



symmetry

Symmetric and Asymmetric Distributions

Edited by

Emilio Gómez-Déniz

Printed Edition of the Special Issue Published in *Symmetry*

Symmetric and Asymmetric Distributions

Symmetric and Asymmetric Distributions

Theoretical Developments and Applications

Editor

Emilio Gómez-Déniz

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin



Editor

Emilio Gómez-Déniz
Department of Quantitative
Methods and TIDES Institute,
University of Las Palmas de
Gran Canaria
Spain

Editorial Office

MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Symmetry* (ISSN 2073-8994) (available at: https://www.mdpi.com/journal/symmetry/special_issues/Symmetric_Asymmetric_Distributions_Theoretical_Developments_Applications).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. <i>Journal Name</i> Year , Article Number, Page Range.

ISBN 978-3-03936-646-0 (Hbk)

ISBN 978-3-03936-647-7 (PDF)

© 2020 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license, which allows users to download, copy and build upon published articles, as long as the author and publisher are properly credited, which ensures maximum dissemination and a wider impact of our publications.

The book as a whole is distributed by MDPI under the terms and conditions of the Creative Commons license CC BY-NC-ND.

Contents

About the Editor	vii
Preface to “Symmetric and Asymmetric Distributions”	ix
Emilio Gómez-Déniz, Yuri A. Iriarte, Enrique Calderín-Ojeda and Héctor W. Gómez Modified Power-Symmetric Distribution Reprinted from: <i>Symmetry</i> 2019, 11, 1410, doi:10.3390/sym11111410	1
Fábio V. J. Silveira, Frank Gomes-Silva, Cícero C. R. Brito, Moacyr Cunha-Filho, Felipe R. S. Gusmão and Sílvio F. A. Xavier-Júnior Normal- <i>G</i> Class of Probability Distributions: Properties and Applications Reprinted from: <i>Symmetry</i> 2019, 11, 1407, doi:10.3390/sym11111407	17
Héctor J. Gómez, Diego I. Gallardo and Osvaldo Venegas Generalized Truncation Positive Normal Distribution Reprinted from: <i>Symmetry</i> 2019, 11, 1361, doi:10.3390/sym11111361	35
Tiago M. Magalhães, Diego I. Gallardo and Héctor W. Gómez Skewness of Maximum Likelihood Estimators in the Weibull Censored Data Reprinted from: <i>Symmetry</i> 2019, 11, 1351, doi:10.3390/sym11111351	53
Guillermo Martínez-Flórez, Inmaculada Barranco-Chamorro, Heleno Bolfarine and Héctor W. Gómez Flexible Birnbaum–Saunders Distribution Reprinted from: <i>Symmetry</i> 2019, 11, 1305, doi:10.3390/sym11101305	63
Neveka M. Olmos, Osvaldo Venegas, Yolanda M. Gómez and Yuri Iriarte An Asymmetric Distribution with Heavy Tails and Its Expectation–Maximization (EM) Algorithm Implementation Reprinted from: <i>Symmetry</i> 2019, 11, 1150, doi:10.3390/sym11091150	81
Yolanda M. Gómez, Emilio Gómez-Déniz, Osvaldo Venegas, Diego I. Gallardo and Héctor W. Gómez An Asymmetric Bimodal Distribution with Application to Quantile Regression Reprinted from: <i>Symmetry</i> 2019, 11, 899, doi:10.3390/sym11070899	95
Barry C. Arnold, Héctor W. Gómez, Héctor Varela and Ignacio Vidal Univariate and Bivariate Models Related to the Generalized Epsilon–Skew–Cauchy Distribution Reprinted from: <i>Symmetry</i> 2019, 11, 794, doi:10.3390/sym11060794	107
V. J. García, M. Martel–Escobar and F. J. Vázquez–Polo A Note on Ordering Probability Distributions by Skewness Reprinted from: <i>Symmetry</i> 2018, 10, 286, doi:10.3390/sym10070286	121

About the Editor

Emilio Gómez-Déniz (M.S. in Mathematics, M.S. in Economics, and Ph.D. in Economics) is professor of Mathematics at the University of Las Palmas de Gran Canaria (Spain), where he has taught and conducted research for more than 30 years. He is also member of the Institute of Tourism and Sustainable Economic Development (TIDES). His main lines of research focus on distribution theory, Bayesian statistics, robustness, and Bayesian applications in economics, with emphasis in actuarial statistics, tourism, education, sports, electronic engineering, health economics, etc. He has published more than 150 articles, most of them with JCR impact factor according to Web of Science and his work has been cited more than 1000 times on Google Scholar with an h-index of 16 and an i10 index of 35. He is currently the editor or member of the editorial board of the Spanish Journal of Statistics (SJS), Chilean Journal of Statistics, Mathematical Problems in Engineering, Risks, Risk Magazine, and Journal of Financial and Risk Management (Reviewer Board) and Statistical Methodology and Mathematical Reviews in the past. He has also acted and regularly acts as reviewer for more than 100 international journals. His academic leadership is further reflected through collaborative research with more than 50 leading international scientists worldwide. He is the author of many books on mathematics education and research, including "Actuarial Statistics. Theory and Applications", Prentice Hall (in Spanish). He has organized and participated in many international conferences and seminars. In addition, he has visited, as guest researcher and speaker, the University of Melbourne, the University of Kuwait, the University of Cantabria, the University of Antofagasta (Chile), the University of Barcelona, and the University of Granada, among others. In 2008, he was awarded the prestigious IV Julio Castelo Matrán International Insurance Award, sponsored by the Mapfre Foundation and also recognized by his university for excellence in teaching.

Preface to "Symmetric and Asymmetric Distributions"

This Special Issue contains nine chapters selected after a comprehensive peer review process. Each chapter exclusively adheres to the topic specified in this Special Issue. The Associate Editor wants to especially thank all the authors who made this Special Issue possible, and all the anonymous reviewers for their altruistic effort and help in reviewing and for their excellent suggestions and critical reviews of the submitted manuscripts. All the chapters are written in the format of a research article for scholarly journals, i.e., title, abstract, and development, with an extensive bibliography at the end of the paper. All of them include real applications that will undoubtedly be useful for other researchers and graduate students who conduct similar research. I express my gratitude to MPDI for publishing this book, to Ms. Celina Si, Section Managing Editor, for her work and patience, and last but not least to my family for their support and cooperation.

Emilio Gómez-Déniz

Editor

Modified Power-Symmetric Distribution

Emilio Gómez-Déniz ¹, Yuri A. Iriarte ², Enrique Calderín-Ojeda ³ and Héctor W. Gómez ^{2,*}

¹ Department of Quantitative Methods in Economics and TIDES Institute, University of Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Spain; emilio.gomez-deniz@ulpgc.es

² Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta 1240000, Chile; yuri.iriarte@uantof.cl

³ Centre for Actuarial Studies, Department of Economics, The University of Melbourne, Melbourne, VIC 3010, Australia; enrique.calderin@unimelb.edu.au

* Correspondence: hector.gomez@uantof.cl

Received: 2 October 2019; Accepted: 13 November 2019; Published: 15 November 2019

Abstract: In this paper, a general class of modified power-symmetric distributions is introduced. By choosing as symmetric model the normal distribution, the modified power-normal distribution is obtained. For the latter model, some of its more relevant statistical properties are examined. Parameters estimation is carried out by using the method of moments and maximum likelihood estimation. A simulation analysis is accomplished to study the performance of the maximum likelihood estimators. Finally, we compare the efficiency of the modified power-normal distribution with other existing distributions in the literature by using a real dataset.

Keywords: maximum likelihood; kurtosis; power-normal distribution

1. Introduction

Over the last few years, the search for flexible probabilistic families capable of modeling different levels of bias and kurtosis has been an issue of great interest in the field of distributions theory. This was mainly motivated by the seminal work of Azzalini [1]. In that paper, the probability density function (pdf) of a skew-symmetric distribution was introduced. The expression of this density is given by

$$g(z; \lambda) = 2f(z)G(\lambda z), \quad z, \lambda \in \mathbb{R}, \quad (1)$$

where $f(\cdot)$ is a symmetric pdf about zero; $G(\cdot)$ is an absolutely continuous distribution function, which is also symmetric about zero; and λ is a parameter of asymmetry. For the case where $f(\cdot)$ is the standard normal density (from now on, we reserve the symbol ϕ for this function), and $G(\cdot)$ is the standard normal cumulative distribution function (henceforth, denoted by Φ), the so-called skew-normal (\mathcal{SN}) distribution with density

$$\phi_Z(z; \lambda) = 2\phi(z)\Phi(\lambda z), \quad z, \lambda \in \mathbb{R}, \quad (2)$$

is obtained. We use the notation $Z \sim \mathcal{SN}(\lambda)$ to denote the random variable Z with pdf given by Equation (2). A generalization of the \mathcal{SN} distribution is introduced by Arellano-Valle et al. [2] and Arellano-Valle et al. [3]; they study Fisher's information matrix of this generalization. For further details about the \mathcal{SN} distribution, the reader is referred to Azzalini [4]. Martínez-Flórez et al. [5] used generalizations of the \mathcal{SN} distribution to extend the Birnbaum-Saunders model, and Contreras-Reyes and Arellano-Valle [6] utilized the Kullback–Leibler divergence measure to compare the multivariate normal distribution with the skew-multivariate normal.

One of the main limitations of working with the family given by Equation (1) is that the information matrix could be singular for some of its particular models (see Azzalini [1]). This might

lead to some difficulties in the estimation, due to the asymptotic convergence of the maximum likelihood (ML) estimators. To overcome this issue, some authors (see Chiogna [7] or Arellano-Valle and Azzalini [8]) have used a reparametrization of the \mathcal{SN} model to obtain a nonsingular information matrix. However, this methodology cannot be extended to all type of skew-symmetric models which suffers of this convergence problem. On the other hand, the family of power-symmetric (\mathcal{PS}) distributions does not have this problem of singularity in the information matrix (see, Pewsey et al. [9]). The pdf of this family of distribution is given by

$$\varphi_F(z; \alpha) = \alpha f(z) \{F(z)\}^{\alpha-1}, \quad z \in \mathbb{R}, \alpha \in \mathbb{R}^+. \quad (3)$$

where $F(\cdot)$ is itself a cumulative distribution function (cdf) and α is the shape parameter. For the particular case that $F(\cdot) = \Phi(\cdot)$, the power-normal (\mathcal{PN}) distribution is obtained, with density given by

$$f(z; \alpha) = \alpha \phi(z) \{\Phi(z)\}^{\alpha-1}, \quad z \in \mathbb{R}, \alpha \in \mathbb{R}^+. \quad (4)$$

For some references where this family is discussed, the reader is referred to Lehmann [10], Durrans [11], Gupta and Gupta [12], and Pewsey et al. [9], among other papers. Other extensions of this model are given in Martínez-Flórez et al. [13], where a multivariate version from the model is introduced; also, Martínez-Flórez et al. [14] carried out applications by using regression models; finally, Martínez-Flórez et al. [15] examined the exponential transformation of the model, and Martínez-Flórez et al. [16] examined a version of the model doubly censored with inflation in a regression context. Truncations of the \mathcal{PN} distribution were considered by Castillo et al. [17].

In this paper, a modification in the pdf of the \mathcal{PS} probabilistic family is implemented to increase the degree of kurtosis. This methodology is later used to explain datasets that include atypical observations. Usually, this methodology is accomplished by increasing the number of parameters in the model.

The paper is organized as follows. In Section 2, first, we introduce the modified power symmetric distribution. Then, the particular case of the modified power normal distribution is derived. Some of the most relevant statistical properties of this model, including moments and kurtosis coefficient, are presented. Next, in Section 3, some methods of estimation are discussed. Later, a simulation study is provided to illustrate the behavior of the shape parameter. A numerical application where the modified power normal distribution is compared to the \mathcal{SN} and \mathcal{PN} distributions is given in Section 4. Finally, Section 5 concludes the paper.

2. Genesis and Properties of Modified Power-Normal Distribution

In this section, we introduce a new family of probability distributions. The idea is to make a transformation to a given probability density, as the skew-symmetric or power-symmetric distributions does. As there exists a certain resemblance between our formula (Equation (6)) and the formula for the power-symmetric distributions (Equation (3)), we agree to name these new distributions as modified power-symmetric (\mathcal{MPS}) distributions. From the standard normal distribution, we obtain the so-called Modified Power-Normal (\mathcal{MPN}) distribution. The main parameters and properties of this particular distribution will be studied throughout this work.

2.1. Probability Density Function

Definition 1. Let Z be a continuous and symmetric random variable with cdf $G(z; \boldsymbol{\eta})$ and pdf $g(z; \boldsymbol{\eta})$, where $\boldsymbol{\eta}$ denotes a vector of parameters. We say that, a random variable, X , follows a \mathcal{MPS} distribution, denoted as $X \sim \mathcal{MPS}(\boldsymbol{\eta}, \alpha)$, if its cdf is given by

$$F(x; \boldsymbol{\eta}, \alpha) = \frac{[1 + G(x; \boldsymbol{\eta})]^\alpha - 1}{2^\alpha - 1}, \quad (5)$$

and its pdf is given by

$$f(x; \eta, \alpha) = \frac{\alpha}{2^\alpha - 1} g(x; \eta) [1 + G(x; \eta)]^{\alpha-1}. \quad (6)$$

where $x \in \mathbb{R}$ and $\alpha > 0$.

Remark 1. In the case $\alpha = 1$, the transformation given by Equation (6) is the identity. That is, the MPN distribution for $\alpha = 1$ always provides the input probability density function.

Therefore, we proceed to examine the MPN distribution, whose cdf is provided by

$$F(x; \mu, \sigma, \alpha) = \frac{\left[1 + \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^\alpha - 1}{2^\alpha - 1}, \quad (7)$$

and whose pdf is given by

$$f(x; \mu, \sigma, \alpha) = \frac{\alpha}{(2^\alpha - 1)\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) \left[1 + \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha-1}, \quad (8)$$

where $x \in \mathbb{R}$, $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter, and $\alpha > 0$ is the shape parameter. Hereafter, this will be denoted as $X \sim \mathcal{MPN}(\mu, \sigma, \alpha)$. Figure 1 depicts some different shapes of the pdf of this model, for selected values of the parameter α with $\mu = -1, 1$ and $\sigma = 1$. The MPN class of distributions is applicable for the change point problem, due to its favorable properties (see Maciak et al. [18]); moreover, the MPN model can be utilized in calibration (see Pešta [19]).

Remark 2. Here, $\mu \in \mathbb{R}$ and $\sigma > 0$ are location and scale parameters of the MPN distribution, respectively. For the particular case $\alpha = 1$, these are not only location and scale parameters but also the mean and standard deviation of the standard normal distribution.

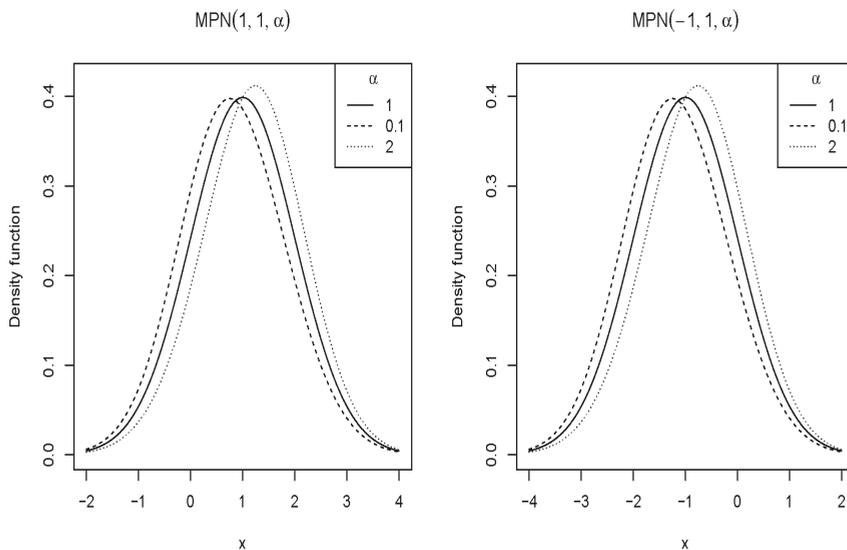


Figure 1. Plot of the pdf of MPN distribution for selected values of the parameters.

2.2. Statistical Properties

2.2.1. Shape of the Density

The \mathcal{MPN} distribution exhibits a bell-shaped form, which can be symmetric or positively or negatively skewed depending on the value of the parameter α . Now, we derive some analytical expressions that are useful to obtain approximations of modal values and inflection points of this model. In the following, it will be assumed that $\mu = 0$ and $\sigma = 1$.

Proposition 1. *The pdf of $X \sim \mathcal{MPN}(0, 1, \alpha)$ has a local maximum at $(x_1, f(x_1; \alpha))$ and two inflection points at $(x_2, f(x_2; \alpha))$ and $(x_3, f(x_3; \alpha))$, respectively, where x_1 is the root of the equation*

$$x^* = \frac{(\alpha - 1)\phi(x^*)}{1 + \Phi(x^*)}, \tag{9}$$

and x_2 and x_3 are two solutions of the equation

$$1 = \left(-x + \frac{(\alpha - 1)\phi(x)}{1 + \Phi(x)}\right)^2 - \frac{(\alpha - 1)\phi(x)}{1 + \Phi(x)} \left(x + \frac{\phi(x)}{1 + \Phi(x)}\right). \tag{10}$$

Proof. The proof consists of simple derivatives of the function f . From the equation (8), we calculate

$$\begin{aligned} \frac{\partial}{\partial x} f(x; \alpha) &= \frac{\alpha}{2^\alpha - 1} \phi(x) [1 + \Phi(x)]^{\alpha - 1} \left(-x + \frac{(\alpha - 1)\phi(x)}{1 + \Phi(x)}\right). \\ \frac{\partial^2}{\partial x^2} f(x; \alpha) &= \frac{\alpha}{2^\alpha - 1} \phi(x) [1 + \Phi(x)]^{\alpha - 1} \left\{ \left(-x + \frac{(\alpha - 1)\phi(x)}{1 + \Phi(x)}\right)^2 - \left[1 + \frac{(\alpha - 1)\phi(x)}{1 + \Phi(x)}\right] \right. \\ &\quad \left. \times \left(x + \frac{\phi(x)}{1 + \Phi(x)}\right) \right\}. \end{aligned}$$

By setting Equations (9) and (10) to be equal to zero, the results are obtained after some algebra. Figure 2 displays the graph of the first derivative of $f(\cdot)$, where it is observed that the maximum exists and it is unique. Therefore, the \mathcal{MPN} distribution is unimodal. □

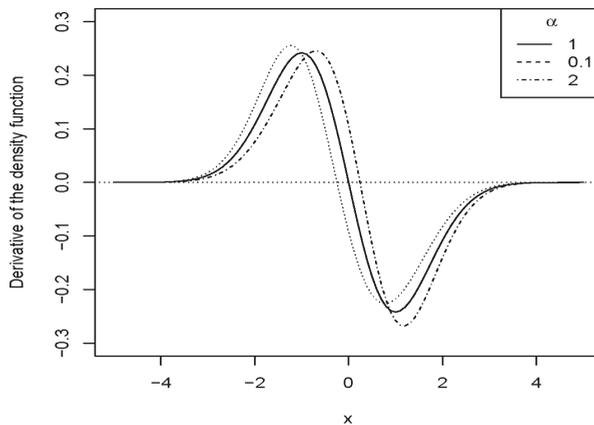


Figure 2. Plot of the first derivative of \mathcal{MPN} distribution for selected values of the parameters.

Remark 3. The solutions of Equations (9) and (10) can be numerically obtained by using the built-in function “uniroot” in the software package R. Table 1 below illustrates some approximations of the roots x_1 , x_2 , and x_3 , and the corresponding figures of the pdf evaluated at these values.

Table 1. Approximations of the roots of Equations (9) and (10) for some values of α , and the corresponding figures of the pdf of the \mathcal{MPN} evaluated at these roots.

α	x_1	x_2	x_3	$f(x_1; \alpha)$	$f(x_2; \alpha)$	$f(x_3; \alpha)$
0.5	-0.136	-1.135	0.886	0.397	0.239	0.241
1.0	0.000	-1.000	1.000	0.399	0.242	0.242
2.0	0.243	-0.691	1.173	0.412	0.261	0.251
3.0	0.435	-0.414	1.299	0.433	0.282	0.266
4.0	0.586	-0.203	1.396	0.457	0.298	0.284
5.0	0.706	-0.041	1.475	0.481	0.316	0.301

2.2.2. Moments

Proposition 2. The r th moments of $X \sim \mathcal{MPN}(0, 1, \alpha)$ for $r = 1, 2, 3, \dots$, are given by

$$\mathbb{E}(X^r) = \frac{\alpha}{2^\alpha - 1} a_r(\alpha), \tag{11}$$

where $a_r(\alpha)$ is defined as

$$a_r(\alpha) = \int_0^1 [\Phi^{-1}(u)]^r (1 + u)^{\alpha-1} du. \tag{12}$$

Here, $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution.

Proof. By using the change of variable $u = \Phi(x)$, it follows that

$$\begin{aligned} \mathbb{E}(X^r) &= \int_{-\infty}^{\infty} x^r \frac{\alpha}{2^\alpha - 1} \phi(x) [1 + \Phi(x)]^{\alpha-1} dx \\ &= \frac{\alpha}{2^\alpha - 1} \int_0^1 [\Phi^{-1}(u)]^r (1 + u)^{\alpha-1} du \\ &= \frac{\alpha}{2^\alpha - 1} a_r(\alpha). \quad \square \end{aligned}$$

Corollary 1. The mean and variance of X are given by

$$\begin{aligned} \mathbb{E}(X) &= \frac{\alpha}{2^\alpha - 1} a_1(\alpha) \quad \text{and} \\ \mathbb{V}ar(X) &= \frac{\alpha}{2^\alpha - 1} \left(a_2(\alpha) - \frac{\alpha}{2^\alpha - 1} a_1^2(\alpha) \right), \end{aligned}$$

respectively.

Corollary 2. The skewness (β_1) and kurtosis (β_2) coefficients are, respectively, given by

$$\begin{aligned} \beta_1 &= \frac{a_3(\alpha) - \frac{3\alpha}{2^\alpha - 1} a_1(\alpha) a_2(\alpha) + \frac{2\alpha^2}{(2^\alpha - 1)^2} a_1^3(\alpha)}{\left(\frac{\alpha}{2^\alpha - 1} \right)^{3/2} [a_2(\alpha) - \frac{\alpha}{2^\alpha - 1} a_1^2(\alpha)]^{3/2}} \quad \text{and} \\ \beta_2 &= \frac{a_4(\alpha) - \frac{4\alpha}{2^\alpha - 1} a_1(\alpha) a_3(\alpha) + \frac{6\alpha^2}{(2^\alpha - 1)^2} a_1^2(\alpha) a_2(\alpha) - \frac{3\alpha^3}{(2^\alpha - 1)^3} a_1^4(\alpha)}{\frac{\alpha}{2^\alpha - 1} [a_2(\alpha) - \frac{\alpha}{2^\alpha - 1} a_1^2(\alpha)]^2}. \end{aligned}$$

Remark 4. Observe that the integral in Equation (12) can be numerically approximated by using the built-in function integrate available in the software package R. Below, in Table 2, some approximations of the mean and

variance for the MPN distribution for different values of α are displayed. Figure 3 illustrates the behavior of the $\mathbb{E}(X)$ and $\text{Var}(X)$ of the MPN distribution for different values of α . It is observable that when α grows, the mean increases and the variance decreases.

Figure 4 displays the curves associated with the coefficients of skewness (left panel) and kurtosis (right panel) of the MPN and PN distributions. It is shown that, depending on the values of α , the MPN distribution exhibits equal, greater, or lesser values for these coefficients compared to the PN model. In general, the MPN distribution has a smaller range of skewness than the PN distribution. On the other hand, when $\alpha < 13.05$, the MPN distribution has a greater kurtosis coefficient than the PN model.

Table 2. Approximations of $\mathbb{E}(X)$ and $\text{Var}(X)$ of the MPN distribution for different values of α .

α	$\mathbb{E}(X)$	$\text{Var}(X)$
0.5	-0.097	1.006
1.0	0.000	1.000
5.0	0.659	0.770
10.0	1.119	0.521
100.0	2.247	0.218

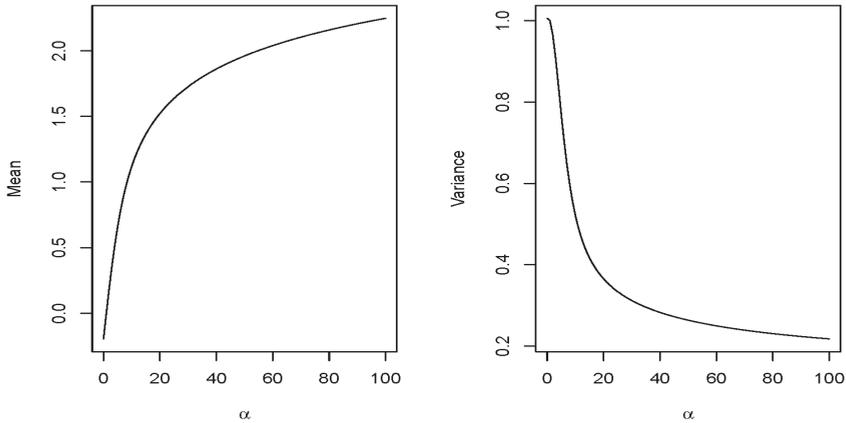


Figure 3. Plot of the $\mathbb{E}(X)$ and $\text{Var}(X)$ of the MPN distribution.

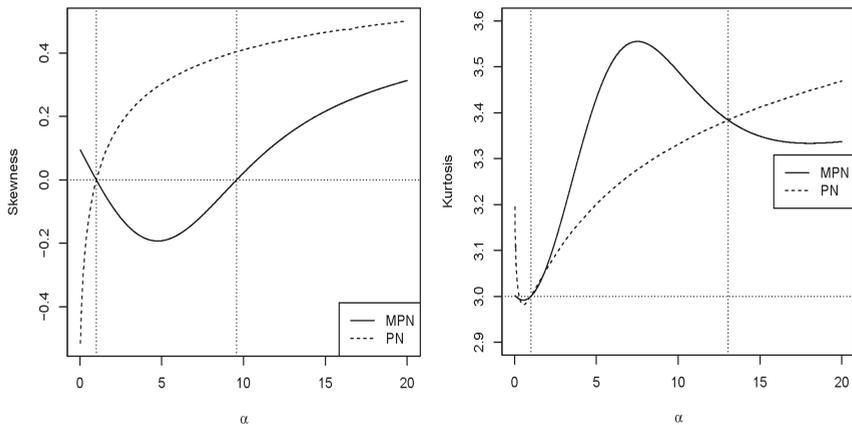


Figure 4. Graphs of the skewness and kurtosis coefficients for the MPN and PN distributions.

2.2.3. Stochastic Ordering

Stochastic ordering is an important tool to compare continuous random variables. It is well-known that random variable X_1 is smaller than random variable X_2 in stochastic ordering ($X_1 \leq_{st} X_2$) if $F_{X_1}(x) \geq F_{X_2}(x)$ for all x , and in likelihood ratio order ($X_1 \leq_{lr} X_2$) if $f_{X_1}(x)/f_{X_2}(x)$ decreases with x . Using Theorem 1.C.1 and Theorem 2.A.1 of Shaked and Shanthikumar [20], the above stochastic orders hold according to the following implications,

$$X_1 \leq_{lr} X_2 \Rightarrow X_1 \leq_{st} X_2. \tag{13}$$

The proposition shows that the members of the \mathcal{MPN} family can be stochastically ordered according to parameters values.

Proposition 3. Let $X_1 \sim \mathcal{MPN}(0, 1, \alpha_1)$ and $X_2 \sim \mathcal{MPN}(0, 1, \alpha_2)$. If $\alpha_1 > \alpha_2$, then $X_1 \leq_{lr} X_2$ and, therefore, $X_1 \leq_{st} X_2$.

Proof. From the quotient of both densities, it follows that

$$\frac{f_{X_2}(x; \alpha_2)}{f_{X_1}(x; \alpha_1)} = \frac{\alpha_2}{\alpha_1} \left(\frac{2^{\alpha_1} - 1}{2^{\alpha_2} - 1} \right) [1 + \Phi(x)]^{\alpha_2 - \alpha_1},$$

is non-decreasing if and only if $\mu'(x) \geq 0$ for $x \in (-\infty, \infty)$, where

$$\mu(x) = [1 + \Phi(x)]^{\alpha_2 - \alpha_1}.$$

After some calculations, it is shown that

$$\mu'(x) = (\alpha_2 - \alpha_1)\phi(x)[1 + \Phi(x)]^{\alpha_2 - \alpha_1 - 1}.$$

It is straightforward that for $\alpha_1 > \alpha_2$, then $\mu'(x) < 0$ for $x \in (-\infty, \infty)$. Therefore, $f_{X_2}(x; \alpha_2)/f_{X_1}(x; \alpha_1)$ is decreasing in x , and consequently $X_1 \leq_{lr} X_2$. The other implication follows immediately from (13). \square

3. Inference

In this section, parameters estimation for the \mathcal{MPN} distribution is discussed by using the method of moments and ML estimation. Additionally, a simulation analysis is carried out to illustrate the behavior of the ML estimators.

3.1. Method of Moments

The following proposition illustrates the derivation of the moment estimates of the \mathcal{MPN} distribution.

Proposition 4. Let x_1, \dots, x_n be a random sample obtained from the random variable $X \sim \mathcal{MPN}(\mu, \sigma, \alpha)$, then the moment estimates $\hat{\theta}_M = (\hat{\mu}_M, \hat{\sigma}_M, \hat{\alpha}_M)$ for $\theta = (\mu, \sigma, \alpha)$ are given by

$$\hat{\sigma}_M = \frac{S_{\underline{x}}}{\sqrt{\frac{\hat{\alpha}_M}{2^{\hat{\alpha}_M} - 1} \left(a_2(\hat{\alpha}_M) - \frac{\hat{\alpha}_M}{2^{\hat{\alpha}_M} - 1} a_1^2(\hat{\alpha}_M) \right)}}, \quad \hat{\mu}_M = \bar{x} - \hat{\sigma}_M \frac{\hat{\alpha}_M}{2^{\hat{\alpha}_M} - 1} a_1(\hat{\alpha}_M) \tag{14}$$

and

$$\frac{a_3(\hat{\alpha}_M) - \frac{3\hat{\alpha}_M}{2^{\hat{\alpha}_M} - 1} a_1(\hat{\alpha}_M)a_2(\hat{\alpha}_M) + \frac{2\hat{\alpha}_M^2}{(2^{\hat{\alpha}_M} - 1)^2} a_1^3(\hat{\alpha}_M)}{\left(\frac{\hat{\alpha}_M}{2^{\hat{\alpha}_M} - 1} \right)^{3/2} [a_2(\hat{\alpha}_M) - \frac{\hat{\alpha}_M}{2^{\hat{\alpha}_M} - 1} a_1^2(\hat{\alpha}_M)]^{3/2}} - A_{\underline{x}} = 0, \tag{15}$$

where \bar{x} , $S_{\underline{x}}$ and $A_{\underline{x}}$ denote the sample mean, sample standard deviation and sample Fisher's skewness coefficient respectively.

Proof. As μ and σ are location and scale parameters respectively, the skewness coefficient does not depend on these parameters. Thus, the result in (15) is directly obtained from matching the sample skewness coefficient with population counterpart given in Corollary 2. In addition, by considering that $X = \sigma Z + \mu$, where $Z \sim \mathcal{MPN}(0, 1, \alpha)$, and again by equating sample mean and sample variance to the mean and variance respectively, it follows that

$$\begin{aligned}\bar{x} &= \hat{\sigma}_M \mathbb{E}(X) + \hat{\mu}_M, \\ &= \hat{\sigma}_M \frac{\hat{\alpha}_M}{2^{\hat{\alpha}_M} - 1} a_1(\hat{\alpha}_M) + \hat{\mu}_M,\end{aligned}$$

and

$$\begin{aligned}S_{\underline{x}}^2 &= \hat{\sigma}_M^2 \text{Var}(X) \\ &= \hat{\sigma}_M^2 \frac{\hat{\alpha}_M}{2^{\hat{\alpha}_M} - 1} \left(a_2(\hat{\alpha}_M) - \frac{\hat{\alpha}_M}{2^{\hat{\alpha}_M} - 1} a_1^2(\hat{\alpha}_M) \right),\end{aligned}$$

where $\hat{\alpha}_M$ satisfies expression (15). Then, (14) is obtained by solving the latter equations for $\hat{\mu}_M$ and $\hat{\sigma}_M$, respectively. \square

3.2. Maximum Likelihood Estimation

For a random sample x_1, \dots, x_n derived from the $\mathcal{MPN}(\mu, \sigma, \alpha)$ distribution, the log-likelihood function can be written as

$$\ell(\mu, \sigma, \alpha) = nc(\sigma, \alpha) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 + (\alpha - 1) \sum_{i=1}^n \log \left[1 + \Phi \left(\frac{x_i - \mu}{\sigma} \right) \right], \quad (16)$$

where $c(\sigma, \alpha) = \log(\alpha) - \log(2^\alpha - 1) - \log(\sigma) - \frac{1}{2} \log(2\pi)$.

The score equations are given by

$$n\mu + \sigma(\alpha - 1) \sum_{i=1}^n \kappa(x_i) = n\bar{x}, \quad (17)$$

$$n\sigma^2 + \sigma(\alpha - 1) \sum_{i=1}^n (x_i - \mu)\kappa(x_i) = \sum_{i=1}^n (x_i - \mu)^2, \quad (18)$$

$$\frac{n}{\alpha} + \sum_{i=1}^n \log \left[1 + \Phi \left(\frac{x_i - \mu}{\sigma} \right) \right] = \frac{2^\alpha \log(2)n}{2^\alpha - 1}, \quad (19)$$

where $\kappa(w) = \kappa(w; \mu, \sigma) = \frac{\phi\left(\frac{w-\mu}{\sigma}\right)}{1+\Phi\left(\frac{w-\mu}{\sigma}\right)}$.

Solutions for these Equations (17)–(19) can be obtained by using numerical procedures such as Newton–Raphson algorithm. Alternatively, these estimates can be found by directly maximizing the log-likelihood surface given by (16) and using the subroutine “optim” in the software package [21].

3.3. Simulation Study

To examine the behavior of the proposed approach, a simulation study is carried out to assess the performance of the estimation procedure for the parameters μ , σ , and α in the \mathcal{MPN} model. The simulation analysis is conducted by considering 1000 generated samples of sizes $n = 50, 100$, and 200 from the \mathcal{MPN} distribution. The goal of this simulation is to study the behavior of the ML

estimators of the parameters by using our proposed procedure. To generate $X \sim \mathcal{MPN}(\mu, \sigma, \alpha)$, the following algorithm is used,

1. Step 1: Generate $W \sim Uniform(0, 1)$.
2. Step 2: Compute $X = \mu + \sigma \Phi^{-1} \left[(2^\alpha W - W + 1)^{1/\alpha} - 1 \right]$.

where $\mu \in \mathbb{R}, \sigma > 0, \alpha > 0$ and $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution. For each generated sample of the \mathcal{MPN} distribution, the ML estimates and corresponding standard deviation (SD) were computed for each parameter. As it can be seen in Table 3, the performance of the estimates improves when n and α increases.

Table 3. Maximum likelihood (ML) estimates and standard deviation (SD) for the parameters μ, σ and α of the \mathcal{MPN} model for different generated samples of sizes $n = 50, 100$, and 200 .

$n = 50$					
μ	σ	α	$\hat{\mu}$ (SD)	$\hat{\sigma}$ (SD)	$\hat{\alpha}$ (SD)
0	1	0.1	-0.352478 (0.149214)	0.994441 (0.091321)	0.190243 (0.175202)
		0.5	-0.19534 (0.14501)	0.990622 (0.094550)	0.613052 (0.272096)
		0.8	-0.083183 (0.144587)	0.990286 (0.098669)	0.854338 (0.164924)
		1	-0.009586 (0.141691)	0.995312 (0.0997256)	1.007328 (0.122688)
		5	0.004225 (0.100001)	0.997408 (0.088254)	5.030272 (0.229064)
		10	0.001108 (0.066610)	0.999124 (0.068611)	10.060478 (0.475019)
		100	0.002171 (0.017362)	1.001152 (0.029604)	100.437990 (2.668190)
$n = 100$					
0	1	0.1	-0.351446 (0.104552)	0.998513 (0.070831)	0.180054 (0.111930)
		0.5	-0.19268 (0.101786)	0.997622 (0.068806)	0.576957 (0.223378)
		0.8	-0.08140 (0.099360)	0.997674 (0.069451)	0.830318 (0.152995)
		1	0.002786 (0.097411)	0.996444 (0.069648)	1.002200 (0.088749)
		5	0.002014 (0.099305)	0.996788 (0.085987)	5.023032 (0.221756)
		10	0.002897 (0.046109)	1.000515 (0.050192)	10.032857 (0.339106)
		100	0.000623 (0.012137)	1.000185 (0.019759)	100.168752 (1.866302)
$n = 200$					
0	1	0.1	-0.348177 (0.072732)	0.999433 (0.047548)	0.170978 (0.076165)
		0.5	-0.196617 (0.072015)	0.999142 (0.047896)	0.562935 (0.218890)
		0.8	-0.076657 (0.069510)	0.997719 (0.050718)	0.824700 (0.127661)
		1	0.001158 (0.06877)	0.998408 (0.050586)	1.003651 (0.058344)
		5	-0.000165 (0.053006)	1.000623 (0.044182)	5.005130 (0.115719)
		10	-0.000239 (0.033615)	1.000017 (0.035902)	10.014958 (0.246652)
		100	0.000514 (0.008452)	1.000491 (0.014599)	100.104380 (1.295144)

Fisher’s Information Matrix

Let us now consider $X \sim \mathcal{MPN}(\mu, \sigma, \alpha)$ and $Z = \left(\frac{X-\mu}{\sigma}\right) \sim \mathcal{MPN}(0, 1, \alpha)$. For a single observation x of X , the log-likelihood function for $\theta = (\mu, \sigma, \alpha)$ is given by

$$\ell(\theta) = \log f_X(\theta, x) = c(\sigma, \alpha) - \frac{1}{2\sigma^2}(x - \mu)^2 + (\alpha - 1) \log \left[1 + \Phi \left(\frac{x - \mu}{\sigma} \right) \right].$$

The corresponding first and second partial derivatives of the log-likelihood function are derived in the Appendix A. It can be shown that the Fisher’s information matrix for the \mathcal{MPN} distribution is provided by

$$I_F(\theta) = \begin{pmatrix} I_{\mu\mu} & I_{\mu\sigma} & I_{\mu\alpha} \\ & I_{\sigma\sigma} & I_{\sigma\alpha} \\ & & I_{\alpha\alpha} \end{pmatrix}$$

with the following entries,

$$\begin{aligned}
 I_{\mu\mu} &= \frac{1}{\sigma^2} + \left(\frac{\alpha-1}{\sigma^2}\right)(b_{11} + b_{02}), \\
 I_{\mu\sigma} &= \frac{2}{\sigma^2}\mathbb{E}(Z) - \left(\frac{\alpha-1}{\sigma^2}\right)(b_{01} - b_{21} - b_{12}), \\
 I_{\mu\alpha} &= \frac{1}{\sigma}b_{01}, \\
 I_{\sigma\sigma} &= -\frac{1}{\sigma^2} + \frac{3}{\sigma^2}\mathbb{E}(Z^2) - \left(\frac{\alpha-1}{\sigma^2}\right)(2b_{11} - b_{31} - b_{22}), \\
 I_{\sigma\alpha} &= \frac{1}{\sigma}b_{11}, \\
 I_{\alpha\alpha} &= \frac{1}{\alpha^2} - \frac{2^\alpha \log^2(2)}{(2^\alpha - 1)^2},
 \end{aligned}$$

where $b_{ij} = \mathbb{E}[Z^i \kappa^j(Z; 0, 1)]$ must be numerically computed.

The Fisher's (expected) information matrix can be obtained by computing the expected values of the above expressions. By taking in this matrix, $\alpha = 1$, we have that $Z \sim N(\mu, \sigma)$ and

$$I_F(\mu, \sigma, \alpha = 1) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 & \frac{1}{\sigma}d_{02} \\ 0 & \frac{2}{\sigma^2} & \frac{1}{\sigma}d_{12} \\ \frac{1}{\sigma}d_{02} & \frac{1}{\sigma}d_{12} & 1 - 2\log^2(2) \end{pmatrix},$$

where $d_{ij} = \int_{-\infty}^{\infty} \frac{z^i \phi^j(z)}{1 + \Phi(z)} dz$ must be numerically obtained.

The determinant of $I_F(\mu, \sigma, \alpha = 1)$ is $(2 - 4\log^2(2) - b_{12}^2 - 2b_{02}^2)/\sigma^4 = 0.003357435/\sigma^4 \neq 0$, consequently, the Fisher's information matrix is nonsingular at $\alpha = 1$.

Therefore, for large samples, the ML estimators, $\hat{\theta}$, of θ are asymptotically normal, that is,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} N_3(\mathbf{0}, I(\theta)^{-1}),$$

resulting in the asymptotic variance of the ML estimators $\hat{\theta}$ being the inverse of Fisher's information matrix $I(\theta)$. As the parameters are unknown, the observed information matrix is usually considered, where the unknown parameters are estimated by ML.

4. Application

In this section, a numerical illustration based on a real dataset is presented. The goal of this application is to show empirical evidence that the \mathcal{MPN} yields a better fit to data than the \mathcal{PN} , \mathcal{SN} , and t-student (\mathcal{TS}) with α degrees of freedom distributions. For that reason, we consider a set of 3848 observations of the variable "density" included in the dataset verb "POLLEN5.DA" available at <http://lib.stat.cmu.edu/datasets/pollen.data>. This variable measures a geometric characteristic of a specific type of pollen. This dataset was previously used by Pewsey et al. [9] to compare the \mathcal{PN} and \mathcal{SN} distributions. A summary of some descriptive statistics are displayed in Table 4 below.

Table 4. Summary of descriptive statistics for the pollen density dataset.

Mean	Median	Variance	Skewness	Kurtosis
0.000	−0.030	9.887	0.109	3.193

By using the results derived in Proposition 4, we have computed the moment estimates for the parameters (μ, σ, α) of the MPN distribution, obtaining $(−5.609, 4.576, 11.857)$. Then, by taking these numbers as initial values, the ML estimates are derived. In Table 5, the ML estimates for the parameters of the MPN , PN , SN , and TS distributions. The figures between brackets are the asymptotic standard errors of the estimates obtained by inverting the Fisher’s information matrices for the three models evaluated at their respective ML estimates. Additionally, for each model, the values of the maximum of the log-likelihood function (ℓ_{max}) are reported. The MPN distribution attains the largest value, and consequently provides a better fit to data.

Table 5. Parameter estimates; standard errors (SE); and maximum of the log-likelihood function, ℓ_{max} , for the TS , SN , PN , and MPN corresponding to the pollen density dataset.

Parameters	$TS(SE)$	$SN(SE)$	$PN(SE)$	$MPN(SE)$
μ	−0.010 (0.05)	−2.04 (0.24)	−1.74 (0.68)	−5.73 (0.43)
σ	3.037 (0.05)	3.75 (0.14)	3.69 (0.21)	4.62 (0.14)
α	29.995 (13.01)	0.93 (0.14)	1.77 (0.37)	12.13 (1.21)
ℓ_{max}	−9864.99	−9863.42	−9863.37	−9861.98

To compare the fit achieved by each distribution, the values of several measures of model selection, i.e., Akaike’s information criterion (AIC) (see Akaike [22]) and Bayesian information criterion (BIC) (see Schwarz [23]) are reported in Table 6. A model with lower numbers in these measures of model selection is preferable. It can be seen that the MPN is preferable in terms of these two measures of model validation. In addition, the Kolmogorov–Smirnov test statistics and the corresponding p -values has been included in this table for all the models considered. It can be observed that none of the models is rejected at the usual significance levels. However, the MPN distribution has a higher p -value and is rejected later than the other two models. Alternative methods of model selection to the Kolmogorov–Smirnov test that can be applied here can be found in Jäntschi and Bolboacă [24] and Jäntschi [25]. Furthermore, the histogram associated to the empirical distribution of the variable “density” in the pollen dataset is illustrated in the left hand side of Figure 5. In addition, the densities of TS , SN , PN , and MPN , by using the maximum likelihood estimates of their parameters, have been superimposed. Similarly, on the right hand side of Figure 5, the fit in both tails is shown. It is observable that, for this dataset, the MPN has thicker tails than the other three distributions. Finally, the QQ-plots for each distribution considered have been illustrated in Figure 6. Here, note that the MPN distribution exhibits an almost perfect alignment with the 45° line, and therefore it provides a better fit for extreme quantiles. Finally, Figure 7 displays the profile log-likelihood of μ , σ , and α of the MPN distribution. It is noticeable that the estimates are unique.

