{1 + \gamma^{2/3}s^2}$, and $s = [2/(4 - \pi)]^{1/3}$. The quantity λ is the asymmetry parameter, see [Azzalini \(1985\)](#). For $\mu = 0$ and $\sigma = 1$, we have the standard CSN distribution, denoted by $Y \sim \text{CSN}(0, 1, \gamma)$, whose density is given by

$$f_Y(y) = \frac{2}{\omega} \phi\left(\frac{y - \xi}{\omega}\right) \Phi\left[\lambda\left(\frac{y - \xi}{\omega}\right)\right], y \in \mathbb{R},$$

where $\xi = -\gamma^{1/3}s$ and $\omega = \sqrt{1 + \gamma^{2/3}s^2}$. For inferential purposes, a useful stochastic representation of Y is given by

$$Y = \frac{1}{\sigma_z} \{ \delta |X_0| + (1 - \delta^2)X_1 - \mu_z \}, \quad (2)$$

where $X_i \sim N(0, 1)$, for $i = 0, 1$, are independent and so $H = |X_0|$ follows a half-normal (HN) distribution, denoted by $\text{HN}(0, 1)$.

2.2 THE PROPOSED DISTRIBUTION

Here, we present the CSNBS distribution, which is defined similarly to the usual BS and the SNBS distributions by

$$T = \beta \left[\frac{\alpha Y}{2} + \sqrt{\left(\frac{\alpha Y}{2}\right)^2 + 1} \right]^2, \quad (3)$$

where $Y \sim \text{CSN}(0, 1, \gamma)$, α is the shape parameter, β is the location parameter, and γ is the asymmetry parameter. We use the following notation $T \sim \text{CSNBS}(\alpha, \beta, \gamma)$. The vector $(\alpha, \beta, \gamma)^\top$ is called centered parameter and based on the SN distribution, that is, $(\alpha, \beta, \lambda)^\top$ is named direct parameters. Following the same steps as in the usual BS distribution, we have that its density is given by

$$f_T(t) = 2\phi[a_{t;\mu,\sigma}(\alpha, \beta)] \Phi[\lambda a_{t;\mu,\sigma}(\alpha, \beta)] A_{t;\sigma}(\alpha, \beta), t > 0, \quad (4)$$

where $a_{t;\mu,\sigma}(\alpha, \beta) = \mu_z + \sigma_z a_t(\alpha, \beta)$, $A_{t;\sigma}(\alpha, \beta) = \sigma_z A_t(\alpha, \beta)$, $a_t(\alpha, \beta) = (\sqrt{t/\beta} - \sqrt{\beta/t})/\alpha$, $A_t(\alpha, \beta) = da_t(\alpha, \beta)/dt = t^{-3/2}(t + \beta)/(2\alpha\beta^{1/2})$, and the other quantities are previously defined. Note that for $\gamma = 0$, we have the usual BS distribution. The mean and variance of T (see Appendix A for more details) are given, respectively, by

$$\mathbb{E}(T) = \beta \left(1 + \frac{\alpha^2}{2} \right) \quad \text{and} \quad \text{Var}(T) = (\alpha\beta)^2 \left\{ 1 + \frac{\alpha^2}{4} [2\Delta - 1] \right\},$$

where $\Delta = 2(\pi - 3)(4/\pi^2)\delta^4[1 - (2\delta^2/\pi)]^{-2} + 3$.

The following theorem is very useful to develop both classical and Bayesian approaches since they lead to conditional distributions that allow us to implement, more easily, the EM algorithm, and simplify the Bayesian developments. For the use of standard MCMC software, such as WinBUGS, OpenBUGS, JAGS or Stan, see [Lunn et al. \(2000\)](#), [Lunn et al. \(2009\)](#), [Depaoli et al. \(2016\)](#) and [Carpenter et al. \(2016\)](#).

THEOREM 2.1 Let $T \sim \text{CSNBS}(\alpha, \beta, \gamma)$ as in Equation (3), and Y and H as defined in Equation (2). Then,

- (i) The conditional density of T , given $H = h$, can be expressed as

$$f_{T|H}(t|h) = \phi(\nu_h + a_t(\alpha_\delta, \beta)) A_t(\alpha_\delta, \beta),$$

where $\alpha_\delta = \alpha\sqrt{(1 - \delta^2)/(1 - r^2\delta^2)}$ and $\nu_h = -(\delta(h - r))/\sqrt{1 - \delta^2}$.

- (ii) $f_{H|T}(h|t) = \frac{\phi\left(h|\delta\sqrt{1-r^2\delta^2}\left(a_t(\alpha,\beta) + \frac{r\delta}{\sqrt{1-r^2\delta^2}}\right), 1-\delta^2\right)}{\Phi\left(\lambda\sigma_z\left(a_t(\alpha,\beta) + \frac{r\delta}{\sqrt{1-r^2\delta^2}}\right)\right)}$, $h > 0$. Moreover,

$$\mathbb{E}(H|T = t) = \eta_t + W_\Phi\left(\frac{\eta_t}{\tau}\right)\tau \quad \text{and} \quad \mathbb{E}(H^2|T = t) = \eta_t^2 + \tau^2 + W_\Phi\left(\frac{\eta_t}{\tau}\right)\eta_t\tau,$$

where $\eta_t = \delta\sqrt{1 - r^2\delta^2}(a_{t_i}(\alpha, \beta) + (r\delta)/\sqrt{1 - r^2\delta^2})$.

The density in Theorem 2.1 corresponds to the extended Birnbaum-Saunders (EBS) discussed in [Leiva et al. \(2008\)](#) and denoted by $\text{EBS}(\alpha_\delta, \beta, \sigma = 2, \nu_h)$. The proof of Theorem 2.1 is in the Appendix B.

Figures 1-3 present the density of the CSNBS distribution for different values of α , β and γ . From Figure 1, we have that for $\alpha = 0.2$ the density is concentrated around β ($\beta = 1$), and for $\alpha = 0.8$ the density is more asymmetric, with a higher variability. As α increases, the density becomes more flat, positively skewed and more dispersed, as it can be seen in Figure 2, for different values of α , fixing the other parameters. In addition, Figure 3 shows densities more concentrated around β for different values of α and β , with $\gamma = 0.9$. It is also possible to see that for large values of β , the density is more negatively skewed. Note that the distribution tends to be symmetric around β , for $\gamma = 0$ (the usual BS distribution) and/or for small values of α . Positive asymmetry is observed as α increases, β decreases and/or γ assumes positive values. In addition, negative asymmetry is observed as α decreases, β increases and/or γ assumes negative values. Another interesting point is that the CSNBS distribution may be negatively skewed, which is an unusual behavior for positive random variables. This feature makes our distribution a very useful alternative for modeling positive skewed data.

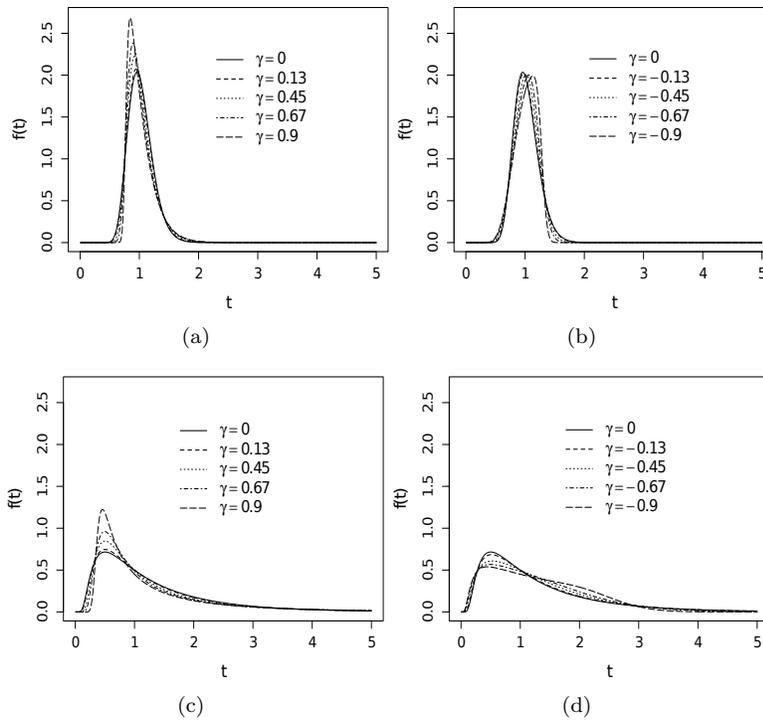


Figure 1. CSNBS density for different values of γ , with $\beta = 1$, $\alpha = 0.2$ (a)-(b) and $\alpha = 0.8$ (c)-(d).

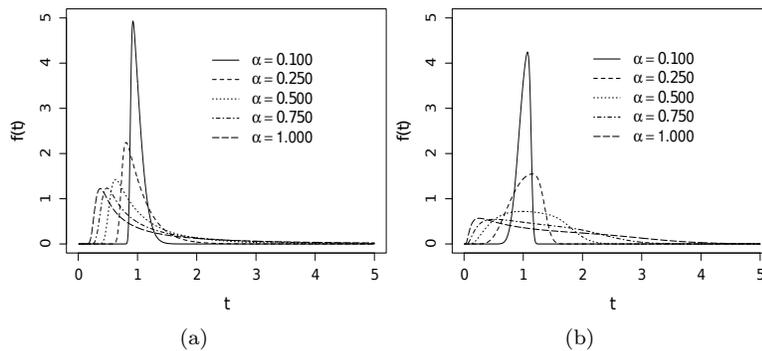


Figure 2. CSNBS density for different values of α , with $\beta = 1$, $\gamma = 0.9$ (a) and $\gamma = -0.9$ (b).

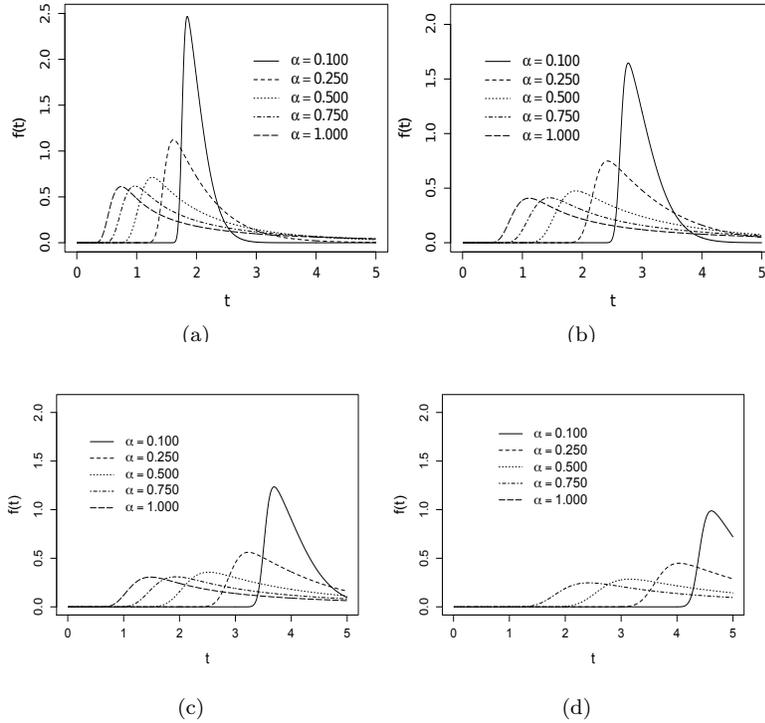


Figure 3. CSNBS density with $\beta = 2$ (a), $\beta = 3$ (b), $\beta = 4$ (c), and $\beta = 5$ (d) for indicated α and $\gamma = 0.9$.

2.3 SOME MOTIVATIONAL REMARKS ON THE PROPOSAL

- (i) It is well known that there is some difficulty in estimating the parameters of the SN distribution by the ML approach, when the asymmetry parameter is close to zero. Some problems seem to persist even if one switched to the Bayesian inference, unless a strongly informative prior is considered, as pointed out by [Arellano and Azzalini \(2008\)](#). The SNBS distribution seems to inherit such problems. Thus, the proposed CSNBS distribution can circumvents these problems, since it is based on the CSN model.
- (ii) When the asymmetry parameter is equals to zero, the Fisher information matrix is singular, even if all parameters are identifiable. This affects the behavior of the empirical distributions of the ML estimators and the Bayesian estimators. To get a direct perception of the problem, we generated 100 samples of size $n = 200$, from the SNBS distribution and for each sample, the ML and Bayesian estimates $(\hat{\alpha}, \hat{\beta}, \hat{\lambda})$ have been computed. In this case, we fix $\alpha = 0.5$, $\beta = 1$ and $\lambda = 1$, which induces a strong positively skewed behavior of the SNBS distribution. Figures 4 and 5 display the corresponding empirical distribution of $\hat{\alpha}$ and $(\hat{\alpha}, \hat{\beta})$, through a histogram and scatter plot, respectively. Moreover, 100 samples of size $n = 200$ are generated from the CSNBS distribution, and the respective ML and Bayesian estimates $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$ have been computed. In this case, we fix $\alpha = 0.5$, $\beta = 1$ and $\gamma = 0.137$, which induces a strong positively skewed behavior of the CSNBS model. The empirical distributions of $\hat{\alpha}$ and $(\hat{\alpha}, \hat{\beta})$ are shown in Figures 6 and 7, respectively. Clearly these empirical distributions are much closer to normality than those in Figures 4 and 5. In fact, it can be shown that the singularity of the expected Fisher information matrix, when the asymmetry parameter is null, no longer occurs.
- (iii) The CP circumvents the problem concerning the existence of an inflection point in the RPLL of this parameter. This can be seen in Figure 8, which refers to the plots of twice the RPLL function for λ , the asymmetry parameter of the SNBS distribution (left panel), and the for γ , the asymmetry parameter of the CSNBS distribution (right panel). The RPLL corresponds to $\ell(\hat{\alpha}(\gamma), \hat{\beta}(\gamma), \gamma) - \ell(\hat{\alpha}(\gamma), \hat{\beta}(\gamma), \hat{\gamma})$, where $\ell(\cdot)$ represents the

log-likelihood function. The respective plots are constructed by considering a random sample of both SNBS and CSBNS distributions, under suitable values of α , β and γ . We can notice a non-quadratic form of the log-likelihood function under the SNBS model, induced by the existence of an inflection point when the asymmetry parameter is very close to zero, making it difficult the obtaining of the ML estimates. However, the log-likelihood function of the CSNBS distribution presents a concave shape. Also, there is no inflection point when the asymmetry parameter is equals zero.

- (iv) The posterior distribution of λ for the SNBS distribution has a non-quadratic form, as it can be seen in Figure 9 (a), and this occurs even if we consider an informative prior. However, the posterior distribution of γ for the CSNBS distribution is well-behaved, presenting a concave shape, as it can be seen in Figure 9 (b).

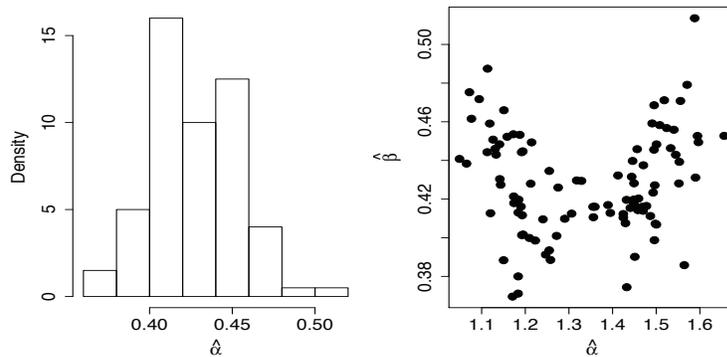


Figure 4. Estimated distributions of the ML estimates when samples of size $n = 200$ are drawn from SNBS; the left panel displays the histogram of $\hat{\alpha}$, the right panel displays the scatter plot of $(\hat{\alpha}, \hat{\beta})$.

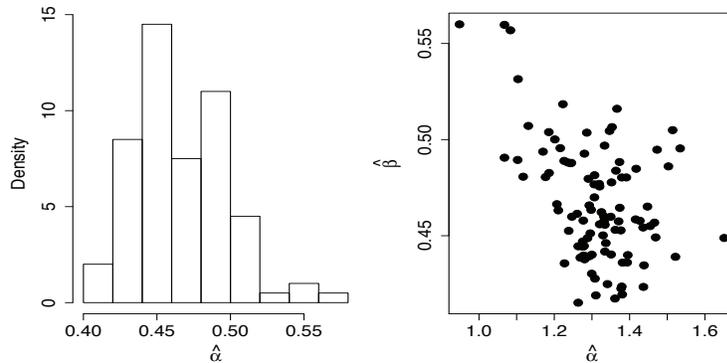


Figure 5. Estimated distributions of the Bayesian estimates when samples of size $n = 200$ are drawn from SNBS distribution; the left panel displays the histogram of $\hat{\alpha}$, the right panel displays the scatter plot of $(\hat{\alpha}, \hat{\beta})$.