Table 6. Akaike’s information criterion (AIC), Bayesian information criterion (BIC), Kolmogorov–Smirnov (KSS) test, and the corresponding p -values for all the models considered.

Criteria	TS	SN	PN	MPN
AIC	19,735.98	19,732.84	19,732.74	19,729.96
BIC	19,754.74	19,751.61	19,751.50	19,748.72
KSS (p -value)	0.014 (0.516)	0.013 (0.559)	0.012 (0.627)	0.010 (0.820)

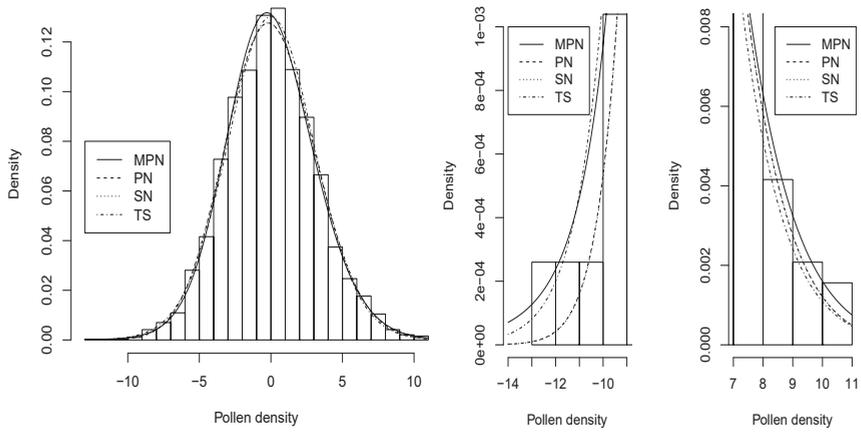


Figure 5. Left panel: Histogram of the empirical distribution and fitted densities by ML superimposed for pollen dataset. Right panel: Plots of the tails for the four models.

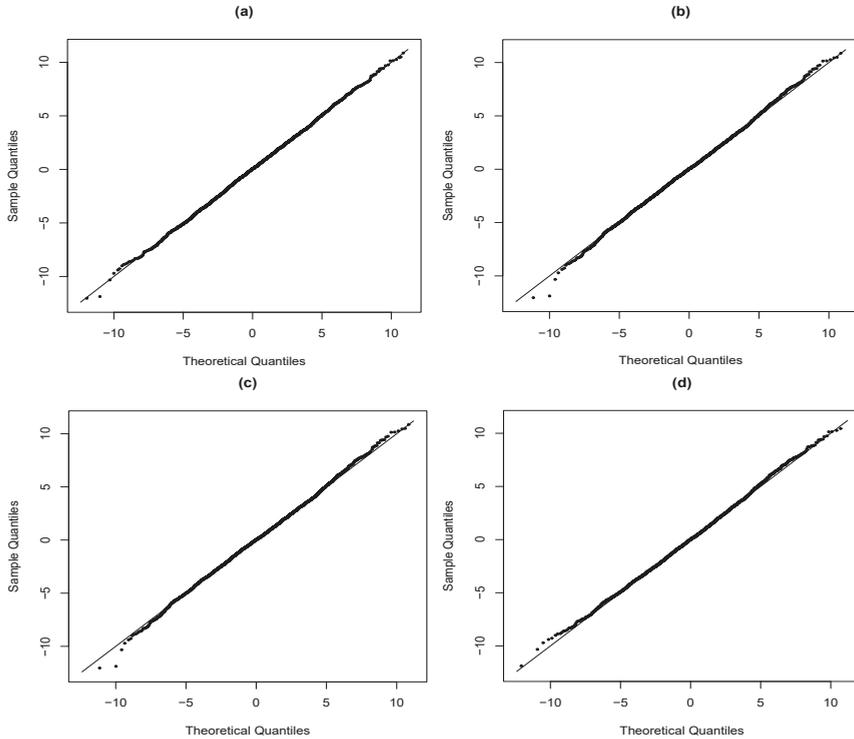


Figure 6. QQ-plots: (a) MPN model; (b) PN model; (c) SN model; (d) TS model.

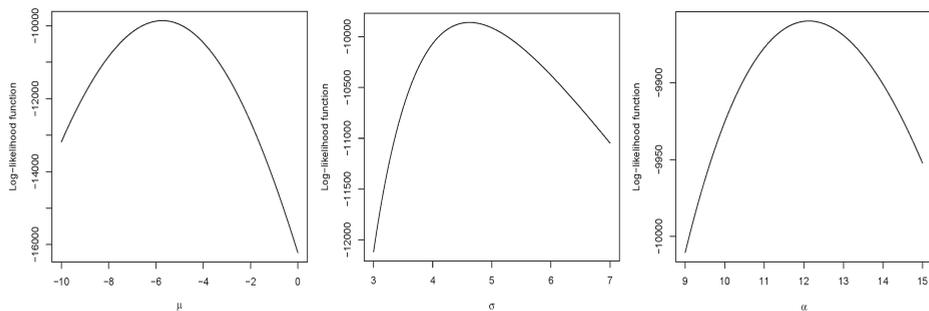


Figure 7. Profile log-likelihood of μ , σ and α for the MPN distribution.

5. Concluding Remarks

In this paper, a modification of the continuous symmetric-power distribution has been introduced. The particular case of the normal distribution the MPN distribution has been examined in detail. This distribution arises by modifying the distribution function of the symmetrical powers family. After carrying out this modification, a more flexible family of probability distributions is obtained, allowing for the kurtosis coefficient to take a certain range of values in the parameter space. For this model, its basic properties, different method of estimation and Fisher's information matrix were studied. By using a real dataset, we showed that the MPN distribution provides a better fit than other existing models in the literature such as the TS , SN , and PN distributions.

Author Contributions: The authors contributed equally to this work.

Funding: This work was partially completed while Héctor W. Gómez visited the Universidad de Las Palmas de Gran Canaria, supported by MINEDUC-UA project, code ANT1855. This research was also funded by (EGD) [Ministerio de Economía y Competitividad, Spain] grant number [ECO2013-47092]; (EGD)[Ministerio de Economía, Industria y Competitividad. Agencia Estatal de Investigación] grant number [ECO2017-85577-P];

Acknowledgments: We also acknowledge the referee's suggestions that helped us to improve this work.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

The first derivatives of $\ell(\theta)$ are given by

$$\begin{aligned}\frac{\partial \ell(\theta)}{\partial \mu} &= \frac{1}{\sigma} \left[\frac{x - \mu}{\sigma} - (\alpha - 1) \kappa(x) \right] \\ \frac{\partial \ell(\theta)}{\partial \sigma} &= -\frac{1}{\sigma} \left[1 - \left(\frac{x - \mu}{\sigma} \right)^2 + (\alpha - 1) \left(\frac{x - \mu}{\sigma} \right) \kappa(x) \right] \\ \frac{\partial \ell(\theta)}{\partial \alpha} &= \frac{1}{\alpha} - \frac{2^\alpha \log(2)}{2^\alpha - 1} + \log \left[1 + \Phi \left(\frac{x - \mu}{\sigma} \right) \right]\end{aligned}$$

The second derivatives of $l(\theta)$ are

$$\begin{aligned}\frac{\partial^2 \ell(\theta)}{\partial \mu^2} &= -\frac{1}{\sigma^2} - \left(\frac{\alpha-1}{\sigma^2}\right) \left[\left(\frac{x-\mu}{\sigma}\right) \kappa(x) + \kappa^2(x) \right] \\ \frac{\partial^2 \ell(\theta)}{\partial \mu \partial \sigma} &= -\frac{2}{\sigma^2} \left(\frac{x-\mu}{\sigma}\right) + \left(\frac{\alpha-1}{\sigma^2}\right) \kappa(x) \left[1 - \left(\frac{x-\mu}{\sigma}\right)^2 - \left(\frac{x-\mu}{\sigma}\right) \kappa(x) \right] \\ \frac{\partial^2 \ell(\theta)}{\partial \mu \partial \alpha} &= -\frac{k(x)}{\sigma} \\ \frac{\partial^2 \ell(\theta)}{\partial \sigma^2} &= \frac{1}{\sigma^2} - \frac{3}{\sigma^2} \left(\frac{x-\mu}{\sigma}\right)^2 + \left(\frac{\alpha-1}{\sigma^2}\right) \left(\frac{x-\mu}{\sigma}\right) \kappa(x) \left[2 - \left(\frac{x-\mu}{\sigma}\right)^2 - \left(\frac{x-\mu}{\sigma}\right) \kappa(x) \right] \\ \frac{\partial^2 \ell(\theta)}{\partial \sigma \partial \alpha} &= -\left(\frac{x-\mu}{\sigma^2}\right) k(x) \\ \frac{\partial^2 \ell(\theta)}{\partial \alpha^2} &= -\frac{1}{\alpha^2} + \frac{2^\alpha \log^2(2)}{(2^\alpha - 1)^2}\end{aligned}$$

References

1. Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
2. Arellano-Valle, R.B.; Gómez, H.W.; Quintana, F.A. A New Class of Skew-Normal Distributions. *Commun. Stat. Theory Methods* **2004**, *33*, 1465–1480. [[CrossRef](#)]
3. Arellano-Valle, R.B.; Gómez, H.W.; Salinas, H.S. A note on the Fisher information matrix for the skew-generalized-normal model. *Stat. Oper. Res. Trans.* **2013**, *37*, 19–28.
4. Azzalini, A. *The Skew-Normal and Related Families*; IMS monographs; Cambridge University Press: New York, NY, USA 2014.
5. Martínez-Flórez, G.; Barranco-Chamorro, I.; Bolfarine, H.; Gómez, H.W. Flexible Birnbaum–Saunders Distribution. *Symmetry* **2019**, *11*, 1305. [[CrossRef](#)]
6. Contreras-Reyes, J.E.; Arellano-Valle, R.B. Kullback–Leibler Divergence Measure for Multivariate Skew-Normal Distributions. *Entropy* **2012**, *14*, 1606–1626. [[CrossRef](#)]
7. Chiogna, M. A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution. *Stat. Methods Appl.* **2005**, *14*, 331–341. [[CrossRef](#)]
8. Arellano-Valle, R.B.; Azzalini, A. The centred parametrization for the multivariate skew-normal distribution. *J. Multivar. Anal.* **2008**, *99*, 1362–1382. [[CrossRef](#)]
9. Pewsey, A.; Gómez, H.W.; Bolfarine, H. Likelihood-based inference for power distributions. *Test* **2012**, *21*, 775–789. [[CrossRef](#)]
10. Lehmann, E.L. The power of rank tests. *Ann. Math. Statist.* **1953**, *24*, 23–43. [[CrossRef](#)]
11. Durrans, S.R. Distributions of fractional order statistics in hydrology. *Water Resour. Res.* **1992**, *28*, 1649–1655. [[CrossRef](#)]
12. Gupta, D.; Gupta, R.C. Analyzing skewed data by power normal model. *Test* **2008**, *17*, 197–210. [[CrossRef](#)]
13. Martínez-Flórez, G.; Arnold, B.C.; Bolfarine, H.; Gómez, H.W. The Multivariate Alpha-power Model. *J. Stat. Plan. Inference* **2013**, *143*, 1236–1247. [[CrossRef](#)]
14. Martínez-Flórez, G.; Bolfarine, H.; Gómez, H.W. Asymmetric regression models with limited responses with an application to antibody response to vaccine. *Biom. J.* **2013**, *55*, 156–172. [[CrossRef](#)] [[PubMed](#)]
15. Martínez-Flórez, G.; Bolfarine, H.; Gómez, H.W. The log alpha-power asymmetric distribution with application to air pollution. *Environmetrics* **2014**, *25*, 44–56. [[CrossRef](#)]
16. Martínez-Flórez, G.; Bolfarine, H.; Gómez, H.W. Doubly censored power-normal regression models with inflation. *Test* **2015**, *24*, 265–286. [[CrossRef](#)]
17. Castillo, N.O.; Gallardo, D.I.; Bolfarine, H.; Gómez, H.W. Truncated Power-Normal Distribution with Application to Non-Negative Measurements. *Entropy* **2018**, *20*, 433. [[CrossRef](#)]

18. Maciak, M.; Peštová, B.; Pešta, M. Structural breaks in dependent, heteroscedastic, and extremal panel data. *Kybernetika* **2018**, *54*, 1106–1121. [[CrossRef](#)]
19. Pešta, M. Total least squares and bootstrapping with application in calibration. *Statistics* **2013**, *47*, 966–991. [[CrossRef](#)]
20. Shaked, M.; Shanthikumar, J.G. *Stochastic Orders*; Springer Science & Business Media: New York, NY, USA, 2007.
21. R Development Core Team. *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
22. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [[CrossRef](#)]
23. Schwarz, G. Estimating the dimension of a model. *Ann. Statist.* **1978**, *6*, 461–464. [[CrossRef](#)]
24. Jäntschi, L.; Bolboacă, S.D. Computation of Probability Associated with Anderson–Darling Statistic. *Mathematics* **2018**, *6*, 88. [[CrossRef](#)]
25. Jäntschi, L. A Test Detecting the Outliers for Continuous Distributions Based on the Cumulative Distribution Function of the Data Being Tested. *Symmetry* **2019**, *11*, 835. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Article

Normal-G Class of Probability Distributions: Properties and Applications

Fábio V. J. Silveira ¹, Frank Gomes-Silva ^{1,*}, Cícero C. R. Brito ², Moacyr Cunha-Filho ¹, Felipe R. S. Gusmão ¹ and Sílvia F. A. Xavier-Júnior ³

¹ Department of Statistics and Informatics, Rural Federal University of Pernambuco, Recife 52171900, Pernambuco, Brazil; fabiovjs@gmail.com (F.V.J.S.); moacyr.cunhafo@ufpe.br (M.C.-F.); felipe556@gmail.com (F.R.S.G.)

² Federal Institute of Education, Science and Technology of Pernambuco, Recife 50740545, Pernambuco, Brazil; ciceroCarlosbrito@yahoo.com.br

³ Department of Statistics, Paraíba State University, Campina Grande 58429500, Paraíba, Brazil; profsilviofernando@gmail.com

* Correspondence: franksinatrags@gmail.com

Received: 28 September 2019; Accepted: 12 November 2019; Published: 15 November 2019

Abstract: In this paper, we propose a novel class of probability distributions called Normal-G. It has the advantage of demanding no additional parameters besides those of the parent distribution, thereby providing parsimonious models. Furthermore, the class enjoys the property of identifiability whenever the baseline is identifiable. We present special Normal-G sub-models, which can fit asymmetrical data with either positive or negative skew. Other important mathematical properties are described, such as the series expansion of the probability density function (pdf), which is used to derive expressions for the moments and the moment generating function (mgf). We bring Monte Carlo simulation studies to investigate the behavior of the maximum likelihood estimates (MLEs) of two distributions generated by the class and we also present applications to real datasets to illustrate its usefulness.

Keywords: probabilistic distribution class; normal distribution; identifiability; maximum likelihood; moments

1. Introduction

For many purposes, statistical distributions are used in a plethora of science fields. They are regularly useful tools to describe natural and social phenomena, providing suitable models which can help dealing with real problems, such as for instance, those concerning the prediction of an event of interest. Recent works have focused attention at formulating and describing new classes of probability distributions, which are defined generally as extensions of widely known models by adding a single or more parameters to the cumulative distribution function (cdf). Hopefully, the new models will provide more flexibility and better fitting to real data. Some examples are [1,2] where a shape parameter is added to the model by exponentiating the cdf. A general method of introducing a parameter to expand a family of distributions was presented by [3]; they applied the method to create a new two-parameter extension of the exponential distribution and a new three-parameter Weibull distribution.

A natural generalization of the Normal pdf was proposed by [4] and perhaps it is the most widely known generalized Normal distribution. The power 2 appearing in the original pdf was replaced by a shape parameter $s > 0$. Therewith, the new pdf becomes:

$$f(x|\mu, \sigma, s) = K \exp \left\{ - \left| \frac{x - \mu}{\sigma} \right|^s \right\},$$

where K is a normalizing constant, which depends on σ and s . One can see that the Laplace distribution is a particular case of the generalized Normal of Nadarajah [4] when $s = 1$.

Azzalini [5] defined a mathematically tractable class that includes strictly (not just asymptotically) the Normal distribution. The general pdf of the class is $2G(\lambda y)f(y)$ for $-\infty < y < \infty$, where $\lambda \in \mathbb{R}$, G is an absolutely continuous cdf, $\frac{d}{dy}G$ and f are pdfs symmetric about 0. Making $G = \Phi$ and $f = \phi$, namely the standard normal cdf and pdf respectively, one gets to the well-known skew-normal distribution, whose pdf is $\phi(y; \lambda) = 2\phi(y)\Phi(\lambda y)$. It is easy to see that $\phi(y; 0) = \phi(y)$, but when $\lambda \neq 0$, the distribution is asymmetric and its coefficient of skewness has the same sign as λ .

A generalization denoted by compressed normal distribution was introduced by [6], whose objective was dealing with negatively skewed data (specifically with human longevity data); in this way, they induced a skew by adding kx to the denominator of the location-scale transformation, that is,

$$t(x) = \frac{x - \mu}{\sigma + kx}$$

and when $k < 0$, the curve presents a negative skew; for $k > 0$, a positive skew occurs.

Classes with one or more additional parameters usually generalize existing classes as particular cases. The McDonald-Weibull distribution [7] is an important sub-model of the McDonald class; it has three extra parameters and includes the Beta-Weibull [8] and the Kumaraswamy-Weibull [9] as special cases.

A technique to derive families of continuous distributions using a pdf as a generator was introduced by [10] and the models emerged from such method are called members of the T-X family. In other words, if $r(t)$ is the pdf of a random variable $T \in [a, b]$, for $-\infty \leq a < b \leq \infty$ and $W(G(x))$ is a function of the cdf $G(x)$ of a random variable X so that:

- $W(G(x)) \in [a, b]$;
- $W(G(x))$ is differentiable and monotonically non-decreasing;
- $W(G(x)) \rightarrow a$ as $x \rightarrow -\infty$ and $W(G(x)) \rightarrow b$ as $x \rightarrow \infty$;

then $F(x) = \int_a^{W(G(x))} r(t)dt$ is the cdf of a new family of distributions.

An example of a T-X family member is the Gompertz-G class [11]; to define its cdf, the chosen functions were $W[G(x)] = -\log[1 - G(x)]$ and $r(t) = \theta e^{\gamma t} e^{-\frac{\theta}{\gamma}(e^{\gamma t} - 1)}$ for $t > 0$, given that $\theta > 0$, $\gamma > 0$. Varying $G(x)$, one can get different sub-models of the class.

The procedure to define a T-X family member is indeed capable to generalize a large number of distributions. Even though it can be regarded as a particular case described by the method of generating classes of probability distributions presented in the recent work of [12]. This new method has a high power of generalization. It consists of creating distribution classes by integrating a cdf, such that the limits of the integration are special functions that satisfy some conditions. Thus, the cdf of the general class is given by:

$$F(x) = \zeta(x) \sum_{j=1}^n \int_{L_j(x)}^{U_j(x)} dH(t) - \nu(x) \sum_{j=1}^n \int_{M_j(x)}^{V_j(x)} dH(t) \quad (1)$$

where H is a cdf, $n \in \mathbb{N}$, $\zeta, \nu : \mathbb{R} \mapsto \mathbb{R}$ and $L_j, U_j, M_j, V_j : \mathbb{R} \mapsto \mathbb{R} \cup \{\pm\infty\}$ are the aforementioned special functions that will be discussed in the next section.

Based on this innovative method, we introduce the Normal-G class of distributions. We consider that this extension will yield good submodels. This paper aims to investigate and compare some of them with other competitive extended probability distributions.

2. The Normal-G Class and Some Mathematical Properties

The method established by [12] states that if $H, \zeta, \nu : \mathbb{R} \mapsto \mathbb{R}$ and $L_j, U_j, M_j, V_j : \mathbb{R} \mapsto \mathbb{R} \cup \{\pm\infty\}$ for $j = 1, 2, 3, \dots, n$ are monotonic and right continuous functions such that:

- (c1) H is a cdf and ζ and ν are non-negative;

- (c2) $\zeta(x), U_j(x)$ and $M_j(x)$ are non-decreasing and $v(x), V_j(x), L_j(x)$ are non-increasing $\forall j = 1, 2, 3, \dots, n$;
- (c3) If $\lim_{x \rightarrow -\infty} \zeta(x) \neq \lim_{x \rightarrow -\infty} v(x)$, then $\lim_{x \rightarrow -\infty} \zeta(x) = 0$ or $\lim_{x \rightarrow -\infty} U_j(x) = \lim_{x \rightarrow -\infty} L_j(x) \forall j = 1, 2, 3, \dots, n$, and $\lim_{x \rightarrow -\infty} v(x) = 0$ or $\lim_{x \rightarrow -\infty} M_j(x) = \lim_{x \rightarrow -\infty} V_j(x) \forall j = 1, 2, 3, \dots, n$;
- (c4) If $\lim_{x \rightarrow -\infty} \zeta(x) = \lim_{x \rightarrow -\infty} v(x) \neq 0$, then $\lim_{x \rightarrow -\infty} U_j(x) = \lim_{x \rightarrow -\infty} V_j(x)$ and $\lim_{x \rightarrow -\infty} M_j(x) = \lim_{x \rightarrow -\infty} L_j(x) \forall j = 1, 2, 3, \dots, n$;
- (c5) $\lim_{x \rightarrow -\infty} L_j(x) \leq \lim_{x \rightarrow -\infty} U_j(x)$ and if $\lim_{x \rightarrow -\infty} v(x) \neq 0$, then $\lim_{x \rightarrow +\infty} M_j(x) \leq \lim_{x \rightarrow +\infty} V_j(x) \forall j = 1, 2, 3, \dots, n$;
- (c6) $\lim_{x \rightarrow +\infty} U_n(x) \geq \sup\{x \in \mathbb{R} : H(x) < 1\}$ and $\lim_{x \rightarrow +\infty} L_1(x) \leq \inf\{x \in \mathbb{R} : H(x) > 0\}$;
- (c7) $\lim_{x \rightarrow +\infty} \zeta(x) = 1$;
- (c8) $\lim_{x \rightarrow +\infty} v(x) = 0$ or $\lim_{x \rightarrow +\infty} M_j(x) = \lim_{x \rightarrow +\infty} V_j(x) \forall j = 1, 2, 3, \dots, n$ and $n \geq 1$;
- (c9) $\lim_{x \rightarrow +\infty} U_j(x) = \lim_{x \rightarrow +\infty} L_{j+1}(x) \forall j = 1, 2, 3, \dots, n - 1$ and $n \geq 2$;
- (c10) H is a cdf without points of discontinuity or all functions $L_j(x)$ and $V_j(x)$ are constant at the right of the vicinity of points whose image are points of discontinuity of H , being also continuous in that points. Moreover, H does not have any point of discontinuity in the set $\left\{ \lim_{x \rightarrow \pm\infty} L_j(x), \lim_{x \rightarrow \pm\infty} U_j(x), \lim_{x \rightarrow \pm\infty} M_j(x), \lim_{x \rightarrow \pm\infty} V_j(x) \right\}$ for some $j = 1, 2, 3, \dots, n$;

then Equation (1) is a cdf.

Let $n = 1, \zeta(x) = 1, v(x) = 0, L_1(x) = -\infty, U_1(x) = [2G(x) - 1] / (G(x)[1 - G(x)])$ and $H(t) = \Phi(t)$; the function in Equation (1) turns into:

$$F_G(x) = \int_{-\infty}^{\frac{2G(x)-1}{G(x)(1-G(x))}} d\Phi(t), \tag{2}$$

where $G(x)$ is a cdf. Since $v(x) = 0$, there is no need to specify $M_1(x)$ and $V_1(x)$. The conditions (c1), (c7), (c8) and (c10) are straightforward; clearly (c4), (c5) and (c9) do not need to be verified in this case. Given that $G(x)$ is non-decreasing:

$$\begin{aligned} x_1 < x_2 &\Rightarrow G(x_1) \leq G(x_2) \Rightarrow \frac{1}{1 - G(x_1)} \leq \frac{1}{1 - G(x_2)} \\ &\Rightarrow \frac{1}{1 - G(x_1)} - \frac{1}{G(x_1)} \leq \frac{1}{1 - G(x_2)} - \frac{1}{G(x_2)} \\ &\Rightarrow \frac{2G(x_1) - 1}{G(x_1)(1 - G(x_1))} \leq \frac{2G(x_2) - 1}{G(x_2)(1 - G(x_2))} \Rightarrow U_1(x_1) \leq U_1(x_2), \end{aligned}$$

so $U_1(x)$ is non-decreasing, as well as $\zeta(x)$; and since $L_1(x)$ is non-increasing, (c2) is true. Considering that $U_1(x) = 1/[1 - G(x)] - 1/G(x)$, it is easy to verify that $\lim_{x \rightarrow -\infty} U_1(x) = -\infty = \lim_{x \rightarrow -\infty} L_1(x)$; and since $\lim_{x \rightarrow -\infty} v(x) = 0$, (c3) is satisfied. The condition (c6) is also true because $\lim_{x \rightarrow +\infty} U_1(x) = +\infty = \sup\{x \in \mathbb{R} : H(x) < 1\}$ and $\lim_{x \rightarrow +\infty} L_1(x) = -\infty = \inf\{x \in \mathbb{R} : H(x) > 0\}$.

Therefore, according to the method exposed above, Equation (2) is a cdf and, from now on, we will denote it by Normal-G class of probability distributions. The new cdf can be viewed as a composed function of $G(x)$, which will be referred as parent distribution or baseline; in agreement with [12], if the baseline is continuous (discrete), then the Normal-G will generate a continuous (discrete) distribution, whose support will be the same as $G(x)$. It is worth remarking that the proposed class demands no additional parameters other than the ones of the parent distribution.

Although the Normal-G class has been defined as a composed function of a single $G(x)$, it is possible to formulate classes that depend on more than one baseline; see [12] for further details.

We can rewrite Equation (2) as:

$$F_G(x) = \int_{-\infty}^{\frac{2G(x)-1}{G(x)[1-G(x)]}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt, \quad (3)$$

and since $\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$, and $\Phi(x) = \int_{-\infty}^x \phi(t) dt$, we get to:

$$F_G(x) = \Phi \left(\frac{2G(x) - 1}{G(x)[1 - G(x)]} \right). \quad (4)$$

In case of continuous $G(x)$, we can take the derivative of Equation (4) with respect to x :

$$f_G(x) = \phi \left(\frac{2G(x) - 1}{G(x)[1 - G(x)]} \right) \frac{1 - 2G(x)[1 - G(x)]}{G(x)^2[1 - G(x)]^2} g(x). \quad (5)$$

The expression in Equation (5) is the pdf of the class Normal-G, whose hazard rate function (hrf) is given by:

$$\tau_G(x) = \frac{\phi \left(\frac{2G(x)-1}{G(x)[1-G(x)]} \right)}{1 - \Phi \left(\frac{2G(x)-1}{G(x)[1-G(x)]} \right)} \left[\frac{1 - 2G(x)[1 - G(x)]}{G(x)^2[1 - G(x)]^2} g(x) \right].$$

Many distributions presented in the statistical literature undergo the problem of non-identifiability. One cannot assume that the parameters of a non-identifiable model will be uniquely determined from a set of observed random variables; in other words, inferences on the parameters may not be reliable. As the Theorem 1 states, the Normal-G class is exempt from this problem, whenever the parent distribution G satisfies the property of identifiability.

Theorem 1. *If the cdf F_G belongs to the Normal-G class and the cdf G is identifiable, then F_G is identifiable.*

Proof of Theorem 1. Given that $0 < G(x|\xi_j) < 1$ for $j = 1, 2$, where ξ_j is a parametric vector and assuming that $F_G(x|\xi_1) = F_G(x|\xi_2)$, we have:

$$\Phi \left(\frac{2G(x|\xi_1) - 1}{G(x|\xi_1)[1 - G(x|\xi_1)]} \right) = \Phi \left(\frac{2G(x|\xi_2) - 1}{G(x|\xi_2)[1 - G(x|\xi_2)]} \right).$$

Since the function Φ is injective, we can write:

$$\frac{1}{1 - G(x|\xi_1)} - \frac{1}{G(x|\xi_1)} = \frac{1}{1 - G(x|\xi_2)} - \frac{1}{G(x|\xi_2)}$$

$$\frac{G(x|\xi_1) - G(x|\xi_2)}{[1 - G(x|\xi_1)][1 - G(x|\xi_2)]} = \frac{G(x|\xi_1) - G(x|\xi_2)}{-G(x|\xi_1)G(x|\xi_2)}$$

If $G(x|\xi_1) \neq G(x|\xi_2)$, then:

$$[1 - G(x|\xi_1)][1 - G(x|\xi_2)] = -G(x|\xi_1)G(x|\xi_2) \quad (6)$$

The left-hand side of Equation (6) is necessarily positive for almost all $x \in \mathbb{R}$, whereas the right-hand side is negative, a contradiction. Thereby, $G(x|\xi_1) = G(x|\xi_2) \Rightarrow \xi_1 = \xi_2$. \square

2.1. Special Normal-G Sub-Models

Here we present two distributions from the Normal-G class.

2.1.1. The Normal-Weibull Distribution

Weibull is one of the most used models to describe natural phenomena and failure of several kinds of components. It is extensively used in survival analysis and reliability. In recent times, many authors have focused on new extensions for it, such as [13,14]. The two-parameter Weibull cdf is given by $G_W(x|k, \lambda) = 1 - e^{-(x/\lambda)^k}$ for $x \geq 0$, where $k, \lambda > 0$. Replacing the baseline G in Equation (4) by G_W , we get to the Normal-Weibull cdf, namely:

$$F_{NW}(x) = \Phi \left[\frac{e^{(x/\lambda)^k} - 2}{1 - e^{-(x/\lambda)^k}} \right], \tag{7}$$

for $x \geq 0$. Using Equation (5) to write the corresponding pdf, we have:

$$f_{NW}(x) = \phi \left[\frac{e^{(x/\lambda)^k} - 2}{1 - e^{-(x/\lambda)^k}} \right] \left(\frac{kx^{k-1}}{\lambda^k} \right) \frac{1 - 2 \left[1 - e^{-(x/\lambda)^k} \right] e^{-(x/\lambda)^k}}{e^{-(x/\lambda)^k} \left[1 - e^{-(x/\lambda)^k} \right]^2}. \tag{8}$$

Plots of pdf and hrf of the Normal-Weibull distribution for different values of the parameters are portrayed in Figure 1. The different shapes of the hrf curve evince the flexibility of the model. Particularly for $k = 1$, the Weibull distribution is equivalent to an Exponential distribution, so the hrf is constant; in contrast, the Normal-Exponential model has an increasing hrf in some left-bounded interval.

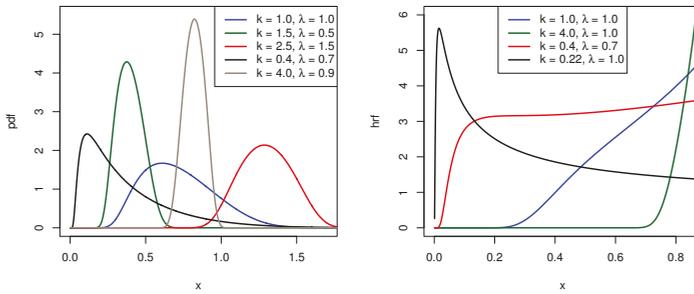


Figure 1. Plots of pdf and hrf for the Normal-Weibull distribution.

In Figure 2, the vertical axis shows the range of values of Pearson’s moment coefficient of skewness, which depends on the parameters k and λ . We can see in the graph that the Normal-Weibull distribution is also able to fit data with either positive or negative skew.

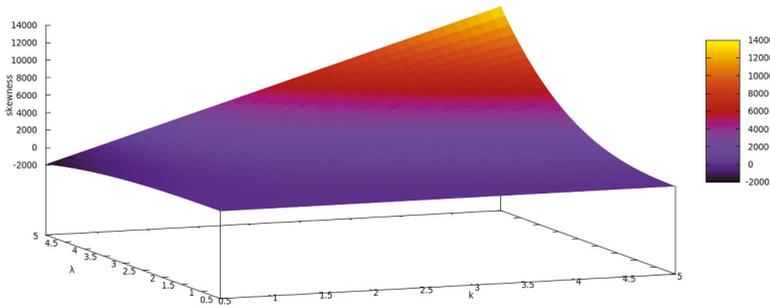


Figure 2. Skewness of the Normal-Weibull distribution.

2.1.2. The Normal-Log-Logistic Distribution

The Log-logistic distribution is commonly applied to reliability and oftentimes it works well as a lifetime model. Its cdf is given by $G_{LL}(x|\alpha, \beta) = 1 - [1 + (x/\alpha)^\beta]^{-1}$ for $x \geq 0$, where $\alpha, \beta > 0$. The Normal-log-logistic cdf is easily obtained replacing the parent distribution G in Equation (4) by G_{LL} . Thus:

$$F_{NLL}(x) = \Phi \left[\left(\frac{x}{\alpha} \right)^\beta - \left(\frac{x}{\alpha} \right)^{-\beta} \right], \tag{9}$$

for $x \geq 0$. Taking the derivative of Equation (9) with respect to x , we get to the pdf:

$$f_{NLL}(x) = \phi \left[\left(\frac{x}{\alpha} \right)^\beta - \left(\frac{x}{\alpha} \right)^{-\beta} \right] \left[1 + \left(\frac{x}{\alpha} \right)^{2\beta} \right] \beta \alpha^\beta x^{-\beta-1}. \tag{10}$$

Figure 3 shows plots of pdf and hrf for different values of α and β . It is worth noting that the Normal-log-logistic distribution may have a decreasing hrf of early failure. It is also possible for the hrf to be increasing or unimodal.

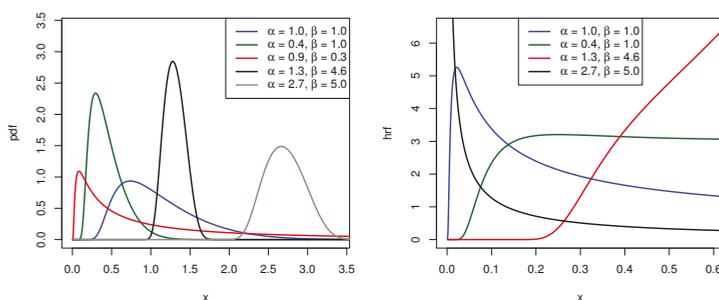


Figure 3. Plots of pdf and hrf for the Normal-log-logistic distribution.

Pearson’s moment coefficient of skewness for the Normal-log-logistic distribution is depicted in Figure 4.

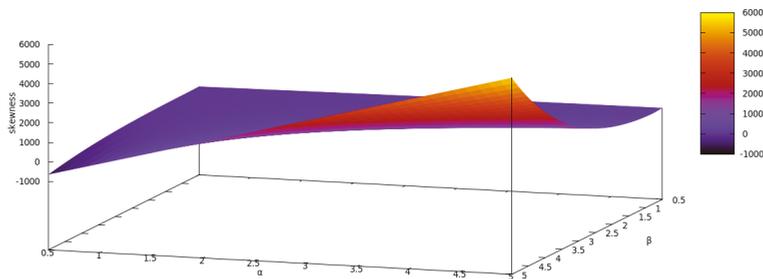


Figure 4. Skewness of the Normal-log-logistic distribution.

2.2. Series Representation

The normal cdf is related to the error function erf as follows:

$$\Phi(z) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{z}{\sqrt{2}} \right) \right], \tag{11}$$

where $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$. Provided that $\operatorname{erf}(z/\sqrt{2})$ can be linearly represented by:

$$\begin{aligned} \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right) &= \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n \cdot (z/\sqrt{2})^{2n+1}}{n!(2n+1)} \\ &= \sqrt{\frac{2}{\pi}} \cdot \sum_{n=0}^{\infty} \left(-\frac{1}{2}\right)^n \frac{z^{2n+1}}{n!(2n+1)}, \end{aligned} \tag{12}$$

replacing Equation (12) in Equation (11), we obtain:

$$\Phi(z) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \left(-\frac{1}{2}\right)^n \frac{z^{2n+1}}{n!(2n+1)}. \tag{13}$$

Now, considering $|G(x)| < 1$, we can write:

$$\frac{2G(x) - 1}{G(x)[1 - G(x)]} = \frac{2G(x) - 1}{G(x)} \cdot \frac{1}{1 - G(x)} = \left(2 - \frac{1}{G(x)}\right) \sum_{k=0}^{\infty} G(x)^k \tag{14}$$

and replacing z of the right member of Equation (13) by the expression in Equation (14), we have:

$$\begin{aligned} \Phi\left(\frac{2G(x) - 1}{G(x)[1 - G(x)]}\right) &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1/2)^n}{n!(2n+1)} \left[\left(2 - \frac{1}{G(x)}\right) \sum_{k=0}^{\infty} G(x)^k\right]^{2n+1} \\ &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1/2)^n}{n!(2n+1)} \underbrace{\left(2 - \frac{1}{G(x)}\right)^{2n+1}}_{A1} \underbrace{\left[\sum_{k=0}^{\infty} G(x)^k\right]^{2n+1}}_{A2}. \end{aligned} \tag{15}$$

The right member of Equation (15) has two factors, namely, A1 and A2, that can be rewritten as power series. Concerning to A1, the binomial theorem allows us to write:

$$\begin{aligned} \left(2 - \frac{1}{G(x)}\right)^{2n+1} &= \sum_{j=0}^{2n+1} \binom{2n+1}{j} 2^{2n+1-j} \left(-\frac{1}{G(x)}\right)^j \\ &= \sum_{j=0}^{2n+1} \binom{2n+1}{j} (-1)^j \cdot 2^{2n+1-j} \cdot G(x)^{-j} \\ &= \sum_{j=0}^{2n+1} \delta_j \cdot G(x)^{-j} \end{aligned} \tag{16}$$

It is a known result related to power series raised to powers that:

$$\left[\sum_{k=0}^{\infty} a_k G(x)^k\right]^N = \sum_{k=0}^{\infty} c_k G(x)^k, \tag{17}$$

where $c_0 = a_0^N$, $c_k = \frac{1}{ka_0} \sum_{s=1}^k (sN - k + s)a_s c_{k-s}$ for $k \geq 1$ and $N \in \mathbb{N}$. Setting $N = 2n + 1$ and $a_k = 1$ for all $k \geq 0$, we get to the expression A2 in Equation (15) and we can use the result in Equation (17) to write as follows:

$$\left[\sum_{k=0}^{\infty} G(x)^k\right]^{2n+1} = \sum_{k=0}^{\infty} c_k \cdot G(x)^k, \tag{18}$$

such that $c_0 = 1$, $c_k = \frac{1}{k} \sum_{s=1}^k (s[2n + 1] - k + s)c_{k-s}$ for $k \geq 1$ and $2n + 1 \in \mathbb{N}$. Now replacing A1 and A2 of the Equation (15) by the right members of the Equations (16) and (18) respectively, we obtain the result below:

$$\begin{aligned} \Phi\left(\frac{2G(x) - 1}{G(x)[1 - G(x)]}\right) &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1/2)^n}{n!(2n + 1)} \cdot \sum_{j=0}^{2n+1} \delta_j \cdot G(x)^{-j} \cdot \sum_{k=0}^{\infty} c_k \cdot G(x)^k \\ &= \frac{1}{2} + \sum_{n=0}^{\infty} \sum_{j=0}^{2n+1} \sum_{k=0}^{\infty} \underbrace{\binom{2n+1}{j} \frac{(-1)^{n+j} \cdot 2^{n+1-j}}{n!(2n+1)\sqrt{2\pi}} c_k}_{\eta_{j,n,k}} \cdot G(x)^{k-j} \\ &= \frac{1}{2} + \sum_{n,k=0}^{\infty} \sum_{j=0}^{2n+1} \eta_{j,n,k} \cdot G(x)^{k-j}. \end{aligned} \tag{19}$$

The Fubini’s theorem on differentiation allows us to write the derivative of Equation (19) with respect to x as follows:

$$f_G(x) = \sum_{n,k=0}^{\infty} \sum_{j=0}^{2n+1} \eta_{j,n,k} \cdot \underbrace{(k-j)g(x)G(x)^{k-j-1}}_{g_{k-j}(x)}. \tag{20}$$

Since $g_{k-j}(x)$ is the pdf of a random variable of the exponentiated family, as described in [15,16], one can say that (20) is the Normal-G pdf (5) expressed as a linear combination of pdfs of exponentiated distributions. Such useful property is typically found and detailed in works on new classes of distributions; see for instance: [17–20].

2.3. Quantile Function

By inverting Equation (4), the quantile function associated with the Normal-G class is obtained. For simplification, let us write $v = F_G(x)$. From Equation (4) we have:

$$\Phi^{-1}(v) = \frac{2G(x) - 1}{G(x)[1 - G(x)]} \Rightarrow \Phi^{-1}(v)G(x)^2 + [2 - \Phi^{-1}(v)]G(x) - 1 = 0,$$

that is, a quadratic equation for $G(x)$, that admits the following two solutions:

$$\frac{\Phi^{-1}(v) - 2 - \sqrt{4 + \Phi^{-1}(v)^2}}{2\Phi^{-1}(v)} \quad \text{and} \quad \frac{\Phi^{-1}(v) - 2 + \sqrt{4 + \Phi^{-1}(v)^2}}{2\Phi^{-1}(v)}.$$

If the first solution above is picked, then $G(x)$ might assume values lesser than 0 (see $v = 0.95$ for example). On the other hand, the second one allows us to verify that $0 < G(x) < 1$ is valid for all $x \in \mathbb{R}$. Finally, we can write the quantile function of Equation (4) as follows:

$$Q_F(v) = Q_G\left[\frac{\Phi^{-1}(v) - 2 + \sqrt{4 + \Phi^{-1}(v)^2}}{2\Phi^{-1}(v)}\right], \tag{21}$$

such that $Q_G(\cdot)$ is the quantile function of the baseline G . A uniform random number generator and (21) make the simulation of random variables following (3) quite simple. Namely, if $Z \sim \mathcal{U}(0,1)$, then $Q_F(Z)$ follows a Normal-G distribution.

2.4. Raw Moments, Incomplete Moments and Moment Generating Function

Provided that X follows a Normal-G distribution, the r th raw moment of X is $E(X^r) = \int_{-\infty}^{\infty} x^r f_G(x) dx$, where $f_G(x)$ is given in Equation (20) and $r \in \mathbb{Z}_+^*$. Using Fubini's theorem to change the order of integration and series, we have:

$$E(X^r) = \sum_{n,k=0}^{\infty} \sum_{j=0}^{2n+1} \eta_{j,n,k} \int_{-\infty}^{\infty} x^r g_{k-j}(x) dx \quad (22)$$

$$= \sum_{n,k=0}^{\infty} \sum_{j=0}^{2n+1} \eta_{j,n,k} E(Y_{k-j}^r) \quad (23)$$

where Y_{k-j} follows the exponentiated distribution whose pdf is $g_{k-j}(x)$ shown in Equation (20).

Despite the upper infinity limit in the sums, expressions like Equation (23) are intractable. According to [21], one can get fairly accurate results truncating each infinite sum by 20; they used numerical routines to compute accurately similar expressions for the moments of some Kumaraswamy generalized distributions.

The r th moment can also be represented in terms of the quantile function of the baseline. Defining $u = G_{k-j}(x)$ and replacing x in Equation (22) by $Q_G(u^{1/(k-j)})$, we have:

$$E(X^r) = \sum_{n,k=0}^{\infty} \sum_{j=0}^{2n+1} \eta_{j,n,k} \int_0^1 \left[Q_G \left(u^{\frac{1}{k-j}} \right) \right]^r du.$$

The r th incomplete moment of X is given by the following expression:

$$T_r(z) = \int_{-\infty}^z x^r f_G(x) dx = \sum_{n,k=0}^{\infty} \sum_{j=0}^{2n+1} \eta_{j,n,k} T_r^*(z) \quad (24)$$

where $T_r^*(z)$ is the r th incomplete moment of Y_{k-j} . One can also write Equation (24) in terms of the quantile function of G :

$$T_r(z) = \sum_{n,k=0}^{\infty} \sum_{j=0}^{2n+1} \eta_{j,n,k} \int_0^{[G(x)]^{k-j}} \left[Q_G \left(u^{\frac{1}{k-j}} \right) \right]^r du.$$

The mgf is a function associated with a random variable, whose moments can be straightforwardly derived using it. It is also useful to check whether two functions of random variables are equal since there is a bijection between pdfs and mgfs (when they exist). The mgf $M_X(t)$ of X is the expected value of e^{tX} , where $t \in (-\iota, \iota)$, $\iota > 0$. Given that $M_{Y_{k-j}}(t)$ is the mgf of Y_{k-j} , on the lines of Equation (23), we can write:

$$\begin{aligned} M_X(t) &= \sum_{n,k=0}^{\infty} \sum_{j=0}^{2n+1} \eta_{j,n,k} \int_{-\infty}^{\infty} e^{tx} g_{k-j}(x) dx \\ &= \sum_{n,k=0}^{\infty} \sum_{j=0}^{2n+1} \eta_{j,n,k} M_{Y_{k-j}}(t). \end{aligned}$$

2.5. Estimation and Inference

Attractive asymptotic properties, such as efficiency and consistency, are some of the reasons that make the maximum likelihood method the most usually applied method of parametric point estimation. The MLEs are the points that maximize the likelihood function over the domain of the parameter space. Since the logarithmic function is increasing, performing the maximization of the log-likelihood function, besides being a more convenient task, also provides the MLEs.

Given that $\xi = (\xi_1, \dots, \xi_r)^\top$ is the $r \times 1$ parametric vector of a random variable X that follows a Normal-G distribution, $G(x|\xi) = G_\xi(x)$ is the baseline, $g(x|\xi) = g_\xi(x)$ is its corresponding pdf and $\mathbf{X} = (x_1, \dots, x_m)$ is a complete random sample of size m from X , then the log-likelihood function is:

$$\begin{aligned} \ell(\xi|\mathbf{X}) = & \sum_{j=1}^m \log \phi \left(\frac{2G_\xi(x_j) - 1}{G_\xi(x_j) [1 - G_\xi(x_j)]} \right) + \sum_{j=1}^m \log \left(1 - 2G_\xi(x_j) + 2G_\xi^2(x_j) \right) \\ & - 2 \sum_{j=1}^m \log G_\xi(x_j) - 2 \sum_{j=1}^m \log [1 - G_\xi(x_j)] + \sum_{j=1}^m \log g_\xi(x_j). \end{aligned} \tag{25}$$

Thanks to powerful functions available within the software for statistical computing, it is possible to use numerical methods to maximize (25); for this purpose, R [22] brings the function `optim` in package `stats`.

The MLEs can also be obtained by solving the system of equations $U(\xi|\mathbf{X}) = \mathbf{0}_r$, where $U(\xi|\mathbf{X}) = \nabla_\xi \ell(\xi|\mathbf{X}) = (u_i)_{1 \leq i \leq r}$ is the score vector, such that:

$$\begin{aligned} u_i = & \sum_{j=1}^m \frac{[G_\xi(x_j) - 1]^4 - G_\xi^4(x_j)}{G_\xi^3(x_j)[1 - G_\xi(x_j)]^3} \cdot \frac{\partial}{\partial \xi_i} G_\xi(x_j) + \sum_{j=1}^m \frac{4G_\xi(x_j) - 2}{1 - 2G_\xi(x_j) + 2G_\xi^2(x_j)} \cdot \frac{\partial}{\partial \xi_i} G_\xi(x_j) \\ & - 2 \sum_{j=1}^m \frac{1}{G_\xi(x_j)} \cdot \frac{\partial}{\partial \xi_i} G_\xi(x_j) + 2 \sum_{j=1}^m \frac{1}{1 - G_\xi(x_j)} \cdot \frac{\partial}{\partial \xi_i} G_\xi(x_j) + \sum_{j=1}^m \frac{1}{g_\xi(x_j)} \cdot \frac{\partial}{\partial \xi_i} g_\xi(x_j) \end{aligned}$$

and $\mathbf{0}_r$ is a $r \times 1$ vector of zeros.

The information matrix $J(\xi|\mathbf{X})$ is essential to construct confidence intervals and to test hypotheses on ξ . The expectation of $J(\xi|\mathbf{X})$ is the expected Fisher information matrix \mathcal{I}_ξ and under certain conditions of regularity, $\sqrt{m}(\hat{\xi} - \xi)$ follows approximately a multivariate normal distribution $N_r(\mathbf{0}_r, \mathcal{I}_\xi^{-1})$. The expression for $J(\xi|\mathbf{X})$ is presented in Appendix A.

3. Numerical Analysis

3.1. Simulation Study

We used the free software R version 3.4.4 [22] to carry out the Monte Carlo simulation study; the number of replications was 10,000. The pseudo-random samples were generated via Von Neumann's acceptance-rejection method [23]. This simple procedure requires the corresponding pdf $y = f(x)$, a minorant and a majorant for x and a majorant for y ; it is not necessary to implement the quantile function in this case. Four sample sizes, namely $n = 50, 100, 200$ and 500 , and five different values for the vector of parameters were considered. For each scenario, we calculated the bias and the mean squared error (MSE) as follows:

$$\text{Bias}_i = \frac{1}{10000} \sum_{j=1}^{10000} (\hat{\xi}_{ij} - \xi_i), \quad \text{MSE}_i = \frac{1}{10000} \sum_{j=1}^{10000} (\hat{\xi}_{ij} - \xi_i)^2$$

where ξ_i is the i -th element of the vector of parameters $\xi = (\xi_1, \dots, \xi_r)^\top$ and $\hat{\xi}_{ij}$ is the estimate for ξ_i at the j -th replication. The log-likelihood function was maximized using the technique of simulated annealing, available by the `optim` subroutine, for which the user has to pass a vector ξ_0 of initial values. At first, we took $\xi_0 = \mathbf{1}_r$, namely a $r \times 1$ vector of ones, then we run one single replication considering sample size $n = 50$; the obtained estimates from this procedure were assigned to ξ_0 and used in all of the aforementioned scenarios.

The results for both parameters of the Normal-Weibull density (8), shown in Table 1, indicate that the estimates are fairly close to the actual values. Moreover, as it would be expected, the bigger the sample size, the smaller the MSEs.

Table 1. Bias and MSE of the estimates under the maximum likelihood method for the Normal-Weibull model.

Actual Value			Bias		MSE	
<i>n</i>	<i>k</i>	λ	\hat{k}	$\hat{\lambda}$	\hat{k}	$\hat{\lambda}$
50	1.0	1.7	0.02707850	−0.00948110	0.00822299	0.00659991
	0.5	2.0	0.01446186	−0.02037125	0.00209051	0.03647395
	3.0	0.5	0.07878343	−0.00096778	0.07352625	0.00006357
	0.9	4.0	0.02586182	−0.02752101	0.00675452	0.04497248
	7.1	5.8	0.19086399	−0.00540407	0.41402178	0.00153179
100	1.0	1.7	0.01306919	−0.00453981	0.00377883	0.00332042
	0.5	2.0	0.00726917	−0.01412373	0.00095786	0.01838681
	3.0	0.5	0.03914672	−0.00057766	0.03400929	0.00003204
	0.9	4.0	0.01167774	−0.01403219	0.00305903	0.02273089
	7.1	5.8	0.08363335	−0.00279451	0.18835856	0.00077190
200	1.0	1.7	0.00651588	−0.00253820	0.00181409	0.00166703
	0.5	2.0	0.00358578	−0.00678178	0.00045681	0.00923355
	3.0	0.5	0.01901041	−0.00029362	0.01628677	0.00001604
	0.9	4.0	0.00658567	−0.00745541	0.00148059	0.01138373
	7.1	5.8	0.03656102	−0.00066316	0.09041610	0.00038519
500	1.0	1.7	0.00317837	−0.00127234	0.00071164	0.00066754
	0.5	2.0	0.00195165	−0.00609800	0.00017967	0.00370983
	3.0	0.5	0.00748033	−0.00008804	0.00636008	0.00000641
	0.9	4.0	0.00297109	−0.00200533	0.00057744	0.00455810
	7.1	5.8	0.01427116	−0.00045889	0.03550063	0.00015444

The results given in Table 2 suggest that the estimates of the parameters of the Normal-log-logistic model (10) have similar behavior of those shown in Table 1, that is to say, the biases are quite small and the MSE decreases as the sample size increases.

Table 2. Bias and MSE of the estimates under the maximum likelihood method for the Normal-log-logistic model.

Actual Value			Bias		MSE	
<i>n</i>	α	β	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$
50	2.7	5.0	0.00054404	0.13424700	0.00109551	0.20695579
	0.4	1.2	0.00024567	0.03275857	0.00041864	0.01195999
	6.0	2.5	−0.00002168	0.06835325	0.02161847	0.05192641
	4.0	3.4	0.00185659	0.08997153	0.00521251	0.09539035
	1.0	8.0	0.00010181	0.21835854	0.00005868	0.53164800
100	2.7	5.0	0.00012046	0.06377031	0.00055694	0.09509012
	0.4	1.2	0.00005101	0.01519026	0.00021246	0.00547655
	6.0	2.5	0.00191769	0.03297888	0.01099844	0.02386620
	4.0	3.4	−0.00004923	0.04701171	0.00263820	0.04439236
	1.0	8.0	0.00009827	0.11057419	0.00002980	0.24582360
200	2.7	5.0	0.00015504	0.03299212	0.00028047	0.04585339
	0.4	1.2	0.00006551	0.00866507	0.00010677	0.00265802
	6.0	2.5	−0.00106277	0.01625314	0.00553891	0.01145699
	4.0	3.4	−0.00002407	0.02295567	0.00133064	0.02123514
	1.0	8.0	0.00000852	0.05090923	0.00001503	0.11715066
500	2.7	5.0	0.00021558	0.01327469	0.00011277	0.01789692
	0.4	1.2	−0.00008941	0.00435633	0.00004284	0.00104215
	6.0	2.5	−0.00042358	0.00648252	0.00222639	0.00447192
	4.0	3.4	−0.00007789	0.00855609	0.00053513	0.00826466
	1.0	8.0	0.00003954	0.01663711	0.00000604	0.04559331

3.2. Applications

The first data to be considered is related to the soil fertility influence and the characterization of the biologic fixation of N₂ for the *Dimorphandra wilsonii* Rizz growth. It was originally studied by [24] and it also figures in the work of [25]. For 128 plants, the phosphorus concentration in the leaves was quantified. Here are the numbers: 0.22, 0.17, 0.11, 0.10, 0.15, 0.06, 0.05, 0.07, 0.12, 0.09, 0.23, 0.25, 0.23, 0.24, 0.20, 0.08, 0.11, 0.12, 0.10, 0.06, 0.20, 0.17, 0.20, 0.11, 0.16, 0.09, 0.10, 0.12, 0.12, 0.10, 0.09, 0.17, 0.19, 0.21, 0.18, 0.26, 0.19, 0.17, 0.18, 0.20, 0.24, 0.19, 0.21, 0.22, 0.17, 0.08, 0.08, 0.06, 0.09, 0.22, 0.23, 0.22, 0.19, 0.27, 0.16, 0.28, 0.11, 0.10, 0.20, 0.12, 0.15, 0.08, 0.12, 0.09, 0.14, 0.07, 0.09, 0.05, 0.06, 0.11, 0.16, 0.20, 0.25, 0.16, 0.13, 0.11, 0.11, 0.11, 0.08, 0.22, 0.11, 0.13, 0.12, 0.15, 0.12, 0.11, 0.11, 0.15, 0.10, 0.15, 0.17, 0.14, 0.12, 0.18, 0.14, 0.18, 0.13, 0.12, 0.14, 0.09, 0.10, 0.13, 0.09, 0.11, 0.11, 0.14, 0.07, 0.07, 0.19, 0.17, 0.18, 0.16, 0.19, 0.15, 0.07, 0.09, 0.17, 0.10, 0.08, 0.15, 0.21, 0.16, 0.08, 0.10, 0.06, 0.08, 0.12, 0.13. Table 3 brings some descriptive statistics.

Table 3. Descriptive statistics for soil fertility dataset.

n	mean	median	min	max	Variance	Skewness	Kurtosis
128	0.14078	0.13	0.05	0.28	0.00296	0.45438	−0.64478

We fitted the Normal-Weibull distribution (NW) (7) to the soil fertility dataset and compared it to the fits of Weibull (W), Exponentiated Weibull (ExpW) [1], Marshall-Olkin Extended Weibull (MOEW) [26], Kumaraswamy-Weibull (KwW) [9], Beta-Weibull (BW) [8] and McDonald-Weibull (McW) [7]. The function `goodness.fit` of the R package `AdequacyModel` provides, besides the MLEs and the standard errors (SE), some criteria for model selection (AIC, CAIC, BIC and HQIC); they are shown in Table 4.

Table 4. Fitted distributions to the soil fertility dataset (estimates and information criteria).

Distribution	Parameters	Estimates (SE)	AIC	CAIC	BIC	HQIC
NW	<i>k</i>	0.8398477 (0.0445182)	−395.7584	−395.6624	−390.0544	−393.4408
	<i>λ</i>	0.2049909 (0.0074017)				
W	<i>k</i>	2.8185566 (0.1919639)	−385.6297	−385.5337	−379.9256	−383.3121
	<i>λ</i>	0.1584836 (0.0052564)				
ExpW	<i>k</i>	1.5321145 (0.5023377)	−387.4361	−387.2426	−378.8801	−383.9598
	<i>λ</i>	0.0939938 (0.0374090)				
	<i>a</i>	3.5076974 (2.6763009)				
MOEW	<i>k</i>	3.9300962 (0.2426080)	−377.9475	−377.754	−369.3914	−374.4711
	<i>λ</i>	8.9163819 (4.5940983)				
	<i>a</i>	0.0031628 (0.0007240)				
KwW	<i>k</i>	1.1503912 (0.3443931)	−384.1491	−383.8239	−372.7409	−379.5139
	<i>λ</i>	0.1953371 (0.1291154)				
	<i>a</i>	3.3444607 (1.5352029)				
	<i>b</i>	7.5480698 (10.206142)				
BW	<i>k</i>	0.8477957 (0.2166409)	−385.7589	−385.4337	−374.3508	−381.1237
	<i>λ</i>	0.3304922 (0.4395169)				
	<i>a</i>	9.0436364 (4.5271059)				
	<i>b</i>	15.211970 (22.481984)				
McW	<i>k</i>	5.6665646 (8.3928707)	−384.6657	−384.1739	−370.4055	−378.8717
	<i>λ</i>	0.5912941 (0.5124852)				
	<i>a</i>	13.441193 (23.051917)				
	<i>b</i>	14.363802 (18.264058)				
	<i>c</i>	0.0870787 (0.1234075)				

Information criteria may be used as relative goodness-of-fit measures, such that the lowest values will characterize the best fitted models. In this sense, the Normal-Weibull distribution outperforms the other ones.

Figure 5 shows the histogram of soil fertility data and the fitted densities with the three lowest values of AIC among the distributions in the first column of Table 4. Although the Normal-Weibull and Exponentiated Weibull curves appear to be very close, the blue one (NW) seems to be closer to the histogram.

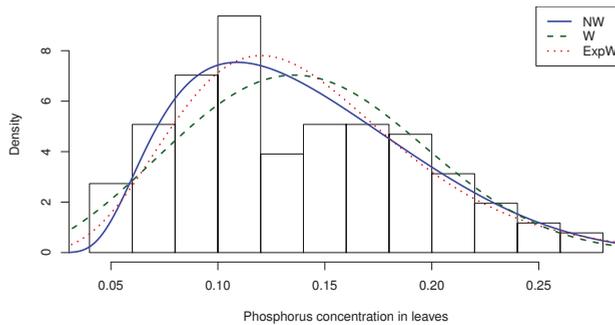


Figure 5. Histogram of soil fertility dataset and fitted densities.

The modified versions of Anderson-Darling (A^*) and Cramér-von Mises (W^*) statistics (more details in [27]) are typically used to investigate the quality of fit of probabilistic models. Table 5 brings these statistics concerning the fitted models to soil fertility data.

Table 5. Goodness-of-fit test statistics.

Distribution	A^*	W^*
NW	0.454008	0.079841
W	1.156994	0.207118
ExpW	0.784451	0.138403
MOEW	1.123759	0.183128
KwW	0.907239	0.163617
BW	0.750593	0.130501
McW	0.758296	0.137509

The measures portrayed in Table 5 represent the difference between the empirical distribution function and the real underlying cdf; hence we will consider that the models with lower values of A^* and W^* fit the data better. Therefore, once again the Normal-Weibull distribution beats the competing models.

The second application concerns to a dataset representing waiting times (in seconds) between 65 successive eruptions of water through a hole in the cliff at the coastal town of Kiama (New South Wales, Australia), known as the Blowhole. These data can be obtained in [17,28]. Here are they: 83, 51, 87, 60, 28, 95, 8, 27, 15, 10, 18, 16, 29, 54, 91, 8, 17, 55, 10, 35, 47, 77, 36, 17, 21, 36, 18, 40, 10, 7, 34, 27, 28, 56, 8, 25, 68, 146, 89, 18, 73, 69, 9, 37, 10, 82, 29, 8, 60, 61, 61, 18, 169, 25, 8, 26, 11, 83, 11, 42, 17, 14, 9, 12. Table 6 provides descriptive statistics.

Table 6. Descriptive statistics for eruption dataset.

n	Mean	Median	min	max	Variance	Skewness	Kurtosis
64	39.82812	28	7	169	1139.097	1.54641	2.77108

We fitted the Normal-log-logistic distribution (NLL) (9) to the eruption dataset and compared it to the fits of Log-logistic (LL), Exponentiated Log-logistic (ExpLL), Beta-log-logistic (BLL),

Kumaraswamy-log-logistic (KwLL) and Gompertz-log-logistic (GoLL); the four latter along the lines of [1,8,9,11] respectively. Table 7 brings the MLEs, SEs and information criteria.

Table 7. Fitted distributions to the eruption dataset (estimates and information criteria).

Distribution	Parameters	Estimates (SE)	AIC	CAIC	BIC	HQIC
NLL	α	28.71747 (2.751091)	587.5681	587.7649	591.8859	589.2691
	β	0.568200 (0.042998)				
LL	α	28.27831 (3.203986)	597.1497	597.3464	601.4674	598.8506
	β	1.969345 (0.198878)				
ExpLL	α	7.394859 (6.479904)	597.3629	597.7629	603.8396	599.9144
	β	1.461528 (0.197256)				
	a	4.572569 (4.470086)				
BLL	α	7.445103 (10.72863)	596.186	596.864	604.8215	599.588
	β	0.484528 (0.223626)				
	a	17.28664 (13.82502)				
	b	9.285354 (9.566756)				
KwLL	α	2.107772 (5.325557)	596.68	597.358	605.3156	600.082
	β	0.511324 (0.130629)				
	a	12.14489 (12.25210)				
	b	11.42477 (8.749787)				
GoLL	α	5.667167 (1.680598)	591.7172	592.3952	600.3528	595.1192
	β	4.348435 (1.450980)				
	a	0.035617 (0.017111)				
	b	0.234894 (0.100666)				

Since the Normal-log-logistic fitted model presents the smallest values of AIC, CAIC, BIC and HQIC compared to the fits of the other distributions, selecting it rather than the others is a reasonable decision in this case.

In Figure 6 the histogram of eruption data and the fitted densities with the three lowest values of AIC among the distributions in the first column of Table 7 are depicted. By a visual comparison, the three curves are apparently good approximations to the histogram, but the Normal-log-logistic seems to explain the behavior of the data more accurately.

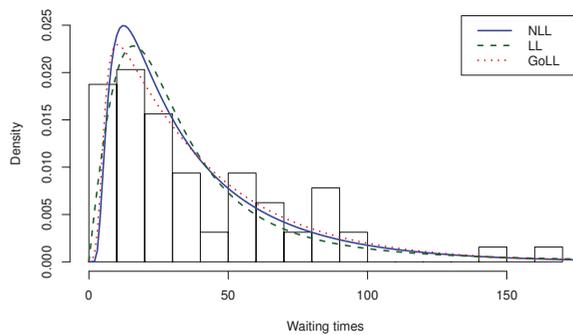


Figure 6. Histogram of eruption dataset and fitted densities.

Table 8 provides the values of A^* and W^* of the distributions in the first column of Table 7. These statistics suggest that GoLL and NLL models fit the eruption dataset very closely. Nonetheless, in order to pick a more parsimonious model, one should prefer the NLL, since it has fewer parameters than GoLL.

Table 8. Goodness-of-fit tests.

Distribution	A*	W*
NLL	0.612291	0.0803799
LL	1.019129	0.1413872
ExpLL	1.138218	0.1617136
BLL	0.837211	0.1141264
KwLL	0.818072	0.1118931
GoLL	0.605822	0.0805111

It is worth mentioning that [17] proposed the new class Exponentiated Kumaraswamy-G and fitted one of its submodels (with Weibull as baseline) to the same eruption dataset. It presented $A^* = 0.7594$ and $W^* = 0.1037$, whereas NLL presented lower values of these statistics as one can check in Table 8.

4. Concluding Remarks

Based on the method of generating classes of probability distributions presented by [12], we introduce a new class called Normal-G. It has the advantage of demanding no additional parameters besides the baseline ones. We demonstrate that the proposed class generates identifiable sub-models as long as the parent distribution is identifiable. The pdf of the class can be written as a linear combination of pdfs of exponentiated distributions; it allows us to easily derive the raw moments, the incomplete moments and the moment generating function.

We bring Monte Carlo simulation studies to attest the good performance of the MLEs of two distributions generated by the class and to illustrate its usefulness, applications to real datasets are made. The fitted models are compared to other competitive distributions regarding the Anderson-Darling and the Cramér-von Mises statistics, as well as commonly used information criteria as goodness-of-fit measures. The general results indicate that the Normal-G outperforms the other distributions in comparison. The new class is powerful and provides parsimonious models, which may hopefully interest practitioners of statistics, soil science, oceanography and other fields.

Author Contributions: All of the authors contributed relevantly to this research article.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The information matrix mentioned in Section 2.5 is given by $J(\xi|\mathbf{X}) = -\nabla_{\xi} \nabla_{\xi}^{\top} \ell(\xi|\mathbf{X}) = -(u_{ih})_{1 \leq i \leq r, 1 \leq h \leq r}$, where:

$$\begin{aligned} u_{ih} = & \sum_{j=1}^m \left[\frac{2G_{\xi}(x_j) - 3}{G_{\xi}^4(x_j)} - \frac{2G_{\xi}(x_j) + 1}{[1 - G_{\xi}(x_j)]^4} \right] \cdot \frac{\partial}{\partial \xi_i} G_{\xi}(x_j) \cdot \frac{\partial}{\partial \xi_h} G_{\xi}(x_j) \\ & + \sum_{j=1}^m \frac{[1 - G_{\xi}(x_j)]^4 - G_{\xi}^4(x_j)}{G_{\xi}^3(x_j)[1 - G_{\xi}(x_j)]^3} \cdot \frac{\partial^2}{\partial \xi_i \partial \xi_h} G_{\xi}(x_j) \\ & + \sum_{j=1}^m \frac{8[1 - G_{\xi}(x_j)]G_{\xi}(x_j)}{1 - 2G_{\xi}(x_j) + 2G_{\xi}^2(x_j)} \cdot \frac{\partial}{\partial \xi_i} G_{\xi}(x_j) \cdot \frac{\partial}{\partial \xi_h} G_{\xi}(x_j) \\ & + \sum_{j=1}^m \frac{4G_{\xi}(x_j) - 2}{1 - 2G_{\xi}(x_j) + 2G_{\xi}^2(x_j)} \cdot \frac{\partial^2}{\partial \xi_i \partial \xi_h} G_{\xi}(x_j) \\ & + \sum_{j=1}^m \frac{2}{G_{\xi}^2(x_j)} \cdot \frac{\partial}{\partial \xi_i} G_{\xi}(x_j) \cdot \frac{\partial}{\partial \xi_h} G_{\xi}(x_j) - \sum_{j=1}^m \frac{2}{G_{\xi}(x_j)} \cdot \frac{\partial^2}{\partial \xi_i \partial \xi_h} G_{\xi}(x_j) \\ & + \sum_{j=1}^m \frac{2}{[1 - G_{\xi}(x_j)]^2} \cdot \frac{\partial}{\partial \xi_i} G_{\xi}(x_j) \cdot \frac{\partial}{\partial \xi_h} G_{\xi}(x_j) + \sum_{j=1}^m \frac{2}{1 - G_{\xi}(x_j)} \cdot \frac{\partial^2}{\partial \xi_i \partial \xi_h} G_{\xi}(x_j) \\ & - \sum_{j=1}^m \frac{1}{g_{\xi}^2(x_j)} \cdot \frac{\partial}{\partial \xi_i} g_{\xi}(x_j) \cdot \frac{\partial}{\partial \xi_h} g_{\xi}(x_j) + \sum_{j=1}^m \frac{1}{g_{\xi}(x_j)} \cdot \frac{\partial^2}{\partial \xi_i \partial \xi_h} g_{\xi}(x_j) . \end{aligned}$$

References

- Mudholkar, G.S.; Srivastava, D.K.; Freimer M. The exponentiated Weibull family: A reanalysis of the bus motor failure data. *Technometrics* **1995**, *37*, 436–445. [\[CrossRef\]](#)
- Gupta, R.D.; Kundu D. Generalized Exponential Distributions. *Aust. N. Z. J. Stat.* **1999**, *41*, 173–188. [\[CrossRef\]](#)
- Marshall, A.W.; Olkin I. A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika* **1997**, *84*, 641–652. [\[CrossRef\]](#)
- Nadarajah, S. A Generalized Normal Distribution. *J. Appl. Stat.* **2005**, *32*, 685–694. [\[CrossRef\]](#)
- Azzalini, A. A Class of Distributions which includes the Normal ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
- Robertson, H.T.; Allison, D.B. A Novel Generalized Normal Distribution for Human Longevity and other Negatively Skewed Data. *PLoS ONE* **2012**, *7*, e37025. [\[CrossRef\]](#)
- Cordeiro, G.M.; Hashimoto, E.M.; Ortega, E.M.M. The McDonald Weibull Model. *Statistics* **2014**, *48*, 256–278. [\[CrossRef\]](#)
- Famoye, F.; Lee, C.; Olumolade, O. The Beta-Weibull distribution. *J. Stat. Theory Appl.* **2005**, *4*, 121–136.
- Cordeiro, G.M.; Castro, M. A new family of generalized distributions. *J. Stat. Comput. Simul.* **2010**, *81*, 883–898. [\[CrossRef\]](#)
- Alzaatreh, A.; Lee, C.; Famoye, F. A new method for generating families of continuous distributions. *Metron* **2013**, *71*, 63–79. [\[CrossRef\]](#)
- Alizadeh, M.; Cordeiro, G.M.; Pinho, L.G.B.; Ghosh, I. The Gompertz-G family of distributions. *J. Stat. Theory Pract.* **2017**, *11*, 179–207. [\[CrossRef\]](#)
- Brito, C.R.; Rêgo, L.C.; Oliveira, W.R.; Gomes-Silva F. Method for Generating Distributions and Classes of Probability Distributions: The Univariate Case. *Hacet. J. Math. Stat.* **2019**, *48*, 897–930.
- Xie, M.; Tang, Y.; Goh T.N. A modified Weibull extension with bathtub-shaped failure rate function. *Reliab. Eng. Syst. Safe.* **2002**, *76*, 279–285. [\[CrossRef\]](#)
- Bebbington, M.; Lai, C.D.; Zitikis, R. A flexible Weibull extension. *Reliab. Eng. Syst. Safe.* **2007**, *92*, 719–726. [\[CrossRef\]](#)
- Tahir, M.H.; Nadarajah, S. Parameter induction in continuous univariate distributions: Well-established G families. *An. Acad. Bras. Ciênc.* **2015**, *87*, 539–568. [\[CrossRef\]](#)

16. Cordeiro, G.M.; Ortega, E.M.M.; Cunha, D.C.C. The Exponentiated Generalized Class of Distributions. *J. Data Sci.* **2013**, *11*, 1–27.
17. Silva, R.; Gomes-Silva F.; Ramos, M.; Cordeiro, G.; Marinho, P.; Andrade, T.A.N. The Exponentiated Kumaraswamy-G Class: General Properties and Application. *Rev. Colomb. Estad.* **2019**, *42*, 1–33. [[CrossRef](#)]
18. Cakmakyapan, S.; Ozel, G. The Lindley Family of Distributions: Properties and Applications. *Hacet. J. Math. Stat.* **2017**, *46*, 1113–1137. [[CrossRef](#)]
19. Barreto-Souza, W.; Cordeiro, G.M.; Simas, A.B. Some Results for Beta Fréchet Distribution. *Commun. Stat. Theory Methods* **2011**, *40*, 798–811. [[CrossRef](#)]
20. Huang, S.; Oluyede, B.O. Exponentiated Kumaraswamy-Dagum distribution with applications to income and lifetime data. *J. Stat. Dist. Appl.* **2014**, *1*, 1–20. [[CrossRef](#)]
21. Cordeiro, G.M.; Bager, R.S.B. Moments for Some Kumaraswamy Generalized Distributions. *Comm. Statist. Theory Methods* **2015** *44*, 2720–2737. [[CrossRef](#)]
22. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018.
23. Von Neumann, J. Various techniques used in connection with random digits. In *Applied Mathematics Series 12*; National Bureau of Standards: Washington, DC, USA, 1951; pp. 36–38.
24. Fonseca, M.B. A influência da fertilidade do solo e caracterização da fixação biológica de N_2 para o crescimento de *Dimorphandra wilsonii* Rizz. Master's Thesis, Federal University of Minas Gerais, Belo Horizonte, Brazil, 2007.
25. Silva, R.B.; Bourguignon, M.; Dias, C.R.B.; Cordeiro, G.M. The compound class of extended Weibull power series distributions. *Comput. Stat. Data Anal.* **2013**, *58*, 352–367. [[CrossRef](#)]
26. Cordeiro, G.M.; Lemonte, A.J. On the Marshall–Olkin extended Weibull distribution. *Stat. Pap.* **2013**, *54*, 333–353. [[CrossRef](#)]
27. Chen, G.; Balakrishnan, N. A general purpose approximate goodness-of-fit test. *J. Qual. Technol.* **1995**, *27*, 154–161. [[CrossRef](#)]
28. StatSci.org. Available online: <http://www.statsci.org/data/oz/kiama.html> (accessed on 21 October 2019).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Generalized Truncation Positive Normal Distribution

Héctor J. Gómez ^{1,*}, Diego I. Gallardo ² and Osvaldo Venegas ¹

¹ Departamento de Ciencias Matemáticas y Físicas, Facultad de Ingeniería, Universidad Católica de Temuco, Temuco 4780000, Chile; ovenegas@uct.cl

² Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Copiapó 1530000, Chile; diego.gallardo@uda.cl

* Correspondence: hgomez@uct.cl; Tel.: +56-45-2205410

Received: 27 September 2019; Accepted: 29 October 2019; Published: 3 November 2019

Abstract: In this article we study the properties, inference, and statistical applications to a parametric generalization of the truncation positive normal distribution, introducing a new parameter so as to increase the flexibility of the new model. For certain combinations of parameters, the model includes both symmetric and asymmetric shapes. We study the model's basic properties, maximum likelihood estimators and Fisher information matrix. Finally, we apply it to two real data sets to show the model's good performance compared to other models with positive support: the first, related to the height of the drum of the roller and the second, related to daily cholesterol consumption.

Keywords: truncation; half-normal distribution; likelihood

1. Introduction

The half-normal (HN) distribution is a very important model in the statistical literature. Its density function has a closed-form and its cumulative distribution function (cdf) depends on the cdf of the standard normal model (or the error function), which is implemented in practically all mathematical and statistical software. Pewsey [1,2] provides the maximum likelihood (ML) estimation for the general location-scale HN distribution and its asymptotic properties. Wiper et al. [3] and Khan and Islam [4] perform analysis and applications for the HN model from a Bayesian framework. Moral et al. [5] also present the *hnp* R package, which produces half-normal plots with simulated envelopes using different diagnostics from a range of different fitted models. The HN model is also presented in the stochastic representation of the skew-normal distribution in Azzalini [6,7] and Henze [8]. In recent years this distribution has been used to model positive data, and it is becoming an important model in reliability theory despite the fact that it accommodates only decreasing hazard rates. Some of the generalizations of this distribution can be found in Cooray and Ananda [9], Olmos et al. [10], Cordeiro et al. [11], Gómez and Bolfarine [12], among others.

In particular, we focused on the extension of Cooray and Ananda [9]. The authors provided a motivation related to static fatigue life to consider the transformation $Z = \sigma Y^{1/\alpha}$, where $Y \sim HN(1)$. This model was named the generalized half-normal (GHN) distribution. An alternative way to extend the HN model was introduced by Gómez et al. [13] considering a normal distribution with mean and standard deviation μ and σ , respectively, truncated to the interval $(0, +\infty)$ and considering the reparametrization $\lambda = \mu/\sigma$. This model was named the truncated positive normal (TPN) distribution with density function given by

$$f(z; \sigma, \lambda) = \frac{1}{\sigma \Phi(\lambda)} \phi\left(\frac{z}{\sigma} - \lambda\right), \quad z, \sigma > 0, \lambda \in \mathbb{R}, \quad (1)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and cdf of the standard normal models, respectively. We use $TPN(\sigma, \lambda)$ to refer to a random variable (r.v.) with density function as in Equation (1). Note that $TPN(\sigma, 0) \equiv HN(\sigma)$.

In this work we consider a similar idea to that used in Cooray and Ananda [9] to extend the TPN model including the transformation $Y = \sigma X^{1/\alpha}$, where $X \sim TPN(1, \lambda)$. We will refer to this distribution as the generalized truncation positive normal (GTPN).

The rest of the manuscript is organized as follows. Section 2 is devoted to study of some important properties of the model, as well as its moments, quantile and hazard functions and its entropy. In Section 3 we perform an inference and present the Fisher information matrix for the proposed model. Section 4 discusses the selection model in nested and non-nested models for the GTPN distribution. In Section 5 we carry out a simulation study in order to study properties of the ML estimators in finite samples for the proposed distribution. Section 6 presents two applications to real data-sets to illustrate that the proposed model is competitive versus other common models for positive data in the literature. Finally, in Section 7, we present some concluding remarks.

2. Model Properties

In this section we introduce the main properties of the GTPN model such as density, quantile and hazard functions, moments, among others.

2.1. Stochastic Representation and Particular Cases

As mentioned previously, we say that a r.v. Z has $GTPN(\sigma, \lambda, \alpha)$ distribution if $Z = \sigma Y^{1/\alpha}$, where $Y \sim TPN(1, \lambda)$. By construction, the following models are particular cases for the GTPN distribution:

- $GTPN(\sigma, \lambda, \alpha = 1) \equiv TPN(\sigma, \lambda)$.
- $GTPN(\sigma, \lambda = 0, \alpha) \equiv GHN(\sigma, \alpha)$.
- $GTPN(\sigma, \lambda = 0, \alpha = 1) \equiv HN(\sigma)$.

Figure 1 summarizes the relationships among the GTPN and its particular cases. We highlight that $\lambda = 0$ and $\alpha = 1$ are within the parametric space (not on the boundary). Therefore, to decide between the GTPN versus the TPN, GHN, or HN distributions we can use classical hypothesis tests such as the likelihood ratio test (LRT), score test (ST), or gradient test (GT).

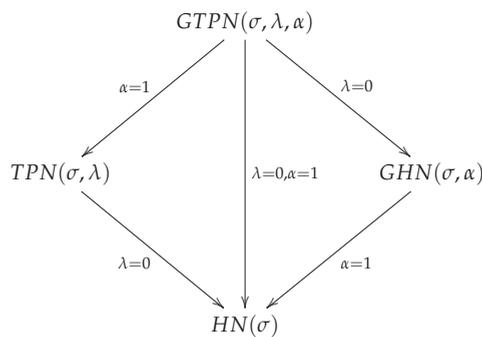


Figure 1. Particular cases for the GTPN distribution.

2.2. Density, Cdf and Hazard Functions

Proposition 1. For $Z \sim GTPN(\sigma, \lambda, \alpha)$, the density function is given by

$$f(z; \sigma, \lambda, \alpha) = \frac{\alpha}{\sigma^\alpha \Phi(\lambda)} z^{\alpha-1} \phi\left(\left(\frac{z}{\sigma}\right)^\alpha - \lambda\right), \quad z \geq 0. \tag{2}$$

where $\sigma, \alpha > 0$ and $\lambda \in \mathbb{R}$

Proof. Considering the stochastic representation discussed in Section 2.1, we have that $z = g(y) = \sigma y^{1/\alpha}$ and

$$f_Z(z) = f_Y(g^{-1}(z)) \left| \frac{dg^{-1}(z)}{dz} \right| = \frac{1}{\Phi(\lambda)} \phi(g^{-1}(z) - \lambda) \frac{\alpha z^{\alpha-1}}{\sigma^\alpha}.$$

Replacing $g^{-1}(z) = (z/\sigma)^\alpha$ the result is obtained. \square

Proposition 2. For $Z \sim \text{GTPN}(\sigma, \lambda, \alpha)$, the cdf and hazard function are given by

$$F(z; \sigma, \lambda, \alpha) = \frac{\Phi\left(\left(\frac{z}{\sigma}\right)^\alpha - \lambda\right) + \Phi(\lambda) - 1}{\Phi(\lambda)} \quad \text{and} \quad h(z) = \frac{f(z)}{1 - F(z)} = \frac{\alpha z^{\alpha-1} \phi\left(\left(\frac{z}{\sigma}\right)^\alpha - \lambda\right)}{\sigma^\alpha \left[1 - \Phi\left(\left(\frac{z}{\sigma}\right)^\alpha - \lambda\right)\right]},$$

respectively, for all $z \geq 0$.

Figure 2 shows the density and hazard functions for the $\text{GTPN}(\sigma = 1, \lambda, \alpha)$ model, considering some combinations for (λ, α) . Note that the GTPN model can assume decreasing and unimodal shapes for the density function and decreasing or increasing shapes for the hazard function.

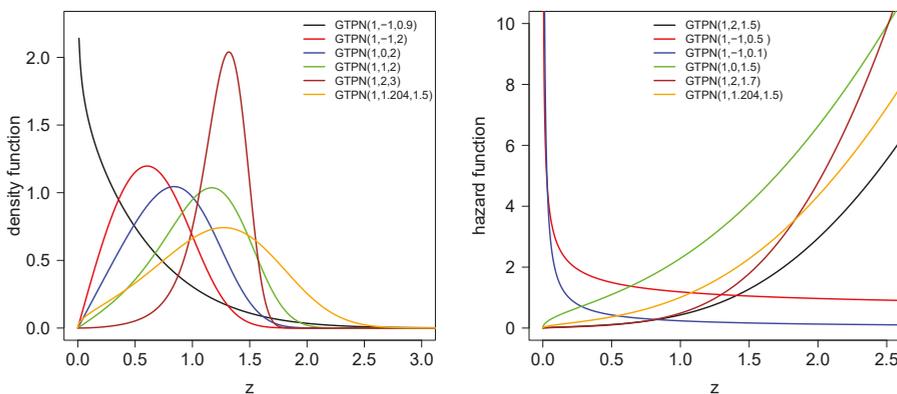


Figure 2. Density and hazard functions for the $\text{GTPN}(\sigma = 1, \lambda, \alpha)$ model with different combinations of λ and α .

2.3. Mode

Proposition 3. The mode of the $\text{GTPN}(\sigma, \lambda, \alpha)$ model is attained:

1. at $z = \frac{\sigma}{2^{1/\alpha}} \left(\lambda + \sqrt{\lambda^2 + 4 \left(1 - \frac{1}{\alpha}\right)} \right)^{1/\alpha}$ whenever $\alpha \geq 1$ or $0 < \alpha < 1$ and $\lambda > 0$,
2. at $z = 0$ in otherwise.

Proof. Let $l = \log(f)$, where f is the density function defined in (2), of a direct computation we have

$$\frac{\partial l}{\partial z} = l_z = -\frac{1}{z} \left[\alpha \left(\frac{z}{\sigma}\right)^{2\alpha} - \alpha \lambda \left(\frac{z}{\sigma}\right)^\alpha - (\alpha - 1) \right].$$

Note that l_z vanishes when $\alpha \left(\frac{z}{\sigma}\right)^{2\alpha} - \alpha \lambda \left(\frac{z}{\sigma}\right)^\alpha - (\alpha - 1) = 0$. Using the auxiliary variable $w = \left(\frac{z}{\sigma}\right)^\alpha$ the last equation is rewritten as follows:

$$\alpha w^2 - \alpha \lambda w - (\alpha - 1) = 0. \tag{3}$$

In the rest of the proof we use the discriminant of the quadratic equation in Equation (3), which is given by $\Delta = \alpha^2 \lambda^2 + 4\alpha(\alpha - 1) = \alpha^2[\lambda^2 + 4(1 - \frac{1}{\alpha})]$ and its zeros are given by: $w = (\lambda \pm \sqrt{\lambda^2 + 4(1 - \frac{1}{\alpha})})/2$.

If $\alpha \geq 1$ then $\Delta \geq \lambda^2$. In consequence, the mode is attained at $z_1 = \frac{\sigma}{2^{1/\alpha}} \left(\lambda + \sqrt{\lambda^2 + 4 \left(1 - \frac{1}{\alpha}\right)} \right)^{1/\alpha}$.

If $0 < \alpha < 1$ then $\Delta < \lambda^2$. Here, two cases may occur. The first when $0 < \Delta < \lambda^2$, in which case if $\lambda > 0$ its mode is attained at $z = \frac{\sigma}{2^{1/\alpha}} \left(\lambda + \sqrt{\lambda^2 + 4 \left(1 - \frac{1}{\alpha}\right)} \right)^{1/\alpha}$, since $l_{zz} < 0$, and if $\lambda < 0$, then the zeros of Equation (3) are negative, implying that function l is strictly decreasing. Its mode is therefore attained at zero. The other case is when $\Delta < 0$, then we have that $\alpha w^2 - \alpha \lambda w - (\alpha - 1) > 0$ for all $w \geq 0$, implying that $l_z < 0$ for all $z \geq 0$. Therefore, l is strictly decreasing and thus its mode is zero. \square

Remark 1. Note that $\alpha \geq 1$ or $\lambda > 0$ implies that the mode of the GTPN model is attached in a positive value.

2.4. Quantiles

Proposition 4. The quantile function for the GTPN(σ, λ, α) is given by

$$Q(p) = \sigma \left[\Phi^{-1}(1 - (1 - p)\Phi(\lambda)) + \lambda \right]^{\frac{1}{\alpha}}.$$

Proof. Follows from a direct computation, applying the definition of quantile function. \square

Corollary 1. The quartiles of the GTPN distribution are

1. First quartile $\sigma[\Phi^{-1}(1 - \frac{3}{4}\Phi(\lambda)) + \lambda]^{\frac{1}{\alpha}}$.
2. Median $\sigma[\Phi^{-1}(1 - \frac{\Phi(\lambda)}{2}) + \lambda]^{\frac{1}{\alpha}}$.
3. Third quartile $\sigma[\Phi^{-1}(1 - \frac{1}{4}\Phi(\lambda)) + \lambda]^{\frac{1}{\alpha}}$.