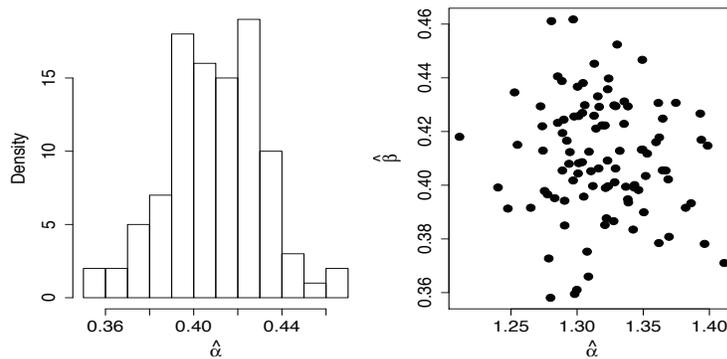


Figure 6. Estimated distributions of the ML estimates when samples of size $n = 200$ are drawn from CSNBS distribution; the left panel displays the histogram of $\hat{\alpha}$, the right panel displays the scatter plot of $(\hat{\alpha}, \hat{\beta})$.

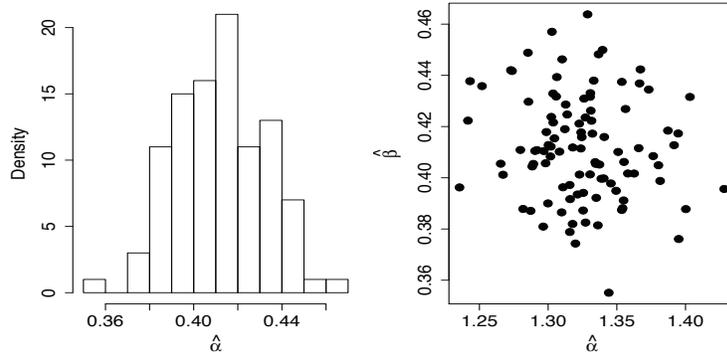


Figure 7. Estimated distributions of the Bayesian estimates when samples of size $n = 200$ are drawn from CSNBS distribution; histogram of $\hat{\alpha}$ (left) and scatter plot of $(\hat{\alpha}, \hat{\beta})$ (right).

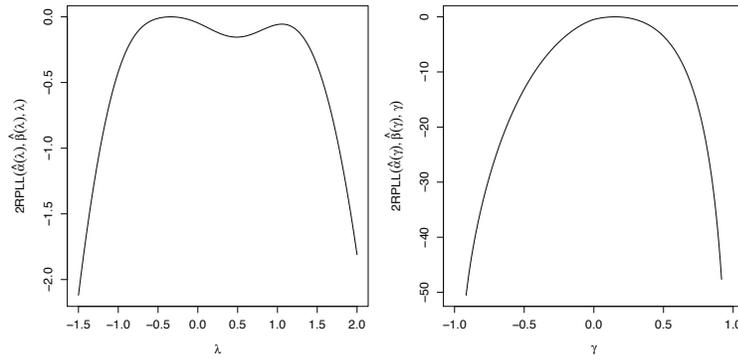


Figure 8. Twice the relative profiled log-likelihood function for the asymmetry parameter of the SNBS (left) and CSNBS (right) distributions.

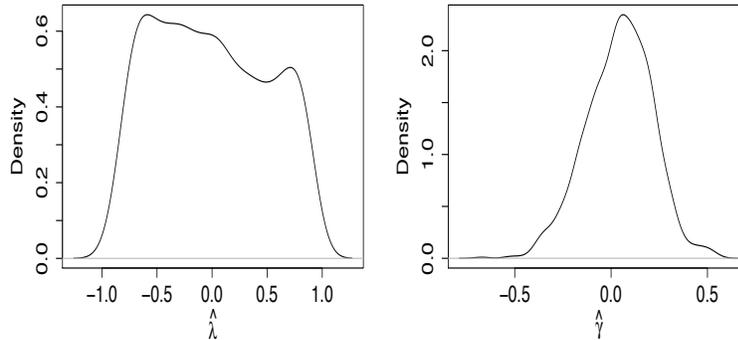


Figure 9. Posterior distribution of λ for the SNBS distribution (left) and of γ for the CSNBS distribution (right).

3. ESTIMATION AND INFERENCE

3.1 GENERAL CONTEXT

We present the ML estimation, based on the expectation conditional maximization (ECM) algorithm as in [Meng and Rubin \(1993\)](#), and the Bayesian approach, through MCMC algorithms. Let $T \sim \text{CSNBS}(\alpha, \beta, \gamma)$ and then, recall that, from [Theorem 2.1](#), we have $T|(H = h) \sim \text{EBS}(\alpha_\delta, \beta, \sigma = 2, \nu_h)$, where $H = |X_0| \sim \text{HN}(0, 1)$, $\alpha_\delta = \alpha\sqrt{(1 - \delta^2)/(1 - r^2\delta^2)}$ and $\nu_h = -\delta(h - r)/\sqrt{1 - \delta^2}$. In [Appendix B](#), we present some results that are useful for obtaining the ML estimators. For both methods, we consider a random sample T_1, \dots, T_n from $T \sim \text{SNBS}(\alpha, \beta, \gamma)$, where $\boldsymbol{\theta} = (\alpha, \beta, \gamma)^\top$.

3.2 THE ECM ALGORITHM AND ML ESTIMATION

Here, we discuss the ML estimation through the ECM algorithm. The log-likelihood function for $\boldsymbol{\theta}$ is given by $\ell(\boldsymbol{\theta}|\mathbf{t}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}|t_i)$, where

$$\ell_i(\boldsymbol{\theta}|t_i) = \log(2) + \log \{ \phi [a_{t_i;\mu,\sigma}(\alpha, \beta)] \} + \log \{ \Phi [\lambda a_{t_i;\mu,\sigma}(\alpha, \beta)] \} + \log [A_{t_i;\sigma}(\alpha, \beta)], \quad (5)$$

and $a_{t_i;\mu,\sigma}(\alpha, \beta)$ and $A_{t_i;\sigma}(\alpha, \beta)$ are given in Equation (4). Instead of considering the direct maximization of Equation (5), we obtain the ML estimates through the ECM algorithm, since it allows for a more tractable optimization process. In this case, we need to work with the so-called augmented likelihood function. Also, instead of working with $\boldsymbol{\theta}^* = (\alpha, \beta, \gamma)^\top$, we estimate $\boldsymbol{\theta} = (\alpha, \beta, \delta)^\top$, where δ is defined in Equation (1). Then, we recover γ through the invariance principle related to the ML estimators. This is performed since the related expressions (both analytically and computationally) are more tractable for $\boldsymbol{\theta}$.

Recall that, From Theorem 2.1, we have $T_i|H_i = h_i \stackrel{\text{IND}}{\sim} \text{EBS}(\alpha_\delta, \beta, \sigma = 2, \nu_{h_i})$ and $H_i \stackrel{\text{IND}}{\sim} \text{HN}(0, 1); i = 1, \dots, n$, where ‘‘IND’’ denotes ‘‘independent’’, $\alpha_\delta = \alpha\sqrt{(1 - \delta^2)/(1 - r^2\delta^2)}$ and $\nu_{h_i} = -(\delta(h_i - r))\sqrt{1 - \delta^2}$. Then, defining $\mathbf{t}_c = (\mathbf{t}^\top, \mathbf{h}^\top)^\top$, with $\mathbf{t} = (t_1, \dots, t_n)^\top$ and $\mathbf{h} = (h_1, \dots, h_n)^\top$, the augmented log-likelihood function can be written as

$$\begin{aligned} \ell(\boldsymbol{\theta}|\mathbf{t}_c) &= \sum_{i=1}^n \log[f_{T|H}(t_i|h_i)] + \sum_{i=1}^n f_H(h_i) \\ &= c - \frac{\delta^2}{2(1 - \delta^2)} \sum_{i=1}^n h_i^2 + \frac{r\delta^2}{(1 - \delta^2)} \sum_{i=1}^n h_i - \frac{nr^2\delta^2}{2(1 - \delta^2)} \\ &\quad + \frac{\delta\sqrt{1 - r^2\delta^2}}{1 - \delta^2} \sum_{i=1}^n h_i a_{t_i}(\alpha, \beta) - \frac{r\delta\sqrt{1 - r^2\delta^2}}{1 - \delta^2} \sum_{i=1}^n a_{t_i}(\alpha, \beta) - \frac{1 - r^2\delta^2}{2(1 - \delta^2)} \sum_{i=1}^n a_{t_i}^2(\alpha, \beta) \\ &\quad + \frac{n}{2} \log(1 - r^2\delta^2) + \sum_{i=1}^n \log(t_i + \beta) - \frac{n}{2} \log(1 - \delta^2) - n \log(\alpha) - \frac{n}{2} \log(\beta). \end{aligned}$$

For a current value of $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}$, the E-step requires the evaluation of $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \text{E}[\ell(\boldsymbol{\theta}|\mathbf{t}_c)|\mathbf{t}, \hat{\boldsymbol{\theta}}]$, where the expectation is taken with respect to the conditional distribution $H|(T = t)$, evaluated at $\hat{\boldsymbol{\theta}}$. For a estimate of $\boldsymbol{\theta}$ at r -th iteration, say $\boldsymbol{\theta}^{(r)} = (\alpha^{(r)}, \beta^{(r)}, \delta^{(r)})^\top$, consider $\hat{h}_i = \text{E}[H_i|\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}, t_i]$ and $\hat{h}_i^2 = \text{E}[H_i^2|\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}, t_i]$, given in Theorem 2.1, that is,

$$\hat{h}_i = \hat{\eta}_{t_i} + W_\Phi \left(\frac{\hat{\eta}_{t_i}}{\hat{\tau}} \right) \hat{\tau} \quad \text{and} \quad \hat{h}_i^2 = \hat{\eta}_{t_i}^2 + \hat{\tau}^2 + W_\Phi \left(\frac{\hat{\eta}_{t_i}}{\hat{\tau}} \right) (\hat{\eta}_{t_i} \hat{\tau}), \quad (6)$$

respectively, where $\hat{\eta}_{t_i} = \hat{\delta}\sqrt{1 - r^2\hat{\delta}^2} (a_{t_i}(\hat{\alpha}, \hat{\beta}) + r\hat{\delta}/\sqrt{1 - r^2\hat{\delta}^2})$, $\hat{\tau} = \sqrt{1 - \hat{\delta}^2}$ and $W_\Phi(z) = \phi(z)/\Phi(z)$, $z \in \mathbb{R}$. Then, let $\boldsymbol{\theta}^{(r)} = (\alpha^{(r)}, \beta^{(r)}, \delta^{(r)})^\top$ be the estimate of $\boldsymbol{\theta}$ at the k -th iteration. By considering Equation (6), we have that the augmented log-likelihood function becomes $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) = \text{E}[\ell(\boldsymbol{\theta}|\mathbf{t}_c)|\mathbf{t}, \hat{\boldsymbol{\theta}}^{(r)}]$, where

$$\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) &= c - \frac{\delta^{2(r)}}{2(1-\delta^{2(r)})} \sum_{i=1}^n \widehat{h}_i^{2(r)} + \frac{r\delta^{2(r)}}{(1-\delta^{2(r)})} \sum_{i=1}^n \widehat{h}_i^{(r)} - \frac{nr^2\delta^{2(r)}}{2\delta^{2(r)}} \\
&+ \frac{\delta^{(r)}\sqrt{1-r^2\delta^{2(r)}}}{\alpha^{(r)}(\delta^{2(r)})} \sum_{i=1}^n \widehat{h}_i^{(r)} a_{t_i}(1, \beta^{(r)}) - \frac{r\delta^{(r)}\sqrt{1-r^2\delta^{2(r)}}}{\alpha^{(r)}(1-\delta^{2(r)})} \\
&\times \sum_{i=1}^n a_{t_i}(1, \beta^{(r)}) - \frac{1-r^2\delta^{2(r)}}{2\alpha^{2(r)}(1-\delta^{2(r)})} \sum_{i=1}^n [a_{t_i}(1, \beta^{(r)})]^2 + \frac{n}{2} \log(1-\delta^{2(r)}) \\
&+ \sum_{i=1}^n \log(t_i + \beta^{(r)}) - \frac{n}{2} \log(1-\delta^{2(r)}) - n \log(\alpha^{(r)}) - \frac{n}{2} \log(\beta^{(r)}).
\end{aligned}$$

Hence, the ECM algorithm corresponds to iterate the following steps:

E-step: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(r)}$, compute \widehat{h}_i and \widehat{h}_i^2 , for $i = 1, \dots, n$ by using Equation (6);

CM-step 1: Fix $\beta = \widehat{\beta}^{(r)}$ and $\delta = \widehat{\delta}^{(r)}$ and update $\widehat{\alpha}^{(r)}$ through the positive root of $\widehat{\alpha}^2 + \widehat{b}^{(r)}\widehat{\alpha} + \widehat{c}^{(r)} = 0$, where

$$\begin{aligned}
\widehat{b}^{(r)} &= \frac{1}{(1-\widehat{\delta}^{2(r)})} \left[\widehat{\delta}^{(r)}\sqrt{1-r^2\widehat{\delta}^{2(r)}} \frac{1}{n} \sum_{i=1}^n \widehat{h}_i a_{t_i}(1, \widehat{\beta}^{(r)}) - r\widehat{\delta}^{(r)}\sqrt{1-r^2\widehat{\delta}^{2(r)}} \frac{1}{n} \sum_{i=1}^n a_{t_i}(1, \widehat{\beta}^{(r)}) \right], \\
\widehat{c}^{(r)} &= -\frac{(1-r^2\widehat{\delta}^{2(r)})}{(1-\widehat{\delta}^{2(r)})} \frac{1}{n} \sum_{i=1}^n \widehat{h}_i [a_{t_i}(1, \widehat{\beta}^{(r)})]^2,
\end{aligned}$$

that is, $\widehat{\alpha}^{(r+1)} = (-b(r+1) + \sqrt{b^2(r+1) - 4c(r+1)})/2$;

CM-step 2: Fix $\alpha = \widehat{\alpha}^{(r+1)}$ and update $\widehat{\beta}^{(r)}$ and $\widehat{\delta}^{(r)}$ using

$$\widehat{\beta}^{(r+1)} = \underset{\beta}{\operatorname{argmax}} Q\left(\widehat{\alpha}^{(r+1)}, \beta, \widehat{\delta}^{(r)}\right) \quad \text{and} \quad \widehat{\delta}^{(r+1)} = \underset{\delta}{\operatorname{argmax}} Q\left(\widehat{\alpha}^{(r+1)}, \widehat{\beta}^{(r+1)}, \delta\right).$$

The updating of $\widehat{\beta}^{(r+1)}$ and $\widehat{\delta}^{(r+1)}$ needs to be done through some numerical optimization method. In this work we use the function `optim`, available on the R software (see [R Development Core Team, 2017](#)), considering the L-BFGS-B optimization algorithm (see [Byrd et al., 1995](#)). Also, we start the ECM algorithm with initial values, say, $\widehat{\alpha}^{(0)}$, $\widehat{\beta}^{(0)}$ and $\widehat{\delta}^{(0)}$, using: $\widehat{\alpha}^{(0)} = [2(s/v) - 1]^{1/2}$ and $\widehat{\beta}^{(0)} = (sv)^{1/2}$, where $s = (1/n) \sum_{i=1}^n t_i$ and $v = [(1/n) \sum_{i=1}^n 1/t_i]^{-1}$, as in [Vilca et al. \(2011\)](#). After obtaining $\widehat{\alpha}^{(0)}$ and $\widehat{\beta}^{(0)}$, we can define $z_i = (1/\widehat{\alpha}^{(0)})[(t_i/\widehat{\beta}^{(0)})^{1/2} - (\widehat{\beta}^{(0)}/t_i)^{1/2}]$, for $i = 1, \dots, n$, which are observations related to a CSN distribution. Thus, $\widehat{\delta}^{(0)}$ can be obtained by maximizing (numerically) the log-likelihood function of a SN distribution with respect to δ , which is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \left[\log(2) + \log(\sigma_z) + \log[\phi(\mu_z + \sigma_z y_i)] + \log\{\Phi[\lambda(\mu_z + \sigma_z y_i)]\} \right].$$

According to [Vilca et al. \(2011\)](#), for ensuring that the true ML estimates are obtained, it is recommended to run the ECM algorithm using a range of different starting values and checking whether all of them result in similar estimates. The steps of the

ECM algorithm are repeated until a suitable convergence is attained, for example, using $\|\boldsymbol{\theta}^{(r)} - \boldsymbol{\theta}^{(r-1)}\| < \varepsilon$, with $\varepsilon > 0$. It is worthwhile to mention, under certain regularity conditions, that $\widehat{\boldsymbol{\theta}}$ converges in distribution to $N_3(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}})$, as $n \rightarrow \infty$. We approximate $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}$ by $I^{-1}(\boldsymbol{\theta})$, where $I(\boldsymbol{\theta}) = -\ddot{\ell}$, $\ddot{\ell} = [\ddot{\ell}_{\theta_1\theta_2}]$, $\theta_1, \theta_2 = \alpha, \beta$ or γ is the Hessian matrix, and $\ddot{\ell}_{\theta_1\theta_2} = \ddot{\ell}_{\theta_2\theta_1} = \partial^2 \ell(\boldsymbol{\theta}) / \partial \theta_1 \partial \theta_2 = \sum_{i=1}^n \partial^2 \ell_i(\boldsymbol{\theta}) / \partial \theta_1 \partial \theta_2$. The second derivatives of $\ell_i(\boldsymbol{\theta})$ are provided in Appendix C. The approximate standard errors (SE) of $\widehat{\boldsymbol{\theta}}$ can be estimated with the square roots of the diagonal elements of $I^{-1}(\boldsymbol{\theta})$, replacing $\boldsymbol{\theta}$ by $\widehat{\boldsymbol{\theta}}$.