2.5. Central Moments

Proposition 5. Let $Z \sim \text{GTPN}(\sigma, \lambda, \alpha)$ and $r = 1, 2, \dots$. The r -th non-central moment is given by

$$\mu_r = \mathbb{E}(Z^r) = \frac{(\sigma \lambda^{1/\alpha})^r b_r(\lambda, \alpha)}{2\Phi(\lambda)\sqrt{\pi}},$$

where $b_r(\lambda, \alpha) = \sum_{k=0}^{\infty} \binom{r/\alpha}{k} \left(\frac{\lambda}{2}\right)^{-k} \Gamma\left(\frac{k+1}{2}, \frac{\lambda^2}{2}\right)$ and $\binom{r/\alpha}{k} = \frac{1}{k!} \prod_{n=1}^{k-1} (r/\alpha - n)$ is the generalized binomial coefficient. When $r/\alpha \in \mathbb{N}$, the sum in $b_r(\lambda, \alpha)$ stops at r/α .

Proof. Considering the stochastic representation of the GTPN model in Section 2.1, it is immediate that $\mathbb{E}(Z^r) = \sigma^r \mathbb{E}(Y_{\frac{r}{\alpha}}^{\frac{r}{\alpha}})$, where $Y \sim \text{TPN}(1, \lambda)$. This expected value can be computed using Proposition 2.2 in Gómez et al. [13]. \square

Remark 2. When $r/\alpha \in \mathbb{N}$, Closed Forms Can Be Obtained for μ_r

Figure 3 illustrates the mean, variance, skewness, and kurtosis coefficients for the GTPN($\sigma = 1, \lambda, \alpha$) model for some combinations of its parameters.

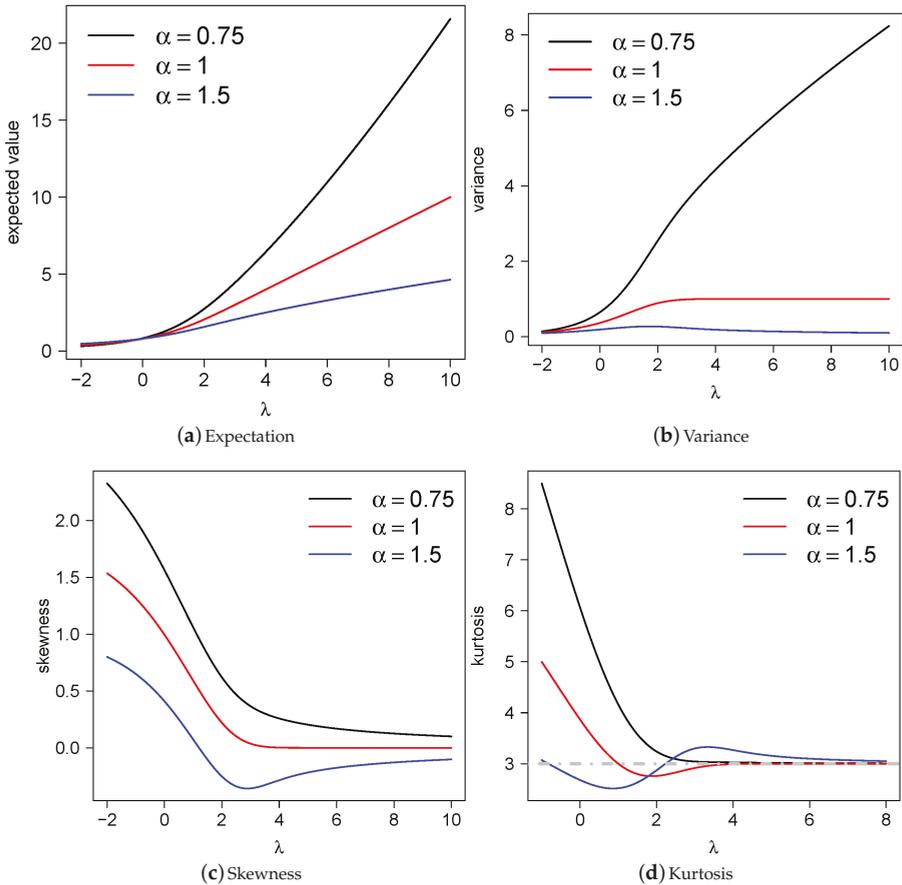


Figure 3. Plots of the (a) expectation, (b) variance, (c) skewness and (d) kurtosis for GTPN($\sigma = 1, \lambda, \alpha$) for $\alpha \in \{0.75, 1, 1.5\}$ as a function of λ . In (d), the dashed line represents the kurtosis of the normal distribution.

2.6. Bonferroni and Lorenz Curves

In this subsection we present the Bonferroni and Lorenz curves (see Bonferroni [14]). These curves have applications not only in economics to study income and poverty, but also in medicine, reliability, etc. The Bonferroni curve is defined as

$$B(p) = \frac{1}{p\mu} \int_0^q zf(z)dz, \quad 0 \leq p < 1.$$

where $\mu = E(Z)$, $q = F^{-1}(p)$. The Lorenz curve is obtained by the relation $L(p) = pB(p)$. Particularly, it can be checked that for the GTPN model the Bonferroni curve is given by

$$B(p) = \frac{1}{p\mu} \left[E(Z) - \frac{\sigma}{\Phi(\lambda)\sqrt{2\pi}} \sum_{k=0}^{\infty} \binom{\frac{1}{\alpha}}{k} \lambda^{\frac{1}{\alpha} - k} 2^{\frac{k-1}{2}} \Gamma\left(\frac{k+1}{2}, \frac{((\frac{q}{\sigma})^\alpha - \lambda)^2}{2}\right) \right].$$

These curves serve as graphic methods for analysis and comparison, e.g., the inequality of non-negative distributions. See, for example, for a more detailed discussion [15].

Figure 4 shows the Bonferroni curve for the GTPN($\sigma = 1, \lambda, \alpha$) model, considering different values for λ and α .

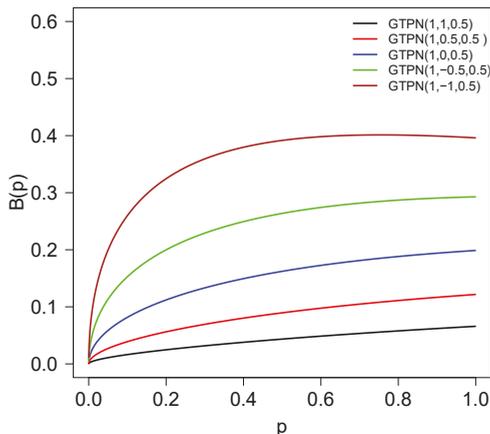


Figure 4. Bonferroni curve for the generalized truncation positive normal (GTPN)(σ, λ, α) model.

2.7. Shannon Entropy

Shannon entropy (see Shannon [16]) measures the amount of uncertainty for a random variable. It is defined as:

$$S(Z) = -\mathbb{E}(\log f(Z)).$$

Therefore, it can be checked that the Shannon entropy for the GTPN model is

$$S(Z) = \log\left(\frac{\sqrt{2\pi\sigma}\Phi(\lambda)}{\alpha}\right) - (\alpha - 1)\mathbb{E}(\log(Z)) + \frac{\mu^{2\alpha}}{2\sigma^{2\alpha}} - \frac{\lambda\mu^\alpha}{\sigma^\alpha} + \frac{\lambda^2}{2},$$

Figure 5 shows the entropy curve for the GTPN($\sigma = 1.5, \lambda, \alpha$) model, considering different values for λ and α . We note that this function is increasing in λ and α . where $\mathbb{E}(\log(Z)) = \int_0^\infty \log(z)f(z;\sigma, \lambda, \alpha)dz$. For $\alpha = 1$, the Shannon entropy is reduced to

$$S(Z) = \log\left(\sqrt{2\pi\sigma}\Phi(\lambda)\right) + \frac{\mu^2}{2\sigma^2} - \frac{\lambda\mu}{\sigma} + \frac{\lambda^2}{2},$$

which corresponds to the Shannon entropy for the TPN model; and for $\alpha = 1$ and $\lambda = 0$, $S(Z)$ is reduced to

$$S(Z) = \log(\sqrt{\pi\sigma}) + \frac{\mu^2}{2\sigma^2},$$

which corresponds to the Shannon entropy for the HN distribution.

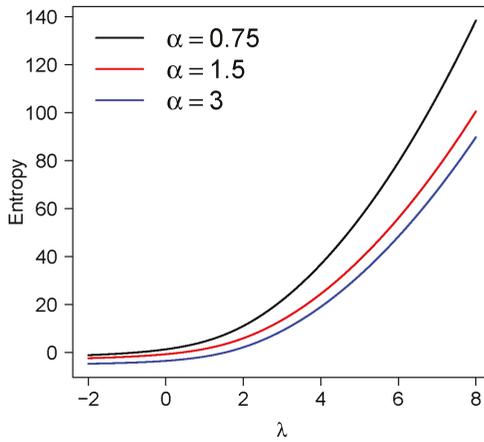


Figure 5. Entropy for the $GTPN(\sigma = 1, \alpha, \lambda)$ model.

3. Inference

In this section we discuss the ML method for parameter estimation in the GTPN model.

3.1. Maximum Likelihood Estimators

For a random sample z_1, \dots, z_n from the $GTPN(\sigma, \lambda, \alpha)$ model, the log-likelihood function for $\theta = (\sigma, \lambda, \alpha)$ is given by

$$\ell(\theta) = n \log(\alpha) + (\alpha - 1) \sum_{i=1}^n \log(z_i) - n\alpha \log(\sigma) - n \log(\Phi(\lambda)) - \frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \left(\left(\frac{z_i}{\sigma} \right)^\alpha - \lambda \right)^2.$$

Therefore, the score assumes the form $S(\theta) = (S_\sigma(\theta), S_\lambda(\theta), S_\alpha(\theta))$, where

$$\begin{aligned} S_\sigma(\theta) &= -\frac{n\alpha}{\sigma} + \sum_{i=1}^n \left(\left(\frac{z_i}{\sigma} \right)^\alpha - \lambda \right) \alpha \left(\frac{z_i}{\sigma} \right)^{\alpha-1} \frac{z_i}{\sigma^2}, \\ S_\lambda(\theta) &= -n\zeta(\lambda) + \sum_{i=1}^n \left(\left(\frac{z_i}{\sigma} \right)^\alpha - \lambda \right) \quad \text{and} \\ S_\alpha(\theta) &= \frac{n}{\alpha} + \sum_{i=1}^n \log z_i - n \log(\sigma) - \sum_{i=1}^n \left(\left(\frac{z_i}{\sigma} \right)^\alpha - \lambda \right) \left(\frac{z_i}{\sigma} \right)^\alpha \log \left(\frac{z_i}{\sigma} \right), \end{aligned}$$

where $\zeta(\lambda) = \frac{\phi(\lambda)}{\Phi(\lambda)}$ is the negative of the inverse Mills ratio. The ML estimators are obtained by solving the equation $S(\theta) = \mathbf{0}_3$, where $\mathbf{0}_p$ denotes a vector of zeros with dimension p . This equation has the following solution for λ

$$\hat{\lambda}(\hat{\sigma}, \hat{\alpha}) = \frac{\sum_{i=1}^n \left(\frac{z_i}{\hat{\sigma}} \right)^{2\hat{\alpha}} - n}{\sum_{i=1}^n \left(\frac{z_i}{\hat{\sigma}} \right)^{\hat{\alpha}}}. \tag{4}$$

Replacing Equation (4) in $S_\lambda(\theta) = 0$ and $S_\alpha(\theta) = 0$, the problem is reduced to two equations. The solution of this problem needs to be solved by numerical methods such as Newton-Raphson. Below we discuss initial values for the vector θ to initialize the algorithm.

3.2. Initial Point to Obtain the Maximum Likelihood Estimators

In this subsection, we discuss the initial points for the iterative methods to find the ML estimators in the GTPN distribution.

3.2.1. A Naive Point Based on the HN Model

In Section 2 we discuss that $GTPN(\sigma, \lambda = 0, \alpha = 1) \equiv HN(\sigma)$. Based on this fact, and considering that the ML estimator for σ in the HN distribution has a closed-form, we can consider as an initial point $\theta_{naive} = \left(\sqrt{\frac{1}{n} \sum_{i=1}^n z_i^2}, 0, 1 \right)$.

3.2.2. An Initial Point Based on Centiles

Let $q_t, t = 1, \dots, 99$, the t -th sample centile based on z_1, \dots, z_n . An initial point to θ can be obtained by matching q_u, q_{50} and q_{100-u} , with $u \in \{1, 2, \dots, 48, 49\}$, with their respective theoretical counterparts. Defining $p = u/100$, the equations obtained are

$$q_u = \sigma \left[\Phi^{-1}(1 - (1 - p)\Phi(\lambda)) + \lambda \right]^{\frac{1}{\alpha}}, \quad q_{50} = \sigma \left[\Phi^{-1}(1 - 0.5\Phi(\lambda)) + \lambda \right]^{\frac{1}{\alpha}} \quad \text{and}$$

$$q_{100-u} = \sigma \left[\Phi^{-1}(1 - p\Phi(\lambda)) + \lambda \right]^{\frac{1}{\alpha}}.$$

The solutions for σ and α are

$$\tilde{\sigma} = \tilde{\sigma}(\tilde{\lambda}) = \frac{q_{100-u}}{\left[\Phi^{-1}(1 - p\Phi(\tilde{\lambda})) + \tilde{\lambda} \right]^{1/\tilde{\alpha}(\tilde{\lambda})}} \quad \text{and} \quad \tilde{\alpha} = \tilde{\alpha}(\tilde{\lambda}) = \frac{\log \left(\frac{\Phi^{-1}(1 - (1-p)\Phi(\tilde{\lambda})) + \tilde{\lambda}}{\Phi^{-1}(1 - p\Phi(\tilde{\lambda})) + \tilde{\lambda}} \right)}{\log \left(\frac{q_u}{q_{100-u}} \right)},$$

where $\tilde{\lambda}$ is obtained from the non-linear equation

$$\tilde{\sigma}(\tilde{\lambda}) \left[\Phi^{-1}(1 - 0.5\Phi(\tilde{\lambda})) + \tilde{\lambda} \right]^{1/\tilde{\alpha}(\tilde{\lambda})} = q_{50}.$$

Therefore, the initial point based on this method is given by $\theta_{cent} = (\tilde{\sigma}, \tilde{\lambda}, \tilde{\alpha})$.

3.3. An Initial Point Based on the Method of Moments

A more robust initial point can be obtained using the method of moments. The equations to solve are $\mu_r = \bar{z}^r, r = 1, 2, 3$. The solution for σ is

$$\tilde{\sigma}^* = \tilde{\sigma}^*(\tilde{\lambda}^*, \tilde{\alpha}^*) = \frac{2\sqrt{\pi}\bar{z}\Phi(\tilde{\lambda}^*)}{(\tilde{\lambda}^*)^{1/\tilde{\alpha}^*} b_1(\tilde{\lambda}^*, \tilde{\alpha}^*)}.$$

The solution for λ and α (say $\tilde{\lambda}^*$ and $\tilde{\alpha}^*$, respectively) are obtained from the non-linear equations

$$\frac{(\tilde{\sigma}^*(\lambda, \alpha)\lambda^{1/\alpha})^2 b_2(\alpha, \lambda)}{2\Phi(\lambda)\sqrt{\pi}} = \bar{z}^2 \quad \text{and} \quad \frac{(\tilde{\sigma}^*(\lambda, \alpha)\lambda^{1/\alpha})^3 b_3(\alpha, \lambda)}{2\Phi(\lambda)\sqrt{\pi}} = \bar{z}^3.$$

Therefore, the initial point based on this method is given by $\theta_{mom} = (\tilde{\sigma}^*, \tilde{\lambda}^*, \tilde{\alpha}^*)$.

3.4. Fisher Information Matrix

The Fisher information (FI) matrix for the GTPN distribution is given by $IF(\theta) = (I_{ab})_{a,b \in \{\sigma, \lambda, \alpha\}}$. Consider the notation

$$T_{jk} = T_{jk}(\lambda) = \int_0^\infty [\log(w)]^k w^j \phi(w - \lambda) dw, \quad j \in \{1, 2\}, \quad k \in \{0, 1, 2\}.$$

Therefore,

$$\begin{aligned} I_{\sigma\sigma} &= -\frac{\alpha}{\sigma^2} + \frac{\alpha(2\alpha + 1)}{\sigma^2\Phi(\lambda)}T_{20} - \frac{\lambda\alpha(\alpha + 1)}{\sigma^2\Phi(\lambda)}T_{10}, \\ I_{\sigma\lambda} &= \frac{\alpha}{\sigma\Phi(\lambda)}T_{10}, \\ I_{\sigma\alpha} &= \frac{1}{\sigma} + \frac{1}{\sigma\Phi(\lambda)}[\lambda(T_{10} + T_{11}) - T_{20} - 2T_{21}], \\ I_{\lambda\lambda} &= 1 - \lambda\zeta(\lambda) - \xi^2(\lambda), \\ I_{\lambda\alpha} &= -\frac{1}{\alpha\Phi(\lambda)}T_{11}, \\ I_{\alpha\alpha} &= \frac{1}{\alpha^2} + \frac{2}{\alpha^2\Phi(\lambda)}T_{22} - \frac{\lambda}{\alpha^2\Phi(\lambda)}T_{12} \end{aligned}$$

We observe that $T_{jk}(\lambda)$ is a continuous function and $\lim_{|\lambda| \rightarrow +\infty} T_{jk}(\lambda) = 0$, then $T_{jk}(\lambda) < +\infty$. Note that for $\lambda = 0$ and $\alpha = 1$ this matrix is reduced to

$$FI(\sigma, 0, 1) = \begin{pmatrix} \frac{2}{\sigma^2} & \frac{1}{\sigma}\sqrt{\frac{2}{\pi}} & -\frac{4}{\sigma}\widehat{T}_{21} \\ \cdot & 1 - \frac{2}{\pi} & -2\widehat{T}_{11} \\ \cdot & \cdot & 1 + 4\widehat{T}_{22} \end{pmatrix},$$

where $\widehat{T}_{jk} = T_{jk}(0)$. Additionally, we note that

$$\det(FI(\sigma, 0, 1)) = \frac{2}{\sigma^2} \left[\frac{(\pi - 3)}{\pi}(1 + 4\widehat{T}_{22} + 4\widehat{T}_{21}) - 4(\widehat{T}_{11} - \sqrt{\frac{2}{\pi}}\widehat{T}_{21})^2 \right] \approx \frac{0.1021506}{\sigma^2} > 0, \quad \forall \sigma > 0,$$

where $\det(\cdot)$ denotes the determinant operator. Therefore, the FI matrix for the reduced model (HN) is invertible.

4. Model Discrimination

In this section we discuss some techniques to discriminate among the GTPN distribution and other models.

4.1. GTPN versus Submodels

An interesting problem to solve is the discrimination between GTPN and the three submodels represented in Figure 1. In other words, we are interested in testing the following hypotheses:

- $H_0^{(1)} : \alpha = 1$ versus $H_1^{(1)} : \alpha \neq 1$ (TPN versus GTPN distribution).
- $H_0^{(2)} : \lambda = 0$ versus $H_1^{(2)} : \lambda \neq 0$ (GHN versus GTPN distribution).
- $H_0^{(3)} : (\alpha, \lambda) = (1, 0)$ versus $H_1^{(3)} : (\alpha, \lambda) \neq (1, 0)$ (HN versus GTPN distribution).

The three hypotheses can be tested considering the LRT, ST, and GT. Below we present the statistics for the three tests considered and for the three hypotheses of interest.

4.1.1. Likelihood Ratio Test

The statistic for the LRT (say SLR) to tests $H_0^{(j)}$, $j = 1, 2, 3$, is defined as

$$\text{SLR}_j = 2 \left[\ell \left(\hat{\sigma}, \hat{\lambda}, \hat{\alpha} \right) - \ell \left(\hat{\sigma}_{0j}, \hat{\lambda}_{0j}, \hat{\alpha}_{0j} \right) \right],$$

where $\hat{\sigma}_{0j}$, $\hat{\lambda}_{0j}$ and $\hat{\alpha}_{0j}$ denote the ML estimators for σ , λ and α restricted to $H_0^{(j)}$, $j = 1, 2, 3$. Under $H_0^{(j)}$, $j = 1, 2$, $\text{SLR}_j \sim \chi_{(1)}^2$ and under $H_0^{(3)}$, $\text{SLR}_3 \sim \chi_{(2)}^2$, where $\chi_{(p)}^2$ denotes the Chi-squared distribution with p degrees of freedom. For $H_0^{(3)}$, we obtain

$$\hat{\sigma}_{03} = \sqrt{\frac{1}{n} \sum_{i=1}^n z_i^2}, \quad \hat{\lambda}_{03} = 0 \quad \text{and} \quad \hat{\alpha}_{03} = 1,$$

whereas to test $H_0^{(1)}$ and $H_0^{(2)}$ the ML estimators under the null hypotheses need to be computed numerically. However, in both cases the problem is reduced to a unidimensional maximization. For details see Cooray and Ananda [9] and Gómez et al. [13], respectively.

4.1.2. Score Test

The statistic for the ST (say SR) to test $H_0^{(j)}$, $j = 1, 2, 3$, is defined as

$$\text{SR}_j = \left[S \left(\hat{\theta}_{0j} \right) \right]^\top \left[IF \left(\hat{\theta}_{0j} \right) \right]^{-1} S \left(\hat{\theta}_{0j} \right),$$

where $\hat{\theta}_j$ is the ML estimator under $H_0^{(j)}$. Under $H_0^{(j)}$, $j = 1, 2$, $\text{SR}_j \sim \chi_{(1)}^2$ and under $H_0^{(3)}$, $\text{SR}_3 \sim \chi_{(2)}^2$.

4.1.3. Gradient Test

The statistic for the GT (say ST) to tests $H_0^{(j)}$, $j = 1, 2, 3$, is defined as

$$\text{ST}_j = S \left(\hat{\theta}_j \right) \left(\hat{\theta} - \hat{\theta}_j \right).$$

Again, under $H_0^{(j)}$, $j = 1, 2$, $\text{ST}_j \sim \chi_{(1)}^2$ and under $H_0^{(3)}$, $\text{ST}_3 \sim \chi_{(2)}^2$. After some algebraic manipulations, we obtain that

$$\text{ST}_1 = (\hat{\alpha} - 1) \left\{ n \left[1 - \log \left(\hat{\sigma}_{01} \right) \right] + \sum_{i=1}^n \left[\log z_i - \left(\left(\frac{z_i}{\hat{\sigma}_{01}} \right) - \hat{\lambda}_{01} \right) \left(\frac{z_i}{\hat{\sigma}_{01}} \right) \log \left(\frac{z_i}{\hat{\sigma}_{01}} \right) \right] \right\}.$$

$$\text{ST}_2 = \hat{\lambda} \left[-n \sqrt{\frac{2}{\pi}} + \sum_{i=1}^n \left(\frac{z_i}{\hat{\sigma}_{02}} \right)^{\hat{\alpha}_{02}} \right].$$

$$\text{ST}_3 = (\hat{\alpha} - 1) \left\{ n \left[1 - \log \left(\hat{\sigma}_{03} \right) \right] + \sum_{i=1}^n \left[\log z_i - \left(\frac{z_i}{\hat{\sigma}_{03}} \right)^2 \log \left(\frac{z_i}{\hat{\sigma}_{03}} \right) \right] \right\} + \hat{\lambda} \left[-n \sqrt{\frac{2}{\pi}} + \sum_{i=1}^n \left(\frac{z_i}{\hat{\sigma}_{03}} \right) \right].$$

4.2. Non-Nested Models

The comparison of non-nested models can be performed based on the AIC criteria (Akaike [17]), where the model with a lower AIC is preferred. However, in practice we can have a set of inappropriate models for a certain data set. For this reason, we also need to perform a goodness-of-fit validation. This can be performed, for instance, based on the quantile residuals (QR). For more details see Dunn and Smyth [18]. These residuals are defined as

$$QR_i = G(z_i; \hat{\psi}), \quad i = 1, \dots, n,$$

where $G(\cdot; \hat{\psi})$ is the cdf of the specified distribution evaluated in the estimator for ψ . If the model is correctly specified, such residuals are a random sample from the standard normal distribution. This can be assessed using, for instance, the Anderson–Darling (AD), Cramer–Von–Mises (CVM) and Shapiro–Wilks (SW) tests. A discussion of these tests can be seen in Yazici and Yocalan [19].

5. Simulation

In this section we present a Monte Carlo (MC) simulation study in order to illustrate the behavior of the ML estimators. We consider three sample sizes: $n = 50, 150$ and 300 ; two values for σ : 1 and 2 ; two values for λ : 3 and 4 ; and three values for α : $0.8, 1$ and 2 . For each combination of n, σ, λ and α , we draw $10,000$ samples of size n from the $GTPN(\sigma, \lambda, \alpha)$ model. To simulate a value from this distribution, we consider the following scheme:

1. Simulate $X \sim Uniform(0, 1)$.
2. Compute $Y = \Phi^{-1}(1 + (X + 1)\Phi(\lambda)) + \lambda$.
3. Compute $Z = \sigma Y^{\frac{1}{\alpha}}$.

For each sample generated, ML estimators were computed numerically using the Newton–Raphson algorithm. Table 1 presents means and standard deviations for each parameter in each case. Notice that bias and standard deviations are reduced as the sample size increases, suggesting that the ML estimators are consistent.

Table 1. Monte Carlo (MC) simulation study for the maximum likelihood (ML) estimators in the GTPN(σ, λ, α) model in 12 combinations of σ, λ and α . The results summarize the mean and the standard deviation (sd) of the respective estimators obtained in the 10,000 replicates.

True Value	σ	λ	α	$n = 50$			$n = 150$			$n = 300$			$n = 1000$														
				$\hat{\sigma}$	mean	sd																					
1	0.8	1	0.8	1.523	1.573	3.089	1.797	0.987	1.029	1.227	1.029	3.000	1.052	0.875	0.350	1.081	0.614	3.017	0.634	0.830	0.202	1.011	0.233	3.014	0.267	0.803	0.071
			1	1.286	1.105	3.118	1.843	1.227	0.705	1.111	0.705	3.027	1.040	1.083	0.425	1.044	0.444	3.020	0.629	1.031	0.250	1.006	0.185	3.012	0.266	1.004	0.089
	4	2	1.006	0.510	3.220	2.036	2.432	0.309	1.011	0.309	3.019	1.055	2.171	0.850	1.001	0.196	3.022	0.614	2.058	0.481	0.998	0.093	3.013	0.266	2.008	0.177	
			1.489	1.799	4.391	2.010	0.971	0.860	1.087	0.860	4.280	1.199	0.817	0.282	1.017	0.513	4.164	0.786	0.798	0.148	1.003	0.301	4.049	0.413	0.799	0.082	
	2	0.8	1	1.236	1.219	4.439	2.182	1.203	0.631	1.018	0.631	4.295	1.222	1.017	0.343	0.986	0.418	4.187	0.826	0.995	0.191	0.993	0.230	4.055	0.404	0.998	0.099
				2	0.962	0.543	4.642	2.495	2.360	0.321	0.957	0.321	4.323	1.326	2.033	0.696	0.974	0.223	4.168	0.839	2.000	0.382	0.993	0.118	4.045	0.404	2.000
3		2	3.006	3.131	3.157	1.876	0.974	1.996	2.423	1.996	3.018	1.045	0.869	0.340	2.152	1.185	3.023	0.624	0.823	0.194	2.034	0.472	3.007	0.269	0.805	0.072	
			2.512	2.198	3.215	1.942	1.207	1.437	2.237	1.437	3.028	1.077	1.086	0.431	2.094	0.897	3.017	0.642	1.033	0.250	2.009	0.373	3.015	0.269	1.004	0.089	
4		2	1.999	1.039	3.318	2.223	2.417	0.631	2.017	0.631	3.035	1.096	2.171	0.862	2.000	0.405	3.027	0.638	2.061	0.504	1.995	0.186	3.016	0.267	2.006	0.177	
			2.880	3.544	4.562	2.283	0.970	1.714	2.161	1.714	4.326	1.288	0.813	0.282	2.024	1.078	4.194	0.849	0.796	0.160	2.003	0.573	4.053	0.404	0.799	0.079	
	2	2.402	2.434	4.659	2.419	1.176	1.283	2.033	1.283	4.344	1.346	1.015	0.355	1.974	0.835	4.190	0.844	0.995	0.191	1.990	0.465	4.053	0.408	0.998	0.100		
		1.913	1.132	4.802	2.848	2.378	0.668	1.898	0.668	4.393	1.469	2.024	0.708	1.938	0.460	4.198	0.901	1.993	0.391	1.984	0.236	4.047	0.401	2.000	0.198		

6. Applications

In this section we present two real data applications to illustrate the better performance of the GTPN model over other well known models in the literature. For these comparisons we also consider the Weibull (WEI) and the Generalized Lindley (GL, Zakerzadeh [20]) models. The density function of the Weibull distribution is given by

$$f(x; \sigma, \lambda) = \frac{\sigma}{\lambda} x^{\sigma-1} e^{-x^\sigma/\lambda},$$

with $x > 0, \sigma > 0$ and $\lambda > 0$, whereas for the GL model is given by

$$f(x; \alpha, \theta, \gamma) = \frac{\theta^2(\theta x)^{\alpha-1}(\alpha + \gamma x)e^{-\theta x}}{(\gamma + \theta)\Gamma(\alpha + 1)},$$

with $x > 0, \theta > 0, \alpha > 0$ and $\gamma > 0$.

6.1. Application 1

The data set was taken from Laslett [21], and consisted of $n = 115$ heights measured at 1 micron intervals along the drum of a roller (i.e., parallel to the axis of the roller). This was part of an extensive study of the surface roughness of the rollers. A statistical summary of the data set is presented in Table 2.

Table 2. Descriptive statistics of the Laslett data set.

Dataset	n	\bar{X}	S^2	$\sqrt{b_1}$	b_2
Heights measured	115	3.48	0.52	-1.24	6.30

Initially, we calculate the estimators based on the centiles, naive, and moments of the GTPN distribution, which are $\theta_{cent} = (2.326, 2.741, 2.376)$, $\theta_{naive} = (3.557, 0, 1)$, and $\theta_{mom} = (3.075, 1.572, 3.203)$. We used these estimations as initial values in computing the ML estimators for the GTPN model. Results are presented in Table 3. Note the high estimated standard error for the γ parameter in the GL model. In addition, note that, based on the AIC criteria and BIC criteria [22], the GTPN model is preferred (among the fitted models) for this data set. Figure 6 shows the estimated density for each model in this data set, where the GTPN model appears to provide a better fit. Finally, Figure 7 also presents the qq-plot for the QR in the same models and the p -values for the three normality tests discussed in Section 4.2. Results suggest that the GTPN model is an appropriate model for this data set while the rest of models are not.

Table 3. Estimation of the parameters and their standard errors (in parentheses) for the GTPN, TPN, WEI, and GL models for the data set. The AIC and BIC criteria are also included.

Estimated	GTPN	TPN	WEI	GL
σ	2.354(0.379)	0.720(0.047)	5.855(0.435)	-
λ	2.512(0.495)	4.842(0.333)	3.744(0.062)	4.093(0.539)
α	2.227(0.401)	-	-	13.467(1.902)
γ	-	-	-	15.528 (29.638)
AIC	238.43	254.63	251.74	308.67
BIC	246.67	260.12	257.23	316.90

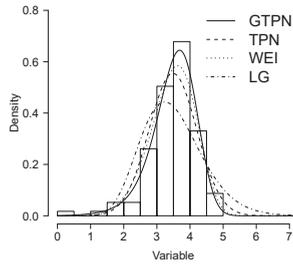


Figure 6. Fit of the distributions for the Laslett data set.

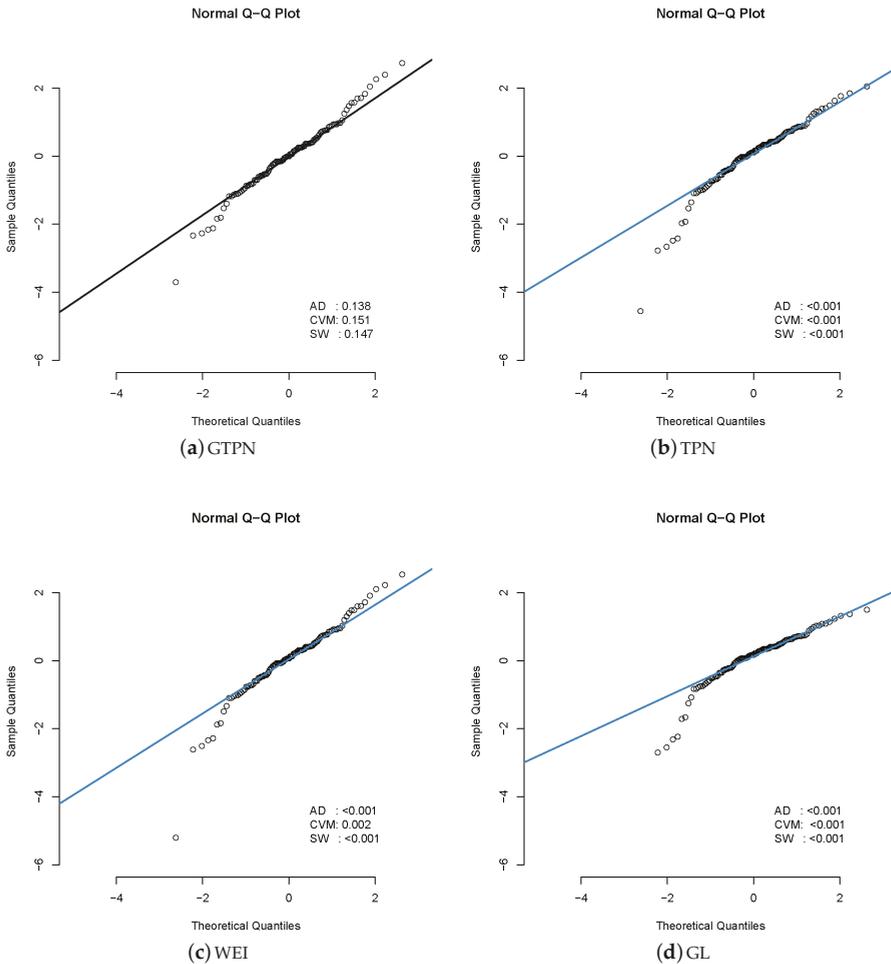


Figure 7. QR for the fitted models in the Laslett data set. The p - values for the Anderson–Darling (AD), Cramer–Von–Mises (CVM) and Shapiro–Wilks (SW) normality tests are also presented to check if the RQ came from the standard normal distribution.

6.2. Application 2

The data set to be investigated was taken from Nierenberg et al. [23], and is related to a study on plasma retinol and betacarotene levels from a sample of $n = 315$ subjects. More specifically, the response variable observed is grams of cholesterol consumed per day. Descriptive statistics for the variable are provided in Table 4.

Initially, we calculate the estimators based on the centiles, naive, and moments of the GTPN distribution, which are $\theta_{cent} = (2.326, 2.741, 2.376)$, $\theta_{naive} = (275.960, 0, 1)$, and $\theta_{mom} = (0.003, -86.024, 1.854)$. We used these estimations as initial values in computing the ML estimators for the GTPN model. Table 5 summarizes the fit for this data set. As in last application, we noted the high estimated standard error for the γ parameter in the GL model. Again, based on the AIC criteria and BIC criteria, the preferred model is the GTPN. Figure 8 shows the estimated density for each model in the cholesterol data set. The GTPN model appears to provide a better fit. For this data set, we also present the hypothesis tests for the three particular models of the GTPN distribution discussed in Section 4.1. Specifically, for $H_0^{(1)} : \alpha = 1$ versus $H_1^{(1)} : \alpha \neq 1$ we obtained p -values $< 0.0001, 0.0157,$ and < 0.0001 for the SLR, SR, and ST tests, respectively. Therefore, with a significance of 5% we preferred the GTPN over the TPN model. For $H_0^{(2)} : \lambda = 0$ versus $H_1^{(2)} : \lambda \neq 0$ we obtained p -values $< 0.0001, 0.0008$ and < 0.0001 for the SLR, SR and ST tests, respectively. Hence, with a significance of 5% we preferred the GTPN over the GHN model. Finally, $H_0^{(3)} : (\alpha, \lambda) = (1, 0)$ versus $H_1^{(3)} : (\alpha, \lambda) \neq (1, 0)$ we obtained p -values < 0.0001 for the three tests. Therefore, with a significance of 5% we preferred the GTPN over the HN model. Finally, Figure 9 also presents the qq-plot for the QR in the fitted models and the p -values for the three normality tests. Results suggest that the GTPN model is appropriate for this data set while the rest of the models are not.

Table 4. Descriptive statistics for the cholesterol data set.

Data Set	n	\bar{X}	S^2	$\sqrt{b_1}$	b_2
Heights measured	315	242.46	17421.79	1.47	6.34

Table 5. Estimated parameters and their standard errors (in parentheses) for the GTPN, TPN, WEI, and GL models for the cholesterol data set. The AIC and BIC criteria are also presented.

Estimated	GTPN	TPN	WEI	GL
σ	0.040(0.027)	151.106(8.734)	1.964(0.080)	-
λ	7.888(0.459)	1.455(0.138)	274.717(8.353)	0.016(0.001)
α	0.240(0.013)	-	-	2.831(0.293)
γ	-	-	-	2.067 (6.720)
AIC	3874.92	3943.25	3905.68	3879.07
BIC	3886.17	3950.75	3913.18	3890.42

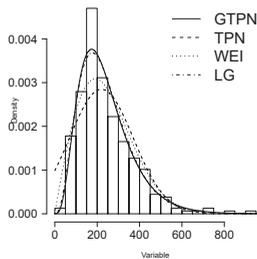


Figure 8. Fit of the distributions for the cholesterol data set.

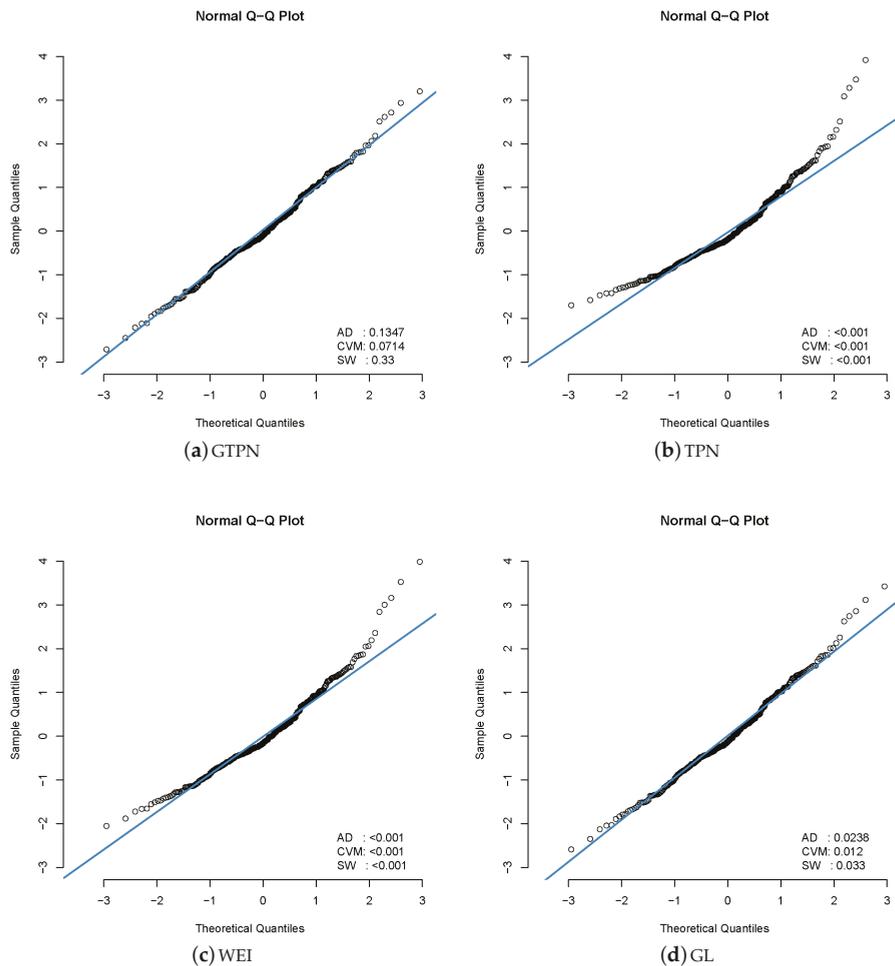


Figure 9. RQ for the fitted models in the cholesterol data set. The p – values for the AD, CVM, and SW normality tests are also presented to check if the QR came from the standard normal distribution.

7. Conclusions

In this work we introduce a new distribution for positive data named the GTPN model. This new distribution includes as particular cases three models well-known in the literature: the TPN, GHN, and HN models. The basic properties of the model and ML estimation were studied. We performed a simulation study in finite samples and two real data applications, showing the good performance of the model compared with other usual models in the literature.

Author Contributions: All of the authors contributed significantly to this research article.

Funding: The research of H.J.G. was supported by Proyecto de Investigación de Facultad de Ingeniería. Universidad Católica de Temuco. UCT-FDI012019. The research of O.V. was supported by Vicerrectoría de Investigación y Postgrado de la Universidad Católica de Temuco, Proyecto FEQUIP 2019-INRN-03.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pewsey, A. Large-sample inference for the general half-normal distribution. *Commun. Stat. Theory Methods* **2002**, *31*, 1045–1054. [[CrossRef](#)]
2. Pewsey, A. Improved likelihood based inference for the general half-normal distribution. *Commun. Stat. Theory Methods* **2004**, *33*, 197–204. [[CrossRef](#)]
3. Wiper, M. P.; Girón, F. J.; Pewsey, A. Objective Bayesian inference for the half-normal and half-t distributions. *Commun. Stat. Theory Methods* **2008**, *37*, 3165–3185. [[CrossRef](#)]
4. Khan, M.A.; Islam, H.N. Bayesian analysis of system availability with half-normal life time. *Math. Sci.* **2012**, *9*, 203–209. [[CrossRef](#)]
5. Moral, R.A.; Hinde, J.; Demétrio, C.G.B. Half-normal plots and overdispersed models in R: The hnp package. *J. Stat. Softw.* **2017**, *81*, 1–23. [[CrossRef](#)]
6. Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
7. Azzalini, A. Further results on a class of distributions which includes the normal ones. *Statistics* **1986**, *12*, 199–208.
8. Henze, N. A probabilistic representation of the skew-normal distribution. *Scand. J. Stat.* **1986**, *13*, 271–275.
9. Cooray, K.; Ananda, M.M.A. A generalization of the half-normal distribution with applications to lifetime data. *Commun. Stat. Theory Methods* **2008**, *10*, 195–224. [[CrossRef](#)]
10. Olmos, N.M.; Varela, H.; Gómez, H.W.; Bolfarine, H. An extension of the half-normal distribution. *Stat. Pap.* **2012**, *53*, 875–886. [[CrossRef](#)]
11. Cordeiro, G.M.; Pescim, R.R.; Ortega, E.M.M. The Kumaraswamy generalized half-normal distribution for skewed positive data. *J. Data Sci.* **2012**, *10*, 195–224.
12. Gómez, Y.M.; Bolfarine, H. Likelihood-based inference for power half-normal distribution. *J. Stat. Theory Appl.* **2015**, *14*, 383–398. [[CrossRef](#)]
13. Gómez, H.J.; Olmos, N.M.; Varela, H.; Bolfarine, H. Inference for a truncated positive normal distribution. *Appl. Math. J. Chin. Univ.* **2015**, *33*, 163–176. [[CrossRef](#)]
14. Bonferroni, C.E. *Elementi di Statistica Generale*; Libreria Seber: Firenze, Italy, 1930.
15. Arcagnia, A.; Porro, F. The Graphical Representation of Inequality. *Rev. Colomb. Estad.* **2014**, *37*, 419–436. [[CrossRef](#)]
16. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
17. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Auto. Contr.* **1974**, *19*, 716–723. [[CrossRef](#)]
18. Dunn, P.K.; Smyth, G.K. Randomized Quantile Residuals. *J. Comput. Graph. Stat.* **1996**, *5*, 236–244.
19. Yazici, B.; Yocalan, S. A comparison of various tests of normality. *J. Stat. Comput. Simul.* **2007**, *77*, 175–183. [[CrossRef](#)]
20. Zakerzadeh, H.; Dolati, A. Generalized Lindley Distribution. *J. Math. Ext.* **2009**, *3*, 1–17.
21. Laslett, G.M. Kriging and Splines: An Empirical Comparison of Their Predictive Performance in Some Applications. *J. Am. Stat. Assoc.* **1994**, *89*, 391–400. [[CrossRef](#)]
22. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [[CrossRef](#)]
23. Nierenberg, D.W.; Stukel, T.A.; Baron, J.A.; Dain, B.J.; Greenberg, E.R. Determinants of plasma levels of beta-carotene and retinol. *Am. J. Epidemiol.* **1989**, *130*, 511–521. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Article

Skewness of Maximum Likelihood Estimators in the Weibull Censored Data

Tiago M. Magalhães ¹, Diego I. Gallardo ² and Héctor W. Gómez ^{3,*}

¹ Department of Statistics, Institute of Exact Sciences, Federal University of Juiz de Fora, Juiz de Fora 36000-000, Brazil; tiago.magalhaes@ice.ufjf.br

² Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Copiapó 1530000, Chile; diego.gallardo@uda.cl

³ Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta 1240000, Chile

* Correspondence: hector.gomez@uantof.cl

Received: 29 September 2019; Accepted: 21 October 2019; Published: 1 November 2019

Abstract: In this paper, we obtain a matrix formula of order $n^{-1/2}$, where n is the sample size, for the skewness coefficient of the distribution of the maximum likelihood estimators in the Weibull censored data. The present result is a nice approach to verify if the assumption of the normality of the regression parameter distribution is satisfied. Also, the expression derived is simple, as one only has to define a few matrices. We conduct an extensive Monte Carlo study to illustrate the behavior of the skewness coefficient and we apply it in two real datasets.

Keywords: maximum likelihood estimates; type I and II censoring; skewness coefficient; Weibull censored data

1. Introduction

In its first appearance, the Weibull distribution [1] claimed its wide applicability. Survival analysis, reliability engineering, and extreme value theory are some of its applicability. To amplify the relevance of the Weibull, a regression structure is added to one of the parameters, i.e., the behavior of the distribution may be explained from covariates (explanatory variables) and unknown parameters to be estimated from observable data.

In statistical inference, it is often desirable to test if there are regression parameters statistically significant and the Wald test is commonly performed. Under standard regularity conditions, the null distribution of the Wald statistic is asymptotically chi-squared, a consequence of the maximum likelihood estimators (MLE) distribution. Therefore, the Wald test must be avoided if the sample size is not large enough, because the distribution of the MLE will be poorly approximated by the normal distribution.

Preventing the complexity of the statistical tests, the skewness coefficient (say γ) of the distribution of the MLE is an easy way to verify if the approximation to normality is adequate. A value of γ far from zero indicates a departure from the normal distribution. Pearson's standardized third cumulant defined by $\gamma = \kappa_3/\kappa_2^{3/2}$, where κ_r is the r th cumulant of the distribution, is the most well-known measure of skewness. When $\gamma > 0$ ($\gamma < 0$) the distribution is positively (negatively) skewed and will have a longer (shorter) right tail and a shorter (longer) left tail. If the distribution is symmetrical, γ equals zero. However, there are in [2] (Exercise 3.26) asymmetrical distributions with as many zero-odd order central moments as desired, so, the value of γ must be interpreted with some caution.

In the statistical literature, there is not a closed-form for the skewness coefficient of γ of the MLE in several regression models. Ref. [3] obtained a general $n^{-1/2}$ γ expression (say γ_1) for the distribution of the MLE, where n is the sample size. Following [3], several works have been developed in order

to obtain the γ_1 coefficient. In the first, Ref. [4] determined its expression for the class of generalized linear models and, the last one, Ref. [5] defined the γ_1 for the varying dispersion beta regression model and showed that this coefficient for the distribution of the MLE of the precision parameter is relatively large in small to moderate sample sizes. This paper is the first focused on a censored model.

In this work, we derive the γ_1 coefficient of the distribution of the MLE of the linear parameters in the Weibull censored data, assuming σ known, as $\sigma = 1/2$ and 1, the Rayleigh and exponential models, respectively. We discuss the situation when σ is unknown, however, it can be replaced by a consistent estimator, and then we can turn back to the original situation. This type of procedure was performed, for instance, by [4].

The remainder of the paper is organized as follows. Section 2 defines the Weibull censored data. In Section 3, we obtain a simple matrix expression, of order $n^{-1/2}$, for the skewness coefficients of the distributions of the MLEs of the linear regression parameters. In Section 4, some Monte Carlo simulations are performed. Two applications are presented in Section 5. Concluding remarks are offered in Section 6.

2. The Weibull Censored Data

We say that a continuous random variable T has Weibull distribution with scale parameter θ and shape parameter σ , or $T \sim \text{WE}(\theta, \sigma)$, if its probability density function (pdf) is given by

$$f(t; \theta, \sigma) = \frac{1}{\sigma \theta^{1/\sigma}} t^{1/\sigma - 1} \exp \left\{ - (t/\theta)^{1/\sigma} \right\}, \quad (1)$$

with $t > 0$, $\sigma > 0$ and $\theta > 0$. From (1), we can observe two particular distributions: the exponential and the Rayleigh, where $\sigma = 1$ and $\sigma = 1/2$, respectively. In lifetime data, there is the censoring restriction, i.e., if T_1, \dots, T_n are a random sample from (1), instead of T_i , we observe, under right censoring, $t_i = \min(T_i, L_i)$, where L_i is the censoring time, independent of T_i , $i = 1, \dots, n$. In this work, we consider an hybrid censoring scheme, where the study is finalized when a pre-fixed number, $r \leq n$, out of n observations have failed, as well as when a prefixed time, say $L_1 = \dots = L_n = L$, has been reached. The type I censoring is a particular case for $r = n$ and the type II censoring appears when $L_1, \dots, L_n = +\infty$. Additionally, we add the non-informative censoring assumption, i.e., the random variables L_i does not depend on θ . Under this setup, the log-likelihood function has the form

$$L(\theta, \sigma) = \left(\sigma \theta^{1/\sigma} \right)^{-r} \exp \left\{ \left(\frac{1}{\sigma} - 1 \right) A_1 - \frac{1}{\theta^{1/\sigma}} A_2 \right\},$$

where $r = \sum_{i=1}^n \delta_i$, $A_1 = \sum_{i=1}^n \delta_i \log t_i$, $A_2 = \sum_{i=1}^n t_i^{1/\sigma}$, $\delta_i = 1$, if $T_i \leq L_i$ and $\delta_i = 0$, otherwise. Usually, the regression modeling considers the distribution of $Y_i = \log(T_i)$ instead of T_i . The distribution of Y_i is of the extreme value form with pdf given by

$$f(y_i; x_i) = \frac{1}{\sigma} \exp \left\{ \frac{y_i - \mu_i}{\sigma} - \exp \left(\frac{y_i - \mu_i}{\sigma} \right) \right\}, \quad -\infty < y_i < \infty, \quad (2)$$

where $\mu_i = \log \theta_i$. The regression structure can be incorporated in (2) by making $\theta_i = \exp(x_i^\top \beta)$, where β is a p -vector of unknown parameters and x_i is a vector of regressors related to the i th observation. From this moment, we assume that σ is known, then, the log-likelihood function derived from (2) is given by

$$\ell(\beta) = \sum_{i=1}^n \left[\delta_i \left(-n \log \sigma + \frac{y_i - \mu_i}{\sigma} \right) - \exp \left(\frac{y_i - \mu_i}{\sigma} \right) \right].$$

The total score function and the total Fisher information matrix for β are, respectively, $\mathbf{U}_\beta = \sigma^{-1} \mathbf{X}^\top \mathbf{W}^{1/2} \mathbf{v}$ and $\mathbf{K}_{\beta\beta} = \sigma^{-2} \mathbf{X}^\top \mathbf{W} \mathbf{X}$, where $\mathbf{X} = (x_1, \dots, x_n)^\top$, the model matrix, assuming

rank $(\mathbf{X}) = p$, $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$, $w_i = \mathbb{E} \left[\exp \left(\frac{y_i - \mu_i}{\sigma} \right) \right]$ and $\mathbf{v} = (v_1, \dots, v_n)^\top$, $v_i = \left\{ -\delta_i + \exp \left(\frac{y_i - \mu_i}{\sigma} \right) \right\} w_i^{-1/2}$. It can be observed that the value of w_i depends on the mechanism of censoring. That means $w_i = q \times \left(1 - \exp \left\{ -L_i^{1/\sigma} \exp(-\mu_i/\sigma) \right\} \right) + (1 - q) \times (r/n)$, where $W_{(r)}$ denotes the r th order statistic from W_1, \dots, W_n and $q = \mathbb{P} \left(W_{(r)} \leq \log L_i \right)$. Note that $q = 1$ and $q = 0$ for types I and II censoring, respectively. The proof is presented in the Appendix A. The MLE of β , $\hat{\beta}$, is the solution of $\mathbf{U}_\beta = \mathbf{0}$. The $\hat{\beta}$ can not be expressed in closed-form. It is typically obtained by numerically maximizing the log-likelihood function using a Newton or quasi-Newton nonlinear optimization algorithm. Under mild regularity conditions and in large samples,

$$\hat{\beta} \sim N_p \left(\beta, \mathbf{K}_{\beta\beta}^{-1} \right),$$

approximately.

3. Skewness Coefficient

As discussed, the skewness coefficient is simple way to verify whether the approximation to normality is adequate. The model presented in (2) does not has a closed-form for this coefficient. The alternative is to apply the [3] result. These authors derived an approximation of order $\mathcal{O}(n^{-2})$ for the third cumulant of the MLE of the a -th regressor, i.e.,

$$\kappa_3(\hat{\beta}_a) = \mathbb{E} \left\{ [\hat{\beta}_a - \mathbb{E}(\hat{\beta}_a)]^3 \right\},$$

$a = 1, \dots, p$, which can be expressed as

$$\kappa_3(\hat{\beta}_a) = \sum' \kappa^{a,b} \kappa^{a,c} \kappa^{a,d} m_{bc}^{(d)}, \tag{3}$$

where $m_{bc}^{(d)} = 5\kappa_{bc}^{(d)} - (\kappa_{cd}^{(b)} + \kappa_{bd}^{(c)} + \kappa_{bcd})$, $a = 1, \dots, p$. Here, \sum' represents the summation over all combinations of parameters and over all the observations. From (3), after some algebra, we can express the third cumulant of the distribution of $\hat{\beta}$ for the Weibull censored data as

$$\kappa_3(\hat{\beta}) = -\sigma^{-3} \mathbf{P}^{(3)} (\mathbf{W} + 3\sigma \mathbf{W}') \mathbf{1}, \tag{4}$$

where $\mathbf{W}' = \text{diag}(w'_1, \dots, w'_n)$, $w'_i = -\sigma^{-1} L_i^{1/\sigma} \exp\{-L_i^{1/\sigma} \exp(-\mu_i/\sigma) - \mu_i/\sigma\}$, $\mathbf{P} = \mathbf{K}_{\beta\beta}^{-1} \mathbf{X}^\top = \sigma^2 (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top$, $\mathbf{P}^{(3)} = \mathbf{P} \odot \mathbf{P} \odot \mathbf{P}$, \odot represents a direct product of matrices and $\mathbf{1}$ is a n -dimensional vector of ones. Finally, by (4) and the Fisher information matrix, the asymmetry coefficient of the distribution of $\hat{\beta}$ to order $n^{-1/2}$ is given by

$$\gamma_1(\hat{\beta}) = -\sigma^{-3} \mathbf{P}^{(3)} (\mathbf{W} + 3\sigma \mathbf{W}') \mathbf{1} \odot \left\{ \text{diag} \left(\mathbf{K}_{\beta\beta}^{-1} \right) \mathbf{1} \right\}^{-3/2}, \tag{5}$$

for type II censoring, $\mathbf{W}' = \mathbf{0}$, then (5) reduces to $\gamma_1(\hat{\beta}) = -\sigma^{-3} \mathbf{P}^{(3)} \mathbf{W} \mathbf{1} \odot \left\{ \text{diag} \left(\mathbf{K}_{\beta\beta}^{-1} \right) \mathbf{1} \right\}^{-3/2}$. More details about the involved expressions are presented in Appendix A. The study of asymptotic properties of the Weibull censored data was the goal of many papers. Refs. [6,7] derived the Bartlett and the Bartlett-type correction factors for likelihood ratio and score tests, respectively, for the exponential censored data. Ref. [8] generalized these previous for the Weibull censored data and also derived the Bartlett-type correction factors for the gradient test. Ref. [9] presented the asymptotic expansions up to order $n^{-1/2}$ of the non null distribution functions of the likelihood ratio, Wald, Rao score and gradient statistics also for the censored exponential data. The result in expression (5) can be incorporated in this gallery.

4. Simulation Study

In this section, we compare the sample skewness coefficient (ρ) and the $n^{-1/2}$ skewness coefficients evaluated in the true and estimated parameters ($\hat{\gamma}_1^*$ and $\hat{\gamma}_1$, respectively) of the distributions of the MLEs in the Weibull censored model. To draw the data, we consider three values for σ : 0.5, 1 and 3; five sample sizes: 20, 30, 40, 60 and 100; three values for the percent of censoring C: 10%, 25% and 50%; and two number of regressors p : 3 and 5, where we consider two vectors for β in each case: $(-2, 0.5, 1)$ and $(1, -0.75, 0.5)$ for $p = 3$ and $(-2, 0.5, 1, -0.3, -0.5)$ and $(1, -0.75, 0.5, -1, 0.8)$ for $p = 5$. For each combination of σ , β , % of censoring and sample size we considered 20,000 Monte Carlo replicates. Each vector of covariates x_i considers an intercept term and the $p - 1$ remaining covariates were drawn independently from the standard normal distribution. Values from the Weibull model are drawn considering the inverse transformation method. Therefore, the greater $n \times C/100$ values were censored at the observed $(1 - C/100)$ -th quantile (a type II censoring scheme). For each sample, we considered the jackknife estimator for σ , say $\hat{\sigma}_J$. Therefore, the computation of $\hat{\gamma}_1^*$ and $\hat{\gamma}_1$ was performed considering (β, σ) and $(\hat{\beta}, \hat{\sigma}_J)$, the true and estimated parameters, respectively. Additionally, ρ is computed based on the 20,000 (marginal) skewness coefficient for the components of $\hat{\beta}$. Table 1 summarizes the case $\beta = (-2, 0.5, 1)$ (with $p = 3$ regressors) and $C = 10\%$. The main conclusions are the following:

Table 1. The $n^{-1/2}$ and sample skewness coefficients of the distributions of the MLEs in the Weibull censored data with $p = 3$ regressors and $\beta = (-2, 0.5, 1)$.

C	σ	n	$\hat{\beta}_0$			$\hat{\beta}_1$			$\hat{\beta}_2$		
			ρ	$\hat{\gamma}_1^*$	$\hat{\gamma}_1$	ρ	$\hat{\gamma}_1^*$	$\hat{\gamma}_1$	ρ	$\hat{\gamma}_1^*$	$\hat{\gamma}_1$
10%	0.5	20	-0.235	-0.080	-0.095	0.052	0.207	0.197	0.071	0.245	0.234
		30	-0.169	0.009	0.001	0.037	0.040	0.041	0.097	0.212	0.204
		40	-0.114	-0.013	-0.019	0.038	0.107	0.104	0.075	0.196	0.193
		60	-0.139	-0.082	-0.083	0.011	0.061	0.061	0.033	0.171	0.170
		100	-0.059	-0.055	-0.055	0.054	0.074	0.073	-0.005	0.087	0.087
	1.0	20	-0.198	0.017	0.004	-0.008	0.225	0.197	0.068	0.298	0.284
		30	-0.219	-0.101	-0.107	0.110	0.200	0.192	0.216	0.304	0.299
		40	-0.210	0.004	-0.002	0.083	0.140	0.137	0.109	0.185	0.182
		60	-0.147	-0.005	-0.010	0.085	0.143	0.137	0.057	0.173	0.171
		100	-0.092	-0.013	-0.015	0.019	0.048	0.047	0.116	0.132	0.131
	3.0	20	-0.232	0.006	0.005	0.094	0.143	0.131	0.022	0.093	0.087
		30	-0.178	-0.041	-0.040	0.004	0.054	0.048	-0.007	0.147	0.136
		40	-0.185	0.005	0.003	0.002	0.024	0.023	0.064	0.156	0.151
		60	-0.128	-0.020	-0.020	0.044	0.049	0.047	0.073	0.124	0.120
		100	-0.117	-0.028	-0.028	0.084	0.100	0.095	0.039	0.077	0.075

- The terms $\hat{\gamma}_1^*$ and $\hat{\gamma}_1$ are closer in all the considered combinations, suggesting that $\hat{\gamma}_1$ approaches $\hat{\gamma}_1^*$ in a reasonable way, even when the sample size is small.
- In general terms, $\hat{\gamma}_1$ approaches well ρ for $\hat{\beta}_1$ and $\hat{\beta}_2$. However, for $\hat{\beta}_0$ the terms seem discrepant even for $n = 100$.
- Considering the 90 cases for $p = 3$, ρ ranges from $(-0.245, 0.255)$, $(-0.429, 0.340)$ and $(-0.819, 1.181)$ for C 10%, 25% and 50%, respectively. For $p = 5$, ρ ranges from $(-0.373, 0.252)$, $(-0.402, 0.198)$ and $(-0.787, 0.495)$ for C 10%, 25% and 50%, respectively. This suggest that a higher percentage of censorship produce a higher skewness in the MLE estimators for the components of β .
- Considering the 90 cases for $p = 3$, ρ ranges from $(-0.819, 1.181)$, $(-0.363, 0.867)$, $(-0.351, 0.411)$, $(-0.305, 0.346)$ and $(-0.273, 0.255)$, for $n = 20, 30, 40, 60$ and 100, respectively. For $p = 5$, ρ ranges

from $(-1.015, 0.740)$, $(-0.529, 0.426)$, $(-0.372, 0.413)$, $(-0.318, 0.320)$ and $(-0.225, 0.243)$ for $n = 20, 30, 40, 60$ and 100 , respectively. This suggests that, as expected, when n increases the skewness coefficient of the MLE estimators for the components of β will be more symmetric.

Results suggest that, even with a moderate percentage of censored observations and small sample sizes, the distribution of the MLE for the components of β in the Weibull censored model are closer to the symmetry. The combinations of β , p and C not seem to affect the results. A simulation study showing this finding was omitted for the sake of brevity.

5. Applications

In this section we illustrate with two real dataset the application of the estimated skewness coefficient for the MLE estimators in the Weibull censored regression model. All the routine was performed in the statistical software R, [10]. Codes can be found in the personal website from the first author https://www.ufjf.br/tiago_magalhaes/downloads/.

5.1. Smokers Dataset

This dataset is related to a clinical trial on the effectiveness of triple-combination pharmacotherapy for tobacco dependence treatment conducted by the Cancer Institute of New Jersey and Robert Wood Johnson Foundation. The trial recruited 127 smokers 18 years or older with predefined medical illnesses from the local community. The outcome were the time (in days) to first relapse (return to smoking). The study lasted 182 days (26 weeks). Therefore, the times are subject to a censoring type I (32% of times were censored). We only considered the 113 patients where such observed time was positive (non-zero). Other measures were assigned randomly treatment group with levels combination or patch only (grp), age in years at time of randomization (age) and employment (full-time or non-full-time). We consider that $\text{time}_i \sim \text{WE}(\theta_i; \sigma)$, where $\log \theta_i = X_i^T \beta$, $\beta = (\beta_{\text{intercept}}, \beta_{\text{grp}}, \beta_{\text{age}}, \beta_{\text{employment}})^T$ and

$$X_i^T = (1, \text{grp}_i, \text{age}_i, \text{employment}_i)$$

We estimated $\hat{\sigma}_j = 1.617008$ based on the jackknife method, which was used as known in all the computations. Table 2 shows the parameters estimates, their standard errors and the estimated skewness coefficients and Figure 1 shows the estimated density function based on 1000 bootstrap samples for the coefficients related to the covariates grp, age and employment. Note that the estimated skewness for all parameters were closer to zero, suggesting a symmetric distribution for the estimators which is corroborated by the estimated density based on the bootstrap.

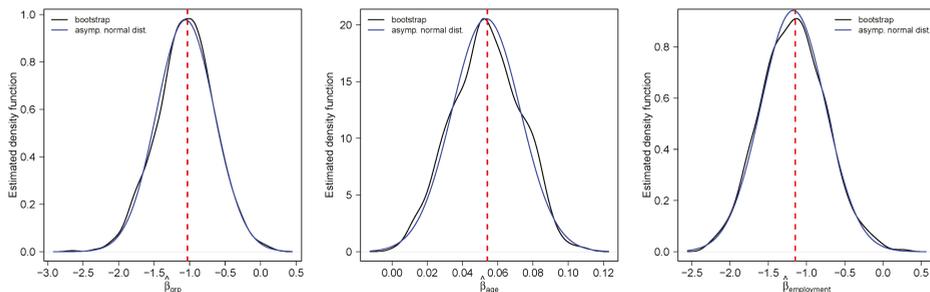


Figure 1. Estimated density function based on 1000 bootstrap samples and the asymptotic distribution for $\hat{\beta}_{\text{grp}}$ (left panel), $\hat{\beta}_{\text{age}}$ (center panel) and $\hat{\beta}_{\text{employment}}$ (right panel). The red line denotes the estimated parameter.

Table 2. Estimates for parameters and skewness coefficient in smokers dataset.

Parameter	Estimate	s.e.	γ_1
$\beta_{\text{intercept}}$	3.1690	0.8136	-0.0478
β_{grp}	-1.0303	0.3694	-0.0529
β_{age}	0.0541	0.0167	0.1251
$\beta_{\text{employment}}$	-1.1460	0.3935	-0.0753

5.2. Insulating Fluids Dataset

This dataset was presented in [11] on insulating fluids and it is related an accelerated test performed in order to determine the relationship between time (in minutes) to breakdown and voltage (in kilovolts). The authors assumed a regression structure based on the Weibull model and a common censoring time at $L = 200$ (type I censoring), i.e., $\text{time}_i \sim \text{WEI}(\theta_i, \sigma)$, where $\log \theta_i = \mathbf{X}_i^\top \beta$, $i = 1, \dots, 76$, $\mathbf{X}_i^\top = (\beta_{\text{Intercept}}, \beta_{\text{log-voltage}})$. We estimated $\hat{\sigma}_j = 1.296704$ based on the jackknife method, which was used as known. Table 3 shows the estimates, standard errors and estimated skewness coefficient for the MLE estimators and Figure 2 shows the estimated density function for $\hat{\beta}_{\text{Intercept}}$ and $\hat{\beta}_{\text{log-voltage}}$. Newly, the estimated skewness for both parameters are closer to zero, suggesting a symmetric distribution for the estimators as also suggest the estimated density based on bootstrap.

Table 3. Estimates for parameters and skewness coefficient in insulating fluids dataset.

Parameter	Estimate	s.e.	γ_1
$\beta_{\text{intercept}}$	20.4342	1.8772	0.1451
$\beta_{\text{log-voltage}}$	-0.5311	0.0557	-0.1517

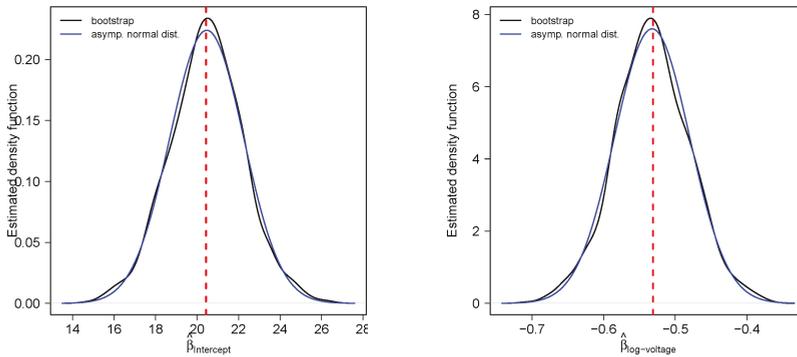


Figure 2. Estimated density function based on 1000 bootstrap samples and the asymptotic distribution for $\hat{\beta}_{\text{Intercept}}$ (left panel) and $\hat{\beta}_{\text{log-voltage}}$ (right panel). The red line denotes the estimated parameter.

6. Concluding Remarks

Since its beginning, the Weibull distribution and regression showed how it is important. In the frequentist context, this model depends strongly on the asymptotic properties of the MLE. Here, we presented an expression of the skewness that, in practical applications, can be used as an indicator of departure from the normal distribution of the MLE. Although the expression (3) entails a great deal of algebra, the final expression (5) of the skewness of the MLE distribution has a very nice form only

involving simple operations on diagonal matrices and can be easily implemented into a statistical software, for instance, R, [10].

Author Contributions: All the authors contributed significantly to the present paper.

Funding: The research of D.I. Gallardo was supported by Grant from FONDECYT (Chile) 11160670. H.W. Gómez was supported by Grant SEMILLERO UA-2019 (Chile).

Acknowledgments: We also acknowledge the referee’s suggestions that helped us to improve this work.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

In this Section we provided some additional details related to the computation of the manuscript.

Appendix A.1. W’s Quantities

In order to compute $w_i = E \left[\exp \left(\frac{y_i - \mu_i}{\sigma} \right) \right]$, $i = 1, \dots, n$, we first study the case type I censoring. Note that

$$\exp \left(\frac{y_i - \mu_i}{\sigma} \right) = \begin{cases} \exp \left(\frac{y_i - \mu_i}{\sigma} \right), & \text{if } y_i \leq \log L_i \\ \exp \left(\frac{\log L_i - \mu_i}{\sigma} \right), & \text{otherwise} \end{cases}$$

Therefore,

$$\begin{aligned} w_i &= \int_{-\infty}^{\log L_i} \frac{1}{\sigma} \exp \left(\frac{2(y_i - \mu_i)}{\sigma} - \exp \left(\frac{y_i - \mu_i}{\sigma} \right) \right) dy_i + \exp \left(\frac{\log L_i - \mu_i}{\sigma} \right) P(T_i > L_i) \\ &= 1 - \exp \left(-L_i^{1/\sigma} e^{-\mu_i/\sigma} \right) \left(1 + L_i^{1/\sigma} e^{-\mu_i/\sigma} \right) + L_i^{1/\sigma} e^{-\mu_i/\sigma} \exp \left(-L_i^{1/\sigma} e^{-\mu_i/\sigma} \right) \\ &= 1 - \exp \left(-L_i^{1/\sigma} e^{-\mu_i/\sigma} \right). \end{aligned}$$

Direct computation also shows that

$$\begin{aligned} v_i &= E \left[\exp \left(\frac{2(y_i - \mu_i)}{\sigma} \right) \right] \\ &= \int_{-\infty}^{\log L_i} \frac{1}{\sigma} \exp \left(\frac{3(y_i - \mu_i)}{\sigma} - \exp \left(\frac{y_i - \mu_i}{\sigma} \right) \right) dy_i + \exp \left(\frac{2(\log L_i - \mu_i)}{\sigma} \right) P(T_i > L_i) \\ &= 2 - \exp \left(-L_i^{1/\sigma} e^{-\mu_i/\sigma} \right) \left[2 + 2L_i^{1/\sigma} e^{-\mu_i/\sigma} + L_i^{2/\sigma} e^{-2\mu_i/\sigma} \right] + L_i^{2/\sigma} e^{-2\mu_i/\sigma} \exp \left(-L_i^{1/\sigma} e^{-\mu_i/\sigma} \right) \\ &= 2 \left\{ 1 - \exp \left(-L_i^{1/\sigma} e^{-\mu_i/\sigma} \right) \left[1 + L_i^{1/\sigma} e^{-\mu_i/\sigma} \right] \right\} \\ &= 2 \{w_i + \sigma w'_i\}. \end{aligned}$$

On the other hand, for the type II censoring note that $W_i = \exp \left(\frac{y_i - \mu_i}{\sigma} \right) \sim E(1)$. By [12], we have that

$$W_{(i)} \stackrel{D}{=} \sum_{j=1}^i \frac{Z_j}{n - j + 1}, \tag{A1}$$

where $W_{(i)}$ is the i th order statistic from W_1, \dots, W_n , $\stackrel{D}{=}$ denotes “equal in distribution” and Z_1, \dots, Z_n are independent and identically distributed $E(1)$ random variables. Therefore, $\mathbb{E}(W_{(j)}) = \text{Var}(W_{(j)}) = \sum_{k=1}^j (n - k + 1)^{-1}$ and

$$W_i = \begin{cases} W_{(1)}, & \text{with probability } 1/n \\ \vdots \\ W_{(r-1)}, & \text{with probability } 1/n \\ W_{(r)}, & \text{with probability } (n - r + 1)/n \end{cases}$$

Therefore,

$$w_i = \mathbb{E}(W_i) = \frac{1}{n} \sum_{j=1}^{r-1} \sum_{k=1}^j (n - k + 1)^{-1} + \frac{(n - r + 1)}{n} \sum_{k=1}^r (n - k + 1)^{-1}.$$

With some manipulations, we obtain that $\mathbb{E}(W_i) = r/n$. Also, we have that $\mathbb{E}(W_{(j)}^2) = \mathbb{E}(W_{(j)}) + \mathbb{E}^2(W_{(j)})$ and

$$V_i = \begin{cases} W_{(1)}^2, & \text{with probability } 1/n \\ \vdots \\ W_{(r-1)}^2, & \text{with probability } 1/n \\ W_{(r)}^2, & \text{with probability } (n - r + 1)/n \end{cases}$$

Therefore,

$$v_i = w_i + \frac{1}{n} \sum_{j=1}^{r-1} \left[\sum_{k=1}^j (n - k + 1)^{-1} \right]^2 + \frac{(n - r + 1)}{n} \left[\sum_{k=1}^r (n - k + 1)^{-1} \right]^2.$$

Algebraic manipulations shows that

$$v_i = \frac{1}{n} \left[r + \sum_{k=1}^r \frac{2(r - k) + 1}{n - k + 1} \right].$$

Finally, as the hybrid scheme can be seen as a mixture between type I and II censoring, we obtain directly that

$$w_i = q \times \left(1 - \exp \left(-L_i^{1/\sigma} e^{-\mu_i/\sigma} \right) \right) + (1 - q) \times (r/n),$$

$$v_i = q \times 2 \left\{ 1 - \exp \left(-L_i^{1/\sigma} e^{-\mu_i/\sigma} \right) \left[1 + L_i^{1/\sigma} e^{-\mu_i/\sigma} \right] \right\} + (1 - q) \times \frac{1}{n} \left[r + \sum_{k=1}^r \frac{2(r - k) + 1}{n - k + 1} \right],$$

where q is the mixing probability given by $q = \mathbb{P}(W_{(r)} \leq \log L)$. By (A1), $W_{(r)}$ has hypoexponential [13] distribution with vector of parameters $\lambda = (\lambda_1, \dots, \lambda_n)$, where $\lambda_j = (n - j + 1)^{-1}$. Therefore,

$$q = 1 - \sum_{j=1}^n \frac{L^{-\lambda_j}}{P_j},$$

where $P_j = \prod_{k=1, k \neq j}^n (k - j) / (n - j + 1)$.

Appendix A.2. Derivatives and Cumulants

Let Y_1, \dots, Y_n a random sample from Weibull censored data, the logarithm of the likelihood function is given by

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^n \left\{ \delta_i \left[-n \log \sigma + \frac{y_i - \mu_i}{\sigma} \right] - \exp \left(\frac{y_i - \mu_i}{\sigma} \right) \right\}. \quad (\text{A2})$$

The first four derivatives of (A2) can be expressed, respectively, for

$$\begin{aligned} \frac{\partial}{\partial \beta_r} \ell(\boldsymbol{\beta}) &= \frac{1}{\sigma} \sum_{i=1}^n \left\{ -\delta_i + \exp \left(\frac{y_i - \mu_i}{\sigma} \right) \right\} x_{ri}; \\ \frac{\partial^2}{\partial \beta_r \partial \beta_s} \ell(\boldsymbol{\beta}) &= -\frac{1}{\sigma^2} \sum_{i=1}^n \exp \left(\frac{y_i - \mu_i}{\sigma} \right) x_{ri} x_{si}; \\ \frac{\partial^3}{\partial \beta_r \partial \beta_s \partial \beta_t} \ell(\boldsymbol{\beta}) &= \frac{1}{\sigma^3} \sum_{i=1}^n \exp \left(\frac{y_i - \mu_i}{\sigma} \right) x_{ri} x_{si} x_{ti}. \end{aligned}$$

The second- to third-order cumulants are

$$\begin{aligned} \kappa_{rs} &= -\frac{1}{\sigma^2} \sum_{i=1}^n w_i x_{ri} x_{si}; \quad \kappa_{r,s} = -\kappa_{rs} = \frac{1}{\sigma^2} \sum_{i=1}^n w_i x_{ri} x_{si}; \\ \kappa_{rst} &= \frac{1}{\sigma^3} \sum_{i=1}^n w_i x_{ri} x_{si} x_{ti}; \quad \kappa_{r's}^{(t)} = -\frac{1}{\sigma^2} \sum_{i=1}^n w'_i x_{ri} x_{si} x_{ti}; \end{aligned}$$

where $w_i = E \left\{ \exp \left(\frac{y_i - \mu_i}{\sigma} \right) \right\}$,

$$\begin{aligned} w_i &= 1 - \exp \left\{ -L_i^{1/\sigma} \exp(-\mu_i/\sigma) \right\}, \\ w'_i &= -\frac{1}{\sigma} L_i^{1/\sigma} \exp \left\{ -L_i^{1/\sigma} \exp(-\mu_i/\sigma) - \mu_i/\sigma \right\}, \end{aligned}$$

It can be observed that $w'_i = 0$ for type II censoring.

References

1. Weibull, W. A statistical distribution function of wide applicability. *J. Appl. Mech.* **1951**, *18*, 293–297.
2. Kendall, M.G.; Stuart, A. *The Advanced Theory of Statistic*, 4th ed.; Griffin: London, UK, 1977.
3. Bowman, K.O.; Shenton, L.R. Asymptotic skewness and the distribution of maximum likelihood estimators. *Commun. Stat. Theory Methods* **1998**, *27*, 2743–2760.
4. Cordeiro, H.H.; Cordeiro, G.M. Skewness for parameters in generalized linear models. *Commun. Stat. Theory Methods* **2001**, *30*, 1317–1334.
5. Magalhães, T.M.; Botter, D.A.; Sandoval, M.C.; Pereira, G.H.A.; Cordeiro, G.M. Skewness of maximum likelihood estimators in the varying dispersion beta regression model. *Commun. Stat. Theory Methods* **2019**, *48*, 4250–4260.
6. Cordeiro, G.M.; Colosimo, E.A. Improved likelihood ratio tests for exponential censored data. *J. Stat. Comput. Sim.* **1997**, *56*, 303–315.
7. Cordeiro, G.M.; Colosimo, E.A. Corrected score tests for exponential censored data. *Stat. Probab. Lett.* **1999**, *44*, 365–373.
8. Magalhães, T.M.; Gallardo, D.I. Bartlett and Bartlett-type corrections for Weibull censored data. *Stat. Oper. Res. Trans.* **2019**, Submitted.

9. Lemonte, A.J. Non null asymptotic distribution of some criteria in censored exponential regression. *Commun. Stat. Theory Methods* **2014**, *43*, 3314–3328.
10. R Development Core Team. *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
11. Nelson, W.; Meeker, W.Q. Theory for optimum accelerated censored life tests for Weibull and extreme value distributions. *Technometrics* **1978**, *20*, 171–177.
12. Renyi, A. On the theory of order statistics. *Acta Math. Hung.* **1953**, *4*, 191–231.
13. Devianto, D.; Oktasari, L.; Maiyastri. Some properties of hypoexponential distribution with stabilizer constant. *Appl. Math. Sci.* **2015**, *142*, 7063–7070.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Article

Flexible Birnbaum–Saunders Distribution

Guillermo Martínez-Flórez ¹, Inmaculada Barranco-Chamorro ², Heleno Bolfarine ³ and Héctor W. Gómez ^{4,*}

¹ Departamento de Matemáticas y Estadística, Facultad de Ciencias Básicas, Universidad de Córdoba, Córdoba 230027, Colombia; guillermomartinez@correo.unicordoba.edu.co

² Departamento de Estadística e Investigación Operativa, Universidad de Sevilla, 41000 Sevilla, Spain; chamorro@us.es

³ Departamento de Estatística, IME, Universidade de São Paulo, São Paulo 01000, Brazil; hbolfar@ime.usp.br

⁴ Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta 1240000, Chile

* Correspondence: hector.gomez@uantof.cl

Received: 5 September 2019; Accepted: 4 October 2019; Published: 16 October 2019

Abstract: In this paper, we propose a bimodal extension of the Birnbaum–Saunders model by including an extra parameter. This new model is termed flexible Birnbaum–Saunders (FBS) and includes the ordinary Birnbaum–Saunders (BS) and the skew Birnbaum–Saunders (SBS) model as special cases. Its properties are studied. Parameter estimation is considered via an iterative maximum likelihood approach. Two real applications, of interest in environmental sciences, are included, which reveal that our proposal can perform better than other competing models.

Keywords: flexible skew-normal distribution; skew Birnbaum–Saunders distribution; bimodality; maximum likelihood estimation; Fisher information matrix

1. Introduction

The BS distribution was originally introduced in [1] to model the fatigue in the lifetime of certain materials. During the last decades, mainly due to its good properties, the use of this model spread out to other fields, such as economics and environmental sciences. In these applied scenarios, quite often, departures of the BS model are found, and therefore it is necessary to introduce some improvements. In this paper, we focus on those situations in which extra asymmetry or bimodality are present in our data, and a generalization of the BS model should be considered to deal with these issues. To reach this end, a flexible BS model is introduced. Our proposal is based on the flexible skew-normal distribution introduced in [2], and includes, as particular cases, the BS and skew BS distribution. Next, we briefly describe the key aspects that properly combined result in the flexible BS model. These are asymmetry, bimodality and main features of the basic BS model.

1.1. Asymmetry

Earlier results on asymmetric models started with the pioneering works by [1,3,4]. This topic regained interest with the study in [5], which from a Bayesian point of view developed a new asymmetric model which was later studied in depth by Azzalini [6], from a classical point of view. Azzalini model was termed the skew-normal distribution. Following Azzalini's method, a general family of asymmetric models termed skew-symmetric models appeared in the literature. The following lemma, originally presented in [6], can be considered as the starting point for the development of these asymmetric models.

Lemma 1. Let f_0 be a probability density function (pdf) which is symmetric around zero, and G a cumulative distribution function (cdf) such that G' exists and is a symmetric pdf around zero. Then

$$f_Z(z; \lambda) = 2f_0(z)G(\lambda z), \quad z \in \mathbb{R}, \tag{1}$$

is a pdf for $\lambda \in \mathbb{R}$.

Equation (1) provides the skew version of $f_0(\cdot)$ with skewing function $G(\cdot)$ and λ the skewness parameter. If $f_0(\cdot) = \phi(\cdot)$ and $G(\cdot) = \Phi(\cdot)$, the pdf and cdf, respectively, of the $N(0, 1)$ distribution, then the skew-normal is obtained, whose pdf is

$$f_Z(z) = 2\phi(z)\Phi(\lambda z), \quad z \in \mathbb{R}, \lambda \in \mathbb{R}. \tag{2}$$

Other examples of skew models are: skew-t, skew-Cauchy, skew-elliptical, and generalized skew-elliptical. We highlight that all of them are unimodal distributions.

1.2. Bimodality

Another fundamental result in our proposal will be the following lemma, which was given in Gómez et al. [2]. These authors extended (1), by introducing a parameter δ in f_0 , in such a way that for certain values of δ the resulting distribution is bimodal.

Lemma 2. Let f be a symmetric pdf around zero, F the corresponding cdf and G an absolutely continuous cdf such that G' exists and is symmetric around zero. Then

$$g(z; \delta, \lambda) = c_\delta f(|z| + \delta)G(\lambda z), \quad z \in \mathbb{R}, \lambda, \delta \in \mathbb{R} \tag{3}$$

is a pdf and $c_\delta^{-1} = 1 - F(\delta)$.

Taking $f(\cdot) = \phi(\cdot)$ and $G(\cdot) = \Phi(\cdot)$, in (3), the flexible skew-normal (FSN) model was obtained and studied in detail in [2]. There, it was proved that the FSN model can be bimodal for certain values of δ . Notice that the FSN model is obtained by adding an extra parameter, δ , to the skew-normal distribution proposed in [6]. That is a random variable (rv) Z follows a FSN distribution, $Z \sim FSN(\delta, \lambda)$, if its pdf is given by

$$f(z; \delta, \lambda) = c_\delta \phi(|z| + \delta)\Phi(\lambda z), \quad z \in \mathbb{R}, \lambda, \delta \in \mathbb{R} \tag{4}$$

where ϕ and Φ are the pdf and cdf of the $N(0, 1)$ distribution, respectively, and $c_\delta^{-1} = 1 - \Phi(\delta)$.

Other recent proposals in the contemporary literature dealing with bimodality are the extended two-pieces skew-normal model (ETN), introduced in [7] and the uni-bi-modal asymmetric power normal model given in [8] whose properties are based on results given in [9,10]. Applications of interest in Economics are given in [11]. All these references show the interest in the latest years for modelling bimodality.

1.3. BS Model

The BS or fatigue life distributions was proposed for modelling survival time data and material lifetime subject to stress in [12,13]. This model is asymmetric and only fits positive data. The pdf of a BS distribution is given by

$$f_T(t) = \phi(a_t) \frac{t^{-3/2}(t + \beta)}{2\alpha\sqrt{\beta}}, \quad t > 0, \tag{5}$$

where

$$a_t = a_t(\alpha, \beta) = \frac{1}{\alpha} \left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}} \right), \quad (6)$$

$\alpha > 0$ is a shape parameter, $\beta > 0$ is a scale parameter and the median of this distribution. (5) is denoted as $T \sim BS(\alpha, \beta)$. It is well known that α is the parameter that controls asymmetry. Specifically, (5) becomes more asymmetric as α increases and symmetric around β as α gets close to zero. It can be seen in [13] that (5) can be obtained as the distribution of the random variable

$$T = \beta \left[\frac{\alpha}{2} Z + \sqrt{\left(\frac{\alpha}{2} Z \right)^2 + 1} \right]^2, \quad (7)$$

where $Z \sim N(0, 1)$.

The BS model has been applied to a variety of practical situations. However, quite often, although the data suggest a BS distribution, some deficiencies are observed in the fitted BS model. This problem has motivated an increasing interest in its generalizations. We highlight that, recently, this model was extended by [14] to the family of elliptical distributions, this is known in the literature as the generalized Birnbaum–Saunders (GBS) distribution. Later, [15] proposed an extension based on the elliptical asymmetric distributions, known as the doubly generalized Birnbaum–Saunders model. On the other hand, [16] presents the asymmetric BS distribution with five parameters called the extended Birnbaum–Saunders (EBS) distribution. Other types of extensions are the asymmetric epsilon-Birnbaum–Saunders model given in [17], models in [18] based on the slash-elliptical family of distributions, and the generalized modified slash Birnbaum–Saunders (GMSBS) proposed in [19], which is based on [20].

In these extensions, we find that the asymmetric BS models previously cited, such as [15,21], are designed to fit data with greater or smaller asymmetry (or kurtosis) than that of the ordinary BS model, but they are not appropriate for fitting bimodal data. On the other hand, we highlight that the extension given in [21], which can become bimodal for certain combination of parameters is unable to capture bimodality unless it is accentuated enough.

Therefore there exists a real need for an asymmetric model, based on the BS distribution, and able to fit data presenting bimodal features, which is not uncommon in the literature. So the present paper presents a flexible BS distribution able to model skewness and to fit data with and without bimodality.

The paper is organized as follows. Section 2 is devoted to the development of an asymmetric uni-bimodal BS model. Its properties are studied in depth. Specifically, a closed expression for the cumulative distribution function (cdf) is given in terms of the cdf of a bivariate normal distribution. Some of the models proposed in [15,22] are obtained as particular cases. The shape and bimodality of the distribution are studied. It is shown that this model is closed under a change of scale and reciprocity. Survival and hazard functions are also obtained. Section 3 deals with moments derivation and iterative maximum likelihood estimation methods for the new model. Section 4 is devoted to real data applications of interest in environmental sciences. The first one deals with a bimodal situation in which our proposal performs better than other BS models and a mixture of normal distributions. The second one is taken from [16], where the extended BS model was proposed as the best for this dataset. It is shown that the FBS outperforms the extended BS model.