3.3 BAYESIAN INFERENCE

Next, we present the developments related to the Bayesian inference through MCMC algorithms. We present the prior and the respective posterior distributions, along with suitable MCMC algorithms to sample from the respective marginal posterior distributions of interest. Consider both original and augmented likelihood functions (in order to compare them). The first of them is given by

$$L(\boldsymbol{\theta}|\mathbf{t}) = \prod_{i=1}^n 2\phi [a_{t_i;\mu,\sigma}(\alpha, \beta)] \Phi [\lambda a_{t_i;\mu,\sigma}(\alpha, \beta)] A_{t_i;\sigma}(\alpha, \beta).$$

We assume the following prior distributions: $\alpha \sim \text{gamma}(r_\alpha; \lambda_\alpha)$, $\beta \sim \text{gamma}(r_\beta; \lambda_\beta)$ and $\gamma \sim U(a; b)$, mutually independent, where $\text{gamma}(r, \lambda)$ stands for a gamma distribution such that $E(\alpha) = r/\lambda$ and $\text{Var}(\alpha) = r/\lambda^2$ and $U(a; b)$ stands for a continuous uniform distribution over the interval $[a, b]$. Combining the likelihood function with the prior distribution, we have that the joint posterior distribution is given by

$$\pi(\boldsymbol{\theta}|\mathbf{t}) \propto \alpha^{r_\alpha-1} \beta^{r_\beta-1} \exp[-(\alpha\lambda_\alpha + \beta\lambda_\beta)] \prod_{i=1}^n \phi [a_{t_i;\mu,\sigma}(\alpha, \beta)] \Phi [\lambda a_{t_i;\mu,\sigma}(\alpha, \beta)] A_{t_i;\sigma}(\alpha, \beta),$$

and the respective full conditional distributions, given by

$$\pi(\alpha|\beta, \gamma, \mathbf{t}) \propto \alpha^{r_\alpha-1} \exp(-\alpha\lambda_\alpha) \prod_{i=1}^n \phi [a_{t_i;\mu,\sigma}(\alpha, \beta)] \Phi [\lambda a_{t_i;\mu,\sigma}(\alpha, \beta)] A_{t_i;\sigma}(\alpha, \beta),$$

$$\pi(\beta|\alpha, \gamma, \mathbf{t}) \propto \beta^{r_\beta-1} \exp(-\beta\lambda_\beta) \prod_{i=1}^n \phi [a_{t_i;\mu,\sigma}(\alpha, \beta)] \Phi [\lambda a_{t_i;\mu,\sigma}(\alpha, \beta)] A_{t_i;\sigma}(\alpha, \beta),$$

$$\pi(\gamma|\alpha, \beta, \mathbf{t}) \propto \prod_{i=1}^n \phi [a_{t_i;\mu,\sigma}(\alpha, \beta)] \Phi [\lambda a_{t_i;\mu,\sigma}(\alpha, \beta)] A_{t_i;\sigma}(\alpha, \beta).$$

In addition, the augmented likelihood function is given by

$$L(\boldsymbol{\theta}|\mathbf{t}_c) = \prod_{i=1}^n \sqrt{2/\pi} \phi [\nu_{h_i} + a_{t_i}(\alpha, \beta)] A_{t_i}(\alpha, \beta) \exp\left(-\frac{h_i^2}{2}\right).$$

Similarly, combining the augmented likelihood function with the above prior distribution, we have that the posterior distribution is expressed as

$$\pi(\boldsymbol{\theta}, \mathbf{h}|\mathbf{t}) \propto \alpha^{r_\alpha-1} \beta^{r_\beta-1} \prod_{i=1}^n \phi [a_{t_i,h_i}(\alpha, \beta)] A_{t_i}(\alpha, \beta) \exp\left[-\frac{1}{2} (h_i^2 + 2\alpha\lambda_\alpha + 2\beta\lambda_\beta)\right]$$

and the respective full conditional distributions are given by

$$\begin{aligned}\pi(\mathbf{h}|\alpha, \beta, \gamma, \mathbf{t}_c) &\propto \prod_{i=1}^n \phi[a_{t_i, h_i}(\alpha, \beta)] A_{t_i}(\alpha, \beta) \exp\left(-\frac{h_i^2}{2}\right), \\ \pi(\alpha|\beta, \gamma, \mathbf{t}_c) &\propto \alpha^{r_\alpha-1} \prod_{i=1}^n \phi[a_{t_i, h_i}(\alpha, \beta)] A_{t_i}(\alpha, \beta) \exp\left[-\frac{1}{2}(h_i^2 + 2\alpha\lambda_\alpha)\right], \\ \pi(\beta|\alpha, \gamma, \mathbf{t}_c) &\propto \beta^{r_\beta-1} \prod_{i=1}^n \phi[a_{t_i, h_i}(\alpha, \beta)] A_{t_i}(\alpha, \beta) \exp\left[-\frac{1}{2}(h_i^2 + 2\beta\lambda_\beta)\right], \\ \pi(\gamma|\alpha, \beta, \mathbf{t}_c) &\propto \prod_{i=1}^n \phi[a_{t_i, h_i}(\alpha, \beta)] A_{t_i}(\alpha, \beta) \exp\left(-\frac{h_i^2}{2}\right),\end{aligned}$$

where $a_{t_i, h_i}(\alpha, \beta) = \nu_{h_i} + a_{t_i}(\alpha, \beta)$. We can see that both posterior distributions are not analytically tractable. Therefore, some numerical method must be employed to obtain suitable numerical approximations for the respective marginal posterior distributions. The above full conditional distributions do not correspond to known distributions, but they can be simulated through some auxiliary algorithm such as the Metropolis-Hastings, slice sampling or adaptive rejection. All these algorithms can be easily implemented in the R program. In addition, which is the approach pursued here, we can use a general MCMC computational framework, such `OpenBUGS`, see [Lunn et al. \(2009\)](#). In this case, it is necessary to provide the original or the augmented likelihood function, along with the prior distributions, such that the full conditional distributions are simulated through suitable algorithms, following a pre-defined hierarchy available on the `OpenBUGS`. We made all simulations using the R package `R2OpenBUGS`.

4. NUMERICAL ASPECTS

4.1 SIMULATION STUDY I

A simulation study is conducted to assess the behavior of the ECM algorithm, in terms of parameter recovery, and the accuracy of the corresponding SEs, calculated through the observed Fisher information matrix. For that, $N = 1,000$ replications are generated considering $n = 500$ and $\boldsymbol{\theta}^\top = (\alpha, \beta, \gamma) = (0.5, 1.0, 0.67)$, which induces a strong positively skewed behavior of the SNBS distribution. In [Table 1](#) we can see the mean of the estimates ($\hat{\boldsymbol{\theta}}$), the mean of the theoretical (asymptotic) SE ($\text{SE}(\hat{\boldsymbol{\theta}})$) and the empirical SE (SE_{emp}). We can notice that the parameters are well recovered and that the empirical SE are close to the theoretical ones, which indicates that the use of the observed Fisher information matrix, to obtain the corresponding SE, is appropriate.

Table 1. Results of the simulation study I.

	$\hat{\boldsymbol{\theta}}$	$\text{SE}(\hat{\boldsymbol{\theta}})$	SE_{emp}
$\hat{\alpha}$	0.495	0.019	0.021
$\hat{\beta}$	1.003	0.032	0.028
$\hat{\gamma}$	0.667	0.015	0.012

4.2 SIMULATION STUDY II

We consider a total of 30 scenarios, resulting from the combination of the levels of three different sample sizes (n) (10, 50, 200), under $\alpha \in (0.5; 1.5)$, $\beta = 1$ and $\gamma \in (-0.67; -0.45; 0; 0.45; 0.67)$. The sample sizes are chosen in order to verify the properties of the estimators, as consistency, and to compare their behavior, in terms of accuracy. The values of α and β are chosen in order to induce different shapes and small variability, whereas the values of γ induce from null to high positive/negative asymmetry. We calculated the usual statistics to measure the accuracy of the estimates: bias, variance (Var), root mean squared error (RMSE) and absolute value of the relative bias (AVRB). Let θ be the parameter of interest, $\hat{\theta}_r$ be some estimate related to the replica r and $\bar{\hat{\theta}} = (1/R) \sum_{r=1}^R \hat{\theta}_r$. The adopted statistics are: Bias = $\bar{\hat{\theta}} - \theta$, Variance = $(1/R) \sum_{r=1}^R (\hat{\theta}_r - \bar{\hat{\theta}})^2$, RMSE = $((1/R) \sum_{r=1}^R (\hat{\theta}_r - \bar{\hat{\theta}})^2)^{1/2}$, AVR B = $|\bar{\hat{\theta}} - \theta|/|\theta|$.

The usual tools for monitoring the convergence of the MCMC algorithms, see [Gaman and Lopes \(2006\)](#), indicate that a burn-in of 4,000, a thin of 100, simulating a total of 100,000 values, are enough to produce valid MCMC samples of size 1,000 for each parameter. Since the other results are similar (they are omitted here but they are available under request from the authors), we present only those related to the scenario where $\alpha = 0.5$, $\beta = 1$, $\gamma = -0.67$, varying the sample size. We used (< 0.001) to represent positive values (statistics and/or estimates) and (> -0.001) to denote negative values, when they are close to zero. In addition, we refer the Bayesian estimates as “augmented”, when the augmented likelihood function is used, and “original”, whenever the original likelihood function is considered. The selected results can be seen in [Table 2](#). In general, we can see that, as the sample size increases, the estimates obtained by the three approaches tend to the correspondent the respective true values. When $\alpha = 0.5$, the ML estimates are more accurate than the Bayesian ones, especially considering the bias and AVR B metrics. In other scenarios (not shown), when $\alpha = 1.5$, the opposite occurs for all sample sizes. Concerning β and γ , it is possible to notice that, under the smallest sample size ($n = 10$), the ML approach presents more accurate estimates than the Bayesian ones. In addition, for $n = 50$ and $n = 200$, Bayesian estimates, for both parameters, are closer to the respective true values. In conclusion, we can say that all estimators, mainly the Bayesian ones, are consistent, since both bias and RMSE tend to decrease, as the sample size increases. Furthermore, the results indicate (including those not shown here) that the Bayesian approach provided the most accurate estimates. Moreover, we can notice that the original and augmented approaches, performed quite similarly. Therefore, we decide to use the original likelihood function) approach, since it is easier to implement and faster.

4.3 REAL DATA ANALYSIS I

We analyze a data set corresponding to self-efficacy, which is available in the R software and can be accessed from the `EstCRM` package through the command `data(SelfEff)`. A group of 307 pre-service teachers, graduated from various departments in the college of education, are asked to check on a 11 cm line segment with two end points (can not do at all, highly certain can do) using their own judgment for the 10 items that measure teacher self-efficacy on different activities. We take, as response variable, the teacher self-efficacy in the creation of learning environments in which students can effectively express themselves. [Table 3](#) presents some descriptive statistics, including location measures, standard deviation (SD), coefficient of skewness (CS), and kurtosis (CK). We can notice that the distribution is strongly negatively skewed. We fit the CSNBS and BS distributions, using the Bayesian augmented and the ML method, to the data. The results obtained considering the frequentist approach are omitted here but they are available under request from

Table 2. Results of simulation study II with $\gamma = -0.67$.

Parameter	n	Method	Mean	Variance	Bias	RMSE	AVRB
α	10	Augmented	0.577	< 0.001	0.077	0.081	0.154
		Original	0.578	0.001	0.078	0.082	0.156
		ML	0.520	0.071	0.020	0.267	0.040
	50	Augmented	0.511	< 0.001	0.011	0.016	0.022
		Original	0.511	< 0.001	0.011	0.015	0.021
		ML	0.498	0.001	-0.002	0.033	0.004
	200	Augmented	0.502	< 0.001	0.002	0.005	0.004
		Original	0.502	< 0.001	0.002	0.005	0.004
		ML	0.490	< 0.001	-0.010	0.012	0.019
β	10	Augmented	1.006	< 0.001	0.006	0.023	0.006
		Original	1.004	< 0.001	0.004	0.021	0.004
		ML	1.105	0.214	0.105	0.474	0.105
	50	Augmented	0.996	< 0.001	-0.004	0.009	0.004
		Original	0.997	< 0.001	-0.003	0.009	0.003
		ML	1.039	0.018	0.039	0.140	0.039
	200	Augmented	0.999	< 0.001	-0.001	0.005	0.001
		Original	0.999	< 0.001	-0.001	0.005	0.001
		ML	0.997	< 0.001	-0.003	0.004	0.003
γ	10	Augmented	-0.157	0.067	0.513	0.575	0.766
		Original	-0.182	0.054	0.488	0.540	0.728
		ML	-0.603	0.028	0.067	0.179	0.100
	50	Augmented	-0.493	0.059	0.177	0.301	0.264
		Original	-0.505	0.049	0.165	0.276	0.247
		ML	-0.569	0.012	0.101	0.148	0.150
	200	Augmented	-0.614	0.017	0.056	0.142	0.083
		Original	-0.601	0.015	0.069	0.141	0.103
		ML	-0.523	0.002	0.147	0.153	0.220

the authors. The prior distributions are the same used in Section 3. In Table 4, in addition to the posterior expectations (PE), the posterior standard deviations (PSD) and the 95% equi-tailed credibility intervals (CI), we also present the model selection criteria. We consider the usual statistics of model comparison for both frequentist (AIC, BIC) and Bayesian (DIC, EAIC, EBIC and LPLM) see, respectively (Akaike, 1974; Schwarz, 1978; Spiegelhalter et al., 2014). The smaller values of AIC and BIC indicates the model that fits the data better. In addition, the smaller the values of DIC, EAIC, EBIC, the better the model fit, occurring the opposite with the LPML. We can notice that the estimates of α and γ (under the CSNBS model) indicate that the distribution is strongly negatively skewed. Notice also that we have indications that the asymmetry parameter is different from zero, since this value does not belong to the CI. Moreover, the criteria indicated the CSNBS model is the best. Figure 10 (left) presents the histogram of the observations and estimated densities. We can notice that the CSNBS distribution presents an advantage over the BS model. From Figure 10, we can notice that the CSNBS distribution predicts better the observations than the BS distribution. In conclusion, we can say that the CSNBS model is preferable to the BS model.

Table 3. Descriptive statistics for the teacher self-efficacy data.

Mean	Median	Minimum	Maximum	SD	Asymmetry	Kurtosis
9.205	9.700	1.650	10.900	1.365	-1.752	7.781

4.4 REAL DATA ANALYSIS II

We analyze now a data set corresponding to prices of bottles of Barolo wine and discussed in Azzalini (2013). It concerns the price (in euros) of bottles (75 cl) of Barolo wine. The data have been obtained in July 2010 from the websites of four Italian wine resellers, selecting only quotations of Barolo wine, which is produced in the Piedmont region of

Table 4. Posterior expectations (PE), posterior standard deviations (PSD), equi-tailed 95% CI and model selection criteria.

Parameter	PE	PSD	CI _{95%}
CSNBS			
α	0.154	0.002	[0.151; 0.157]
β	8.871	0.016	[8.836; 8.903]
γ	-0.971	0.003	[-0.978; -0.966]
EAIC		1,021.912	
EBIC		1,033.093	
DIC		3,047.154	
LPML		-508.531	
BS			
α	0.205	0.008	[0.190; 0.222]
β	9.016	0.105	[8.815; 9.229]
EAIC		1,252.772	
EBIC		1,260.226	
DIC		3,744.335	
LPML		-632.9564	

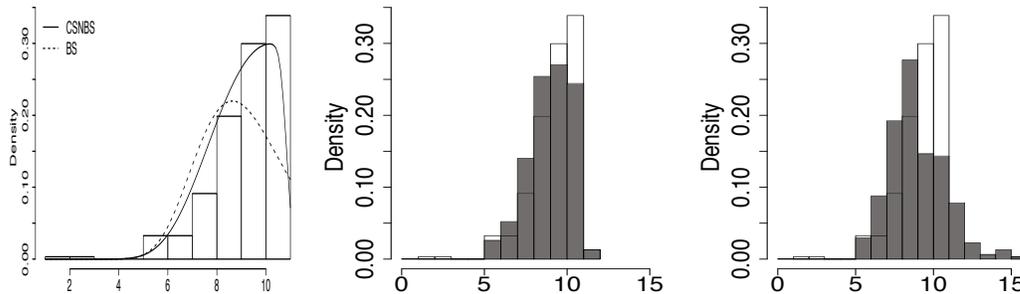


Figure 10. Histogram of the observations and estimated densities (left), histogram of the predicted and observed distributions for the CSNBS (center) and BS (right) models.

Italy. The price does not include the delivery charge. In Table 5 and Figure 11 (left), we present a descriptive analysis. It is possible to see that the distribution is positively skewed and more concentrated in the first class [0,100]. We fit the CSNBS and BS distributions, using the Bayesian augmented and the ML method, to the data. The results obtained considering the frequentist approach are omitted here but they are available under request from the authors. The prior distributions are the same used in Section 3. In Table 6, in addition to the posterior expectations (PE), the posterior standard deviations (PSD) and the 95% equi-tailed CI, we also present the Bayesian criteria. Table 6 shows that the estimates of α and γ (under the CSNBS model) indicate that the distribution of the prices is strongly positively skewed. Notice also that we have indications that the asymmetry parameter is different from zero, since this value does not belong to the CI. Moreover, the criteria indicated the CSNBS model is the best. Also, we construct QQ plots with simulated envelopes. Similar to Vilca et al. (2011), we considered the Bayesian estimates of α and β in $d(\alpha, \beta) = (1/\alpha^2)(T/\beta + \beta/T - 2)$. When $T \sim BS(\alpha, \beta)$, it is known that $d(\alpha, \beta) \sim N(0, 1)$. Since the observations $d(\hat{\alpha}, \hat{\beta})$ are expected to follow a standard normal distribution, under the well fit of the model, the envelopes are simulated from the standard

normal distribution, as described in Atkinson (1985). Similarly, if $T \sim \text{CSNBS}(\alpha, \beta, \gamma)$, thus $d(\alpha, \beta) \sim \text{CSN}(0, 1, \gamma)$. Since the observations $d(\hat{\alpha}, \hat{\beta})$ are expected to follow a CSN distribution, under the well fit of the model, the envelopes are simulated from the CSN distribution. These plots are presented in Figure 11 (lines represent the 5th percentile, the mean, and the 95th percentile of 100 simulated points). From those figures, we conclude that the CSNBS distribution provides a better fit than the BS model. Specifically, from the QQ plot shown in Figure 11 (a), we notice that the observations appear to form a slight upward-facing concave. However, the QQ plot shown in Figures 11 (b) indicate that the CSNBS distribution offers an excellent fit, provided that the majority of observations are inside of the envelope.

Table 5. Descriptive statistics for the prices of bottles of Barolo wine.

Mean	Median	Minimum	Maximum	SD	Asymmetry	Kurtosis
124.617	72	14	1000	37.041	2.903	12.982

Table 6. Posterior expectations (PE), posterior standard deviations (PSD), equi-tailed 95% CI and model selection criteria.

Parameter	PE	PSD	CI _{95%}
CSNBS			
α	0.844	0.037	[0.775; 0.917]
β	89.576	3.911	[82.260; 97.871]
γ	0.690	0.070	[0.541; 0.809]
EAIC		3,437.879	
EBIC		3,449.060	
DIC		10,292.690	
LPML		-1,718.110	
BS			
α	0.858	0.035	[0.794; 0.929]
β	92.444	4.264	[84.778; 101.302]
EAIC		3,474.893	
EBIC		3,482.346	
DIC		10,410.620	
LPML		-1,736.669	

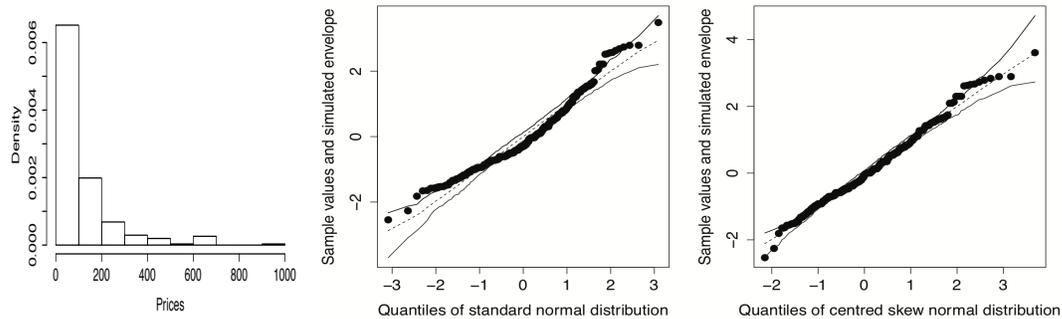


Figure 11. Histogram of the prices of bottles of Barolo wine (left), QQ plots with envelopes for BS (center) and CSNBS (right) distributions for the data of Barolo wine bottle prices.

5. CONCLUDING REMARKS

In this paper, we introduced a new distribution for modeling positive data which can present both positive and negative asymmetry, by combining the Birnbaum-Saunders and the centered skew normal distributions. We developed both maximum likelihood and Bayesian estimation procedures, comparing them through a suitable simulation study. The convergence of the conditional expectation maximization and MCMC algorithms were verified and several factors of interest were compared in the parameter recovery study. In general, as the sample size increases, the results indicated that the Bayesian approach provided the most accurate estimates. In future works we can consider the development of predictive posterior checking to detect the goodness of fit. Furthermore, we suggest the use of Jeffreys-rule prior and independence Jeffreys prior. Other auxiliary algorithms as the Hamiltonian Monte Carlo (see [Homand and Gelman, 2014](#); [Carpenter et al., 2016](#)), adaptive reject sampling and slice sampling (see [Gamerman and Lopes, 2006](#)) can be used and compared. Other family of distributions could be used instead of the centered skew normal distribution, as the scale mixture of the SN distributions, to generate new family of Birnbaum-Saunders-type distributions. Finally, other numerical methods to obtain approximation for the marginal posterior distributions, such as the INLA algorithm, can be considered (see [Rue and Martino, 2009](#)).

6. APPENDIX

APPENDIX A. MOMENTS OF THE CSNBS DISTRIBUTION

THEOREM A.1 Let $T \sim \text{CSNBS}(\alpha, \beta, \gamma)$ and $Y \sim \text{CSN}(0, 1, \gamma)$. If $E[Y^{2(r-j+i)}] < \infty$, then the moments of T are given by

$$E(T^r) = \beta^r \sum_{j=0}^r \binom{2r}{2j} \sum_{i=0}^j E[Y^{2(r-j+i)}] (\alpha/2)^{2(r-j+i)}.$$

Proof of Theorem A.1 From Equation (3), we have that

$$E\left[\left(\frac{T}{\beta}\right)^r\right] = E\left\{\left[\left(\frac{\alpha Y}{2} + \sqrt{\left(\frac{\alpha Y}{2}\right)^2 + 1}\right)^2\right]^r\right\}.$$

From the binomial theorem, that is, $(a + b)^m = \sum_{k=0}^m \binom{m}{k} a^{m-k} b^k$, we have that

$$E\left[\left(\frac{T}{\beta}\right)^r\right] = \sum_{k=0}^{2r} \binom{2r}{k} E\left\{\left[\left(\frac{\alpha Y}{2}\right)^2 + 1\right]^{k/2} \left(\frac{\alpha Y}{2}\right)^{2r-k}\right\}.$$

Considering $k = 2j$, that is, $j = k/2$, it comes that

$$E\left[\left(\frac{T}{\beta}\right)^r\right] = \sum_{j=0}^r \binom{2r}{2j} E\left\{\left[\left(\frac{\alpha Y}{2}\right)^2 + 1\right]^j \left(\frac{\alpha Y}{2}\right)^{2(r-j)}\right\}.$$

From the binomial theorem again, we have

$$\begin{aligned} \mathbb{E} \left[\left(\frac{T}{\beta} \right)^r \right] &= \sum_{j=0}^r \binom{2r}{2j} \mathbb{E} \left\{ \sum_{i=0}^j \binom{j}{i} \left(\frac{\alpha Y}{2} \right)^{2i} \left(\frac{\alpha Y}{2} \right)^{2(r-j)} \right\} \\ &= \sum_{j=0}^r \binom{2r}{2j} \sum_{i=0}^j \binom{j}{i} \mathbb{E} \left[\left(\frac{\alpha Y}{2} \right)^{2(r-j+i)} \right] \\ &= \sum_{j=0}^r \binom{2r}{2j} \sum_{i=0}^j \binom{j}{i} \mathbb{E} \left[Y^{2(r-j+i)} \right] \left(\frac{\alpha}{2} \right)^{2(r-j+i)}. \end{aligned}$$

Therefore,

$$\mathbb{E}(T^r) = \beta^r \sum_{j=0}^r \binom{2r}{2j} \sum_{i=0}^j \binom{j}{i} \mathbb{E} \left[Y^{2(r-j+i)} \right] (\alpha/2)^{2(r-j+i)}. \quad (\text{A1})$$

From Equation (A1), we get

$$\mathbb{E}(T) = \beta \sum_{j=0}^1 \binom{2}{2j} \sum_{i=0}^j \binom{j}{i} \mathbb{E} \left[Y^{2(1-j+i)} \right] (\alpha/2)^{2(1-j+i)}.$$

For $j = 0$, the first term of the sum in Equation (A1) is equal to $\beta \mathbb{E}(Y^2)(\alpha/2)^2$. For $j = 1$, the second term of the sum in Equation (A1) is equal to $\beta [1 + \mathbb{E}(Y^2)(\alpha/2)^2]$. Hence, by adding these two terms, we have

$$\mathbb{E}(T) = \beta [1 + (\alpha^2/2)].$$

Furthermore, from Equation (A1), we have

$$\mathbb{E}(T^2) = \beta^2 \sum_{j=0}^2 \binom{4}{2j} \sum_{i=0}^j \binom{j}{i} \mathbb{E} \left[Y^{2(2-j+i)} \right] (\alpha/2)^{2(2-j+i)}.$$

Developing the above sum in j , we obtain

$$\mathbb{E}(T^2) = \beta^2 \left[1 + \frac{\alpha^4}{2} \Delta + 2\alpha^2 \right].$$

Thus,

$$\begin{aligned} \text{Var}(T) &= \mathbb{E}(T^2) - [\mathbb{E}(T)]^2 \\ &= (\alpha\beta)^2 \left\{ 1 + \frac{\alpha^2}{4} [2\Delta - 1] \right\}. \end{aligned}$$

where $\Delta = \mathbb{E}(Y^4) = 2(\pi - 3)(4/\pi^2)\delta^4[1 - (2\delta^2/\pi)]^{-2} + 3$.

APPENDIX B. THE ECM ALGORITHM

The following result is used in the proof of Theorem 2.1.

Lemma 1. Let $X \sim N(\eta, \tau^2)$, thus $\forall a \in \mathbb{R}$

$$\mathbb{E}(X|X > a) = \eta + \frac{\phi\left(\frac{a-\eta}{\tau}\right)}{1 - \Phi\left(\frac{a-\eta}{\tau}\right)}\tau; \quad \mathbb{E}(X^2|X > a) = \eta^2 + \tau^2 + \frac{\phi\left(\frac{a-\eta}{\tau}\right)}{1 - \Phi\left(\frac{a-\eta}{\tau}\right)}(\eta + a)\tau.$$

Proof of Theorem 2.1

- (i) Since $Y \sim \text{CSN}(0, 1, \gamma)$, using the stochastic representation given by Equation (2), we can define

$$Y = \frac{1}{\sigma_z} \left[\delta H + \sqrt{1 - \delta^2} X_1 - \mu_z \right] = \frac{1}{\alpha} \left[\sqrt{T/\beta} - \sqrt{\beta/T} \right].$$

Therefore,

$$Y|(H = h) = \frac{1}{\alpha} \left(\sqrt{T/\beta} - \sqrt{\beta/T} \right) \Big| (H = h) \sim N(\mu_h, \sigma^2),$$

where $\mu_h = \delta(h - r)/(1 - r^2\delta^2)^{1/2}$ and $\sigma^2 = (1 - \delta^2)/(1 - r^2\delta^2)$. Then,

$$W|(H = h) = -\frac{\mu_h}{\sigma} + \frac{1}{\sigma\alpha} \left(\sqrt{T/\beta} - \sqrt{\beta/T} \right) \Big| (H = h) \sim N(0, 1)$$

and

$$T = \beta \left[\frac{\alpha}{2} (\sigma W + \mu_h) + \sqrt{\left[\frac{\alpha}{2} (\sigma W + \mu_h) \right]^2 + 1} \right].$$

From the above result, the proof is completed.

- (ii) As $f_H(h) = 2\phi(h|0, 1)$, $h > 0$ and

$$\phi(\nu_h + a_t(\alpha, \beta)) = \frac{\sqrt{1 - \delta^2}}{\sqrt{1 - r^2\delta^2}} \phi \left(a_t(\alpha, \beta) \Big| \frac{\delta(h - r)}{\sqrt{1 - r^2\delta^2}}, \frac{1 - \delta^2}{1 - r^2\delta^2} \right).$$

Then, we have

$$\begin{aligned} \phi \left(a_t(\alpha, \beta) \Big| \frac{\delta(h - r)}{\sqrt{1 - r^2\delta^2}}, \frac{1 - \delta^2}{1 - r^2\delta^2} \right) \phi(h|0, 1) &= \phi \left(a_t(\alpha, \beta) \Big| -\frac{r\delta}{\sqrt{1 - r^2\delta^2}}, \frac{1}{1 - r^2\delta^2} \right) \\ &\times \phi \left(h \Big| \delta\sqrt{1 - r^2\delta^2} (a_t(\alpha, \beta) + \frac{r\delta}{\sqrt{1 - r^2\delta^2}}), 1 - \delta^2 \right). \end{aligned}$$

Therefore, the proof of (i) follows directly from that $f_{H|T}(h|t) = f_{T|H}(t|h)f_H(h)/f_T(t)$. To demonstrate (ii)-(iii), notice, for $k = 1, 2$, we have that

$$\mathbb{E} \left[H^k | T \right] = \frac{1}{\Phi \left(\lambda \sigma_z \left(a_t(\alpha, \beta) + \frac{r\delta}{\sqrt{1 - r^2\delta^2}} \right) \right)} \int_0^\infty h^k \phi \{ h | \eta_t, 1 - \delta^2 \} dh = \mathbb{E}(X^k | X > 0).$$

Then, using some properties of the HN distribution from Lemma 1, the proof is completed.

APPENDIX C. THE OBSERVED FISHER INFORMATION MATRIX

The necessary expressions are given below. For the sake of simplicity, we consider the following notation to obtain the necessary expressions, $a_{t_i;\mu,\sigma} = a_{t_i;\mu,\sigma}(\alpha, \beta)$ and $A_{t_i;\sigma} = A_{t_i;\sigma}(\alpha, \beta)$.

$$\begin{aligned} \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_2} &= -\frac{1}{A_{t_i;\sigma}^2} \frac{\partial A_{t_i;\sigma}}{\partial \theta_1} \frac{\partial A_{t_i;\sigma}}{\partial \theta_2} + \frac{1}{A_{t_i;\sigma}} \frac{\partial^2 A_{t_i;\sigma}}{\partial \theta_1 \partial \theta_2} - \frac{1}{2} \frac{\partial^2 a_{t_i;\mu,\sigma}^2}{\partial \theta_1 \partial \theta_2} + \lambda^2 W_{\Phi}'[\lambda a_{t_i;\mu,\sigma}] \frac{\partial a_{t_i;\mu,\sigma}}{\partial \theta_1} \frac{\partial a_{t_i;\mu,\sigma}}{\partial \theta_2} \\ &\quad + \lambda W_{\Phi}[\lambda a_{t_i;\mu,\sigma}] \frac{\partial^2 a_{t_i;\mu,\sigma}}{\partial \theta_1 \partial \theta_2}, \quad \theta_1, \theta_2 = \alpha, \beta \\ \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \theta_3 \partial \gamma} &= -\frac{1}{A_{t_i;\sigma}^2} \frac{\partial A_{t_i;\sigma}}{\partial \theta_3} \frac{\partial A_{t_i;\sigma}}{\partial \gamma} + \frac{1}{A_{t_i;\sigma}} \frac{\partial^2 A_{t_i;\sigma}}{\partial \theta_3 \partial \gamma} - \frac{1}{2} \frac{\partial^2 a_{t_i;\mu,\sigma}^2}{\partial \theta_3 \partial \gamma} + \lambda W_{\Phi}[\lambda a_{t_i;\mu,\sigma}] \frac{\partial^2 a_{t_i;\mu,\sigma}}{\partial \theta_3 \partial \gamma} \\ &\quad + \left\{ \lambda W_{\Phi}'[\lambda a_{t_i;\mu,\sigma}] \frac{\partial \lambda a_{t_i;\mu,\sigma}}{\partial \gamma} + W_{\Phi}[\lambda a_{t_i;\mu,\sigma}] \frac{\partial \lambda}{\partial \gamma} \right\} \frac{\partial a_{t_i;\mu,\sigma}}{\partial \theta_3}, \quad \theta_3 = \alpha, \beta \\ \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \gamma^2} &= -\frac{1}{A_{t_i;\sigma}^2(\alpha, \beta)} \frac{\partial^2 A_{t_i;\sigma}}{\partial \gamma^2} + \frac{1}{A_{t_i;\sigma}(\alpha, \beta)} \frac{\partial^2 A_{t_i;\sigma}}{\partial \gamma^2} - \frac{1}{2} \frac{\partial^2 a_{t_i;\mu,\sigma}^2}{\partial \gamma^2} + W_{\Phi}'[\lambda a_{t_i;\mu,\sigma}] \left(\frac{\partial \lambda a_{t_i;\mu,\sigma}}{\partial \gamma} \right)^2 \\ &\quad + W_{\Phi}[\lambda a_{t_i;\mu,\sigma}] \frac{\partial^2 \lambda a_{t_i;\mu,\sigma}}{\partial \gamma^2}, \end{aligned}$$

where $W_{\Phi}'(x) = -W_{\Phi}(x)[x + W_{\Phi}(x)]$ is the derivative of $W_{\Phi}(x)$ with respect to x , see [Vilca et al. \(2011\)](#), and the other quantities are as before defined.

ACKNOWLEDGEMENTS

The authors would like to thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for the financial support through a Master Scholarship granted to the first author under the guidance of the second and also for Conselho Nacional de Desenvolvimento Científico e Tecnológico, grants numbers 308339/2015-0 and 305336/2017-7, for a research scholarship granted to the second and fourth authors, respectively.