2. Results in Flexible Birnbaum-Saunders

Based on the flexible skew-normal model proposed in [2], we extend the Birnbaum–Saunders. The main idea is to apply (7) with $Z \sim FSN(\delta, \lambda)$ introduced in (4). This new model is called the flexible Birnbaum–Saunders (FBS) distribution whose pdf is given by

$$f(t; \alpha, \beta, \delta, \lambda) = \frac{t^{-3/2}(t + \beta)}{2\alpha\beta^{1/2}(1 - \Phi(\delta))} \phi(|a_t| + \delta) \Phi(\lambda a_t), \quad (8)$$

with a_t defined in (6), $t > 0$, $\alpha > 0$, $\beta > 0$, $\delta \in \mathbb{R}$, $\lambda \in \mathbb{R}$, $\phi(\cdot)$ and $\Phi(\cdot)$ the pdf and cdf of a $N(0, 1)$, respectively. We use the notation $T \sim FBS(\alpha, \beta, \delta, \lambda)$. The inclusion of parameters δ and λ makes our approach more flexible than the extensions previously discussed. λ is a parameter that controls asymmetry (skewness) and δ is a shape parameter related to bimodality of our proposal.

If $\lambda = 0$ then we obtain, as a particular case, the model introduced by [22].

Figures 1 and 2 depict the behaviour of (8) for some values of parameters, illustrating that it can be bimodal for some combinations of them.

2.1. Interpretation of Parameters.

In both figures the values of parameters α and β are fixed. We study the effects of

- (i) λ positive versus λ negative.
- (ii) Increasing $\delta > 0$ in Figure 1. Decreasing $\delta < 0$ in Figure 2.

Figure 1 suggests that, for α and β fixed, if a positive value of δ is considered then we have a unimodal distribution and the peak of the distribution increases when δ increases: $\delta = 0.75$ (red solid line), $\delta = 1.5$ (green dashed line), \dots , $\delta = 3$ (blue dashed dotted line). This happens for positive and negative values of λ .

On the other hand, in Figure 2, we have different situations. This plot suggests that, for α and β fixed, if a negative value of δ is considered then a bimodal distribution can be obtained. For positive λ , if δ decreases: $\delta = -0.75$ (red solid line), $\delta = -1.5$ (green dashed line), \dots , $\delta = -3$ (blue dashed dotted line), then the peaks decrease and bimodality becomes more accentuated. For negative λ , if δ decreases, then main peak increases and bimodality becomes less accentuated.

Also, note in Figures 1 and 2, that in the FBS model the pdf for negative λ is no longer the specular image of plot for positive λ .

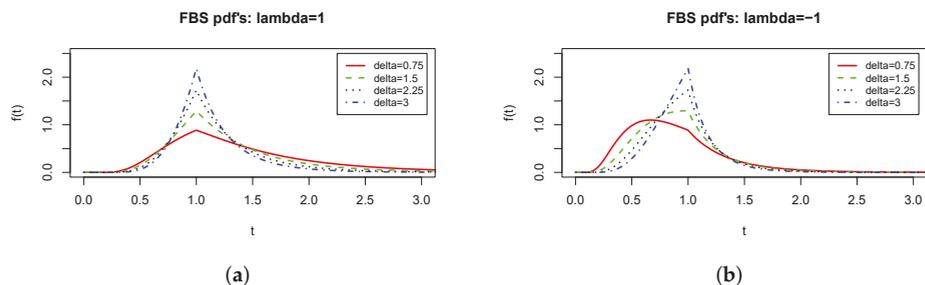


Figure 1. FBS distributions for $\alpha = 0.75$, $\beta = 1$ (both fixed). In (a) $\lambda = 1$ versus (b) $\lambda = -1$. Increasing values of $\delta > 0$: $\delta = 0.75$ (red solid line), 1.5 (green dashed line), 2.25 (black dotted line) and 3.0 (blue dashed and dotted line).

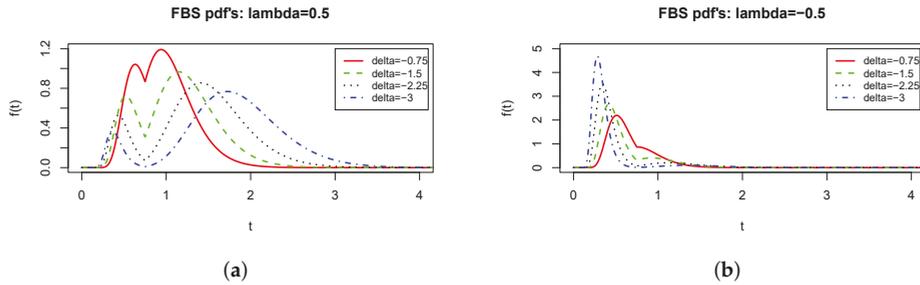


Figure 2. Flexible Birnbaum–Saunders (FBS) distributions for $\alpha = 0.30$, $\beta = 0.75$ (both fixed). In (a) $\lambda = 0.5$ versus (b) $\lambda = -0.5$. Decreasing values of $\delta < 0$: $\delta = -0.75$ (red solid line), -1.5 (green dashed line), -2.25 (black dotted line) and -3.0 (blue dashed and dotted line).

2.2. Properties

Next, important properties of the FBS model are presented. First an explicit expression for the cdf is given in terms of the cdf of a bivariate normal distribution.

Proposition 1. Let $T \sim FBS(\alpha, \beta, \delta, \lambda)$. Then the cdf of T is

$$F_T(t) = \begin{cases} c_\delta \Phi_{BN_\lambda} \left(\frac{\lambda\delta}{\sqrt{1+\lambda^2}}, a_t - \delta \right), & \text{if } 0 < t < \beta \\ c_\delta \left[\Phi_{BN_\lambda} \left(\frac{\lambda\delta}{\sqrt{1+\lambda^2}}, -\delta \right) + \Phi_{BN_\lambda} \left(-\frac{\lambda\delta}{\sqrt{1+\lambda^2}}, a_t + \delta \right) - \Phi_{BN_\lambda} \left(-\frac{\lambda\delta}{\sqrt{1+\lambda^2}}, \delta \right) \right], & \text{if } t \geq \beta, \end{cases} \tag{9}$$

where $\Phi_{BN_\lambda}(x, y)$ is the cdf of a bivariate normal distribution, with mean vector $\mu' = (0, 0)$ and covariance matrix

$$\Omega_\lambda = \begin{pmatrix} 1 & \rho_\lambda \\ \rho_\lambda & 1 \end{pmatrix} \quad \text{where } \rho_\lambda = -\frac{\lambda}{\sqrt{1+\lambda^2}}. \tag{10}$$

Proof. It can be seen in Appendix A. \square

Next some particular cases of interest for λ and δ parameters are discussed. Results about the shape of $f_T(\cdot)$ are included.

2.2.1. Effect of λ .

Corollary 1. Let $T \sim FBS(\alpha, \beta, \delta, \lambda)$. If $\lambda = 0$ then the cdf of T is

$$F_T(t) = \begin{cases} \frac{c_\delta}{2} \Phi(a_t - \delta), & \text{if } 0 < t < \beta \\ \frac{c_\delta}{2} \{ \Phi(a_t + \delta) + 1 - 2\Phi(\delta) \}, & \text{if } t \geq \beta. \end{cases} \tag{11}$$

Proof. If $\lambda = 0$ then ρ_λ , defined in (10), is equal to zero, and since in the bivariate normal distribution uncorrelation implies independence, we have that

$$\Phi_{BN_{\lambda=0}}(x, y) = \Phi(x) \Phi(y), \quad \forall (x, y).$$

Taking into account that $\Phi(0) = 1/2$ and $\Phi(-\delta) = 1 - \Phi(\delta)$, we have that (9) reduces to (11). \square

Result in Corollary 1 corresponds to the model studied in [22].

2.2.2. Effect of δ .

Corollary 2. Let $T \sim FBS(\alpha, \beta, \delta, \lambda)$. If $\delta = 0$ then F_T reduces to $F_T(t) = 2\Phi_{BN_\lambda}(0, a_t)$, for $t > 0$.

Corollary 2 is a particular case of models studied in [15].

2.2.3. Shape of $f_T(\cdot)$.

Proposition 2. Let $T \sim FBS(\alpha, \beta, \delta, \lambda)$. Then the pdf given in (8) is nondifferentiable at $t = \beta$.

Proof. It follows from (8), by noting that if $t = \beta$ then $a_t = 0$ and the absolute value function is not differentiable at zero. \square

Proposition 3. Let $T \sim FBS(\alpha, \beta, \delta, \lambda)$. The pdf given in (8) can be bimodal. The modes are the solution of the following non-linear equations.

1. $0 < t_1^* < \beta$ solution of

$$a_{t_1} = \delta + \lambda \frac{\phi(\lambda a_{t_1})}{\Phi(\lambda a_{t_1})} + \frac{a_{t_1}''}{\{a_{t_1}'\}^2}. \tag{12}$$

2. $t_2^* > \beta$ solution of

$$a_{t_2} = -\delta + \lambda \frac{\phi(\lambda a_{t_2})}{\Phi(\lambda a_{t_2})} + \frac{a_{t_2}''}{\{a_{t_2}'\}^2}. \tag{13}$$

With a_t given in (6), a_t' and a_t'' the first and second derivatives of a_t with respect to t , respectively.

Proof. It is given in Appendix A. \square

Comments on the use of (12) and (13) are included in Appendix A, Remark A1.

Remark 1. Equations obtained in (12) and (13) are similar to those we have in the skew normal and BS model.

1. Let $Z \sim SN(\lambda)$, $\lambda \in \mathbb{R}$. Then Z is unimodal and the mode, z^* , is given by the solution of the non-linear equation

$$z = \lambda \frac{\phi(\lambda z)}{\Phi(\lambda z)}.$$

2. Let $T \sim BS(\alpha, \beta)$, $\alpha, \beta > 0$. Then T is unimodal and the mode, t^* , is given by the solution of the non-linear equation

$$-a_t \{a_t'\}^2 + a_t'' = 0.$$

Next it is shown that the p -th quantile of T can be given in terms of the p th quantile of the $FSN(\delta, \lambda)$. Also it is proved that the FBS model is closed under change of scale and reciprocity.

Theorem 1. Let $T \sim FBS(\alpha, \beta, \delta, \lambda)$, with $\alpha, \beta \in \mathbb{R}^+$ and $\delta, \lambda \in \mathbb{R}$. Then

- (i) Let t_p be the p th quantile of T , $0 < p < 1$.

$$t_p = \beta \left(\frac{\alpha}{2} z_p + \sqrt{\left(\frac{\alpha}{2} z_p \right)^2 + 1} \right)^2 \tag{14}$$

where z_p denotes the p th quantile of $Z \sim FSN(\delta, \lambda)$.

- (ii) $kT \sim FBS(\alpha, k\beta, \delta, \lambda)$ for $k > 0$.
- (iii) $T^{-1} \sim FBS(\alpha, \beta^{-1}, \delta, -\lambda)$.

Proof. It can be seen in Appendix A. \square

2.2.4. Lifetime Analysis

The BS model is commonly used to explain survival and material resistance data. The survival and risk (or hazard) functions are important indicators in such fields. For the FBS model these functions are given next.

Proposition 4. Let $T \sim FBS(\alpha, \beta, \delta, \lambda)$ with $\alpha, \beta \in \mathbb{R}^+$ and $\delta, \lambda \in \mathbb{R}$. Then

- (i) The survival function is $S(t) = P[T > t] = 1 - F_T(t)$ with $F_T(\cdot)$ given in (9).
- (ii) The hazard function, $r(t) = f(t)/S(t)$, is

$$r(t) = \begin{cases} \frac{c_\delta a'_t \phi(|a_t| + \delta) \Phi(\lambda a_t)}{1 - c_\delta \Phi_{BN_\lambda}\left(\frac{\lambda \delta}{\sqrt{1 + \lambda^2}}, a_t - \delta\right)}, & \text{if } 0 < t < \beta \\ \frac{c_\delta a'_t \phi(|a_t| + \delta) \Phi(\lambda a_t)}{1 - c_\delta \left[\Phi_{BN_\lambda}\left(\frac{\lambda \delta}{\sqrt{1 + \lambda^2}}, -\delta\right) + \Phi_{BN_\lambda}\left(-\frac{\lambda \delta}{\sqrt{1 + \lambda^2}}, a_t + \delta\right) - \Phi_{BN_\lambda}\left(-\frac{\lambda \delta}{\sqrt{1 + \lambda^2}}, \delta\right) \right]}, & \text{if } t \geq \beta \end{cases}$$

with $\Phi_{BN_\lambda}(\cdot)$ the cdf of the bivariate normal given in Proposition 1.

In Figure 3, the hazard function for those pdf's considered in Figures 1 and 2 are plotted. These graphs show that, for the FBS distribution, the hazard function admits a variety of shapes, which is interesting from an applied point of view.

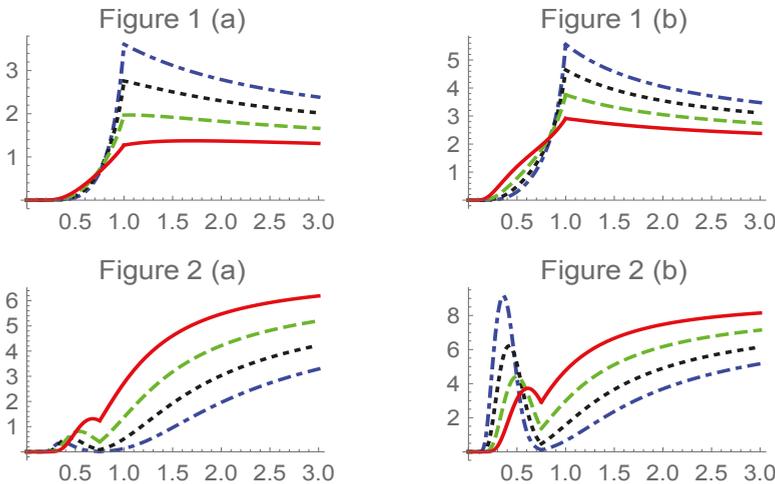


Figure 3. Hazard function of the FBS distribution for plots corresponding to Figure 1 (a), (b): $\alpha = 0.75$, $\beta = 1$ (both fixed), $\delta = 0.75$ (red solid line), 1.5 (green dashed line), 2.25 (black dotted line) and 3.0 (blue dashed dotted line), in (a) $\lambda = 1$ versus (b) $\lambda = -1$. For plots corresponding to Figure 2 (a), (b): $\alpha = 0.30$, $\beta = 0.75$ (both fixed), $\delta = -0.75$ (red solid line), -1.5 (green dashed line), -2.25 (black dotted line) and -3.0 (blue dashed dotted line), in (a) $\lambda = 0.5$ versus (b) $\lambda = -0.5$.

Remark 2. More complicated hazard functions than the traditional ones are obtained when we are dealing with models with complex structure, as it happens with the FBS. For instance, in Figure 3, we have two situations:

1. $r(t)$ corresponding to Figure 1a,b. These are, first, quickly increasing, later decreasing more slowly or even in a flat way. It can be applied in practical situations in which the risk of failure increases quickly until certain point in which its behaviour becomes flatter. As [23] points out, the flat area is very interesting in survival analysis and reliability contexts.

2. $r(t)$ corresponding to Figure 2a,b are increasing-decreasing-increasing. This kind of hazard functions has been recently introduced and discussed in literature, due to its interest in reliability of systems, see for instance [23] or [24] (and references therein). In plot for Figure 2b, $r(t)$ is (quickly) increasing—or (quickly) decreasing. On the other hand, for Figure 2a the initial effect increasing—decreasing is less accentuated.

3. Moments and Maximum Likelihood Estimation

Moments of the FBS model can be obtained from the moments of the flexible skew-normal model given in [2]. The following results present important properties relating those distributions, and the expressions for the first moment and variance in the FBS model.

Theorem 2. Let $T \sim FBS(\alpha, \beta, \delta, \lambda)$ and $Z \sim FSN(\delta, \lambda)$. Then $\mathbb{E}(T^r)$, $r = 0, 1, \dots$, always exists. Moreover

$$\mathbb{E}(T^r) = \frac{\beta^r}{4^r} \sum_{k=0}^{2r} \binom{2r}{k} \mathbb{E} \left[(\alpha Z)^k (\alpha^2 Z^2 + 4)^{\frac{2r-k}{2}} \right], \quad r = 0, 1, \dots \tag{15}$$

Proof. From (7), we can write

$$T = \frac{\beta}{4} \left\{ \alpha Z + (\alpha^2 Z^2 + 4)^{1/2} \right\}^2.$$

Taking expectation of the r th-power of T

$$\mathbb{E}(T^r) = \frac{\beta^r}{4^r} \mathbb{E} \left[\left\{ \alpha Z + (\alpha^2 Z^2 + 4)^{1/2} \right\}^{2r} \right]. \tag{16}$$

From (16), note that for $r = 0, 1, \dots$, $\mathbb{E}(T^r)$ exists if and only if $\mathbb{E}(Z^{2r})$ exists. On the other hand, it can be seen in [2] that $\mathbb{E}(Z^{2r})$ always exists, and therefore $\mathbb{E}(T^r)$ too.

Finally note that (15) is the result of applying the binomial formula to (16). \square

Next, explicit expressions for the expected value and variance of $T \sim FBS(\alpha, \beta, \delta, \lambda)$ are given. In these expressions, $\kappa_j = \mathbb{E}_{FSN} \left(\frac{Z^j}{2} \sqrt{\alpha^2 Z^2 + 4} \right)$ with $Z \sim FSN(\delta, \lambda)$.

Theorem 3. Let $T \sim FBS(\alpha, \beta, \delta, \lambda)$. Then

$$\mathbb{E}(T) = \beta \left[1 + \alpha \kappa_1 + \frac{\alpha^2}{2} c_\delta \left\{ (1 + \delta^2)(1 - \Phi(\delta)) - \delta \phi(\delta) \right\} \right], \tag{17}$$

$$\begin{aligned} \mathbb{E}(T^2) = & \beta^2 \left[\frac{7\alpha^4 c_\delta}{16} \left((3 + 6\delta^2 + \delta^4)(1 - \Phi(\delta)) - \delta(5 + \delta^2)\phi(\delta) \right) + \alpha^3 \kappa_3 + 2\alpha \kappa_1 + 1 \right] \\ & + 2\alpha^2 \beta^2 c_\delta \left((1 + \delta^2)(1 - \Phi(\delta)) - \delta \phi(\delta) \right) \end{aligned}$$

and

$$\begin{aligned} Var(T) = & \beta^2 \left[\frac{7\alpha^4 c_\delta}{16} \left((3 + 6\delta^2 + \delta^4)(1 - \Phi(\delta)) - \delta(5 + \delta^2)\phi(\delta) \right) + \alpha^3 \kappa_3 - \alpha^2 \kappa_1 + 1 \right] \\ & - \frac{\alpha^2 \beta^2 c_\delta}{4} \left((1 + \delta^2)(1 - \Phi(\delta)) - \delta \phi(\delta) \right) \left(\alpha^2 c_\delta \left((1 + \delta^2)(1 - \Phi(\delta)) - \delta \phi(\delta) \right) + 4(1 - \alpha \kappa_1) \right) \end{aligned}$$

where $\kappa_j = \mathbb{E}_{FSN} \left(\frac{Z^j}{2} \sqrt{\alpha^2 Z^2 + 4} \right)$ and $Z \sim FSN(\delta, \lambda)$.

Proof. These results follows from Theorem 2 and the expressions of κ_j , which have been computed by using the results for moments of $Z \sim FSN(\delta, \lambda)$ obtained in [2].

As illustration, note that for the case $r = 1$, (15) reduces to

$$\begin{aligned} \mathbb{E}(T) &= \frac{\beta}{4} \left[\mathbb{E}(\alpha^2 Z^2 + 4) + 2\mathbb{E} \left((\alpha Z) \sqrt{\alpha^2 Z^2 + 4} \right) + \mathbb{E}(\alpha^2 Z^2) \right] \\ &= \beta \left[1 + \alpha \kappa_1 + \frac{\alpha^2}{2} \mathbb{E}(Z^2) \right], \end{aligned}$$

it can be seen in [2] that $\mathbb{E}(Z^2) = c_\delta(1 + \delta^2) (1 - \Phi(\delta)) - \delta\phi(\delta)$, and so (17) is obtained. \square

3.1. Maximum Likelihood Estimators

Parameter estimation in the BS model has been the topic of interest in many papers. Among others, we mentioned [25–27]. To estimate the parameters in the usual BS model, the modified moment method (MME) and maximum likelihood (MLE) are commonly used. To start the maximum likelihood approach moment estimators are used which are given by

$$\hat{\beta}_M = \sqrt{sr}, \quad \hat{\alpha}_M = \sqrt{2 \left(\sqrt{\frac{s}{r}} - 1 \right)}.$$

where $s = \frac{1}{n} \sum_{i=1}^n t_i$ and $r = \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{t_i} \right)^{-1}$ are the sample (arithmetic) and harmonic mean, respectively. Relevant aspects of this distribution such as its robustness with respect to parameter estimation and $O(n^{-1})$ bias corrections for MLEs, can be seen in [25–27].

In the following, we discuss MLE estimation for the FBS model in depth. Thus, given n observations independent and identically distributed, T_1, T_2, \dots, T_n , with $T_i \sim FBS(\alpha, \beta, \delta, \lambda)$, the log-likelihood function for the parameter vector $\theta = (\alpha, \beta, \delta, \lambda)'$ is given by

$$\begin{aligned} \ell(\theta) &= -n \left(\log(\alpha) + \frac{1}{2} \log(\beta) + \log(1 - \Phi(\delta)) \right) - \frac{3}{2} \sum_{i=1}^n \log(t_i) + \sum_{i=1}^n \log(t_i + \beta) \\ &\quad - \frac{1}{2} \sum_{i=1}^n (a_{t_i}^2 + 2\delta|a_{t_i}| + \delta^2) + \sum_{i=1}^n \log(\Phi(\lambda a_{t_i})). \end{aligned} \tag{18}$$

To maximize $l(\theta)$ in θ , consider the first derivatives of $l(\theta)$ with respect to α, β, δ and λ , denoted as $\dot{l}_\alpha, \dot{l}_\beta, \dot{l}_\delta$ and \dot{l}_λ , respectively. From $\dot{l}_\alpha = 0, \dot{l}_\beta = 0, \dot{l}_\delta = 0$ and $\dot{l}_\lambda = 0$, the likelihood equations are given by

$$-n + \sum_{i=1}^n a_{t_i}^2 - \delta \sum_{i=1}^n |a_{t_i}| - \lambda \sum_{i=1}^n a_{t_i} \frac{\phi(\lambda a_{t_i})}{\Phi(\lambda a_{t_i})} = 0 \tag{19}$$

$$-\frac{n}{2\beta} + \sum_{i=1}^n \frac{1}{t_i + \beta} + \frac{1}{2\alpha\beta^{3/2}} \sum_{i=1}^n \left(\text{sgn}(a_{t_i}) (|a_{t_i}| + \delta) - \lambda \frac{\phi(\lambda a_{t_i})}{\Phi(\lambda a_{t_i})} \right) \frac{t_i + \beta}{\sqrt{t_i}} = 0 \tag{20}$$

$$\delta - \frac{\phi(\delta)}{1 - \Phi(\delta)} = -\frac{1}{n} \sum_{i=1}^n |a_{t_i}| \tag{21}$$

$$\sum_{i=1}^n a_{t_i} \frac{\phi(\lambda a_{t_i})}{\Phi(\lambda a_{t_i})} = 0 \tag{22}$$

in which $\text{sgn}(\cdot)$ denotes the *sign* function.

The solution to the previous system of equations must be obtained by iterative methods such as the Newton-Raphson or quasi-Newton procedures, which can be implemented using the statistical software R, [28].

As initial estimates of α and β can be proposed the estimates of these parameters obtained in the basic BS model, denoted as $\hat{\alpha}_0$ and $\hat{\beta}_0$. These estimates can be plugged into (21) and (22) to obtain preliminar estimates of δ and λ , $\hat{\delta}_0$ and $\hat{\lambda}_0$, and so, start the recursion.

3.2. Expected and Observed Information Matrices

Recall that, the Fisher information matrix is given by

$$I(\theta) = (j_{i,j})_{i,j=\alpha,\beta,\delta,\lambda},$$

which entries are equal to minus the second partial derivatives of the log-likelihood function given in (18) with respect to the parameters in the model. They are denoted as $j_{\alpha\alpha} = -\frac{\partial^2}{\partial\alpha^2}l(\theta)$, and so on. So we have

$$\begin{aligned} j_{\lambda\alpha} &= -\frac{n}{\alpha^2} + \frac{1}{\alpha^2} \sum_{i=1}^n (3a_{t_i}^2 + 2\delta|a_{t_i}|) + \frac{\lambda}{\alpha^2} \sum_{i=1}^n a_{t_i}w_i (2 + \lambda a_{t_i}B_i), \\ j_{\beta\alpha} &= -\frac{1}{\alpha^3\beta^2} \sum_{i=1}^n \left(\frac{\beta^2 - t_i^2}{t_i} \right) + \frac{1}{2\alpha^2\beta^{3/2}} \sum_{i=1}^n (\delta \operatorname{sgn}(a_{t_i}) + \lambda w_i(-1 + \lambda a_{t_i}B_i)) \left(\frac{t_i + \beta}{\sqrt{t_i}} \right), \\ j_{\beta\beta} &= -\frac{n}{2\beta^2} + \sum_{i=1}^n \frac{1}{t_i + \beta} + \frac{1}{\alpha^2\beta^3} \sum_{i=1}^n t_i + \frac{1}{4\alpha\beta^{5/2}} \sum_{i=1}^n (\delta \operatorname{sgn}(a_{t_i}) - \lambda w_i) \left(\frac{3t_i + \beta}{\sqrt{t_i}} \right) \\ &\quad + \frac{\lambda^2}{4\alpha^2\beta^{3/2}} \sum_{i=1}^n w_i B_i \left(\frac{t_i + \beta}{\sqrt{t_i}} \right)^2, \\ j_{\delta\alpha} &= -\frac{1}{\alpha} \sum_{i=1}^n |a_{t_i}|, \quad j_{\delta\beta} = -\frac{1}{2\alpha^2\beta^{3/2}} \sum_{i=1}^n \operatorname{sgn}(a_{t_i}) \left(\frac{t_i + \beta}{\sqrt{t_i}} \right), \\ j_{\lambda\alpha} &= \frac{1}{\alpha} \sum_{i=1}^n a_{t_i}w_i (1 - \lambda a_{t_i}B_i), \quad j_{\lambda\beta} = \frac{1}{2\alpha\beta^{3/2}} \sum_{i=1}^n w_i (1 + \lambda a_{t_i}B_i) \left(\frac{t_i + \beta}{\sqrt{t_i}} \right), \\ j_{\delta\delta} &= n(w_\delta(\delta - w_\delta) + 1), \quad j_{\delta\alpha} = j_{\delta\beta} = j_{\lambda\delta} = 0, \quad j_{\lambda\lambda} = \sum_{i=1}^n a_{t_i}^2 w_i B_i \end{aligned}$$

where $w = \frac{\phi(\lambda a_t)}{\Phi(\lambda a_t)}$, $w_\delta = \phi(\delta)/(1 - \Phi(\delta))$ and $B = \lambda a_t + w$.

The Fisher (expected) information matrix would be obtained by computing the expected values of the above second derivatives. Taking in this matrix $\delta = \lambda = 0$, that is, $T \sim BS(\alpha, \beta)$, and, using results in [21], we have

$$I(\theta) = \begin{pmatrix} \frac{2}{\alpha^2} & 0 & -\frac{1}{\alpha}\sqrt{\frac{2}{\pi}} & 0 \\ 0 & \alpha^{-2}\beta^{-2} \left(1 + \frac{\alpha q(\alpha)}{\sqrt{2\pi}} \right) & 0 & \frac{1}{\sqrt{2\pi}} \frac{1}{\alpha\beta^{3/2}} A_1(t) \\ -\frac{1}{\alpha}\sqrt{\frac{2}{\pi}} & 0 & 1 - \frac{2}{\pi} & 0 \\ 0 & \frac{1}{\sqrt{2\pi}} \frac{1}{\alpha\beta^{3/2}} A_1(t) & 0 & \frac{2}{\pi} \end{pmatrix},$$

where $A_1(t) = \mathbb{E} \left(\frac{t_i + \beta}{\sqrt{t_i}} \right)$, $q(\alpha) = \alpha\sqrt{\frac{2}{\pi}} - \frac{\pi \exp(\frac{2}{\alpha^2})}{2} \operatorname{erfc}(\frac{2}{\alpha})$, with $\operatorname{erfc}(\cdot)$ the error function, i.e., $\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-t^2) dt$, see [29].

It can be shown that $|I(\theta)| \neq 0$, so the Fisher information matrix is not singular at $\delta = \lambda = 0$.

Hence, for large samples, the MLE, $\hat{\theta}$, of θ is asymptotically normal, that is,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} N_4(\mathbf{0}, I(\theta)^{-1}),$$

resulting that the asymptotic variance of the MLE, $\hat{\theta}$, is the inverse of Fisher information matrix $I(\theta)$. Since the parameters are unknown, usually the observed information matrix is considered where the unknown parameters are estimated by ML.

Asymptotic confidence intervals for the parameters in the FBS model can be obtained from these results.

4. Numerical Illustrations

The numerical illustrations introduced next are aimed to show that the FBS model can be an alternative to modelling unimodal or bimodal data from different areas. First illustration is related to nickel content in soil samples analyzed at the Mining Department (Departamento de Minas) of Universidad de Atacama, Chile. We start by showing that both, BS and skew-BS (SBS) models are not able to capture bimodality present in this data set. Thus, the FBS model turned out to be a good option, to fit the data even better than a mixture of two normal distributions, which is another competing alternative to fit bimodal data. Second illustration is related to air pollution in New York city in USA, which was previously analyzed in [16,30]. In this case, it is shown that FBS model again provides a better fit than BS and SBS. As competing model the extended Birnbaum-Saunders (EBS) is also considered. Recall that the $EBS(\alpha, \beta, \sigma, \nu, \lambda)$ is a five-parameter model proposed in [16] where the parameter σ affects the kurtosis; ν and λ affect the skewness; and α and β the shape and scale as in the usual BS model. We highlight that, for this dataset, the $FBS(\alpha, \beta, \delta, \lambda)$ model provides a better fit than that given by the $EBS(\alpha, \beta, \sigma, \nu, \lambda)$ in [16] with the merit of using less parameters.

4.1. Nickel Concentration

For illustrative purposes, we apply the FBS model to a data set related to nickel content in soil samples. This data set encompasses 85 observations of the variable concentration of nickel with sample mean = 21.588, sample standard deviation = 16.573, sample asymmetry = 2.392 and sample kurtosis = 8.325, much higher than expected with the ordinary BS distribution.

4.1.1. FBS versus the BS and SBS distributions

To fit the nickel concentration variable, we use the BS, skew BS (SBS) and FBS models. Using function `optim` from the R-package, [28], the following point estimates (and their standard errors) are obtained for each of the three models under consideration

BS model: $\hat{\alpha} = 0.789$ (0.060) and $\hat{\beta} = 16.382$ (1.296).

SBS model: $\hat{\alpha} = 1.073$ (0.201), $\hat{\beta} = 8.841$ (1.998) and $\hat{\lambda} = 1.252$ (0.590).

FBS model: $\hat{\alpha} = 0.870$ (0.104), $\hat{\beta} = 5.072$ (0.763), $\hat{\delta} = -1.520$ (0.282) and $\hat{\lambda} = 1.405$ (0.341).

The bimodal hypothesis can be formally tested as follows

$$H_0 : \delta = 0 \quad \text{versus} \quad H_1 : \delta \neq 0,$$

which is equivalent to compare models SBS versus FBS. Given the nonsingularity of the Fisher information matrix, and since these models are nested, we can consider the likelihood ratio statistics, namely

$$\Lambda_1 = L_{SBS}(\hat{\alpha}, \hat{\beta}, \hat{\lambda}) / L_{FBS}(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\lambda}).$$

It is obtained $-2 \log(\Lambda_1) = 5.618$, which is greater than the 5% chi-square critical value with one degree of freedom (df), which is equal to 3.84. Therefore, the null hypothesis of no-bimodality is rejected at the 5% critical level, leading to the conclusion that FBS model fits better than the unimodal SBS model to the nickel concentration data.

To compare the FBS model with the BS model, consider to test the null hypothesis of a BS distribution versus a FBS distribution, that is

$$H_0 : (\delta, \lambda) = (0, 0) \quad \text{vs} \quad H_1 : (\delta, \lambda) \neq (0, 0)$$

using the likelihood ratio statistics based on the ratio $\Lambda_2 = L_{BS}(\hat{\alpha}, \hat{\beta}) / L_{FBS}(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\lambda})$. After substituting the estimated values, we obtain $-2 \log(\Lambda_2) = 7.628$, which is greater than the 5% chi-square critical value with 2 df, which is 5.99. Therefore the FBS is preferred to BS model for this data set.

4.1.2. FBS versus a Mixture of Normal Distributions

Another model widely applied in such situations of bimodality is the mixture of two normal distributions. The normal mixture model is given by:

$$f(x; \mu_1, \sigma_1, \mu_2, \sigma_2, p) = p f_1(x; \mu_1, \sigma_1) + (1 - p) f_2(x; \mu_2, \sigma_2) \quad (23)$$

where f_j is a normal distribution with parameters (μ_j, σ_j) , $j = 1, 2$ and $0 < p < 1$. (23) is denoted by $MN(\mu_1, \sigma_1, \mu_2, \sigma_2, p)$.

To compare FBS model with the MN model, we propose the Akaike information criterion (AIC), see [31], namely $AIC = -2\hat{\ell}(\cdot) + 2k$, the modified AIC criterion (CAIC), typically called the consistent AIC, namely $CAIC = -2\hat{\ell}(\cdot) + (1 + \log(n))k$ and the Bayesian Information Criterion, BIC, $BIC = -2\hat{\ell}(\cdot) + \log(n)k$, where k is the number of parameters and $\hat{\ell}(\cdot)$ is the log-likelihood function evaluated at the MLEs of parameters. The best model is the one with the smallest AIC or CAIC or BIC.

Now we compare the FBS with $MN(\mu_1, \sigma_1, \mu_2, \sigma_2, p)$. The estimated mixture model is

$$MN(15.348, 6.622, 40.908, 21.960, 0.755)$$

with $AIC = 674.849$, $CAIC = 692.061$ and $BIC = 687.062$. On the other hand, for the FBS model, we have $AIC = 671.859$, $CAIC = 685.628$ and $BIC = 681.630$. According to these criteria, the FBS model provides a better fit to the data of nickel concentration.

4.1.3. FBS versus a Mixture of Log-Normal Distributions

Following reviewer's recommendations, a mixture of two log-normal distributions is also considered. The log-normal mixture model will be given by (23) with f_j the pdf of a log-normal distribution with parameters (μ_j, σ_j) , $j = 1, 2$ and $0 < p < 1$, and it is denoted by $MLN(\mu_1, \sigma_1, \mu_2, \sigma_2, p)$. The estimated mixture model is

$$MLN(2.829, 0.177, 2.8275, 0.877, 0.327)$$

with $AIC = 663.571$, $CAIC = 680.784$ and $BIC = 675.784$. All of them are less than those corresponding to FBS. So, according to these criteria, the mixture of two log-normal distributions provides a better fit to this dataset than the FBS model.

This discussion illustrates that, quite often, the final selection of a model is a matter of choice. FBS model can be considered as appropriate if we want to use a more parsimonious model, and this is better than other BS models and a mixture of two normal distributions. On the other hand, based on AIC, CAIC and BIC, the mixture of two log-normal would be preferred but this model has more parameters than the FBS distribution and may present problems of identifiability. Anyway, the final choice must be properly justified.

Figure 4 depicts maximized likelihoods and empirical cdf for variable nickel concentration revealing that FBS model fitting is quite good.

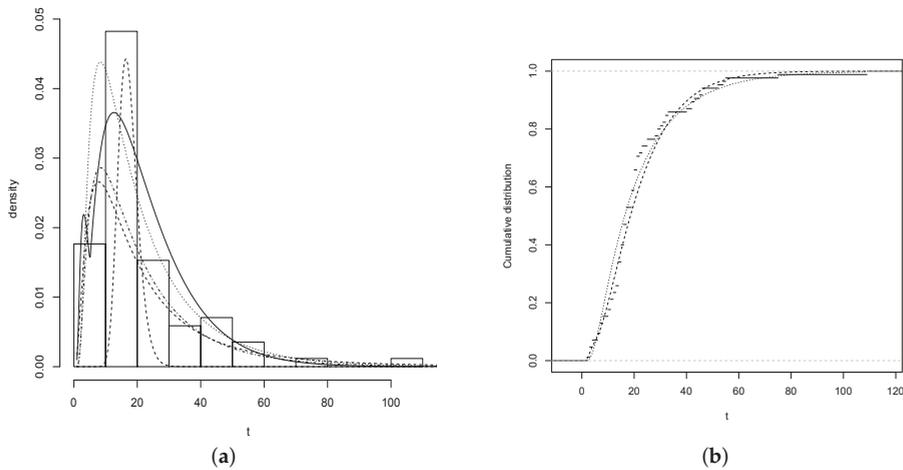


Figure 4. (a) Plots for FBS (solid line), MLN (dashed line), BS (dotted line) and SBS (dotted and dashed line) models. (b) Empirical cdf with estimated FBS cdf (dashed line) and estimated BS cdf (dotted line).

Remark 3. Going through the origin of this data set, the bimodal behavior of the nickel concentration statistical model seems to be due to the fact that the samples were taken according to different lithologies. Lithology classifies according to the physical and chemical elements in rock formation. Mining operations found different lithologies in these samples, as it is depicted in Figure 4.

4.2. Air Pollution

The concentration of average air pollutants has been used in epidemiological surveillance as an indicator of the level of atmospheric contamination. Among its associated adverse effects in humans, diseases such as bronchitis are found. The distribution of this concentration has a bias to the right, and is always positive. It is typically assumed that the data on air pollutant concentrations are uncorrelated and independent and thus they do not require the diurnal or cyclic trend analysis, see [32]. The data set studied in this section corresponds to daily measures of ozone levels (in $ppb = ppm \times 1000$) in the city of New York, USA, from May to September, 1973, collected by the New York State Conservation Department. The sample mean, standard deviation, asymmetry and kurtosis coefficients are given, respectively, by 42.129, 32.987, 1.209 and 1.112.

4.2.1. FBS versus the BS and SBS Distributions

Maximum likelihood estimators, their estimated standard errors (in parenthesis), for the BS, SBS and FBS models, were computed, the results are:

BS model: $\hat{\alpha} = 0.982$ (0.064) and $\hat{\beta} = 28.031$ (2.265).

SBS model: $\hat{\alpha} = 1.270$ (0.235), $\hat{\beta} = 14.831$ (4.019) and $\hat{\lambda} = 1.066$ (0.533).

FBS: $\hat{\alpha} = 5.160$ (0.481), $\hat{\beta} = 78.000$ (0.008), $\hat{\delta} = 3.991$ (0.050) and $\hat{\lambda} = -9.135$ (2.417).

For this data set, the log-likelihood ratio statistics to test BS vs FBS and SBS vs FBS are given by

$$-2 \log(\Lambda_1) = 12.812 \quad \text{and} \quad -2 \log(\Lambda_2) = 5.828,$$

which are greater than the corresponding 5% critical values from the chisquare distribution, which are 5.99 (with 2 df) and 3.84 (with one df), respectively. So, we can conclude that the unimodal FBS model provides a better fit to this dataset than BS and SBS models.

4.2.2. FBS versus the Extended BS (EBS) Model

Fitting the five-parameter EBS model, $EBS(\alpha, \beta, \sigma, \nu, \lambda)$, proposed in [16] as the best for this dataset, and whose point estimates for the parameters and summaries for comparison are equal to

$$EBS(0.780, 0.596, 3.618, -3.539, -0.096),$$

$CAIC = 1111.154$ and $BIC = 1106.154$. On the other hand, for the FBS model, it was obtained $CAIC = 1108.396$ and $BIC = 1104.396$. Then, according to the CAIC and BIC criteria, FBS model presents the best fit to this data set dealing with the daily ozone level concentration in the atmosphere.

Figure 5 depicts the histograms and the fitted density curves for the data set studied and empirical cdf for variable daily ozone level concentration in the atmosphere, while the dashed line corresponds to the cdf for FBS model.

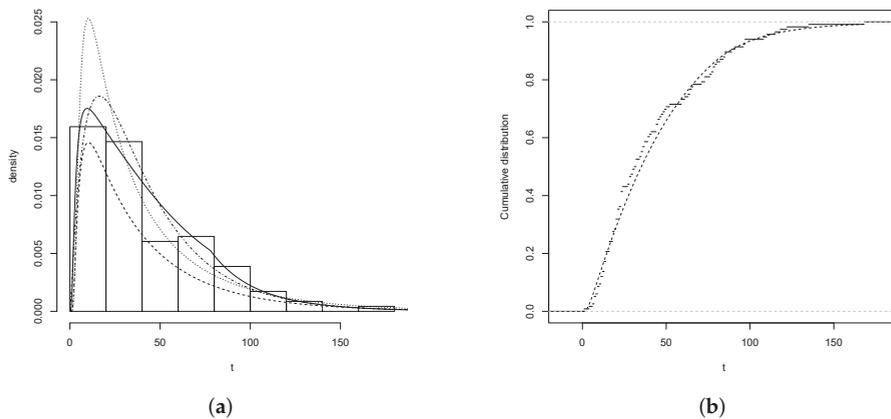


Figure 5. (a) Plots for FBS (solid line), BS (dotted line), SBS (dashed line) and EBS (dotted and dashed line) models. (b) Empirical cdf with estimated FBS cdf (dashed line).

5. Conclusions

We have introduced a new family of distributions able to model skewness, unimodality and bimodality in the BS distribution. We have discussed several of its properties. Explicit expressions for the cdf are given in terms of the cdf of a bivariate normal variable. Non-linear equations to obtain the modes of this distribution are provided. The estimation of parameters is carried out via maximum likelihood. We highlight that the ML equations must be solved by using iterative methods. The information matrix is non-singular and therefore likelihood ratio tests to compare this model with other nested models can be implemented. The interest and flexibility of our proposal is supported with two illustrations to real data in which we show that:

- (i) the FBS model provides consistently better fits than the BS and SBS models (they can be considered relevant precedents of our proposal)
- (ii) the FBS distribution can improve the fit provided by other competing models designed to deal with bimodality (such as a mixture of normal distributions). It can also perform better for unimodal situations in which a generalized BS model with skewness parameters must be applied, such as the EBS model proposed in [16]. We highlight that in both situations FBS provides a better fit with a more parsimonious model (less number of parameters), and the problem of identifiability of mixtures can be circumvented.

Therefore the outcome of these practical demonstrations show that the new family is very flexible and widely applicable.

Author Contributions: All the authors contributed significantly to the present paper.

Funding: The research of G. Martínez-Flórez was supported by Grant Proyecto Universidad de Córdoba: Distribuciones de probabilidad asimétrica bimodal con soporte positivo (Colombia). The research of I. Barranco-Chamorro was supported by Grant CTM2015-68276-R (Spain). H.W. Gómez was supported by Grant SEMILLERO UA-2019 (Chile).

Acknowledgments: The authors would like to thank the editor and the anonymous referees for their comments and suggestions, which significantly improved our manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Proof of Proposition 1 (cdf in the FBS model). From Equation (7), T is a monotonically increasing function of $Z \sim FSN(\delta, \lambda)$. Therefore the cdf of T is given by

$$F_T(t) = F_Z(a_t) \tag{A1}$$

where $F_Z(\cdot)$ denotes the cdf of Z and a_t was given in (6).