REFERENCES

- Achcar, J.A., 1993. Inferences for the Birnbaum-Saunders fatigue life model using Bayesian methods. *Computational Statistics and Data Analysis*, 15, 367-380.
- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716-723.
- Atkinson, A.C., 1985. *Plots, Transformations and Regressions*. Oxford University Press, Oxford, UK.
- Arellano-Valle, R. and Azzalini, A., 2008. The centered parametrization for the multivariate skew-normal distribution. *Journal of Multivariate Analysis*, 99, 1362-1382.
- Azzalini, A., 1985. A class of distribution which includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171-178.
- Azzalini, A., 2013. *The skew-normal and related families*. Cambridge University Press, Cambridge, UK.
- Balakrishnan, N., Leiva, V., Sanhueza, A., and Vilca-Labra, F., 2009. Estimation in the Birnbaum-Saunders distribution based on scale-mixture of normals and the EM-algorithm. *SORT*, 33, 171-192.

- Balakrishnan, N., Saulo, H., and Leao, J., 2017. On a new class of skewed Birnbaum-Saunders models. *Journal of Statistical Theory and Practice*, 11, 573-593.
- Barros, M., Paula, G.A., and Leiva, V., 2008. A New Class of Survival Regression Models with Heavy-Tailed Errors: Robustness and Diagnostics. *Lifetime Data Analysis*, 14, 1-17.
- Birnbaum, Z.W. and Saunders, S.C., 1969a. A new family of life distributions. *Journal of Applied Probability*, 6, 637-652.
- Birnbaum, Z.W. and Saunders, S.C., 1969b. Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability*, 6, 328-347.
- Byrd, R.H., Lu, P., Nocedal, J., and Zhu, C., 1995. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 15, 1190-1208.
- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M.A., Guo, J., Li, P., and Riddell, A., 2016. Stan: A probabilistic programming language. *Journal of Statistical Software*, 20, 1-37.
- Cordeiro, G.M. and Lemonte, A.J., 2011. The β -Birnbaum-Saunders distribution: An improved distribution for fatigue life modeling. *Computational Statistics and Data Analysis*, 55, 116-124.
- Depaoli, S., Clifton, J.P., and Cobb, P. R., 2016. Just another Gibbs sampler (JAGS) flexible software for MCMC implementation. *Journal of Educational and Behavioral Statistics*, 41, 628-649.
- Desmond, A.F., 1985. Stochastic models of failure in random environments. *Canadian Journal of Statistics*, 13, 171-183.
- Desmond, A.F., 1986. On the relationship between two fatigue-life models. *IEEE Transactions on Reliability*, 35, 167-169.
- Diaz-Garcia, J.A. and Leiva, V., 2005. New family of life distributions based on the elliptically contoured distributions. *Journal of Statistical Planning and Inference*, 128, 445-457.
- Engelhardt, M., Bain, L.J., and Wright, F. ., 1981. Inferences on the parameters of the Birnbaum-Saunders fatigue life distribution based on maximum likelihood estimation. *Technometrics*, 23, 251-256.
- Ferreira, M.S., 2013. A study of exponential-type tails applied to Birnbaum-Saunders models. *Chilean Journal of Statistics*, 4, 87-97.
- Gamerman, D. and Lopes, H., 2006. *MCMC-Stochastic Simulation for Bayesian Inference*. Chapman and Hall/CRC, Boca Raton, FL, US.
- Homan, M.D. and Gelman, A., 2014. The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *The Journal of Machine Learning Research*, 15, 1593-1623.
- Leiva, V., Barros, M., Paula, G.A., and Sanhueza, A., 2007. Generalized Birnbaum-Saunders distributions applied to air pollutant concentration. *Environmetrics*, 19, 235-249.
- Leiva, V., Vilca, F., Balakrishnan, N., and Sanhueza, A., 2008. A skewed sinh-normal distribution and its properties and application to air pollution. *Communications in Statistics Theory and Methods*, 39, 426-443.
- Leiva, V., 2016. *The Birnbaum-Saunders Distribution*. Academic Press, New York.
- Lu, M.C. and Chang, D.S., 1997. Bootstrap prediction intervals for the Birnbaum-Saunders distribution. *Microelectronics Reliability*, 37, 1213-1216.
- Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D., 2000. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10, 325-337.
- Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N., 2009. The BUGS project: Evolution, critique and future directions (with discussion). *Statistics in Medicine*, 28, 3049-3082.

- Maehara, R.P. (2018). An extension of Birnbaum-Saunders distributions based on scale mixtures of skew-normal distributions with applications to regression models. PhD thesis, University of Sao Paulo, available at <http://www.teses.usp.br/teses/disponiveis/45/45133/tde-14052018-202935/es.php>.
- Mann, N.R., Singpurwalla, N.D., and Schafer, R.E., 1974. *Methods for Statistical Analysis of Reliability and Life Data*. Wiley, New York.
- Mazucheli, J.M., Menezes, A.F.B., and Dey, S., 2018. The unit-Birnbaum-Saunders distribution with applications. *Chilean Journal of Statistics*, 9, 47-57.
- Meng, X.L. and Rubin, D.B., 1993. Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, 80, 267-278.
- Owen, W.J. and Padgett, W.J., 1999. Accelerated test models for system strength based on Birnbaum-Saunders distributions. *Lifetime Data Analysis*, 5, 133-147.
- Pewsey, A., 2000. Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics*, 27, 859-870.
- R Development Core Team, 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rieck, J.R. and Nedelman, J.R., 1991. A log-linear model for the Birnbaum-Saunders distribution. *Technometrics*, 33, 51-60.
- Rue, H. and Martino, S., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society B*, 71, 319-392.
- Schwarz, G., 1978. Estimating the dimension of a model. *The Annals of Statistics*, 6, 461-464.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and Linde, A., 2014. The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society B*, 76, 485-493.
- Vilca, F. and Leiva, V., 2006. A new fatigue life model based on the family of skew-elliptical distributions. *Communications in Statistics: Theory and Methods*, 35, 229-244.
- Vilca, F., Santana, L., Leiva, V., and Balakrishnan, N., 2011. Estimation of extreme percentiles in Birnbaum-Saunders distributions. *Computational Statistics and Data Analysis*, 55, 1665-1678.
- Wang, M., Sun, X., and Park, C., 2016. Bayesian analysis of BirnbaumSaunders distribution via the generalized ratio-of-uniforms method. *Computational Statistics*, 31, 207-225.
- Xu, A. and Tang, Y., 2011. Bayesian analysis of Birnbaum-Saunders distribution with partial information. *Computational Statistics and Data Analysis*, 55, 2324-2333.

DISTRIBUTION THEORY
RESEARCH PAPER

The Harris extended Lindley distribution for modeling hydrological data

GAUSS M. CORDEIRO¹, M. MANSOOR^{2,*} and SERGE B. PROVOST³

¹Department of Statistics, Federal University of Pernambuco, Recife, PE, Brazil

²Department of Statistics, Government Sadiq Egerton College Bahawalpur, Bahawalpur, Pakistan

³Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Canada

(Received: 17 April 2018 · Accepted in final form: 14 February 2019)

Abstract

We introduce a three-parameter extension of the Lindley distribution, which has as sub-models the Lindley and Marshall-Olkin Lindley distributions. The proposed model turns out to be quite flexible: its probability density function can be decreasing or unimodal and its associated hazard rate may be increasing, decreasing, unimodal or bathtub-shaped. Since this new distribution has a survival function and a hazard rate that can be expressed in closed form, it can readily be simulated and used to analyze censored data. Computable expressions are obtained for certain statistical functions such as its quantile function, ordinary and incomplete moments, moment generating function, order statistics and reliability function. The maximum likelihood method is utilized to obtain estimates of the model parameters and a simulation study is carried out to assess the performance of the corresponding maximum likelihood estimators. Two illustrative examples involving hydrological data sets are presented.

Keywords: Data modeling · Extended distributions · Hazard rate · Maximum likelihood estimation · Monte Carlo simulations · Precipitation data.

Mathematics Subject Classification: Primary 60E05 · Secondary 62E10 · 62N05

1. INTRODUCTION

Lindley (1958) introduced a one-parameter distribution in the context of fiducial and Bayesian statistics, which is obtained as a mixture of exponential(λ) and gamma(2, λ) probability density functions (PDFs), as defined in Equation (2). Aly and Benkherouf (2011) recently proposed a convenient method for adding two parameters to a baseline distribution, which gives rise to what is referred to as the Harris extended (HE) family of distributions. This family includes the baseline distribution itself as a basic exemplar and provides more flexibility for modeling various types of data. This novel approach is based on the probability generating function of a discrete distribution introduced by Harris (1948). In this paper, we define a three-parameter generalization of the Lindley distribution by applying to it the HE generator, the resulting model being named the Harris extended Lindley (HEL) distribution. This distribution is in fact an extension of the Marshall-Olkin

*Corresponding author. Email: mansoor.abbasi143@gmail.com

extended Lindley (MOL) distribution that was proposed by [Ghitany et al. \(2012\)](#), and its additional shape parameter α ought to provide an improved fit related to the MOL distribution. This extra parameter helps in controlling the shape of the HE PDF and enables us to model heavy-tailed distributions which are fairly common in hydrology; see, e.g., [Li et al. \(2013\)](#) and [Ashkar and El Adlouni \(2014\)](#). Moreover, the new distribution has an interesting physical interpretation when α is a positive integer and $0 < \theta < 1$: it is indeed the distribution of the time until failure of a device composed of N serial components having constant failure rate, where N is a random variable which arises from a branching process such as that described in [Harris \(1948\)](#). This distribution can be utilized for modeling purposes in research fields such as hydrology, engineering, insurance, biology and epidemiology wherein skewed positive data are frequently encountered.

One of the most crucial aspects of hydrological data analysis consists in achieving a close fit to the experimental data by employing proper statistical models. The Gumbel, Weibull, gamma, generalized logistic as well as other well-known distributions have been extensively utilized for modeling hydrological observations such as rainfall, flood, precipitation and stream flow data; see, e.g., [Zelenhasic \(1970\)](#), [Chadwick et al. \(2004\)](#), [Heo and Boes \(2011\)](#), [Bhunya et al. \(2012\)](#) and [Kang et al. \(2015\)](#). Yet, there exists a need for developing more flexible statistical models that would be applicable to data sets related to hydrological structures and phenomena or water resource planning and management, and the proposed three-parameter generalization of the Lindley distribution fits the purpose.

Although little attention has been paid to the Lindley distribution, there has recently been a surge of interest in this model, generalizations thereof and related applications. [Nadarajah et al. \(2007\)](#) introduced the exponentiated Lindley distribution as an alternative to the gamma, log-normal, Weibull and exponentiated exponential distributions; see also [Cordeiro et al. \(2016\)](#). Several properties of the Lindley distribution have been studied by [Ghitany et al. \(2008\)](#) who have shown that, for instance, it can provide a better fit than the exponential distribution. [Ghitany et al. \(2011\)](#) studied another two parameter extension of Lindley distribution and called it the weighted Lindley distribution. By making use of the Marshall-Olkin method, [Ghitany et al. \(2012\)](#) introduced and studied another extension of the Lindley model called the Marshall-Olkin extended Lindley (MOL) distribution. [Ghitany et al. \(2013\)](#) introduced a two-parameter power Lindley distribution and discussed its properties. A three-parameter generalization of the Lindley model was introduced by [Mervoci and Sharma \(2014\)](#). This extension, referred to as the beta Lindley (BL) distribution, is generated from the logit of a beta random variable. [Ghitany et al. \(2015\)](#) considered the problem of estimating the stress-strength parameter of the power Lindley distribution. [Mazucheli et al. \(2016\)](#) developed some statistical for testing hypotheses on the parameters of the weighted Lindley distribution. [Alizadeh et al. \(2017\)](#) introduced another extension of the power Lindley distribution.

The objective of this work is to derive the HEL distribution focusing on its probabilistic and statistics aspects, as well as applications in hydrology.

The remainder of the paper is organized as follows. We define the new distribution in Section 2. In Section 3, we provide computable expressions for some of its statistical functions such as its quantile function (QF), ordinary and incomplete moments, mean deviations, moment generating function (MGF) and order statistics. In Section 4, the model parameters are estimated by making use of the maximum likelihood (ML) method and a simulation study is carried out. In Section 5, we illustrate the usefulness of the proposed distribution by modeling two hydrological data sets. Finally, Section 6 offers some concluding remarks.

2. THE HEL DISTRIBUTION

In this section, we provide probabilistic aspects of the HEL distribution. The survival function (SF) and PDF of the distribution introduced by Lindley (1958) are respectively given by

$$\bar{G}_L(x) = \left(\frac{1 + \lambda + \lambda x}{1 + \lambda} \right) e^{-\lambda x}, \quad x > 0, \quad (1)$$

and

$$g_L(x) = \frac{\lambda^2}{\lambda + 1} (1 + x) e^{-\lambda x}, \quad x > 0, \quad (2)$$

where the parameter λ is assumed to be positive. We now describe a technique whereby the so-called Harris extended family of distributions can be generated and apply it to the Lindley distribution. The resulting distribution is referred to as the Harris extended Lindley (HEL) distribution. Let $G(x) = G(x; \xi)$ be a baseline cumulative distribution function (CDF) and

$$\bar{G}(x) = \bar{G}(x; \xi) = 1 - G(x; \xi)$$

be the corresponding SF of a lifetime random variable W , where $\xi = (\xi_1, \dots, \xi_q)$ is a parameter vector of dimension q . Furthermore, let $g(x) = g(x; \xi)$ be the PDF of W . The SF of the HE family is then defined by

$$\bar{F}_{\text{HE}}(x) = \frac{\theta^{1/\alpha} \bar{G}(x)}{[1 - \bar{\theta} \bar{G}(x)^\alpha]^{1/\alpha}}, \quad x > 0, \quad (3)$$

where $\bar{\theta} = 1 - \theta$, the parameters $\theta > 0$ and $\alpha > 0$ being additional shape parameters that allow for greater flexibility. Thereupon, the HE PDF has the form

$$f_{\text{HE}}(x) = \frac{\theta^{1/\alpha} g(x)}{[1 - \bar{\theta} \bar{G}(x)^\alpha]^{1+1/\alpha}}, \quad x > 0.$$

Aly and Benkherouf (2011) pointed out that when $\alpha > 0$ is a positive integer, the HE family can be looked upon as resulting from examining a simple discrete branching process where a particle either splits into $(\alpha + 1)$ identical branches or remains the same during a short interval. Clearly, Equation (3) constitutes a flexible generator for obtaining new parametric distributions from existing ones. For $\theta = 1$, $\bar{F}(x) = \bar{G}(x)$ and $\bar{G}(x)$ is thus a basic exemplar of the distribution. Additionally, the Marshall and Olkin (1997) extended (MOE) family arises from Equation (3) by letting $\alpha = 1$. Accordingly, the HE family can be viewed as a generalization of the MOE family.

The SF of the HEL distribution is defined as

$$\bar{F}(x) = \frac{\theta^{1/\alpha} \bar{G}_L(x)}{[1 - \bar{\theta} \bar{G}_L(x)^\alpha]^{1/\alpha}}, \quad x > 0, \quad (4)$$

for $\alpha > 0$, $\theta > 0$, $\lambda > 0$, where $\bar{G}_L(x)$ is given in Equation (1), with its PDF corresponding

to Equation (4) being

$$f(x) = \frac{\theta^{1/\alpha} \lambda^2 (1+x) e^{-\lambda x}}{(1+\lambda) [1 - \bar{\theta} \bar{G}_L(x)^\alpha]^{1+1/\alpha}}, \quad x > 0. \quad (5)$$

Henceforth, a random variable X having the PDF specified in Equation (5) is denoted by $X \sim \text{HEL}(\theta, \alpha, \lambda)$. This three-parameter PDF has two shape parameters and one scale parameter, and it can be either decreasing or unimodal. The two main special cases of the HEL model are: (i) the MOL distribution in which case $\alpha = 1$; (ii) the Lindley distribution which is obtained by letting $\alpha = \theta = 1$. The hazard rate (HR) associated with HEL model is given by

$$h(x) = \frac{\lambda^2 (1+x)}{(\lambda + 1 + \lambda x)} [1 - \bar{\theta} \bar{G}_L(x)^\alpha]^{-1}, \quad x > 0.$$

This HR can assume the four principal shapes associated with increasing, decreasing, bathtub-shaped or upside-down bathtub-shaped HRs. The HEL model is thus most appropriate to analyze a variety of hydrological and lifetime data sets. We note that there appears to be very few three-parameter distributions in the literature whose HR can take on the four main shapes of an HR. Moreover, the SF and HR of the HEL distribution have closed-form representations. Accordingly, this model can readily be utilized to analyze censored data sets. As well, simulating it is straightforward.

Figures 1 and 2 display some plots of the PDF and HR of the HEL distribution for certain parameter values. Figure 1 indicates that the HEL PDF can be right-skewed and reversed-J shaped. Figure 2 reveals that the HEL HR can be increasing (IFR), decreasing (DFR), upside-down bathtub (UBT) or bathtub-shaped (BT).

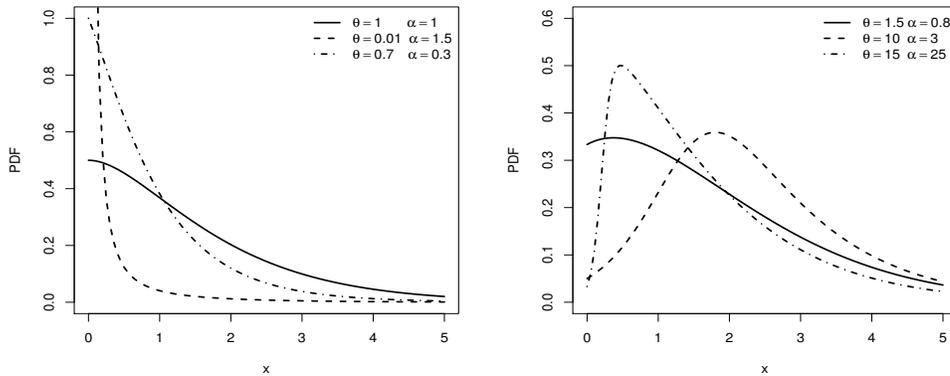


Figure 1. Plots of the HEL PDF for certain parameter values.

Given the functional form of the HEL PDF denoted by $f(x)$, a general representation of the mode that would be expressible in terms of the parameters of the distribution does not appear to be tractable. However, for a specific set of parameters, the command `NSolve[f'[x]==0,x,Reals]` in *Mathematica* can readily be utilized to determine the mode. If the solution happens to be greater than zero, then the PDF has a mode at that point; otherwise, it is strictly decreasing on the positive half-line. The extremum of the HR can be similarly obtained whenever it exists.

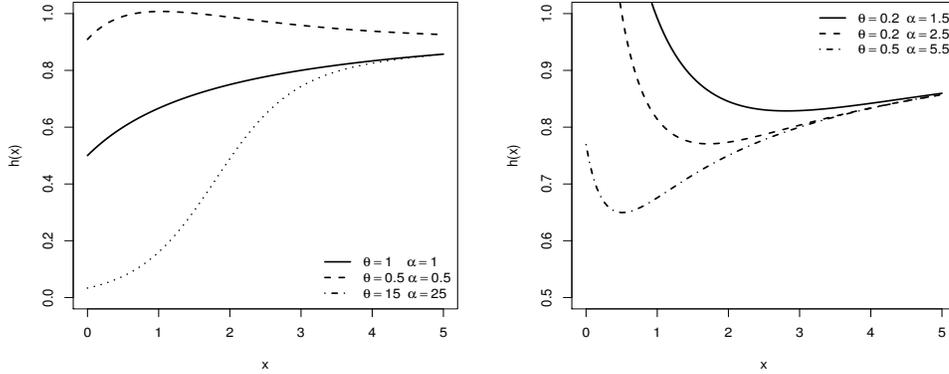


Figure 2. Plots of the HEL HR for certain parameter values.

3. STATISTICAL FUNCTIONS OF THE HEL DISTRIBUTION

In this section, we provide computable representations of certain statistical functions of the HEL distribution. More specifically, we focus, in order, on the quantile function, some useful expansions, the moments, including the incomplete ones, the moment generating function and the order statistics. The derived expressions can be easily evaluated by most symbolic computation software packages such as **Maple**, **Mathematica** and **Matlab**. These platforms can process analytic expressions of great complexity. Whenever available, an explicit representation of a statistical function is preferable to its determination by numerical integration.

The QF of a distribution has numerous uses in both statistical theory and applications. In the case of the HEL distribution, its QF is obtained by inverting the HEL CDF and is given by

$$Q(u) = -1 - \frac{1}{\lambda} - \frac{1}{\lambda} W \left[-(1 + \lambda) \frac{1 - \tau}{e^{1+\lambda}} \right], \quad 0 < u < 1, \tag{6}$$

where $\tau = 1 - (1 - u) [\theta + \bar{\theta}(1 - u)^\alpha]^{-1/\alpha}$ and $W(x)$ is the negative branch of the Lambert W function, see [Corless et al. \(1996\)](#) and [Jodrá \(2010\)](#) for details on its properties. The Lambert function cannot be expressed in terms of elementary functions. However, it is analytically differentiable and integrable and its principal branch satisfies $x = W(x e^x)$, $x \geq -1$. Furthermore, whenever $|x| \leq e^{-1}$, $W(x) = \sum_{n=1}^{\infty} (-n)^{n-1} x^n / n$. Clearly, if U has a uniform distribution in the interval $(0, 1)$, then $X = Q(U)$ has the PDF specified in Equation (5). The Lambert W function is implemented within various scientific libraries, as for example, in the R software (by the `lamW` package), **Mathematica** (by the `ProductLog` function), **Matlab** (by the `lambertw` function) and **Maple** (by the `LambertW` function), thus allowing for efficient evaluation of the QF of the HEL distribution.

Some useful expansions are now provided. Let $g_a(x) = a g(x) \bar{G}(x)^{a-1}$ be the Lehmann type-II-G (LII-G) PDF with power parameter $a > 0$. We demonstrate that the HEL PDF can be expressed as a linear combination of LII-Lindley (LIIL) PDFs. First, for $0 < \theta < 1$, we consider the negative binomial series

$$(1 - z)^{-p} = \sum_{i=0}^{\infty} \frac{\Gamma(p + i)}{\Gamma(p) i!} z^i,$$

which holds for $|z| < 1$ and any real number $p > 0$, where $\Gamma(a) = \int_0^\infty z^{a-1} e^{-z} dz$ is the

complete gamma function. Using this power series in Equation (5), we have

$$f(x) = \theta^{1/\alpha} g_L(x) \sum_{j=0}^{\infty} \bar{\theta}^j \frac{\Gamma(\alpha^{-1} + 1 + j)}{\Gamma(\alpha^{-1} + 1)j!} \bar{G}_L(x)^{j\alpha},$$

where $\bar{G}_L(x)$ and $g_L(x)$ are the SF and PDF of the Lindley distribution as provided by Equations (1) and (2). Note that for $\theta > 1$, we can write

$$f(x) = \theta^{-1} g_L(x) \sum_{j=0}^{\infty} \sum_{\ell=j}^{\infty} (-1)^j \left(\frac{\theta-1}{\theta}\right)^\ell \binom{\ell}{j} \frac{\Gamma(\alpha^{-1} + 1 + \ell)}{\Gamma(\alpha^{-1} + 1)\ell!} \bar{G}_L(x)^{j\alpha}.$$

On combining the last two expressions for $f(x)$ in a single one, we have

$$f(x) = \sum_{j=0}^{\infty} w_j h_{j\alpha+1}(x), \quad (7)$$

where $h_{j\alpha+1}(x) = (j\alpha + 1) g_L(x) \bar{G}_L(x)^{j\alpha}$ is the LIIL PDF with power parameter $j\alpha + 1$ and

$$w_j = w_j(\alpha, \theta) = \begin{cases} \frac{\theta^{1/\alpha} \bar{\theta}^j \Gamma(\alpha^{-1} + 1 + j)}{(j\alpha + 1) \Gamma(\alpha^{-1} + 1) j!}, & 0 < \theta < 1 \\ \frac{(-1)^j \theta^{-1}}{(j\alpha + 1)} \sum_{\ell=j}^{\infty} \left(\frac{\theta-1}{\theta}\right)^\ell \binom{\ell}{j} \frac{\Gamma(\alpha^{-1} + 1 + \ell)}{\Gamma(\alpha^{-1} + 1) \ell!}, & \theta > 1. \end{cases}$$

Equation (7) reveals that the HEL PDF (for any $\theta > 0$) can indeed be expressed as a linear combination of LIIL PDFs. It can also be shown that the HEL PDF can be expressed as a linear combination of gamma PDFs. Given Equations (1) and (2), it follows from the representation of Equation (7) that

$$f(x) = \sum_{j=0}^{\infty} w_j (j\alpha + 1) \left(\frac{\lambda^2}{\lambda + 1}\right) (1 + x) \left(1 + \frac{\lambda x}{1 + \lambda}\right)^{j\alpha} e^{-(j\alpha+1)\lambda x}.$$

On expanding $[1 + \lambda x / (1 + \lambda)]^{j\alpha}$ and using the Taylor series $z^\beta = \sum_{k=0}^{\infty} (\beta)_k (z-1)^k / k!$, where $(\beta)_k = \beta(\beta-1)\cdots(\beta-k+1)$ is the falling factorial, after some algebra, we obtain

$$f(x) = \sum_{i,j=0}^{\infty} v_{i,j} x^i (1+x) e^{-(j\alpha+1)\lambda x}, \quad (8)$$

where $v_{i,j} = (j\alpha + 1) w_j [\lambda^{2+i} / (\lambda + 1)^{i+1}] (j\alpha)_i / i!$ for $i, j = 0, 1, 2, \dots$

Letting $\pi(x; \alpha, \beta) = \beta^\alpha x^{\alpha-1} e^{-\beta x} / \Gamma(\alpha)$ be the gamma PDF with shape parameter $\alpha > 0$ and rate parameter $\beta > 0$, we can then rewrite Equation (8) as

$$f(x) = \sum_{i,j=0}^{\infty} \left[v_{i,j}^{(1)} \pi(x; i+1, (j\alpha + 1)\lambda) + v_{i,j}^{(2)} \pi(x; i+2, (j\alpha + 1)\lambda) \right], \quad (9)$$

where $v_{i,j}^{(1)} = i! v_{i,j} / [(j\alpha + 1)\lambda]^{i+1}$ and $v_{i,j}^{(2)} = (i+1)! v_{i,j} / [(j\alpha + 1)\lambda]^{i+2}$.

Equation (9) indicates that the HEL PDF can also be expressed as a linear combination of gamma PDFs. Thus, this representation can be used to obtain explicit expressions for

the ordinary and incomplete moments and the MGF of the HEL distribution from the corresponding quantities associated with the gamma distribution. Equations (7) and (9) constitute the main results of this section.

Certain of the main characteristics of a distribution such as tendency, dispersion, skewness and kurtosis can be investigated via its moments. We now establish that the ordinary moments of the HEL distribution can be obtained as infinite power series. It follows from Equation (7) that

$$\mu'_r = E(X^r) = \frac{\lambda^2}{1 + \lambda} \sum_{j=0}^{\infty} w_j \int_0^{\infty} x^r (1 + x) \left(1 + \frac{\lambda x}{1 + \lambda}\right)^{j\alpha} e^{-\lambda(j\alpha+1)x} dx,$$

or equivalently

$$\mu'_r = \frac{\lambda^2}{1 + \lambda} \sum_{j=0}^{\infty} w_j \int_0^{\infty} x^r (1 + x) \sum_{i=0}^{\infty} \left(\frac{\lambda}{1 + \lambda}\right)^i x^i \frac{(j\alpha)_i}{i!} e^{-\lambda(j\alpha+1)x} dx.$$

After some algebra, we obtain

$$\mu'_r = \frac{\lambda^2}{1 + \lambda} \sum_{i,j=0}^{\infty} p_{i,j} \frac{\Gamma(r + i + 1)}{[\lambda(j\alpha + 1)]^{r+i+1}} \left(1 + \frac{r + i + 1}{\lambda(j\alpha + 1)}\right), \tag{10}$$

where $p_{i,j} = w_j [(j\alpha)_i / i!] (\lambda / (1 + \lambda))^i$.

Table 1 includes numerical values for the first four ordinary moments of the HEL distribution as evaluated from Equation (10) by truncating the series to 100 terms and computed by numerical integration for some parameter values. We note that the numerical values obtained from both approaches are consistently in close agreement.

Table 1. Ordinary moments of the HEL distribution for certain parameter values with $\lambda = 10$.

μ'_r	$\alpha = 0.5$		$\alpha = 1.5$	
	Numerical	Equation (10)	Numerical	Equation (10)
$\theta = 0.5$				
μ'_1	0.0670906	0.0670905	0.0833919	0.08156687
μ'_2	0.0105268	0.01052653	0.0158697	0.01586975
μ'_3	0.00276376	0.002763106	0.00492889	0.00492885
μ'_4	0.0010382	0.001036676	0.0020813	0.002081299
$\theta = 1.5$				
μ'_1	0.141446	0.1414455	0.127601	0.1276013
μ'_2	0.0364545	0.0364543	0.0295221	0.02952214
μ'_3	0.0132554	0.01325516	0.0098071	0.009807097
μ'_4	0.00616152	0.006160951	0.00425269	0.004252694

The r th incomplete moment of X is given by $m_r(y) = \int_0^y x^r f(x) dx$. On making use of Equation (7) and proceeding as in the case of ordinary moments, we obtain

$$m_r(y) = \frac{\lambda^2}{1 + \lambda} \sum_{j,i=0}^{\infty} w_j \left(\frac{\lambda}{1 + \lambda}\right)^i \frac{(j\alpha)_i}{i!} \int_0^y x^{r+i} (1 + x) e^{-\lambda(j\alpha+1)x} dx. \tag{11}$$

On expressing the integral in Equation (11) in terms of the incomplete gamma function

$\gamma(a, y) = \int_0^y z^{a-1} e^{-z} dz$, we have

$$m_r(y) = \frac{\lambda^2}{1+\lambda} \sum_{i,j=0}^{\infty} K_{i,j} \left\{ \frac{\gamma(r+i+1, (j\alpha+1)\lambda y)}{[(j\alpha+1)\lambda]^{r+i+1}} + \frac{\gamma(r+i+2, (j\alpha+1)\lambda y)}{[(j\alpha+1)\lambda]^{r+i+2}} \right\}, \quad (12)$$

where $K_{i,j} = w_j [\lambda/(1+\lambda)]^i (j\alpha)_i / i!$ for $i, j = 0, 1, \dots$

Bonferroni and Lorenz curves as well as mean deviations can be determined by letting $r = 1$ in Equation (12). The Bonferroni and Lorenz curves are defined (for a given probability π) as $B(\pi) = m_1(q)/(\pi\mu'_1)$ and $L(\pi) = m_1(q)/\mu'_1$, respectively, where $q = Q(\pi)$ may be established from Equation (6). The mean deviations about the mean and about the median are given by $\delta_1 = E(|X - \mu'_1|) = 2\mu'_1 F(\mu'_1) - 2m_1(\mu'_1)$ and $\delta_2 = E(|X - M|) = \mu'_1 - 2m_1(M)$, where the median M and the mean μ'_1 can be evaluated from Equations (6) and (10), respectively. We now provide a general formula for $M(t) = E(e^{tX})$, the MGF of X . The MGF of the gamma PDF with parameters α and β is $(1 - t/\beta)^{-\alpha}$ ($t < \beta$). Then, it follows from Equation (9) that, for $t < \lambda$,

$$M(t) = \sum_{i,j=0}^{\infty} \left[v_{i,j}^{(1)} \left(1 - \frac{t}{(j\alpha+1)\lambda} \right)^{-i-1} + v_{i,j}^{(2)} \left(1 - \frac{t}{(j\alpha+1)\lambda} \right)^{-i-2} \right].$$

The last aspect being discussed in this section is the distribution of order statistics. Order statistics appear in many areas of statistical theory and practice. Suppose X_1, \dots, X_n is a random sample from the HEL distribution and let $X_{i:n}$ denote the i th order statistic. The PDF of $X_{i:n}$ can be expressed as

$$f_{i:n}(x) = K \sum_{k=0}^{n-i} (-1)^k \binom{n-i}{k} f(x) F(x)^{k+i-1}, \quad (13)$$

where $K = 1/B(i, n-i+1)$ and $B(p, q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$ is the beta function.

Consider the following representation available from [Gradshteyn and Ryzhik \(2000\)](#) for a power series raised to a positive integer n :

$$\left(\sum_{j=0}^{\infty} a_j u^j \right)^n = \sum_{j=0}^{\infty} b_{n,j} u^j, \quad (14)$$

where the coefficients $b_{n,j}$, for $n = 1, 2, \dots$ and $j = 1, 2, \dots$, are obtained from the recursive equation

$$b_{n,j} = (j a_0)^{-1} \sum_{m=1}^j [m(n+1) - j] a_m b_{n,j-m},$$

with $b_{n,0} = a_0^n$. On integrating the right-hand side of Equation (7), we can write

$$F(x) = \bar{G}_L(x) \sum_{j=0}^{\infty} w_j \bar{G}_L(x)^{j\alpha},$$

and then making use of Equation (14), we have

$$F(x)^{k+i-1} = \sum_{j=0}^{\infty} t_{k+i-1,j} \bar{G}_L(x)^{j \alpha + k + i - 1},$$

where $t_{k+i-1,j} = (j w_0)^{-1} \sum_{m=1}^j [m(k+i) - j] w_m t_{k+i-1,i-m}$ for $j \geq 1$ and $t_{k+i-1,0} = w_0^{k+i-1}$. Inserting the previous expression for $F(x)^{k+i-1}$ and the representation of Equation (7) of the PDF appearing in Equation (13) gives

$$f_{i:n}(x) = K \sum_{r,j=0}^{\infty} \sum_{k=0}^{n-i} v_{r,j,k} h_{(r+j) \alpha + k + i}(x), \quad (15)$$

where

$$v_{r,j,k} = \frac{(-1)^k (r \alpha + 1) w_r t_{k+i-1,j}}{(r+j) \alpha + k + i} \binom{n-i}{k}.$$

Equation (15) reveals that the PDF of the HEL order statistics can be expressed as a triple linear combination of LIII PDFs. Accordingly, certain mathematical properties of the HEL order statistics could be determined from those of the LIII distribution.

4. PARAMETER ESTIMATION

We now discuss the estimation of the model parameters using the ML method. There exist several approaches for estimating parameters; however, the ML method is the most commonly employed. The ML estimators enjoy several desirable properties and can be utilized in the construction of confidence intervals for the model parameters. They also appear in some test statistics. The normal approximation to the distribution of these estimators follows from large sample distribution theory.