(i) First, we obtain the cdf of $Z \sim FSN(\delta, \lambda)$

It can be seen in Gómez et al. [2], Proposition 4, that the pdf of $Z \sim FSN(\delta, \lambda)$ is

$$f_Z(z) = \begin{cases} c_\delta \phi(z - \delta) \Phi(\lambda z), & \text{if } z < 0 \\ c_\delta \phi(z + \delta) \Phi(\lambda z), & \text{if } z \geq 0. \end{cases}$$

Let us consider the case for $z < 0$

$$F_Z(z) = \int_{-\infty}^z f_Z(t) dt = \int_{-\infty}^z c_\delta \phi(t - \delta) \Phi(\lambda t) dt$$

By making the change of variable $v = t - \delta$, and later, taking into account that $\Phi(\cdot)$ is the cdf of a $N(0, 1)$ distribution, we have that

$$F_Z(z) = c_\delta \int_{-\infty}^{z-\delta} \phi(v) \Phi(\lambda(v + \delta)) dv = c_\delta \int_{-\infty}^{z-\delta} \int_{-\infty}^{\lambda(v+\delta)} \phi(v) \phi(s) ds dv \tag{A2}$$

The integrand in (A2) is the joint pdf of two independent $N(0, 1)$ rv's, (S, V) , i.e.,

$$\begin{pmatrix} S \\ V \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

Note that (A2) can be rewritten as

$$\begin{aligned} F_Z(z) &= c_\delta Pr [S - \lambda V \leq \lambda \delta, V \leq z - \delta] \\ &= c_\delta Pr \left[\frac{S - \lambda V}{\sqrt{1 + \lambda^2}} \leq \frac{\lambda \delta}{\sqrt{1 + \lambda^2}}, V \leq z - \delta \right] \\ &= c_\delta \Phi_{BN\lambda} \left(\frac{\lambda \delta}{\sqrt{1 + \lambda^2}}, z - \delta \right) \end{aligned} \tag{A3}$$

where $\Phi_{BN\lambda}(x, y)$ denotes the cdf of a bivariate normal distribution, with mean vector $\mu' = (0, 0)$ and covariance matrix Ω_λ given in (10).

For $z > 0$, we have that

$$\begin{aligned} F_Z(z) &= \int_{-\infty}^0 f_Z(t) dt + \int_0^z f_Z(t) dt \\ &= F_Z(0) + c_\delta \int_0^z \phi(t + \delta) \Phi(\lambda t) dt. \end{aligned} \tag{A4}$$

From (A3), it follows that

$$F_Z(0) = \lim_{z \rightarrow 0^-} F_Z(z) = c_\delta \Phi_{BN_\lambda} \left(\frac{\lambda\delta}{\sqrt{1+\lambda^2}}, -\delta \right) \tag{A5}$$

On the other hand, proceeding similarly to the previous case (change of variable $v = t + \delta$), it can be proved that

$$\begin{aligned} \int_0^z \phi(t + \delta) \Phi(\lambda t) dt &= Pr \left[\frac{S - \lambda V}{\sqrt{1 + \lambda^2}} \leq -\frac{\lambda\delta}{\sqrt{1 + \lambda^2}}, \delta < V \leq z + \delta \right] \\ &= \Phi_{BN_\lambda} \left(-\frac{\lambda\delta}{\sqrt{1 + \lambda^2}}, z + \delta \right) - \Phi_{BN_\lambda} \left(-\frac{\lambda\delta}{\sqrt{1 + \lambda^2}}, \delta \right) \end{aligned} \tag{A6}$$

Therefore, from (A3)–(A6), we have just proved that the cdf of Z is

$$F_Z(z) = \begin{cases} c_\delta \Phi_{BN_\lambda} \left(\frac{\lambda\delta}{\sqrt{1+\lambda^2}}, z - \delta \right), & \text{if } z < 0 \\ c_\delta \left[\Phi_{BN_\lambda} \left(\frac{\lambda\delta}{\sqrt{1+\lambda^2}}, -\delta \right) + \Phi_{BN_\lambda} \left(-\frac{\lambda\delta}{\sqrt{1+\lambda^2}}, z + \delta \right) - \Phi_{BN_\lambda} \left(-\frac{\lambda\delta}{\sqrt{1+\lambda^2}}, \delta \right) \right], & \text{if } z \geq 0 \end{cases} \tag{A7}$$

(ii) Finally, the expression for the cdf of $T \sim FBS(\alpha, \beta, \delta, \lambda)$ given in (9) follows from (A1) and (A7). □

Proof of Proposition 3 (Modes in the FBS model). Recall that, from (A1), the pdf of T is given by

$$f_T(t) = f_Z(a_t) a_t' = c_\delta a_t' \phi(|a_t| + \delta) \Phi(\lambda a_t)$$

where $f_Z(\cdot)$ denotes the pdf of $Z \sim FSN(\delta, \lambda)$, a_t was given in (6) and

$$a_t' = \frac{\partial}{\partial t} a_t = \frac{t^{-3/2} \beta^{-1/2}}{2\alpha} (t + \beta). \tag{A8}$$

For $a_t < 0$, or equivalently $0 < t < \beta$, consider the first derivative with respect to t of $f_T(\cdot)$ and equating to zero, we have

$$f_T'(t) = c_\delta \frac{\partial}{\partial t} \{ a_t' \phi(-a_t + \delta) \Phi(\lambda a_t) \} = 0 \tag{A9}$$

By using that $\phi'(z) = -z\phi(z)$, it can be proved that (A9) is equivalent to

$$\{ a_t' \}^2 [(\delta - a_t) \Phi(\lambda a_t) + \lambda \phi(\lambda a_t)] + a_t'' \Phi(\lambda a_t) = 0 \tag{A10}$$

Since $a_t' > 0, \forall t > 0 (\beta > 0)$, we have that (A10) is equivalent to (12).

Similarly, for $a_t > 0$, i.e., $t > \beta$, from $f_T'(t) = c_\delta \frac{\partial}{\partial t} \{ a_t' \phi(a_t + \delta) \Phi(\lambda a_t) \} = 0$, (13) is obtained. □

Remark A1 (Comments to Proposition 3). *In order to illustrate the use of Equations (12) and (13) next cases are considered.*

1. Consider the pdf given in Figure 1a, case $\alpha = 0.75, \beta = 1, \lambda = 1, \delta = 0.75$. In this setting there do not exist $t_1^* \in (0, \beta)$ and $t_2^* > \beta$ satisfying (12) and (13), respectively. It can be checked that the distribution is unimodal and the mode is at β .
2. Figure 1b, case $\alpha = 0.75, \beta = 1, \lambda = -1, \delta = 0.75$. There exists $t_1^* \in (0, \beta)$ satisfying (12) and there does not exist $t_2^* > \beta$ satisfying (13). Then the distribution is unimodal and the mode is at t_1^* .
3. Figure 2a,b, in all cases under consideration, there exist $t_1^* \in (0, \beta)$ and $t_2^* > \beta$ satisfying (12) and (13). Then the distribution is bimodal and the modes are t_1^* and t_2^* .

Proof of Theorem 1 (*p*th quantile, change of scale and reciprocity). (i) (14) follows from the fact that (7) is one-to-one function preserving the order from \mathbb{R} to \mathbb{R}^+ .

(ii) Note that the pdf of T can be rewritten as

$$f_T(t) = c_\delta a'_t(\alpha, \beta) \phi(|a_t(\alpha, \beta)| + \delta) \Phi(\lambda a_t(\alpha, \beta)) \quad (\text{A11})$$

with $c_\delta = (1 - \Phi(\delta))^{-1}$, $a_t = a_t(\alpha, \beta)$ given in (6) and

$$a'_t = a'_t(\alpha, \beta) = \frac{\partial}{\partial t} a_t(\alpha, \beta) = \frac{t^{-3/2} \beta^{-1/2}}{2\alpha} (t + \beta). \quad (\text{A12})$$

Let $Y = kT$ with $k > 0$. By applying the Jacobian technique $f_Y(y) = |J| f_T(\frac{y}{k}; \alpha, \beta, \delta, \lambda)$ with $|J| = \frac{1}{k}$. From (6), $a_{y/k}(\alpha, \beta) = a_y(\alpha, k\beta)$, and from (A12)

$$|J| a'_{y/k}(\alpha, \beta) = \frac{y^{-3/2} (k\beta)^{-1/2}}{2\alpha} (y + k\beta) = a'_y(\alpha, k\beta).$$

Therefore

$$f_Y(y) = c_\delta a'_y(\alpha, k\beta) \phi(|a_y(\alpha, k\beta)| + \delta) \Phi(\lambda a_y(\alpha, k\beta)),$$

i.e., $Y \sim FBS(\alpha, k\beta, \delta, \lambda)$.

(iii) Let be $Y = T^{-1}$. In this case $|J| = Y^{-2}$, $a_{y^{-1}}(\alpha, \beta) = -a_y(\alpha, \beta^{-1})$, and $|J| a'_{y^{-1}}(\alpha, \beta) = a'_y(\alpha, \beta^{-1})$.

Therefore

$$f_Y(y) = |J| f_T(y^{-1}; \alpha, \beta, \delta, \lambda) = c_\delta a'_y(\alpha, \beta^{-1}) \phi(|a_y(\alpha, \beta^{-1})| + \delta) \Phi(-\lambda a_y(\alpha, \beta^{-1})),$$

i.e., $Y = T^{-1} \sim FBS(\alpha, \beta^{-1}, \delta, -\lambda)$.

□

References

- Birnbaum, Z.W. Effect of linear truncation on a multinormal population. *Ann. Math. Statist.* **1950**, *21*, 272–279. [\[CrossRef\]](#)
- Gómez, H.W.; Elal-Olivero, D.; Salinas, H.S.; Bolfarine, H. Bimodal Extension Based on the Skew-normal Distribution with Application to Pollen Data. *Environmetrics* **2009**, *22*, 50–62. [\[CrossRef\]](#)
- Lehmann, E.L. The power of rank tests. *Ann. Mathematical Stat.* **1953**, *24*, 23–43. [\[CrossRef\]](#)
- Roberts, C. A Correlation Model Useful in the Study of Twins. *J. Am. Stat. Assoc.* **1966**, *61*, 1184–1190. [\[CrossRef\]](#)
- O'Hagan, A.; Leonard, T. Bayes estimation subject to uncertainty about parameter constraints. *Biometrika* **1976**, *63*, 201–203. [\[CrossRef\]](#)
- Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
- Arnold, B.C.; Gómez, H.W.; Salinas, H.S. On multiple constraint skewed models'. *Statistics* **2009**, *43*, 279–293. [\[CrossRef\]](#)
- Bolfarine, H.; Martínez-Flórez, G.; Salinas, H.S. Bimodal symmetric-asymmetric families. *Commun. Stat. Theory Methods* **2018**, *47*(2), 259–276. [\[CrossRef\]](#)
- Durrans, S.R. Distributions of fractional order statistics in hydrology. *Water Resour. Res.* **1992**, *28*, 1649–1655. [\[CrossRef\]](#)
- Pewsey, A.; Gómez, H.W.; Bolfarine, H. Likelihood-based inference for power distributions. *Test* **2012**, *21*, 775–789. [\[CrossRef\]](#)
- Chan, J.; Choy, S.; Makov, U.; Landsman, Z. Modelling insurance losses using contaminated generalised beta type-II distribution. *ASTIN Bull.* **2018**, *48*, 871–904.
- Birnbaum, Z.W.; Saunders, S.C. Estimation for a family of life distributions with applications to fatigue. *J. Appl. Probab.* **1969**, *6*, 328–347. [\[CrossRef\]](#)

13. Birnbaum, Z.W.; Saunders, S.C. A New Family of Life Distributions. *J. Appl. Probab.* **1969**, *6*, 319–327. [[CrossRef](#)]
14. Díaz-García, J.A.; Leiva-Sánchez, V. A new family of life distributions based on the elliptically contoured distributions. *J. Statist. Plann. Inference* **2005**, *128*, 445–457. [[CrossRef](#)]
15. Vilca-Labra, F.; Leiva-Sanchez, V. A new fatigue life model based on the family of skew-elliptical distributions. *Commun. Stat. Theory Methods* **2006**, *35*, 229–244. [[CrossRef](#)]
16. Leiva, V.; Vilca-Labra, F.; Balakrishnan, N.; Sanhueza, A. A skewed sinh-normal distribution and its properties and application to air pollution. *Commun. Stat. Theory Methods* **2010**, *39*, 426–443. [[CrossRef](#)]
17. Castillo, N.O.; Gómez, H.W.; Bolfarine, H. Epsilon Birnbaum-Saunders distribution family: Properties and inference. *Stat. Pap.* **2011**, *52*, 871–883. [[CrossRef](#)]
18. Gómez, H.W.; Olivares, J.; Bolfarine, H. An extension of the generalized Birnbaum-Saunders distribution. *Stat. Probab. Lett.* **2009**, *79*, 331–338. [[CrossRef](#)]
19. Reyes, J.; Barranco-Chamorro, I.; Gallardo, D.I.; Gómez, H.W. Generalized Modified Slash Birnbaum-Saunders Distribution. *Symmetry* **2018**, *10*, 724. [[CrossRef](#)]
20. Reyes, J.; Barranco-Chamorro, I.; Gómez, H.W. Generalized modified slash distribution with applications. *Commun. Stat. Theory Methods* **2019**, doi:10.1080/03610926.2019.1568484. [[CrossRef](#)]
21. Martínez-Flórez, G.; Bolfarine, H.; Gómez, H.W. An alpha-power extension for the Birnbaum-Saunders distribution. *Statistics* **2014**, *48*, 896–912. [[CrossRef](#)]
22. Olmos, N.M.; Martínez-Flórez, G.; Bolfarine, H. Bimodal Birnbaum-Saunders Distribution with Application to Corrosion Data. *Commun. Stat. Theory Methods* **2017**, *46*, 6240–6257. [[CrossRef](#)]
23. Shakhathreh, M.; Lemonte, A.; Moreno-Arenas, G. The log-normal modified Weibull distribution and its reliability implications. *Reliab. Eng. Syst. Saf.* **2019**, *188*, 6–22. [[CrossRef](#)]
24. Prativiera, F.; Ortega, E.M.M.; Cordeiro, G.M.; Pescim, R.R. A new generalized odd log-logistic flexible Weibull regression model with applications in repairable systems. *Reliab. Eng. Syst. Saf.* **2018**, *176*, 13–26. [[CrossRef](#)]
25. Ng, H.K.T.; Kundu, D.; Balakrishnan, N. Modified moment estimation for the two-parameter Birnbaum-Saunders distribution. *Comput. Statist. Data Anal.* **2003**, *43*, 283–298. [[CrossRef](#)]
26. Barros, M.; Paula, G.A.; Leiva, V. A new class of survival regression models with heavy-tailed errors: Robustness and diagnostics. *Lifetime Data Anal.* **2008**, *14*, 316–332. [[CrossRef](#)]
27. Lemonte, A.; Cribari-Neto, F.; Vasconcellos, K. Improved statistical inference for the two-parameter Birnbaum-saunders distribution. *Comput. Stat. Data Anal.* **2007**, *51*, 4656–4681. [[CrossRef](#)]
28. R Development Core Team. *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2017.
29. Gradshteyn, I.S.; Ryzhik, I.M. *Table of Integrals, Series, and Products*; Academic Press: New York, NY, USA, 2007.
30. Nadarajah, S. A truncated inverted beta distribution with application to air pollution data. *Stoch. Environ. Res. Risk Assess.* **2008**, *22*, 285–289. [[CrossRef](#)]
31. Akaike, H. A new look at statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [[CrossRef](#)]
32. Gokhale, S.; Khare, M. Statistical behavior of carbon monoxide from vehicular exhausts in urban environments. *Environ. Model. Softw.* **2007**, *22*, 526–535. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Article

An Asymmetric Distribution with Heavy Tails and Its Expectation–Maximization (EM) Algorithm Implementation

Neveka M. Olmos ¹, Osvaldo Venegas ^{2,*}, Yolanda M. Gómez ³ and Yuri A. Iriarte ¹

¹ Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta 1240000, Chile; neveka.olmos@uantof.cl (N.M.O.); yuri.iriarte@uantof.cl (Y.A.I.)

² Departamento de Ciencias Matemáticas y Físicas, Facultad de Ingeniería, Universidad Católica de Temuco, Temuco 4780000, Chile

³ Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Copiapó 1530000, Chile; yolanda.gomez@uda.cl

* Correspondence: ovenegas@uct.cl; Tel.: +56-45-2205411

Received: 8 August 2019; Accepted: 2 September 2019; Published: 10 September 2019

Abstract: In this paper we introduce a new distribution constructed on the basis of the quotient of two independent random variables whose distributions are the half-normal distribution and a power of the exponential distribution with parameter 2 respectively. The result is a distribution with greater kurtosis than the well known half-normal and slashed half-normal distributions. We studied the general density function of this distribution, with some of its properties, moments, and its coefficients of asymmetry and kurtosis. We developed the expectation–maximization algorithm and present a simulation study. We calculated the moment and maximum likelihood estimators and present three illustrations in real data sets to show the flexibility of the new model.

Keywords: slashed half-normal distribution; kurtosis; likelihood; EM algorithm

1. Introduction

In recent years, for data with positive support, specifically, lifetime, or reliability, the half-normal (HN) model has been widely used. The probability density function (pdf) is given by

$$f(x; \sigma) = \frac{2}{\sigma} \phi\left(\frac{x}{\sigma}\right) I\{x > 0\},$$

where $\sigma > 0$ is the scale parameter and $\phi(\cdot)$ represents the standard normal pdf. We denote this by writing $X \sim HN(\sigma)$.

Some generalizations for this model are proposed by Cooray and Ananda [1], Cordeiro et al. [2], Bolfarine and Gómez [3] and Gómez and Vidal [4].

Olmos et al. [5] extended the HN distribution by incorporating a kurtosis parameter q , with the purpose of obtaining heavier tails, i.e., it has greater kurtosis than the base model. They called this model the slashed half-normal (SHN) distribution. Its construction is based on considering the quotient of two independent random variables, with random variable $X \sim HN(\sigma)$ in the numerator and the $U \sim U(0, 1)$ in the denominator (See Rogers and Tukey [6] and Mosteller and Tukey [7] for more details). Thus a model is obtained that has more flexible coefficients of asymmetry and kurtosis than the HN model. We say that a random variable T follows a SHN if its pdf is given by

$$f_T(t; \sigma, q) = q \sqrt{\frac{2q}{\pi}} \sigma^q \Gamma((q+1)/2) t^{-(q+1)} G\left(t^2; (q+1)/2, \frac{1}{2\sigma^2}\right), \quad t > 0, \quad (1)$$

where $\sigma > 0$ is a scale parameter, $q > 0$ is a kurtosis parameter, $G(z; a, b) = \int_0^z g(x; a, b) dx$ is the cumulative distribution function (cdf) of the gamma distribution and $g(\cdot; a, b)$ is the pdf of the gamma model with shape and rate parameters a and b , respectively.

Reyes et al. [8] introduced the modified slash (MS) distribution. We say that M has a MS distribution if

$$M = Z/E^{\frac{1}{q}}, \quad (2)$$

the construction of which is based on considering an exponential (Exp) distribution with parameter 2 in the denominator, i.e., they consider that $E \sim \text{Exp}(2)$. The motivation of the selection of the $\text{Exp}(2)$ distribution is given in Reyes et al. [8]. The result of this work shows that the MS model has a greater coefficient of kurtosis and this characteristic is very important for modeling data sets when they contain atypical observations.

The principal goal of this article is to use the idea published by Reyes et al. [8] to construct an extension of the half-normal model with a greater range in the coefficient of kurtosis than the SHN model, in order to use it to model atypical data. This will allow us obtain a new model generated on the basis of a scale mixture between an HN and a Weibull (Wei) distribution.

The rest of the paper is organized as follows: Section 2 contains the representation of this model and we generate the density of the new family, its basic properties and moments, and its coefficients of asymmetry and kurtosis. In Section 3 we make inferences using the moments and maximum likelihood (ML) methods. In Section 4 we implement the expectation–maximization (EM) algorithm. In Section 5 we carry out a simulation study for parameter recovery. We show three illustrations in real datasets in Section 6 and finally in Section 7 we present our conclusions.

2. An Asymmetric Distribution

In this section we introduce the representation, its pdf, and some important properties and graphs to show the flexibility of the new model.

2.1. New Distribution

The representation of this new distribution is

$$T = \frac{X}{Y^{1/q}}, \quad (3)$$

where $X \sim \text{HN}(\sigma)$ and $Y \sim \text{Exp}(2)$ are independent, $\sigma > 0$, $q > 0$. We call the distribution of T the modified slashed half-normal (MSHN) distribution. This is denoted by $T \sim \text{MSHN}(\sigma, q)$.

2.2. Density Function

The following Proposition shows the pdf of the MSHN distribution with scale parameter σ and kurtosis parameter q , generated using the representation given in (3).

Proposition 1. Let $T \sim \text{MSHN}(\sigma, q)$. Then, the pdf of T is given by

$$f_T(t; \sigma, q) = \frac{2q}{\sqrt{2\pi\sigma^2 t^{q+1}}} N\left(\frac{q+1}{2}, \frac{2}{t^q}, \frac{q}{2}, \frac{1}{2\sigma^2}\right), \quad (4)$$

where $t > 0$, $\sigma > 0$, $q > 0$, and $N(\cdot, \cdot, \cdot, \cdot)$ is defined in Lemma 1 in the Appendix A.

Proof. Using the stochastic representation given in (3) and the Jacobian method, we obtain that the density function associated with T is given by

$$f_T(t; \sigma, q) = \frac{4q}{\sqrt{2\pi\sigma^2}} \int_0^\infty w^q \exp\left\{-\left(\frac{t^2 w^2}{2\sigma^2} + 2w^q\right)\right\} dw.$$

Making the change of variable $u = t^2w^2$ we have,

$$f_T(t; \sigma, q) = \frac{2q}{\sqrt{2\pi\sigma^2t^{q+1}}} \int_0^\infty u^{\frac{q-1}{2}} \exp\left\{-\left(\frac{u}{2\sigma^2} + \frac{2u^{q/2}}{t^q}\right)\right\} du.$$

Hence, applying the Lemma 1 as set forth in the Appendix A, we obtain the result. \square

Figure 1 depicts plots of the density of the MSHN distribution for different values of parameter q .

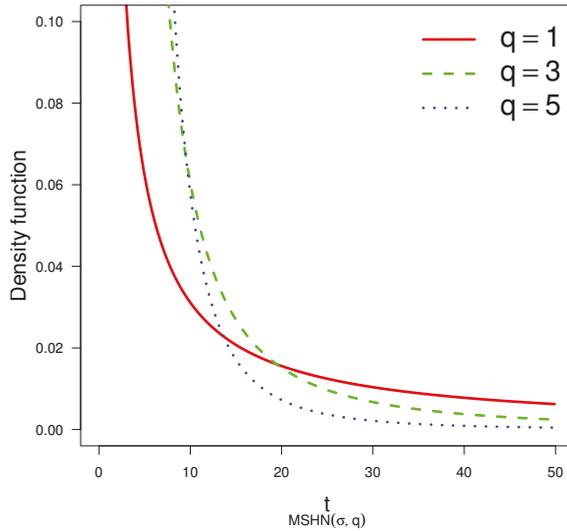


Figure 1. The density function for different values of parameter q and $\sigma = 1$ in the MSHN distribution.

We perform a brief comparison illustrating that the tails of the MSHN distribution are heavier than those of the SHN distribution.

Table 1 shows the tail probability for different values in the SHN and MSHN models. It is immediately apparent that the MSHN tails are heavier than those of the SHN distribution.

Table 1. Tails comparison for different slashed half-normal (SHN) and modified slashed half-normal (MSHN) distributions.

Distribution	$P(T > 3)$	$P(T > 4)$	$P(T > 5)$	$P(T > 6)$	$P(T > 7)$
SHN(1, 0.5)	0.3781	0.3497	0.3239	0.3009	0.2805
MSHN(1, 0.5)	0.5304	0.48289	0.4466	0.4176	0.3936
SHN(1, 1)	0.1777	0.1570	0.1385	0.1224	0.1086
MSHN(1, 1)	0.3678	0.2992	0.2519	0.2173	0.19102
SHN(1, 3)	0.0350	0.0205	0.0120	0.0044	0.0034
MSHN(1, 3)	0.0901	0.0438	0.0238	0.0142	0.0091

2.3. Properties

In this sub-section we study some properties of the MSHN distribution.

Proposition 2. Let $T \sim MSHN(\sigma, q)$, then when $\sigma = q = 1$ the density is

$$f_T(t) = \frac{4}{t^2} \left(\frac{1}{\sqrt{2\pi}} - \frac{2}{t} \exp(2/t^2) \Phi\left(-\frac{2}{t}\right) \right), \quad t > 0, \tag{5}$$

where $\Phi(\cdot)$ is the cdf of the standard normal.

Proof. Using Proposition 1 for $\sigma = q = 1$, we have,

$$f_T(t) = \frac{2}{\sqrt{2\pi t^2}} N\left(1, \frac{2}{z}, \frac{1}{2}, \frac{1}{2}\right) = \frac{2}{\sqrt{2\pi t^2}} \int_0^\infty \exp\left(-\frac{2}{z}x^{1/2} - \frac{1}{2\sigma^2}x\right) dx, \quad t > 0. \tag{6}$$

Changing the variable $x = u^2$ we obtain the result. \square

Proposition 3. If $T|W = w \sim HN\left(\frac{\sigma}{w}\right)$ and $Y^{1/q} = W \sim Wei(q, 1/2)$ then $T \sim MSHN(\sigma, q)$.

Proof. Since the marginal pdf of T is given by

$$f_T(t; \sigma, q) = \int_0^\infty f_{T|W}(t|w) f_W(w) dw = \frac{4q}{\sigma\sqrt{2\pi}} \int_0^\infty w^q e^{-\frac{w^2 t^2}{2\sigma^2} - 2w^q} dw,$$

and using the Lemma 1 in the Appendix A, the result is obtained. \square

Proposition 4. Let $T \sim MSHN(\sigma, q)$. If $q \rightarrow \infty$ then T converges in law to a random variable $T \sim HN(\sigma)$.

Proof. Let $T \sim MSHN(\sigma, q)$ and $T = \frac{X}{Y^{1/q}}$, where $X \sim HN(\sigma)$ and $Y \sim Exp(2)$.

We study the convergence in law of T , since $Y \sim Exp(2)$ then $W = Y^{1/q} \sim Wei(q, 1/2)$, we have that $E(W - 1)^2 = \frac{1}{2^{2/q}}\Gamma(1 + 2/q) - \frac{2}{2^{1/q}}\Gamma(1 + 1/q) + 1$. If $q \rightarrow \infty$ then $E(W - 1)^2 \rightarrow 0$, i.e., we have $W \xrightarrow{\mathcal{P}} 1$ (see Lehmann [9]).

Since $T \sim MSHN(\sigma, q)$, by applying Slutsky’s Lemma (see Lehmann [9]) to $T = \frac{X}{W}$, we have

$$T \xrightarrow{\mathcal{L}} X \sim HN(\sigma), \quad q \rightarrow \infty, \tag{7}$$

that is, for increasing values of q , T converges in law to a $HN(\sigma)$ distribution. \square

Remark 1. Proposition 2 shows us that the $MSHN(1,1)$ distribution has a closed-form expression. Proposition 3 shows that an $MSHN(\sigma, q)$ distribution can also be obtained as a scale mixture of an HN and a Wei distribution. This property is very important since it makes it possible to generate random numbers and implement the EM algorithm. Proposition 4 implies that, if $q \rightarrow \infty$ then the cdf of an $MSHN(\sigma, q)$ model approaches to the cdf of a $HN(\sigma)$ distribution.

2.4. Moments

In this sub-section, the following proposition shows the computation of the moments of a random variable $T \sim MSHN(\sigma, q)$. Hence, it also displays the coefficients of asymmetry and kurtosis.

Proposition 5. Let $T \sim MSHN(\sigma, q)$. Then the r -th moment of T is given by

$$\mu_r = \mathbb{E}(T^r) = \frac{2^{r\left(\frac{1}{q} + \frac{1}{2}\right)}}{\sqrt{\pi}} \sigma^r \Gamma\left(\frac{r+1}{2}\right) \Gamma\left(\frac{q-r}{q}\right), \quad q > r, \tag{8}$$

where $\Gamma(\cdot)$ denotes the gamma function.

Proof. Let $W \sim Wei(q, 1/2)$ and using Proposition 3, we have

$$\mu_r = E(T^r) = E(E(X^r|W^r)) = E\left(\sqrt{\frac{2^r}{\pi}} \Gamma\left(\frac{r+1}{2}\right) \sigma^r W^{-r}\right) = \sqrt{\frac{2^r}{\pi}} \Gamma\left(\frac{r+1}{2}\right) \sigma^r E(W^{-r}),$$

where $E(W^{-r}) = 2^{r/q}\Gamma((q-r)/q)$, $q > r$ is the r -th moment of the inverse Weibull distribution. \square

Corollary 1. Let $T \sim \text{MSHN}(\sigma, q)$. Then the expectation and variance of T are given respectively by

$$E(T) = \frac{2^{\frac{1}{q} + \frac{1}{2}}}{\sqrt{\pi}} \sigma \Gamma\left(\frac{q-1}{q}\right), \quad q > 1, \text{ and}$$

$$\text{Var}(T) = 2\left(\frac{2}{q} + 1\right) \sigma^2 \left[\frac{1}{2} \Gamma\left(\frac{q-2}{q}\right) - \frac{1}{\pi} \Gamma^2\left(\frac{q-1}{q}\right) \right], \quad q > 2.$$

Corollary 2. Let $T \sim \text{MSHN}(\sigma, q)$. Then the coefficients of asymmetry (β_1) and kurtosis (β_2) are given by

$$\beta_1 = \frac{\frac{1}{\sqrt{\pi}} \Gamma\left(\frac{q-3}{q}\right) - \frac{3}{2\sqrt{\pi}} \Gamma\left(\frac{q-1}{q}\right) \Gamma\left(\frac{q-2}{q}\right) + \frac{2}{\sqrt{\pi^3}} \Gamma^3\left(\frac{q-1}{q}\right)}{\left[\frac{1}{2} \Gamma\left(\frac{q-2}{q}\right) - \frac{1}{\pi} \Gamma^2\left(\frac{q-1}{q}\right) \right]^{3/2}}, \quad q > 3, \text{ and}$$

$$\beta_2 = \frac{\frac{3}{4} \Gamma\left(\frac{q-4}{q}\right) - \frac{4}{\pi} \Gamma\left(\frac{q-1}{q}\right) \Gamma\left(\frac{q-3}{q}\right) + \frac{3}{\pi} \Gamma^2\left(\frac{q-1}{q}\right) \Gamma\left(\frac{q-2}{q}\right) - \frac{3}{\pi^2} \Gamma^4\left(\frac{q-1}{q}\right)}{\left[\frac{1}{2} \Gamma\left(\frac{q-2}{q}\right) - \frac{1}{\pi} \Gamma^2\left(\frac{q-1}{q}\right) \right]^2}, \quad q > 4.$$

Remark 2. Figure 2 shows graphs of the coefficients of the MSHN distribution compared with those of the SHN distribution. Note that the MSHN distribution presents higher asymmetry and kurtosis values than the SHN distribution. Furthermore, in both distributions when $q \rightarrow \infty$ the coefficients of asymmetry and kurtosis converge to $\sqrt{2}(4-\pi)(\pi-2)^{-3/2}$ and $(3\pi^2-4\pi-12)(\pi-2)^{-2}$, respectively; they coincide with the coefficients of the HN distribution.

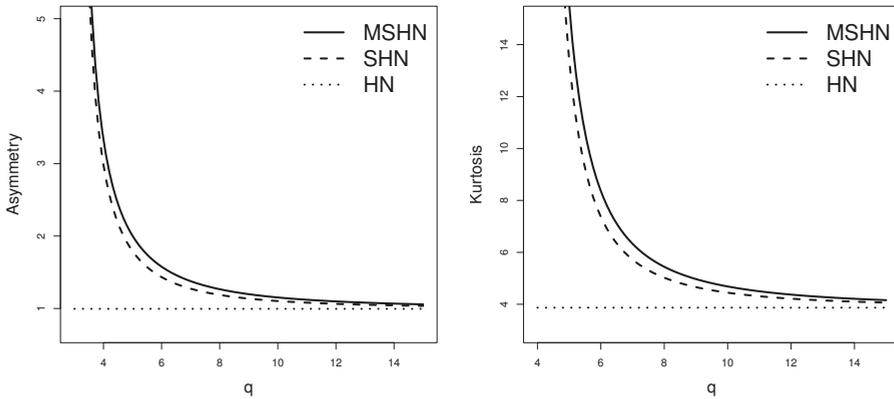


Figure 2. Graph of the coefficients of asymmetry and kurtosis for the MSHN and SHN distributions.

3. Inference

Proposition 6. Let T_1, \dots, T_n be a random sample of size n of the $T \sim \text{MSHN}(\sigma, q)$ distribution. Then for $q > 2$, the moment estimators of σ and q are given by

$$\hat{\sigma}_M = \frac{\sqrt{\pi} \bar{T}}{2^{\frac{1}{q} + \frac{1}{2}} \Gamma\left(\frac{\hat{q}_M - 1}{\hat{q}_M}\right)}, \tag{9}$$

$$\pi \bar{T}^2 \Gamma\left(\frac{\hat{q}_M - 2}{\hat{q}_M}\right) - 2 \bar{T}^2 \Gamma^2\left(\frac{\hat{q}_M - 1}{\hat{q}_M}\right) = 0, \tag{10}$$

where \bar{T} is the mean of the sample and $\overline{T^2}$ is the mean of the sample for the square of the observations.

Proof. From Proposition 5, and considering the first two equations in the moments method, we have

$$\bar{T} = \frac{2^{\frac{1}{q} + \frac{1}{2}}}{\sqrt{\pi}} \sigma \Gamma\left(\frac{q-1}{q}\right) \quad \text{and} \quad \overline{T^2} = 2^{\frac{2}{q}} \sigma^2 \Gamma\left(\frac{q-2}{q}\right).$$

Solving the first equation above for σ we obtain $\hat{\sigma}_M$ given in (9). Substituting $\hat{\sigma}_M$ in the second equation above, we obtain the result given in (10). □

4. Em Algorithm

The EM algorithm (Dempster et al. [10]) is a useful method for ML estimation in the presence of latent variables.

To facilitate the estimation process, we introduce latent variables W_1, \dots, W_n through the following hierarchical representation of the MSHN model:

$$T_i | W_i = w_i \sim HN\left(\frac{\sigma}{w}\right) \quad \text{and} \quad W_i \sim Wei(q, 1/2).$$

In this setting, we have that

$$f_c(w|t) \propto w^q \exp\left\{-\left(\frac{w^2 t^2}{2\sigma^2} + 2w^q\right)\right\}.$$

Therefore, the complete log-likelihood function can be expressed as

$$l_c(\theta|t_c) \propto -n \log(\sigma) - \sum_{i=1}^n \frac{w_i^2 t_i^2}{2\sigma^2} + l_c(q|w_c),$$

where $l_c(q|w_c) = n \log(q) + q \sum_{i=1}^n \log(w_i) - 2 \sum_{i=1}^n w_i^q$.

Letting $\hat{w}_i = \mathbb{E}[W_i|t_i, \theta = \hat{\theta}]$, it follows that the conditional expectation of the complete log-likelihood function has the form

$$Q(\theta|\hat{\theta}) \propto -n \log(\sigma) - \sum_{i=1}^n \frac{\hat{w}_i^2 t_i^2}{2\sigma^2} + Q(q|\hat{\theta}), \tag{11}$$

where $Q(q|\hat{\theta}) = n \log(q) + q S_{1n} - 2 S_{2n,q}$, with $S_{1n} = \sum_{i=1}^n \mathbb{E}[\log(W_i)|t_i]$ and $S_{2n,q} = \sum_{i=1}^n \mathbb{E}[W_i^q|t_i]$.

As both quantities S_{1n} and $S_{2n,q}$ have no explicit forms in the context of the MSHN model, they have to be computed numerically. Thus to compute $Q(q|\hat{\theta})$ we use an approach similar to that of Lee and Xu ([11], Section 3.1), i.e., considering $\{w_r; r = 1, \dots, R\}$ to be a random sample from the conditional distribution $W|(T = t, \theta = \hat{\theta})$, then $Q(q|\hat{\theta})$ can be approximated as

$$Q(q|\hat{\theta}) \approx \frac{1}{R} \sum_{r=1}^R \ell_c(q|w_r).$$

Therefore, the EM algorithm for the MSHN model is given by

- E-step:** Given $\theta = \hat{\theta}^{(k)} = (\hat{\sigma}^{(k)}, \hat{q}^{(k)})^\top$, calculate $\hat{w}_i^{(k)}$, for $i = 1, \dots, n$.
- CM-step I:** Update $\hat{\sigma}^{(k)}$

$$\hat{\sigma}^{2(k+1)} = \frac{S_u^{(k)}}{2},$$

CM-step II: Fix $\alpha = \hat{\sigma}^{(k+1)}$, update $q^{(k)}$ by optimizing $\hat{q}^{(k+1)} = \arg \max_q Q(\hat{\sigma}^{(k+1)}, q | \hat{\theta}^{(k)})$, where $S_u^{(k)} = \frac{1}{n} \sum_{i=1}^n \hat{w}_i^{(k)} t_i$.

The E, CM-I and CM-II steps are repeated until a convergence rule is satisfied, say $|l(\hat{\theta}^{(k+1)}) - l(\hat{\theta}^{(k)})|$ is sufficiently small. Finally, standard errors (SE) can be estimated using the inverse of the observed information matrix.

Remark 3.

- i. For $q \rightarrow \infty$, $\hat{\sigma}$ in M-step reduces to those obtained when the HN distribution is used;
- ii. An alternative to the CM-Steps II is obtained considering the idea in Lin et al. ([12], Section 3), by using the following estimation:
CML-step: Update $q^{(k)}$ by maximizing the constrained actual log-likelihood function, i.e.

$$\hat{q}^{(k+1)} = \arg \max_q \ell(\hat{\sigma}^{(k+1)}, q).$$

5. Simulation

We present a simulation study to assess the performance of the EM algorithm for the parameters σ and q in the MSHN model. We consider 1000 samples of three sample sizes generated from the MSHN model: $n = 30, 50$ and 100 . To generate $T \sim MSHN(\sigma; q)$ the following algorithm was used:

1. Simulate $X \sim N(0, \sigma^2)$ and $Y \sim Exp(2)$.
2. Compute $T = \frac{|X|}{\gamma^{1/q}}$.

For each sample generated, the ML estimates were computed using the EM algorithm. Table 2 shows the mean of the bias estimated for each parameter (bias), its SE and the estimated root of the mean squared error (RMSE). From Table 2, we conclude that the ML estimates are quite stable. The bias is reasonable and diminishes as the sample size is increased. As expected, the terms SE and RMSE are closer when the sample size is increased, suggesting that the SE of the estimators is well estimated.

Table 2. Maximum likelihood (ML) estimations for parameters σ and q of the MSHN distribution. Standard error (SE), root of the mean squared error (RMSE).

True Value		n = 30			n = 50			n = 100			
σ	q	Estimator	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
1	1	σ	0.178	0.430	0.528	0.122	0.320	0.378	0.089	0.206	0.279
		q	0.199	0.422	0.668	0.097	0.219	0.263	0.059	0.138	0.163
	2	σ	0.111	0.355	0.407	0.078	0.258	0.295	0.042	0.172	0.186
		q	1.006	2.500	2.603	0.480	1.105	1.519	0.182	0.458	0.562
	5	σ	0.026	0.277	0.239	0.033	0.222	0.189	0.023	0.159	0.149
		q	2.227	8.833	3.871	2.012	6.743	3.550	1.333	4.092	2.905
2	1	σ	0.284	0.835	0.973	0.192	0.617	0.665	0.104	0.414	0.481
		q	0.168	0.356	0.571	0.094	0.215	0.263	0.058	0.141	0.166
	2	σ	0.294	0.727	0.815	0.122	0.507	0.572	0.074	0.343	0.383
		q	1.210	2.821	2.835	0.465	1.067	1.534	0.174	0.454	0.623
	5	σ	0.057	0.544	0.454	0.066	0.441	0.371	0.044	0.305	0.290
		q	2.456	8.991	3.934	2.089	6.712	3.615	1.545	4.150	3.075
5	1	σ	0.834	2.111	2.548	0.494	1.527	1.826	0.386	1.038	1.233
		q	0.217	0.414	0.740	0.119	0.225	0.287	0.083	0.144	0.174
	2	σ	0.658	1.782	2.065	0.293	1.285	1.475	0.209	0.872	0.966
		q	1.218	2.872	2.836	0.413	1.018	1.414	0.188	0.489	0.694
	5	σ	0.094	1.379	1.160	0.146	1.096	0.950	0.123	0.779	0.731
		q	2.266	8.894	3.880	1.952	6.526	3.557	1.370	4.118	2.948

6. Applications

In this section we provide three applications to real datasets that illustrate the flexibility of the proposed model.

6.1. Application 1

Lyu [13] presents a data set related 104 times with programming in the Centre for Software Reliability (CSR). Some descriptive statistics are: mean = 147.8, variance = 60,071.7, skewness = 3, and kurtosis = 14.6. The moment estimators for the MSHN model were $\hat{\sigma}_M = 74.085$ and $\hat{q}_M = 2.402$, which were used as initial values to compute the ML estimator in Table 3.

For each distribution we report the estimated log-likelihood. To compare the competing models, we consider the Akaike information criterion (AIC) (Akaike [14]) and the Bayesian information criterion (BIC) (Schwarz [15]), which are defined as $AIC = 2k - 2 \log \text{lik}$ and $BIC = k \log(n) - 2 \log \text{lik}$, respectively, where k is the number of parameters in the model, n is the sample size and $\log \text{lik}$ is the maximum value for the log-likelihood function. Table 4 displays the AIC and BIC for each model fitted. Figure 3 presents the histogram of the data fitted with the HN, SHN and MSHN distributions, provided with the ML estimations. The QQ-plot for the MSHN and SHN distributions are presented in Figure 4.

Table 3. ML estimations with the corresponding SE for the models fitted. Half-normal (HN).

Parameters	HN (SE)	SHN (SE)	MSHN (SE)
$\hat{\sigma}$	285.191 (19.774)	20.977 (5.674)	19.874 (4.867)
\hat{q}	-	0.687 (0.118)	0.872 (0.115)
Log-likelihood	-663.411	-605.102	-600.876

Table 4. The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) for each model fitted.

Criterion	HN	SHN	MSHN
AIC	1328.822	1214.204	1205.752
BIC	1331.466	1219.493	1211.041

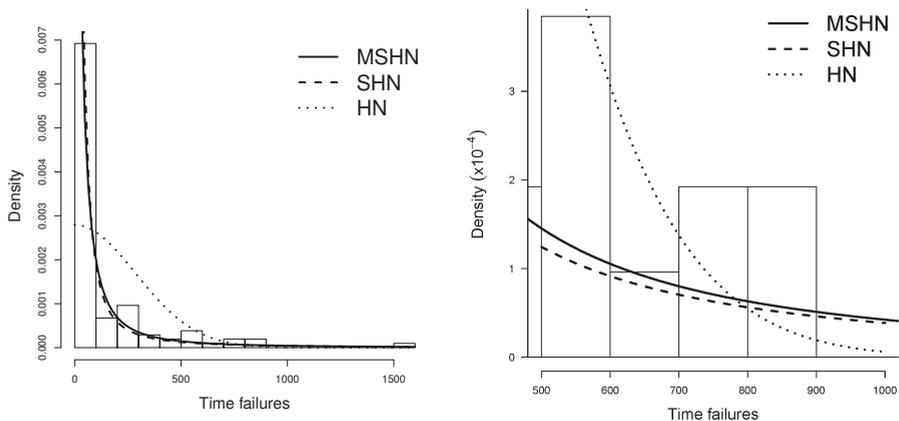


Figure 3. Histogram fitted with the HN, SHN and MSHN distributions provided with the ML estimations.