Let X_1, \dots, X_n be a sample of size n from the HEL distribution whose associated PDF is given in Equation (5). The log-likelihood function $\ell = \ell(\Theta)$ of the vector of parameters $\Theta = (\theta, \alpha, \lambda)^\top$ is given by

$$\ell = \frac{n}{\alpha} \log \theta + n \log \left(\frac{\lambda^2}{1 + \lambda} \right) + \sum_{i=1}^n \log(1 + x_i) - \lambda x_i - \left(1 + \frac{1}{\alpha}\right) \sum_{i=1}^n \log[1 - \theta \bar{G}_L(x)^\alpha]. \quad (16)$$

The ML estimates $\hat{\theta}$, $\hat{\alpha}$ and $\hat{\lambda}$ are determined by maximizing the log-likelihood function of Equation (16) with respect to the parameters θ , α and λ . In general, there is no closed-form representation for these estimates, which are determined in practice the by making use of numerical methods. Equation (16) can be maximized either directly by using the R (`optim` function), SAS (`NLMixed` procedure) or Ox (`MaxBFGS` function), or by solving the nonlinear likelihood equations obtained by equating the partial derivatives of ℓ with respect to each parameter to zero.

The components of the score vector $U(\Theta)$ are expressed as

$$U_\theta = \frac{n}{\alpha\theta} - \left(1 + \frac{1}{\alpha}\right) \sum_{i=1}^n \frac{\bar{G}_L(x)^\alpha}{1 - \theta\bar{G}_L(x)^\alpha},$$

$$U_\alpha = -\frac{n}{\alpha^2} \log \theta - \frac{1}{\alpha^2} \sum_{i=1}^n \frac{\bar{G}_L(x)^\alpha \log \bar{G}_L(x)}{1 - \theta\bar{G}_L(x)^\alpha},$$

$$U_\lambda = \frac{n(2 + \lambda)}{\lambda + \lambda^2} + x_i + \left(1 + \frac{1}{\alpha}\right) \sum_{i=1}^n \left[\frac{\alpha \theta \bar{G}_L(x)^{\alpha-1}}{(1 + \lambda)^2 [1 - \theta\bar{G}_L(x)^\alpha]} \lambda x_i [2 + \lambda + (1 + \lambda)x_i] e^{-\lambda x_i} \right].$$

Setting these equations to zero and solving them simultaneously yields the ML estimates of the model parameters.

We now assess the performance of the ML estimators of the model parameters by means of Monte Carlo simulations. The simulations are replicated 1,000 times with samples of sizes $n = 50, 100, 200$ and the following parameter values: I: $\theta = 0.5, \alpha = 0.5$ and $\lambda = 1$; II: $\theta = 0.1, \alpha = 1.5$ and $\lambda = 1$; III: $\theta = 1.5, \alpha = 0.5$ and $\lambda = 1$; IV: $\theta = 1.5, \alpha = 1.5$ and $\lambda = 1$. Table 2 lists the average bias (Bias) of the ML estimators, mean squared errors (MSE), coverage probabilities (CP) and average widths (AW) of the confidence intervals for the parameters θ, α and λ and the three sample sizes. From these results, we conclude that the ML estimators perform well when it comes to estimating the parameters of the HEL distribution. In general, the biases, MSEs and AWs decrease when the sample size increases. Moreover, the CPs of the confidence intervals are quite close to the 95% nominal level. Thus, the ML estimators and their asymptotic distributional properties can be adopted for constructing approximate confidence intervals for the parameters of the HEL distribution.

5. EMPIRICAL ILLUSTRATIONS WITH HYDROLOGICAL DATA

In this section, we fit the HEL model and some other competing models to two hydrological data sets. We assess how well the HEL distribution performs as compared to the beta-Lindley (BL) studied by [Mervoci and Sharma \(2014\)](#), exponentiated power Lindley (EPL) due to [Ashour and Eltehiwy \(2015\)](#), beta-exponential (BE) proposed by [Nadarajah and Kotz \(2006\)](#), exponentiated Nadarajah and Haghghi (ENH) defined by [Lemonte \(2013\)](#), Harris extended exponential (HEE) discussed by [Pinho et al. \(2015\)](#), exponentiated Weibull (EW) studied by [Mudholkar and Sharivastava \(1993\)](#), power Lindley (PL) introduced by [Ghitany et al. \(2013\)](#), exponentiated Lindley defined by [Nadarajah et al. \(2007\)](#) and Lindley distributions. For each model, we estimated the parameters by the ML method and assessed the goodness-of-fit by means of the Akaike information criterion (AIC), Cramér-von Mises (W), Anderson-Darling (AD), Kolmogrov-Smirnov (KS) and average scaled absolute error (ASAE) statistics. The ASAE is defined as (see [Castilo and Hadi, 2005](#)) $ASAE = (1/n) \sum_{i=1}^n (|x_{(i)} - \hat{x}_{(i)}|) / (x_{(n)} - x_{(1)})$, where $x_{(i)}$ is the observed value of i th order statistic, and $\hat{x}_{(i)}$ is obtained from the QF, $Q(u_i)$, wherein the ML estimates are substituted to the parameters, with $u_i = i/(n+1)$. The ASAE statistic is useful for measuring the accuracy of the fitted model. In general, the smaller values of the above statistics indicate a better fit to the data.

Table 2. Monte Carlo simulation results for the listed statistical indicator.

Parameter	n	Bias	MSE	CP	AW
I					
θ	50	-0.044	0.112	0.92	1.483
	100	-0.037	0.045	0.95	0.979
	200	-0.037	0.033	0.98	0.749
α	50	0.626	1.690	0.96	2.079
	100	0.419	0.429	0.95	0.983
	200	0.314	0.110	0.95	0.799
λ	50	-0.028	0.193	0.93	1.459
	100	-0.042	0.111	0.96	1.154
	200	-0.046	0.079	0.95	0.123
II					
θ	50	0.022	0.007	0.93	0.368
	100	0.012	0.003	0.96	0.232
	200	0.004	0.001	0.95	0.153
α	50	0.621	1.340	0.95	4.809
	100	0.199	0.537	0.95	2.475
	200	0.078	0.167	0.95	1.588
λ	50	0.162	0.293	0.91	2.117
	100	0.080	0.133	0.94	1.436
	200	0.026	0.063	0.95	0.994
III					
θ	50	1.317	0.589	0.98	1.508
	100	0.609	0.371	0.98	1.192
	200	0.288	0.148	0.96	0.506
α	50	1.375	0.473	0.90	1.624
	100	0.563	0.171	0.98	1.270
	200	0.157	0.049	0.95	0.014
λ	50	0.264	0.479	0.91	1.006
	100	0.204	0.278	0.95	0.214
	200	0.199	0.130	0.96	0.102
IV					
θ	50	0.638	3.602	0.90	2.835
	100	0.237	1.276	0.91	1.401
	200	0.141	0.629	0.94	0.038
α	50	-0.003	0.083	0.96	1.156
	100	0.015	0.042	0.96	0.818
	200	-0.001	0.021	0.95	0.571
λ	50	0.117	0.255	0.96	1.977
	100	0.035	0.104	0.96	1.323
	200	0.024	0.055	0.96	0.923

The CDFs of the BL, EPL, BE, ENH, HEE, EW, MOL, PL and EL distributions are given by

$$F_{\text{BL}}(x, a, b, \theta) = I_{1 - (1 + \frac{\theta x}{1 + \theta})e^{-\theta x}}(a, b), \quad x, \theta > 0,$$

$$F_{\text{EPL}}(x, \alpha, \beta, \theta) = \left(1 - \left(1 + \frac{\theta x^\beta}{1 + \theta}\right)e^{-\theta x^\beta}\right)^\alpha, \quad x, \alpha, \beta, \theta > 0,$$

$$F_{\text{BE}}(x, a, b, \lambda) = I_{1 - e^{-\lambda x}}(a, b), \quad x, a, b, \lambda > 0.$$

$$F_{\text{ENH}}(x, \beta, \alpha, \lambda) = \left(1 - e^{1 - (1 + \lambda x)^\alpha}\right)^\beta, \quad x, \beta, \alpha, \lambda > 0,$$

$$F_{\text{HEE}}(x, \beta, k, \lambda) = \frac{\beta^{1/k} e^{-\lambda x}}{[1 - (1 - \beta)e^{-\lambda k x}]^{1/k}}, \quad x, \beta, k, \lambda > 0,$$

$$F_{\text{EW}}(x; c, \alpha, \lambda) = \left(1 - e^{-(x/\lambda)^c}\right)^\alpha, \quad x, c, \alpha, \lambda > 0,$$

$$F_{\text{MOL}}(x, \alpha, \lambda) = \frac{1 - (1 + \lambda)^{-1}[1 + \lambda + \lambda x]e^{-\lambda x}}{1 - (1 - \alpha)(1 + \lambda)^{-1}[1 + \lambda + \lambda x]e^{-\lambda x}}, \quad x, \alpha, \lambda > 0,$$

$$F_{\text{PL}}(x, \beta, \theta) = 1 - \left(1 + \frac{\theta x^\beta}{1 + \theta}\right)e^{-\theta x^\beta}, \quad x, \beta, \theta > 0,$$

$$F_{\text{EL}} = \left[1 - \left(\frac{1 + \theta + \theta x}{1 + \theta}\right)e^{-\theta x}\right]^\alpha, \quad x, \theta > 0,$$

respectively, where $I_z(p, q)$ denotes the incomplete beta function.

First, we consider a data set consisting of s exceedances (rounded to one decimal place) of flood peaks (in m^3/s) of the Wheaton river, which is located in the Yukon Territory, Canada, for the years 1958-1984. The data set is the following: 1.7, 2.2, 14.4, 1.1, 0.4, 20.6, 5.3, 0.7, 1.9, 13.0, 12.0, 9.3, 1.4, 18.7, 8.5, 25.5, 11.6, 14.1, 22.1, 1.1, 2.5, 14.4, 1.7, 37.6, 0.6, 2.2, 39.0, 0.3, 15.0, 11.0, 7.3, 22.9, 1.7, 0.1, 1.1, 0.6, 9.0, 1.7, 7.0, 20.1, 0.4, 2.8, 14.1, 9.9, 10.4, 10.7, 30.0, 3.6, 5.6, 30.8, 13.3, 4.2, 25.5, 3.4, 11.9, 21.5, 27.6, 36.4, 2.7, 64.0, 1.5, 2.5, 27.4, 1.0, 27.1, 20.2, 16.8, 5.3, 9.7, 27.5, 2.5, 27.0. Some summary statistics of these data are: $n = 72$, $\bar{x} = 12.20417$, $s = 12.29722$, coefficient of skewness = 1.47251 and coefficient of kurtosis = 2.88955. The boxplot of these observations displayed in Figure 3(a) indicates that the distribution is right-skewed. The TTT (total time on test) plot (see, e.g., Gill, 1986; Aarset, 1987) of these data is shown in Figure 3(b). It is first convex and then concave, which suggests a bathtub-shaped failure rate. Accordingly, the HEL distribution could, in principle, be appropriate for modeling these data. The ML estimates (with the corresponding standard errors -SEs- in parentheses) as well as the ASAE, AIC, KS, CM and AD statistics are given in Table 3. All five goodness-of-fit statistics indicate that the HEL model provides the best fit. For a visual comparison, the empirical SF (ESF) and estimated SF associated with the HEL model as well as a theoretical versus empirical probability (PP) plot, which compares the empirical CDF of the data with the fitted CDF, are respectively included in Figures 4(a) and 4(b). Clearly, the HEL model closely fits the data distribution.

In this second illustration, the data set, which is freely available on the Korea Meteorological Administration (KMA) website (<http://www.kma.go.kr>), represents the annual maximum daily rainfall amounts in millimeters in Seoul (Korea) during the period 1961-2002. Some summary statistics of these precipitation data are: $n = 128$, $\bar{x} = 144.5991$, $s = 66.17812$, coefficient of skewness = 0.94067 and coefficient of kurtosis = 0.80435. The boxplot of these observations that is displayed in Figure 5(a) indicates that the distribution is right-skewed. The TTT plot appearing in Figure 5(b) suggests an increasing failure

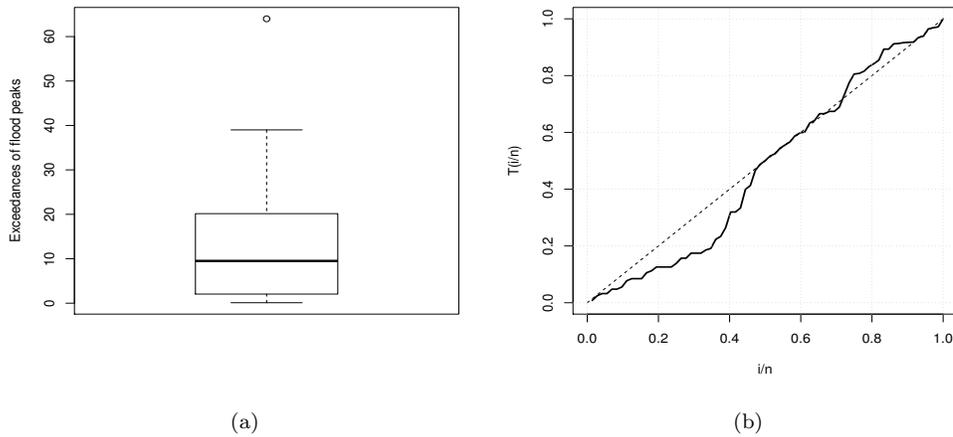


Figure 3. Boxplot (a) and TTT plot (b) for the flood data.

Table 3. ML estimates, SEs (in parentheses) and goodness-of-fit measures for the flood data.

Distribution	Estimates			ASAE	AIC	KS	CM	AD
HEL(θ, α, λ)	0.077 (0.038)	6.135 (2.031)	0.110 (0.014)	0.017	503.194	0.073	0.054	0.338
BL(a, b, θ)	0.556 (0.098)	0.275 (0.241)	0.334 (0.273)	0.020	510.206	0.115	0.126	0.775
EPL(α, β, θ)	0.916 (0.595)	0.730 (0.235)	0.300 (0.279)	0.025	510.425	0.106	0.149	0.857
BE(a, b, λ)	0.812 (0.137)	0.412 (0.290)	0.179 (0.131)	0.023	508.465	0.098	0.122	0.705
ENH(β, α, λ)	0.732 (0.137)	1.675 (0.143)	0.032 (0.032)	0.019	507.850	0.106	0.104	0.632
HEE(β, k, λ)	0.433 (0.193)	5.086 (0.147)	0.071 (0.011)	0.023	506.460	0.078	0.094	0.550
EW(c, α, λ)	1.387 (0.587)	0.519 (0.308)	0.016 (0.036)	0.403	508.050	0.107	0.105	0.642
MOL(α, λ)	0.216 (0.128)	0.090 (0.023)		0.044	522.571	0.175	0.582	4.148
PL(β, θ)	0.700 (0.057)	0.339 (0.056)		0.026	508.444	0.105	0.154	0.877
EL(α, θ)	0.509 (0.077)	0.104 (0.015)		0.021	509.349	0.117	0.135	0.833
L(θ)	0.153 (0.013)			0.044	530.424	0.241	0.819	7.424

rate. The estimates of the parameters of the fitted distributions are listed in Table 4. We note that the HEL model has the lowest ASAE, AIC, KS, CM and AD values, which indicate that it provides the most accurate fit to the data. Furthermore, the ESF and estimated SF and PP plots shown in Figures 6(a) and 6(b) also suggest a close fit to the data distribution.

A likelihood ratio test can be utilized to compare a distribution having additional parameters with some of its sub-models. Accordingly, we made use of the likelihood ratio test to assess the improvement in fit that the HEL distribution produces with respect to

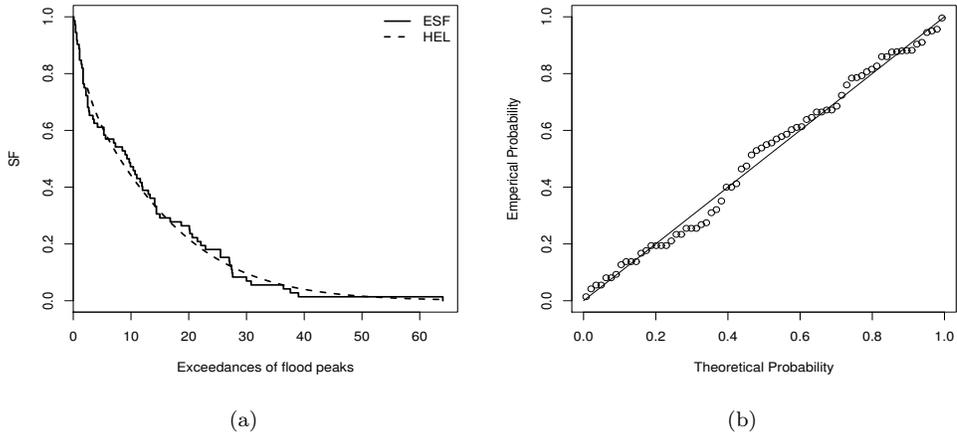


Figure 4. Empirical SF and estimated HEL SF (a) and PP plot (b) for the flood data.

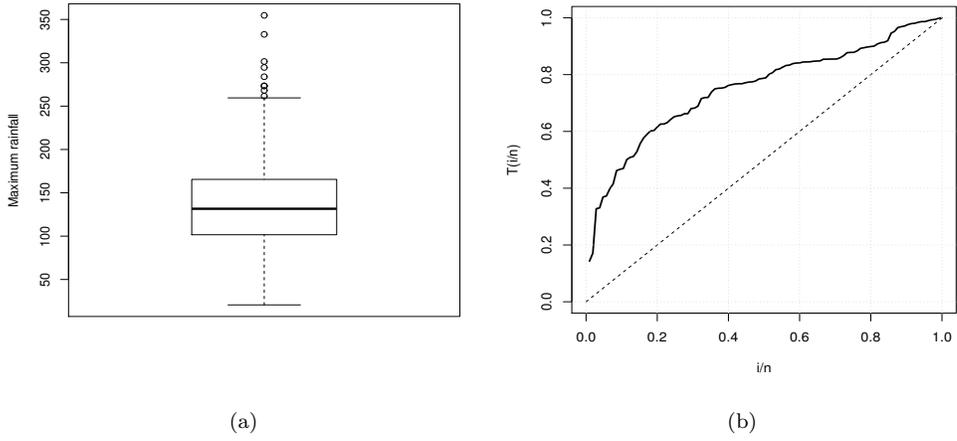


Figure 5. Boxplot (a) and TTT plot (b) for the precipitation data.

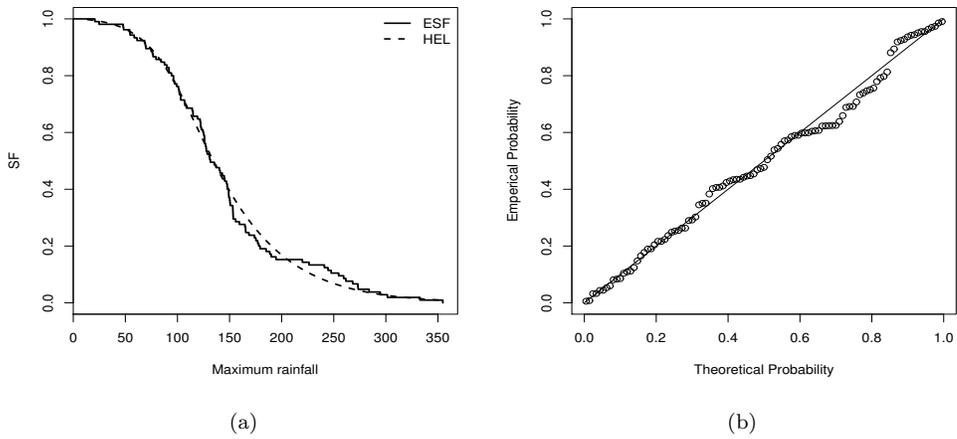


Figure 6. Empirical SF and estimated HEL SF (a) and PP plot (b) for the precipitation data.

Table 4. ML estimates, SEs (in parentheses) and goodness-of-fit measures for the precipitation data.

Distribution	Estimates			ASAE	AIC	KS	CM	AD
HEL(θ, α, λ)	17.443 (9.276)	3.081 (1.069)	0.022 (0.003)	0.019	1165.064	0.077	0.077	0.490
BL(a, b, θ)	2.776 (0.622)	1.117 (0.577)	0.020 (0.007)	0.022	1169.396	0.085	0.144	0.809
EPL(α, β, θ)	1.530 (0.225)	1.318 (0.025)	0.003 (0.004)	0.024	1168.717	0.097	0.150	0.862
BE(a, b, λ)	4.433 (0.685)	1.448 (0.535)	0.012 (0.003)	0.029	1172.022	0.092	0.263	1.412
ENH(β, α, λ)	4.183 (0.687)	1.694 (0.217)	0.006 (0.001)	0.024	1168.620	0.095	0.146	0.837
HEE(β, k, λ)	1.535 (0.299)	1.860 (0.847)	0.008 (0.001)	0.137	1241.535	0.276	2.569	13.078
EW(c, α, λ)	1.411 (0.334)	2.907 (1.519)	98.866 (29.851)	0.433	1168.586	0.093	0.142	0.821
MOL(α, λ)	10.455 (4.118)	0.029 (0.003)		0.032	1171.003	0.103	0.184	1.330
PL(β, θ)	0.014 (0.007)	16.182 (2.037)		1.433	4820.512	0.999	34.999	1631.130
EL(α, θ)	2.871 (0.501)	0.022 (0.002)		0.022	1167.600	0.084	0.146	0.818
L(θ)	0.014 (0.001)			0.584	1199.216	1.187	0.519	6.508

the Lindley and MOL distributions. It is known that, under the null hypothesis,

$$-2 \log \left(\frac{\text{likelihood under the null hypothesis}}{\text{likelihood in the whole parameter space}} \right) \sim \chi^2(d),$$

where, asymptotically, $\chi^2(d)$ follows a chi-square distribution having d degrees of freedom, d being equal to the number of additional parameters in the extended model. Using this result and standard statistical tables, we can obtain critical values for the test statistic. Table 5 includes the likelihood ratio statistics and corresponding p-values for the two data sets. Given the values of these statistics and their associated p-values, we reject the null hypotheses for both data sets and conclude that the HEL model provides a significantly better representation of the distribution of these data than the Lindley or MOL distributions. The 95% bootstrap confidence intervals obtained for the parameters θ, α and λ are given in Table 6.

Table 5. Likelihood ratio statistics and their p-values.

Hypothesis	Flood data	Precipitation data
H ₀ : $\alpha=1$ (MOL)	21.377 (< 0.000)	7.939 (0.005)
H ₁ : $\alpha \neq 1$ (HEL)		
H ₀ : $\alpha=\theta=1$ (L)	31.229 (< 0.000)	38.151 (<0.000)
H ₁ : $\alpha \neq 1, \theta \neq 1$ (HEL)		

Table 6. 95% bootstrap confidence intervals for the parameters θ , α and λ .

Data set	θ	α	λ
Flood data	(0.039, 0.225)	(3.036, 10.429)	(0.087, 0.146)
Precipitation data	(8.243, 20.463)	(1.378, 5.027)	(0.018, 0.031)

Next, we present the concepts of return period, mean deviation about a return level and the r th moment of the order statistics. For a given a data set, the return period can be estimated by $\hat{T} = 1/\bar{F}(x)$, where $\bar{F}(x) = 1 - F(x)$ and $F(x)$ denote the CDF of the distribution. The estimated return periods (\hat{T}) correspond to the return levels (x_T) for each of these two data sets. They are reported in Table 7 and have been computed as $T = 1/\bar{F}(x_T)$, where $\bar{F}(\cdot)$ is as given in Equation (4). The mean deviation about a return level which is the mean of the distances of the values from their return level is given by $\eta = 2x_T F(x_T) - x_T - \mu + 2 \int_{x_T}^{\infty} x f(x) dx$, where $f(\cdot)$ and $F(\cdot)$ denote the HEL PDF and CDF. Table 7 provides the mean deviations about certain values of the return levels (\bar{x}_T) for both the flood and precipitation data sets.

Table 7. Estimated return periods (\hat{T}) and mean deviations about the return levels (η).

Flood data			Precipitation data		
x_T	\hat{T}	η	x_T	\hat{T}	η
140	499147.836	127.800	410	315.215	265.623
100	8350.571	87.802	375.5	160.422	435.000
50	62.48360	38.135	315.5	50.389	172.849
30	10.375	19.949	260	17.693	121.247
10	2.265	9.337	210	7.093	80.513

In order to be able to plan for future emergencies in connection with various hydrological events, it is useful to ascertain some distributional results on certain of the order statistics. To that end, we determine the r th moment, for $r = 1, 2, 3, 4$, of some order statistics for each data sets under the HEL model wherein the parameters are replaced by their ML estimates. Those moments are included in Table 8 for each data set.

Table 8. Some numerical values of $E(X_{i:n}^r)$ for the indicated data set.

Flood data			Precipitation data		
i	r	$E(X_{i:72}^r)$	i	r	$E(X_{i:128}^r)$
1	1	0.097	1	1	21.409
	2	0.019		2	585.869
	3	0.006		3	18628.800
	4	0.002		4	658641.210
20	1	2.868	15	1	77.111
	2	8.962		2	5989.380
	3	30.433		3	468486.450
	4	111.999		4	3.689×10^4
60	1	22.898	30	1	98.427
	2	543.677		2	9719.320
	3	12726.600		3	962824.794
	4	308653.083		4	9.568×10^7

6. CONCLUDING REMARKS

We introduced a three-parameter extension of the Lindley distribution referred to as the Harris extended Lindley (HEL) distribution, which is obtained by applying the Harris extended method to the Lindley distribution. The proposed model has two shape parameters and one scale parameter. It includes as sub-models the Marshall-Olkin Lindley and Lindley distributions. The HEL PDF can be decreasing or unimodal. Moreover, the HEL HR can be increasing, decreasing, unimodal (upside-down bathtub) or bathtub-shaped. We gave explicit expressions for the ordinary and incomplete moments, mean deviations, Bonferroni and Lorenz curves and order statistics associated with the proposed distribution. The estimation of the model parameters was successfully carried out by making use of the maximum likelihood method. In conclusion, the HEL distribution provides a very flexible model for fitting the wide spectrum of positive data sets arising in engineering, survival analysis, hydrology, economics, biology as well as numerous other fields of scientific investigation. All the calculations were performed with the symbolic computing software Mathematica, the code being available from the authors upon request.

ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to the referees, an Associate Editor and the Editor-in-Chief for their valuable comments and suggestions which significantly improved the presentation of the results. The financial support of the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged by the the third author.

REFERENCES

- Aarset, M.V., 1987. How to identify bathtub hazard rate. *IEEE Transactions on Reliability*, 36, 106-108.
- Alizadeh, M., MirMostafaei, S.M.T.K., and Ghosh., I., 2017. A new extension of power Lindley distribution for analyzing bimodal data. *Chilean Journal of Statistics*, 8, 67-86.
- Aly, E.A.A. and Benkherouf, L., 2011. A new family of distributions based on probability generating functions. *Sankhyā B*, 73, 70-80.
- Ashkar, F. and El Adlouni, S., 2014. Correcting confidence intervals for quantiles of a heavy-tailed distribution: Case study of two-parameter kappa distribution. *Journal of Hydrology*, 512, 498-505.
- Ashour, S.K. and Eltehiwy, M.A., 2015. Exponentiated power Lindley distribution. *Journal of Advance Research*, 6, 895-905
- Bhunya, P.K., Singh, R.D., Berndtsson, R., and Panda, S.N., 2012. Flood analysis using generalized logistic models in partial duration series, *Journal of Hydrology*, 420, 59-71.
- Castillo, E., Hadi, A.S., Balakrishnan, N., and Sarabia, J.M., 2005. *Extreme Value and Related Models with Applications in Engineering and Science*, Wiley, Hoboken.
- Chadwick, A., Morfett, J., and Borthwick, M., 2004. *Hydraulics in Civil and Environmental Engineering*, Spon Press, London.
- Cordeiro, G.M., Alizadeh, M., Nascimento, A.D., and Rasekhi, M., 2016. The exponentiated Gompertz generated family of distributions: Properties and applications. *Chilean Journal of Statistics*, 7, 29-50
- Corless, R.M., Gonnet, G.H., Hare, D.E.G., Jeffrey, D.J., and Knuth, D.E., 1996. On the Lambert W function. *Advances in Computational Mathematics*, 5, 329-359.

- Ghitany, M.E., Al-Mutairi, D.K., and Aboukhamseen, S.M., 2015. Estimation of the reliability of a stress-strength system from power Lindley distributions. *Communications in Statistics - Theory and Methods*, 44, 118-136.
- Ghitany M.E., Al-Mutairi D.K., Al-Awadhi, F.A., and Al-Burais M.M., 2012. Marshall-Olkin extended Lindley distribution and its application. *International Journal of Applied Mathematics*, 25, 709-721.
- Ghitany, M.E., Al-Mutairi, D.K., Balakrishnan, N., and Al-Enezi, L.J., 2013. Power Lindley distribution and associated inference. *Computational Statistics and Data Analysis*, 64, 20-33.
- Ghitany, M.E., Alqallaf, F., Al-Mutairi, D.K. and Husain, H.A., 2011. A two-parameter weighted Lindley distribution and its applications to survival data. *Mathematics and Computers in Simulation*, 81, 1190-1201.
- Ghitany, M.E., Atieh, B., and Nadarajah, S., 2008. Lindley distribution and its application. *Mathematics in Computer and Simulation*, 78, 493-506.
- Gill, D.R. (1986). The total time on test plot and the cumulative total time on test statistic for a counting process. *The Annals of Statistics*, 14, 1234-1239.
- Gradshteyn, I.S. and Ryzhik, I.M., 2000. *Tables of Integrals, Series, and Products*, Academic Press, New York.
- Harris, T.E., 1948. Branching processes. *Annals of Mathematical Statistics*, 19, 474-494.
- Heo, J. and Boes, D.C., 2011. Regional flood frequency analysis based on a Weibull model: Part 2. Simulations and applications. *Journal of Hydrology*, 242, 171-182.
- Jodrá, J., 2010. Computer generation of random variables with Lindley or Poisson-Lindley distribution via the Lambert W function. *Mathematics and Computers in Simulation*, 81, 851-859.
- Kang, D., Ko, K., and Huh, J., 2015. Determination of extreme wind values using the Gumbel distribution. *Energy*, 86, 51-58.
- Lemonte, A.J., 2013. A new exponential-type distribution with constant, decreasing, increasing, upside-down bathtub and bathtub-shaped failure rate function. *Computational Statistics and Data Analysis*, 62, 149-170.
- Li, C., Singh, V.P., and Mishra, A.K. 2013. A bivariate mixed distribution with heavy-tailed component and its application to single-site daily rainfall simulation. *Water Resource Research*. 49, 767-789.
- Lindley, D.V., 1958. Fiducial distributions and Bayes theorem. *Journal of Royal Statistical Society B*, 20, 102-107.
- Marshall, A.N. and Olkin, I., 1997. A new method for adding a parameter to a family of distributions with applications to the exponential and Weibull families. *Biometrika*, 84, 641-652.
- Mazucheli, J., Coelho-Barros, E.A., and Louzada, F., 2016. On the hypothesis testing for the weighted Lindley distribution. *Chilean Journal of Statistics*, 7, 17-27.
- Mervoci, F. and Sharma, V.K., 2014. The beta Lindley distribution: Properties and Applications. *Journal of Applied Mathematics*, Art.ID 198951, 10 p.
- Mudholkar, G.S. and Srivastava, D.K., 1993. Exponentiated Weibull family for analyzing bathtub failure data. *IEEE Transactions on Reliability*, 42, 299-302.
- Nadarajah, S., Bakouch, H., and Tahmasbi, R., 2007. A generalized Lindley distribution. *Sankhya B*, 73, 331-359.
- Nadarajah, S. and Kotz, S., 2006. The beta exponential distribution. *Reliability Engineering and System Safety*, 91, 689-697.
- Pinho, L.G.B., Cordeiro, G.M., and Nobre, J.S., 2015. The Harris extended exponential distribution. *Communications in Statistics - Theory and Methods*, 44, 3486-3502.
- Zelenhasic, E.F., 1970. Theoretical probability distributions for flood peaks. *Hydrology papers (Colorado State University)*, ID 42.

INFORMATION FOR AUTHORS

The editorial board of the Chilean Journal of Statistics (ChJS) is seeking papers, which will be refereed. We encourage the authors to submit a PDF electronic version of the manuscript in a free format to Víctor Leiva, Editor-in-Chief of the ChJS (E-mail: chilean.journal.of.statistics@gmail.com). Submitted manuscripts must be written in English and contain the name and affiliation of each author followed by a leading abstract and keywords. The authors must include a "cover letter" presenting their manuscript and mentioning: "We confirm that this manuscript has been read and approved by all named authors. In addition, we declare that the manuscript is original and it is not being published or submitted for publication elsewhere".

PREPARATION OF ACCEPTED MANUSCRIPTS

Manuscripts accepted in the ChJS must be prepared in Latex using the ChJS format. The Latex template and ChJS class files for preparation of accepted manuscripts are available at <http://chjs.mat.utfsm.cl/files/ChJS.zip>. Such as its submitted version, manuscripts accepted in the ChJS must be written in English and contain the name and affiliation of each author, followed by a leading abstract and keywords, but now mathematics subject classification (primary and secondary) are required. AMS classification is available at <http://www.ams.org/mathscinet/msc/>. Sections must be numbered 1, 2, etc., where Section 1 is the introduction part. References must be collected at the end of the manuscript in alphabetical order as in the following examples:

Arellano-Valle, R., (1994). Elliptical Distributions: Properties, Inference and Applications in Regression Models. Unpublished Ph.D. Thesis. Department of Statistics, University of São Paulo, Brazil.

Cook, R.D., (1997). Local influence. In Kotz, S., Read, C.B. and Banks, D.L. (Eds.), Encyclopedia of Statistical Sciences, Update, Vol. 1, Wiley, New York, pp. 380-385.

Rukhin, A.L., (2009). Identities for negative moments of quadratic forms in normal variables. Statistics and Probability Letters 79, 1004-1007.

Stein, M.L., (1999). Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.

Tsay, R.S., Pena, D., and Pankratz, A.E., (2000). Outliers in multivariate time series. Biometrika 87, 789-804.

References in the text must be given by the author's name and year of publication, e.g., Gelfand and Smith (1990). In the case of more than two authors, the citation must be written as Tsay et al. (2000).

COPYRIGHT

Authors who publish their articles in the ChJS automatically transfer their copyright to the Chilean Statistical Society. This enables full copyright protection and wide dissemination of the articles and the journal in any format. The ChJS grants permission to use figures, tables and brief extracts from its collection of articles in scientific and educational works, in which case the source that provides these issues (Chilean Journal of Statistics) must be clearly acknowledged.