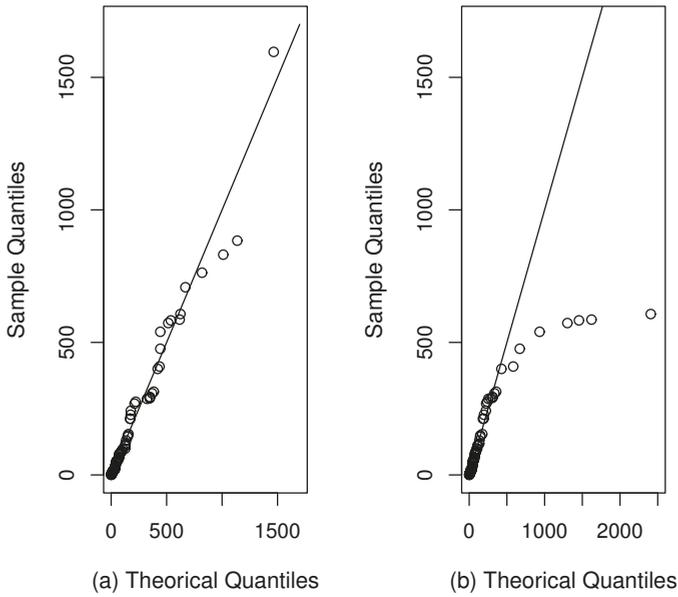


Figure 4. QQ plots: (a) MSHN distribution and (b) SHN distribution.

6.2. Application 2

The second dataset is taken from Von Alven [16], and represents 46 instances of active repairs (in hours) for an airborne communication transceiver. Some descriptive statistics are: mean = 3.607, variance = 24.445, skewness = 2.888, and kurtosis = 11.802.

Initially we computed the moment estimators for the MSHN distribution, obtaining the following estimations: $\hat{\sigma}_M = 2.407$ and $\hat{q}_M = 2.635$. We used these estimations as initial values in computing the ML estimators presented in Table 5. For each distribution we report the estimated log-likelihood.

Table 5. ML estimations with the corresponding SE for the models fitted.

Parameters	HN (SE)	SHN (SE)	MSHN (SE)
$\hat{\sigma}$	6.07 (0.6335)	1.6251 (0.4777)	1.5108 (0.3179)
\hat{q}	-	1.3539 (0.4347)	1.6365 (0.3425)
Log-likelihood	-116.3881	-103.1834	-102.65

Table 6 displays the AIC and BIC for each model fitted. Figure 5 presents the histogram of the data fitted with the HN, SHN and MSHN distributions, provided with the ML estimations.

Table 6. AIC and BIC for each model fitted.

Criterion	HN	SHN	MSHN
AIC	234.7762	210.3668	209.302
BIC	236.6048	214.0241	212.9573

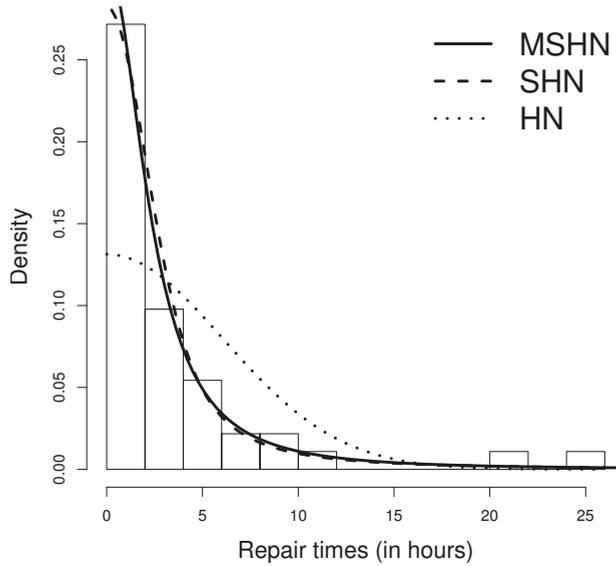


Figure 5. Histogram fitted with the HN, SHN and MSHN distributions provided with the ML estimations.

6.3. Application 3

The third data set (Linhart and Zucchini [17]) represents 31 times of air conditioning system failure of an aeroplane. Some descriptive statistics are: mean = 55.35, variance = 5132.503, skewness = 1.805, and kurtosis = 5.293.

Initially we computed the moment estimators for the MSHN distribution, and obtained the following estimations: $\hat{\sigma}_M = 38.125$ and $\hat{q}_M = 2.743$. We used these estimations as initial values in computing the ML estimators presented in Table 7. For each distribution we report the estimated log-likelihood.

Table 7. ML estimations with the corresponding SE for the models fitted.

Parameters	HN (SE)	SHN (SE)	MSHN (SE)
$\hat{\sigma}$	89.616 (11.381)	13.785 (6.047)	16.148(5.128)
\hat{q}	-	0.859 (0.285)	1.233 (0.251)
Log-likelihood	-161.861	-154.857	-153.954

Table 8 displays the AIC and BIC for each model fitted. Figure 6 presents the histogram of the data fitted with the HN, SHN and MSHN distributions, provided with the ML estimations.

Table 8. AIC and BIC for each model fitted.

Criterion	HN	SHN	MSHN
AIC	325.7224	313.715	311.908
BIC	327.1564	316.583	314.776

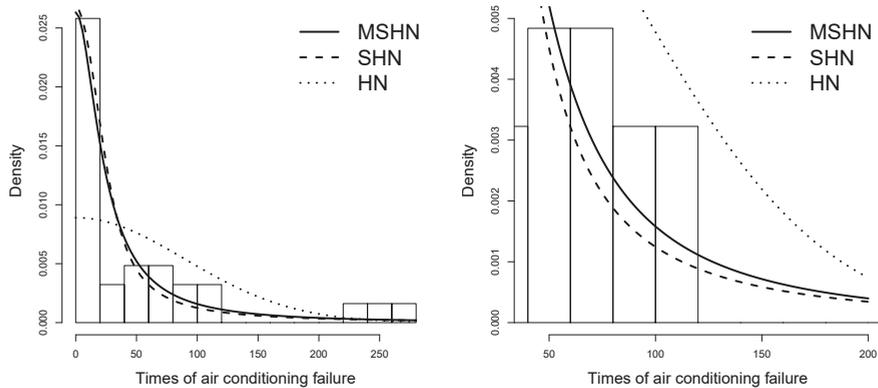


Figure 6. Histogram fitted with the HN, SHN and MSHN distributions provided with the ML estimations.

7. Conclusions

In this paper, we have introduced a new and more flexible model, as it increases kurtosis and contains, as a particular case, the HN distribution. The EM algorithm is implemented, obtaining acceptable results for the maximum likelihood estimators. In applications using real data it performs very well, better than competing models. Some further characteristics of the MSHN distribution are:

- The MSHN distribution has a greater kurtosis than the SHN distribution, as is clearly reflected in Table 1.
- The proposed model has a closed-form expression and presents more flexible asymmetry and kurtosis coefficients than that of the HN model.
- Two stochastic representations for the MSHN model are presented. One is defined as the quotient between two independent random variables: An HN in the numerator and Exp(2) in the denominator. The other shows that the MSHN distribution is a scale mixture of an HN and a Wei distribution.
- Using the mixed scale representation, the EM algorithm was implemented to calculate the ML estimators.
- Results from a simulation study indicate that with a reasonable sample size, an acceptable bias is obtained.
- Three illustrations using real data show that the MSHN model achieves a better fit in terms of the AIC and BIC criteria.

Author Contributions: N.M.O., O.V., Y.M.G. and Y.A.I. contributed significantly to this research article.

Funding: The research of Neveka M. Olmos and Yuri A. Iriarte was supported by SEMILLERO UA-2019. The research of Yolanda M. Gómez was supported by proyecto DIUDA programa de inserción No. 22367 of the Universidad de Atacama.

Acknowledgments: The authors would like to thank the editor and the anonymous referees for their comments and suggestions, which significantly improved our manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Density function of the gamma, exponential and Weibull distributions, respectively, are given by

Gamma distribution:

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x},$$

with, $x > 0, \alpha > 0$ and $\beta > 0$.

Exponential distribution:

$$f(x; \beta) = \frac{1}{\beta} e^{-x/\beta},$$

with, $x > 0$ and $\beta > 0$.

Weibull distribution:

$$f(x; \gamma, \beta) = \frac{\gamma}{\beta} x^{\gamma-1} e^{-x^\gamma/\beta},$$

with, $x > 0, \gamma > 0$ and $\beta > 0$.

In the following, Lemma presents an important result used in the derivation of the pdf for the MSHN distribution.

Lemma A1. Prudnikov et al. [18], Equation (2.3.1.13) For $\gamma > 0, a > 0, r > 0$ and $s > 0$. Then

$$\int_0^\infty x^{\gamma-1} \exp(-ax^r - sx) dx = N(\gamma, a, r, s), \tag{A1}$$

where

$$N = \begin{cases} \sum_{j=0}^{q-1} \frac{(-a)^j}{j! s^{\gamma+rj}} \Gamma(\gamma+rj) {}_{p+1}F_q(1, \Delta(p, \gamma+rj); \Delta(q, 1+j); (-1)^q z), & \text{if } 0 < r < 1 \\ \sum_{h=0}^{p-1} \frac{(-s)^h}{rh! a^{(\gamma+h)/r}} \Gamma\left(\frac{\gamma+h}{r}\right) {}_{q+1}F_p(1, \Delta(q, \frac{\gamma+h}{r}); \Delta(p, 1+h); \frac{(-1)^p}{z}), & \text{if } r > 1 \\ \frac{\Gamma(\gamma)}{(a+s)^\gamma}, & \text{if } r = 1, \end{cases}$$

considering $\gamma = p/q, p \geq 1$ and $q \geq 1$ are coprime integers, where $z = (\frac{p}{s})^p (\frac{a}{q})^q, \Delta(k, a) = \frac{a}{k}, \frac{(a+1)}{k}, \dots, \frac{(a+k-1)}{k}$ and ${}_pF_q(\dots)$ is the generalized hypergeometric function defined by

$${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; x) = \sum_{k=0}^\infty \frac{(a_1)_k (a_2)_k \dots (a_p)_k x^k}{(b_1)_k (b_2)_k \dots (b_q)_k k!}$$

where $(c)_k = c(c+1) \dots (c+k-1)$.

References

1. Cooray, K.; Ananda, M.M.A. A Generalization of the Half-Normal Distribution with Applications to Lifetime Data. *Commun. Stat. Theory Methods* **2008**, *37*, 1323–1337. [CrossRef]
2. Cordeiro, G.M.; Pescim, R.R.; Ortega, E.M.M. The Kumaraswamy Generalized Half-Normal Distribution for Skewed Positive Data. *J. Data Sci.* **2012**, *10*, 195–224.
3. Gómez, Y.M.; Bolfarine, H. Likelihood-based inference for power half-normal distribution. *J. Stat. Theory Appl.* **2015**, *14*, 383–398. [CrossRef]
4. Gómez, Y.M.; Vidal, I. A generalization of the half-normal distribution. *Appl. Math. J. Chin. Univ.* **2016**, *31*, 409–424. [CrossRef]
5. Olmos, N.M.; Varela, H.; Gómez, H.W.; Bolfarine, H. An extension of the half-normal distribution. *Stat. Pap.* **2012**, *53*, 875–886. [CrossRef]
6. Rogers, W.H.; Tukey, J.W. Understanding some long-tailed symmetrical distributions. *Stat. Neerl.* **1972**, *26*, 211–226. [CrossRef]
7. Mosteller, F.; Tukey, J.W. *Data Analysis And Regression*; Addison-Wesley: Reading, MA, USA, 1977.
8. Reyes, J.; Gómez, H.W.; Bolfarine, H. Modified slash distribution. *Statistics* **2013**, *47*, 929–941. [CrossRef]
9. Lehmann, E.L. *Elements of Large-Sample Theory*; Springer: New York, NY, USA, 1999.

10. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. Ser. B* **1977**, *39*, 1–38. [[CrossRef](#)]
11. Lee, S.Y.; Xu, L. Influence analyses of nonlinear mixed-effects models. *Comput. Stat. Data Anal.* **2004**, *45*, 321–341. [[CrossRef](#)]
12. Lin, T.I.; Lee, J.C.; Yen, S.Y. Finite mixture modeling using the skew-normal distribution. *Stat. Sin.* **2007**, *17*, 909–927.
13. Lyu, M. *Handbook of Software Reliability Engineering*; IEEE Computer Society Press: Washington, DC, USA, 1996.
14. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Auto. Contr.* **1974**, *19*, 716–723. [[CrossRef](#)]
15. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [[CrossRef](#)]
16. Von Alven, W.H. *Reliability Engineering by ARINC*; Prentice-Hall, Inc.: Upper Saddle River, NJ, USA, 1964.
17. Linhart, H.; Zucchini, W. *Model Selection*; Wiley Series in Probability and Statistics; Wiley: Hoboken, NJ, USA, 1986.
18. Prudnikov, A.P.; Brychkov, Y.A.; Marichev, O.I. *Integrals and Series*; Gordon & Breach Science Publishers: Amsterdam, The Netherlands, 1986.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Article

An Asymmetric Bimodal Distribution with Application to Quantile Regression

Yolanda M. Gómez ¹, Emilio Gómez-Déniz ², Osvaldo Venegas ^{3,*}, Diego I. Gallardo ¹ and Héctor W. Gómez ⁴

¹ Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Copiapó 1530000, Chile

² Department of Quantitative Methods in Economics and TIDES Institute, University of Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Spain

³ Departamento de Ciencias Matemáticas y Físicas, Facultad de Ingeniería, Universidad Católica de Temuco, Temuco 4780000, Chile

⁴ Departamento de Matemática, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta 1240000, Chile

* Correspondence: ovenegas@uct.cl; Tel.: +56-2205411

Received: 27 May 2019; Accepted: 25 June 2019; Published: 10 July 2019

Abstract: In this article, we study an extension of the sinh Cauchy model in order to obtain asymmetric bimodality. The behavior of the distribution may be either unimodal or bimodal. We calculate its cumulative distribution function and use it to carry out quantile regression. We calculate the maximum likelihood estimators and carry out a simulation study. Two applications are analyzed based on real data to illustrate the flexibility of the distribution for modeling unimodal and bimodal data.

Keywords: asymmetric bimodal distribution; bimodal; maximum likelihood

1. Introduction

It frequently occurs in real life that we find continuous data that are bimodal; these cannot be modeled by known unimodal distributions. It is therefore of interest to investigate more flexible distributions in modes that will be useful for professionals working in different areas of knowledge.

In unimodal distributions, the flexibility is based on the asymmetry and kurtosis of the data. In this context, Azzalini [1] introduced the skew-normal (SN) distribution, with asymmetry parameter λ . It has a probability density function (pdf) given by

$$f(y; \mu, \sigma, \lambda) = \frac{2}{\sigma} \phi\left(\frac{y - \mu}{\sigma}\right) \Phi\left(\lambda \frac{(y - \mu)}{\sigma}\right), \quad y, \mu, \lambda \in \mathbb{R}, \sigma > 0, \quad (1)$$

where ϕ and Φ denote, respectively, the density and cumulative distribution functions of the $N(0, 1)$ distribution. This is denoted as $Y \sim SN(\lambda)$. $SN(0)$ becomes the standard normal distribution.

Bimodal distributions generated from skew distributions can be found in Ma and Genton [2], Kim [3], Lin et al. [4,5], Elal-Olivero et al. [6], Arnold et al. [7], Arnold et al. [8], and Venegas et al. [9], among others. The importance of studying these distributions is based on the fact that they do not have identifiability problems and can be used as alternative parametric models to replace the use of mixtures of distributions that present estimation problems from either the classical or the Bayesian point of view (see McLachlan and Peel [10]; Marin et al. [11]). One difficulty with these distributions is that in general, there is no closed-form expression for their cumulative distribution function (cdf). This makes it more difficult to generate data from these distributions for simulation studies or to carry out quantile regression. Additionally, many such bimodal distributions have complicated expressions for a general quantile (say, the q -th).

A variety of bimodal data sets and appropriate models have been presented by many authors. For example, Cobb et al. [12] used the quartic exponential density presented by Fisher [13] to model crude birth rates data; Rao et al. [14] used a bimodal distribution to analyze fish length data; Famoye et al. [15] used the beta-normal distribution to analyze egg diameter data; Everitt and Hand [16] discussed some mixture distributions for modeling bimodal data; Chatterjee et al. [17] and Weisberg [18] presented two bimodal data sets on the eruption and interruption times of the Old Faithful geyser; Bansal et al. [19] discussed the bimodality of quantum dot size distribution; Famoye et al. [15] cited a variety of bimodal distributions that arise from different areas of science. On the other hand, the sinh Cauchy (SC) distribution is given by

$$f(z; \Lambda) = \frac{\lambda \cosh(z)}{\sigma \pi (1 + \{\lambda \sinh(z)\}^2)},$$

where $\Lambda = (\lambda, \mu, \sigma)$, $z = \frac{y-\mu}{\sigma}$, $z \in \mathbb{R}$, $\mu \in \mathbb{R}$ is a location parameter, $\sigma > 0$ is a scale parameter, and $\lambda > 0$ is a symmetric parameter. The SC distribution produces unimodal and bimodal densities. The disadvantage of the SC distribution is that it is symmetric, which limits it to modeling only symmetric bimodal data. The main objective of this article is therefore to study a bimodal skew-symmetric model with closed cdf, in order to apply it to quantile regression. To do this, we used an extension of the SC distribution that we call the gamma–sinh Cauchy (GSC) distribution, which presents flexibility in its modes and also closed-form expression in its cdf. The GSC distribution belongs to the (gamma-G generator) family introduced by Zografos and Balakrishnan [20]. For any baseline cdf $G(y; \Lambda)$, $x \in \mathbb{R}$, they defined the gamma-G generator by the pdf and cdf given by

$$f(y; \phi, \Lambda) = \frac{g(y; \Lambda)}{\Gamma(\phi)} \{-\log [1 - G(y; \Lambda)]\}^{\phi-1}, \tag{2}$$

and

$$F(y; \phi, \Lambda) = \frac{\gamma(-\log[1 - G(y; \Lambda)], \phi)}{\Gamma(\phi)} = \frac{1}{\Gamma(\phi)} \int_0^{-\log[1 - G(y; \Lambda)]} u^{\phi-1} e^{-u} du, \tag{3}$$

respectively, where $\phi > 0$ is a skewness parameter, Λ is a vector of parameters, $g(y) = \frac{d}{dy}G(y)$, $\gamma(y, a) = \int_0^y t^{a-1} e^{-t} dt$ is the incomplete gamma function, and $\Gamma(a) = \gamma(+\infty, a)$ is the usual gamma function. We remark that in the literature, there are many models that can accommodate bimodal distributions. However, in only a few of them do the parameters have an interpretation in terms of measures of central tendency (mean, median, for instance) or a general q -th quantile. As we will show in Section 3, the main advantage of the GSC is that the location parameter represents the respective q -th quantile under a certain restriction over ϕ , which is very convenient for the use of this model in a quantile regression framework.

The paper is organized as follows. Section 2 develops the GSC distribution, its basic properties, and quantile regression. In Section 3, we perform a small-scale simulation study of the maximum likelihood (ML) estimators for parameters. Two applications to real data are discussed in Section 4, which illustrate the usefulness of the proposed model. Finally, conclusions are given in Section 5.

2. Gamma–Sinh Cauchy Distribution

The GSC distribution is obtained considering G in (2) as the cdf of the SC distribution. The pdf can be written as

$$f(z; \Theta) = \frac{\lambda \cosh(z)}{\sigma \pi \Gamma(\phi) (1 + \{\lambda \sinh(z)\}^2)} \left\{ -\log \left[0.5 - \frac{1}{\pi} \arctan \{\lambda \sinh(z)\} \right] \right\}^{\phi-1}, \tag{4}$$

where $\Theta = (\phi, \lambda, \mu, \sigma)$, and $\phi > 0$ is an asymmetric parameter. We denoted this by $Z \sim GSC(\phi, \lambda, \mu, \sigma)$.

The cdf is given by

$$F(z; \phi, \lambda, \mu, \sigma) = \frac{1}{\Gamma(\phi)} \gamma \left(-\log \left[0.5 - \frac{1}{\pi} \arctan \{ \lambda \sinh(z) \} \right], \phi \right). \tag{5}$$

Particular cases:

1. $\phi = 1$ SC distribution,
2. $\phi = 1, \lambda = 1$ hyperbolic secant distribution (Talacko [21]).

The following proposition states conditions for the symmetry of the GSC distribution.

Proposition 1. *The density of the GSC($\phi, \lambda, \mu, \sigma$) model is symmetric if and only if $\phi = 1$.*

Proof. Without loss of generality, we consider $\mu = 0$ and $\sigma = 1$. For $\phi = 1$, the density of the model is

$$f(y; 1, \lambda, 0, 1) = \frac{\lambda \cosh(y)}{\sigma \pi \Gamma(\phi) (1 + \{ \lambda \sinh(y) \}^2)}.$$

This function is clearly even because $\cosh(y)$ and $\sinh(y)^2$ are even. To prove the reciprocal, we will argue by contradiction. Let $\phi_0 \neq 1$ such that the density is symmetric, i.e. $f(y; \phi_0, \lambda, 0, 1) = f(-y; \phi_0, \lambda, 0, 1), \forall y \in \mathbb{R}, \forall \lambda > 0$. This implies that

$$\left(\frac{\log \left[\frac{1}{2} + \frac{1}{\pi} \arctan \{ \lambda \sinh(y) \} \right]}{\log \left[\frac{1}{2} - \frac{1}{\pi} \arctan \{ \lambda \sinh(y) \} \right]} \right)^{\phi_0 - 1} = 1.$$

From the latter equality, and jointly with the fact that the logarithmic function is injective, we find that $\arctan(\lambda \sinh(y)) = 0, \forall y \in \mathbb{R}$, which implies that $\lambda = 0$, producing a contradiction. \square

The unimodal and bimodal regions for $GSC(\phi, \lambda, \mu, \sigma)$ are illustrated in Figure 1. We can see that for all ϕ , there is λ such that GSC is bimodal. Figure 2 shows the density function for some values of the parameters ϕ and λ , considering the location and scale parameters fixed at 0 and 1, respectively. The distribution assumes symmetric unimodal and bimodal shapes and asymmetric unimodal and bimodal shapes. Figure 3 shows the skewness and kurtosis coefficients for the GSC model under different values of λ and ϕ (such coefficients do not depend on μ and σ). As illustrated previously, the model can assume positive and negative values for the skewness coefficient and can also accommodate kurtosis coefficients lower than, equal to, and greater than the normal model ($<3, =3$ and >3 , respectively).

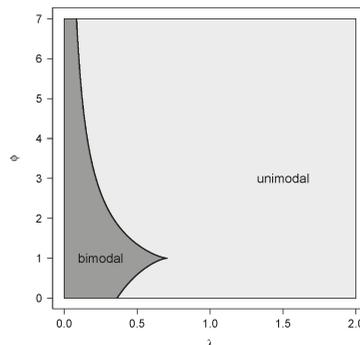


Figure 1. Unimodal and bimodal regions for $GSC(\phi, \lambda, \mu, \sigma)$.

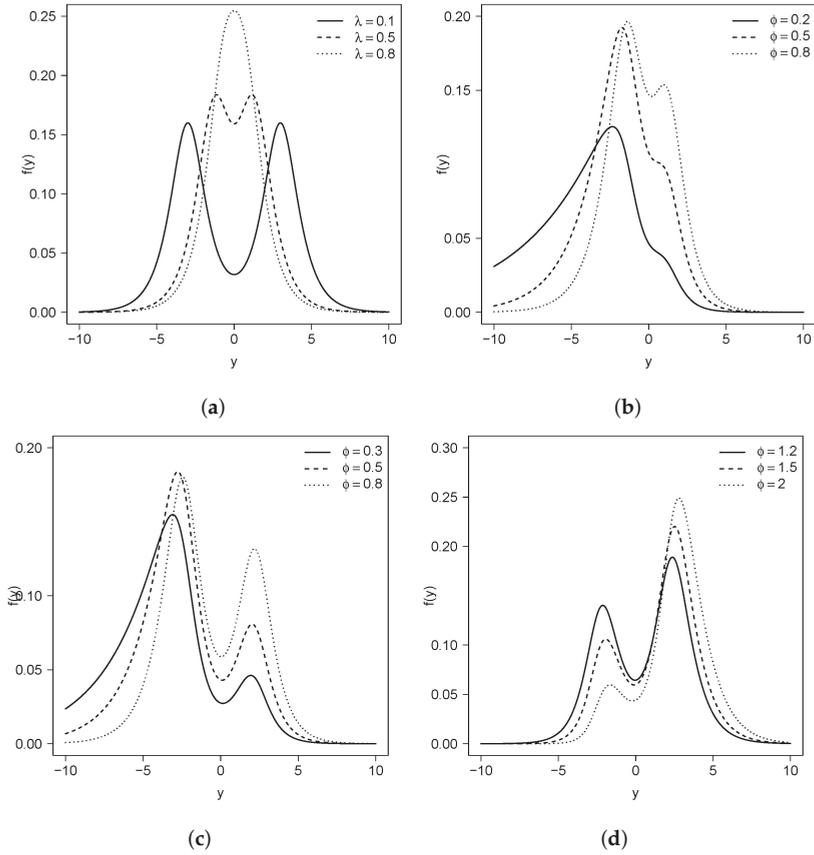


Figure 2. Plots for the gamma–sinh Cauchy (GSC) model for different values of the parameters with $\mu = 0, \sigma = 1$ (a) $\phi = 1$ (b) $\lambda = 0.5$, and (c,d) $\lambda = 0.2$

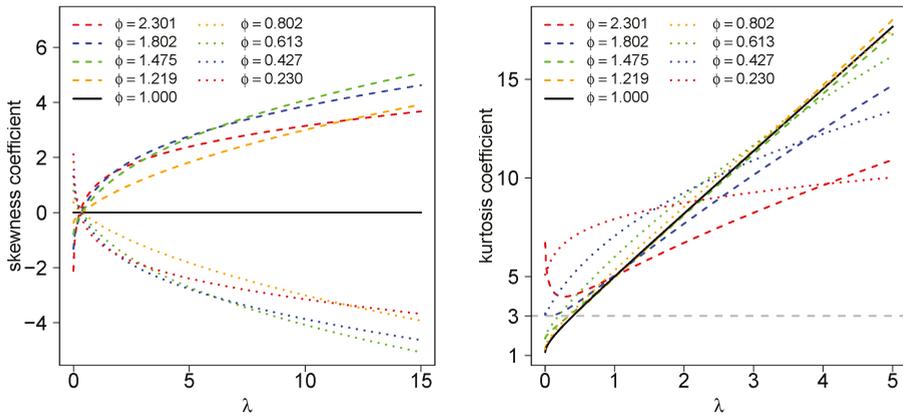


Figure 3. Skewness and kurtosis coefficients for the GSC $(\phi, \lambda, \mu, \sigma)$ model with different values for λ and ϕ .

The GSC Model for Quantile Regression

From Equation (5), it follows that the cdf of the GSC distribution evaluated in μ is given by

$$P(Y \leq \mu; \phi, \lambda, \mu, \sigma) = F(\mu; \phi, \lambda, \mu, \sigma) = \frac{\gamma(\log(2), \phi)}{\Gamma(\phi)} = G(\log(2), \phi), \tag{6}$$

where $G(y, a) = \gamma(y, a) / \Gamma(a)$ corresponds to the cdf of the gamma distribution with shape and scale parameters a and 1, respectively. Note that $G(\log(2), \phi)$ depends only on ϕ (and not on σ or λ). As $G(\log(2), \phi)$ is an increasing function in terms of ϕ and $G(\cdot, \phi) \in (0, 1)$, the equation $G(\log(2), \phi) = q$ has an unique solution for $q \in (0, 1)$, which can also be written as

$$\frac{1}{\Gamma(\phi)} \int_0^{\log(2)} u^{\phi-1} e^{-u} du = q. \tag{7}$$

Equation (7) can be solved numerically. For instance, in R the `uniroot` function can be used. Table 1 shows some values for $\phi(q)$ with different values for q .

Table 1. Some values for $\phi(q)$ in terms of q .

q	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\phi(q)$	2.301	1.802	1.475	1.219	1.000	0.802	0.613	0.427	0.230

For this reason, for a fixed q , if we take $\phi = \phi(q)$ satisfying (7), by (6) the parameter μ directly represents the q -th quantile, allowing regression to be performed conveniently even though μ . Under this setting, a set of available p -covariates, say $\mathbf{x}_i^\top = (x_{i1}, \dots, x_{ip})$, for $i = 1, \dots, n$, can be introduced as follows:

$$\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}, \quad i = 1, \dots, n.$$

This is a convenient property of the GSC distribution because it provides a simple way to performing quantile regression in a model that can be unimodal or bimodal, depending only on parameter λ (because $\phi = \phi(q)$ is considered as fixed in this setting).

As far as we know, there is no model in the literature that is parameterized conveniently in terms of the q -th quantile and can also be unimodal or bimodal. Figure 1 shows that for any $\phi = \phi(q)$ fixed in the GSC, there is an interval $\Lambda_{\phi(q)}^{(1)}$ for λ where the distribution is unimodal and an interval $\Lambda_{\phi(q)}^{(2)}$ where the distribution is bimodal.

3. ML Estimation for the GSC Distribution

3.1. ML Estimation

Consider y_1, y_2, \dots, y_n as a size n random sample from the pdf $GSC(\phi, \lambda, \mu, \sigma)$. Hence, the log-likelihood function is given by

$$\begin{aligned} l(\phi, \lambda, \mu, \sigma) &= n \log(\lambda) + \sum_{i=1}^n \log\{\cosh(z_i)\} - \sum_{i=1}^n \log\left\{1 + \lambda^2 \sinh^2(z_i)\right\} \\ &\quad + (\phi - 1) \sum_{i=1}^n \log\left\{-\log\left[\frac{1}{2} - \frac{1}{\pi} \arctan\{\lambda \sinh(z_i)\}\right]\right\} \\ &\quad - n \log(\sigma) - n \log(\pi) - n \log(\Gamma(\phi)), \end{aligned} \tag{8}$$

where $z_i = \frac{y_i - \mu}{\sigma}$. To compute the ML estimation for $\Theta = (\phi, \lambda, \mu, \sigma)$, (8) must be maximized. That is, we have to solve the following system of equations: $\frac{\partial l}{\partial \phi}$, $\frac{\partial l}{\partial \lambda}$, $\frac{\partial l}{\partial \mu}$, and $\frac{\partial l}{\partial \sigma}$. More precisely, we have to solve

$$\begin{aligned} \sum_{i=1}^n \log(-\log(t_i)) - n\Psi(\phi) &= 0, \\ \sum_{i=1}^n \frac{2\lambda \sinh^2(z_i)}{1 + \lambda^2 \sinh^2(z_i)} + (\phi - 1) \sum_{i=1}^n \frac{\sinh(z_i)}{\pi t_i (1 + \lambda^2 \sinh^2(z_i)) \log(t_i)} &= \frac{n}{\lambda}, \\ \sum_{i=1}^n \frac{\lambda^2 \sinh(2z_i)}{1 + \lambda^2 \sinh^2(z_i)} + \frac{\lambda(\phi - 1)}{\pi} \sum_{i=1}^n \frac{\cosh(z_i)}{t_i (1 + \lambda^2 \sinh^2(z_i)) \log(t_i)} &= \sum_{i=1}^n \tanh(z_i), \\ \sum_{i=1}^n \frac{\lambda^2 z_i \sinh(2z_i)}{1 + \lambda^2 \sinh^2(z_i)} + \frac{\lambda(\phi - 1)}{\pi} \sum_{i=1}^n \frac{z_i \cosh(z_i)}{t_i (1 + \lambda^2 \sinh^2(z_i)) \log(t_i)} - n &= \sum_{i=1}^n z_i \tanh(z_i), \end{aligned}$$

where $t_i = \frac{1}{2} - \frac{1}{\pi} \arctan\{\lambda \sinh(z_i)\}$, and $\Psi(\cdot)$ is the digamma function. The system of equations given above can be solved using numerical procedures such as the Newton–Raphson procedure. An alternative is to use the NumDeriv routine with the R software (R Core Team [22]).

3.2. Simulation Study

In this Section, we present a brief simulation study to assess the performance of MLE in the GSC model. To draw values from the model, we used the inversion method. If $U \sim U(0, 1)$, then

$$Z = \mu + \sigma \sinh^{-1} \left(\lambda^{-1} \tan \left[\pi/2 - \pi \exp \left(-G^{-1}(U; \phi) \right) \right] \right) \sim GSC(\phi, \lambda, \mu, \sigma),$$

where $G^{-1}(\cdot; \phi)$ is the inverse of the cdf of the gamma distribution with shape and rate parameters equal to ϕ and 1, respectively. In all scenarios, we considered $\mu = 0$ and $\sigma = 1$, three values for ϕ —0.75, 1, and 1.5—and three values for λ —0.5, 1, and 2. We also considered three sample sizes: 100, 200, and 500. For each combination of the parameters, we drew 1,000 samples and computed the ML estimates. Table 2 summarizes the results considering the average of the bias (bias), the root of the estimated mean squared error (RMSE), and the 95% coverage probability (CP). Note that in all cases, the bias and RMSE decreased when the sample size was increased, suggesting that the estimators are consistent. Finally, we also remark that the coverage probability converged to the nominal values used for the construction of the confidence intervals when the sample size was increased, suggesting that the normality for the ML estimates is reasonable in sample sizes.

Table 2. Simulation study for the GSC($\phi, \lambda, \mu = 0, \sigma = 1$) model.

ϕ	λ	parameter	$n = 100$			$n = 200$			$n = 500$		
			bias	RMSE	CP	bias	RMSE	CP	bias	RMSE	CP
0.75	0.5	μ	−0.094	0.663	0.846	−0.055	0.578	0.900	−0.014	0.457	0.932
		σ	−0.037	0.392	0.880	−0.014	0.337	0.904	−0.006	0.268	0.933
		λ	−0.031	0.372	0.869	−0.013	0.317	0.905	−0.006	0.252	0.930
	1.0	ϕ	0.044	0.411	0.867	0.023	0.349	0.907	0.006	0.272	0.934
		μ	−0.015	0.607	0.831	0.021	0.532	0.880	0.016	0.428	0.925
		σ	−0.035	0.460	0.848	−0.024	0.396	0.886	−0.009	0.320	0.927
		λ	−0.004	0.544	0.831	−0.010	0.459	0.889	−0.003	0.365	0.927
		ϕ	0.018	0.424	0.843	−0.004	0.363	0.890	−0.006	0.290	0.926
		μ	0.011	0.361	0.934	0.004	0.298	0.944	0.001	0.233	0.944
2.0	σ	−0.011	0.504	0.911	−0.007	0.419	0.932	−0.003	0.333	0.945	
	λ	0.055	0.789	0.928	0.017	0.650	0.939	0.010	0.513	0.945	
	ϕ	−0.007	0.337	0.940	−0.003	0.279	0.949	−0.001	0.220	0.947	

Table 2. Cont.

ϕ	λ	parameter	$n = 100$			$n = 200$			$n = 500$			
			bias	RMSE	CP	bias	RMSE	CP	bias	RMSE	CP	
1.0	0.5	μ	-0.060	0.666	0.846	-0.036	0.580	0.899	-0.012	0.460	0.933	
		σ	-0.044	0.373	0.877	-0.022	0.318	0.915	-0.009	0.253	0.933	
		λ	-0.033	0.366	0.867	-0.018	0.308	0.913	-0.007	0.245	0.936	
	1.0	ϕ	0.045	0.458	0.874	0.023	0.388	0.909	0.007	0.304	0.938	
		μ	-0.052	0.622	0.816	-0.024	0.543	0.881	-0.002	0.428	0.940	
		σ	-0.040	0.431	0.832	-0.024	0.367	0.887	-0.006	0.293	0.936	
	2.0	λ	ϕ	-0.017	0.535	0.811	-0.015	0.448	0.875	0.000	0.354	0.934
			μ	0.060	0.493	0.842	0.027	0.417	0.890	0.005	0.324	0.946
			σ	-0.001	0.357	0.937	0.000	0.291	0.946	-0.001	0.229	0.951
		ϕ	μ	-0.014	0.494	0.896	-0.006	0.413	0.929	0.000	0.328	0.942
			σ	0.033	0.795	0.916	0.017	0.658	0.937	0.011	0.521	0.944
			λ	0.007	0.372	0.942	0.002	0.303	0.947	0.002	0.238	0.949
1.5	0.5	μ	0.015	0.683	0.850	0.014	0.597	0.894	0.008	0.480	0.925	
		σ	-0.045	0.354	0.890	-0.021	0.300	0.926	-0.009	0.239	0.935	
		λ	-0.022	0.366	0.877	-0.014	0.308	0.916	-0.006	0.245	0.937	
	1.0	ϕ	0.026	0.533	0.876	0.008	0.455	0.906	0.002	0.364	0.930	
		μ	-0.134	0.688	0.835	-0.138	0.612	0.856	-0.076	0.492	0.904	
		σ	-0.038	0.413	0.836	-0.025	0.347	0.856	-0.011	0.275	0.896	
	2.0	λ	ϕ	0.091	0.581	0.840	-0.001	0.459	0.860	-0.010	0.360	0.896
			μ	0.211	0.681	0.886	0.166	0.579	0.880	0.076	0.446	0.912
			σ	-0.029	0.383	0.933	-0.009	0.309	0.945	-0.004	0.241	0.948
		ϕ	μ	-0.014	0.472	0.895	-0.007	0.395	0.927	-0.003	0.313	0.942
			σ	0.065	0.784	0.932	0.023	0.646	0.942	0.010	0.510	0.952
			λ	0.070	0.484	0.948	0.023	0.379	0.954	0.009	0.294	0.950

4. Applications

In this section, we carry out two applications to real data, the first using the GSC model without covariates and the second applying quantile regression to uni- and bimodal data.

4.1. Application 1: Without Covariates

The first application reported is for the data set consisting of 1150 heights measured at 1 micron intervals along the drum of a roller (i.e., parallel to the axis of the roller). This was part of an extensive study of roller surface roughness. It is available for downloading at <http://lib.stat.emu.edu/jasadata/laslett>. Table 3 presents summary statistics for the data set where b_1 and b_2 correspond to the sample asymmetry and kurtosis coefficients, respectively.

Table 3. Descriptive statistics for the data set

n	\bar{y}	s^2	b_1	b_2
1150	3.535	0.422	-0.986	4.855

We fitted this using the SN model, the exponentiated sinh Cauchy (ESC) model (see Cooray [23]), and the GSC model. A summary of these fits is presented in Table 4. Based on the AIC criteria, the GSC provided the best fit for the height data set. Figure 4 shows plots of the density functions for the fitted models using the MLEs for SN, ECG and GSC distributions.

Table 4. Maximum likelihood (ML) estimates for the GSC, exponentiated sinh Cauchy (ECG), and skew-normal (SN) models for the roller data set.

Parameter	GSC	ECG	SN
μ	4.1115 (0.0388)	4.0460 (0.0482)	4.2475 (0.0276)
σ	0.2053 (0.0172)	0.1903 (0.0205)	0.9644 (0.0286)
λ	0.5353 (0.0682)	0.5535 (0.0860)	−2.7578 (0.2426)
ϕ	0.3621 (0.0373)	0.3322 (0.0445)	–
log-likelihood	−1053.8	−1056.0	−1071.3
AIC	2115.7	2119.9	2148.7

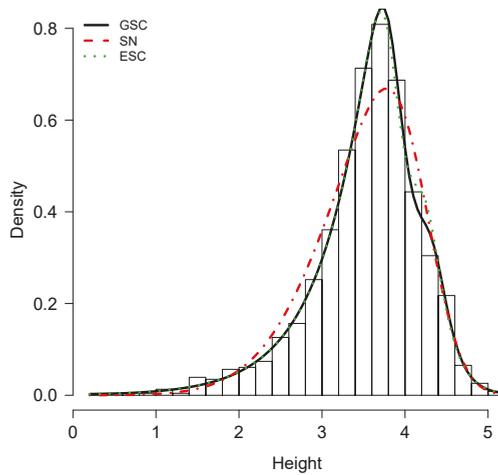


Figure 4. Fitted models for roller data set.

4.2. Data Set 2: Quantile Regression to Bimodal Data

The second application we consider is the Australian data set available in the *sn* package in R. This data set is related to 102 male and 100 female athletes collected at the Australian Institute of Sport. The linear model considered is

$$Bfat_i = \beta_0(q) + \beta_1(q)bmi_i + \beta_2(q)lbi_i + \epsilon_i(q), \quad i = 1, \dots, 202, \tag{9}$$

where $Bfat_i$ is the body fat percentage for the i -th athlete and bmi_i and lbi_i are the covariates body mass index and lean body mass for the i -th athlete, respectively. In addition, $\epsilon_i(q) \sim GSC(0, \sigma, \lambda, \phi(q))$, $\phi(q)$ satisfies Equation (7), and $q \in (0, 1)$ is the fixed quantile that is being modeled. This data set was also analyzed in Martínez-Flórez et al. [24] using a bimodal regression model. However, the authors modeled the mean of the distribution. In our approach, we model the 0.1, 0.25, 0.5, 0.75, and 0.9 percentiles of the distribution, which provides a more informative scenario to explain body fat in terms of the body mass index and lean body mass. Our approach is compared with the skewed Laplace (SKL) and skewed Student-t (SKT) models discussed in Galarza et al. [25], where the authors proposed a flexible model in a quantile regression model context. Table 5 shows the AIC for those models considering different quantiles. We also present the p -value for the Kolmogorov–Smirnov (K–S) test of the hypothesis that the respective quantile residuals came from the standard normal distribution. P -values greater than 5% suggest that with this significance level, the standard normal assumption is reasonable for those residuals, in which case the model would be appropriate for this

data set. Note that based on the AIC criteria, the GSC presents a better fit for this data set, except for the median regression ($q = 0.5$). On the other hand, based on the p -value for the K–S test applied to the quantile residuals, we conclude that GSC, SKL, and SKT are appropriate models for $q = 0.25$ and $q = 0.5$ (the GSC and SKT provide a better fit based on the AIC criteria). However, for $q = 0.1$ and 0.75 , GSC provides the better fit because the p -values are (significantly) greater than 0.05. Finally, for $q = 0.9$ no model seems appropriate, but based on the p -values, the GSC provides a better fit than SKL and SKT distributions.

Table 5. AIC and p -value for K–S test in the ais data set for the GSC, skewed Laplace (SKL), and skewed Student-t (SKT) models and different quantiles.

q	AIC			p -value for K–S test		
	GSC	SKL	SKT	GSC	SKL	SKT
0.10	1156.54	1194.28	1166.74	0.79	0.06	0.02
0.25	1160.72	1172.70	1153.15	0.65	0.71	0.96
0.50	1162.74	1182.66	1161.80	0.32	0.13	0.31
0.75	1159.50	1221.65	1200.87	0.87	<0.001	0.02
0.90	1211.71	1280.45	1253.48	0.02	<0.001	<0.001

Figure 5 shows the regression coefficients for the quantile regression presented in Equation (9) and their respective 95% confidence intervals. Note that body mass index and lean body mass are significant in explaining all the quantiles modeled.

Figure 6 shows the profile density for the q -th quantile of body fat percentage for $q = 0.1$ and $q = 0.75$. Note that the distribution of the 0.1 quantile is unimodal, and the distribution of the 0.9 quantile is bimodal.

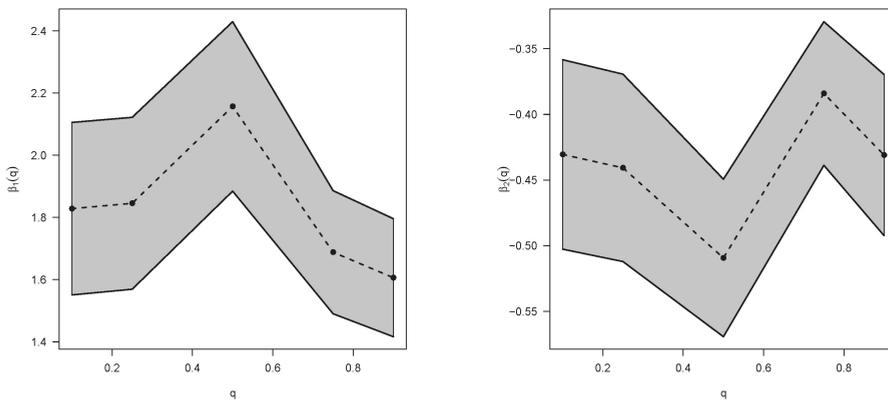


Figure 5. Estimates for regression coefficients (and 95% confidence interval)s for variables bmi (left panel) and lbm (right panel) in different quantile regression models with quantiles equal to 0.1, 0.25, 0.5, 0.75, and 0.9 and response variable Bfat.

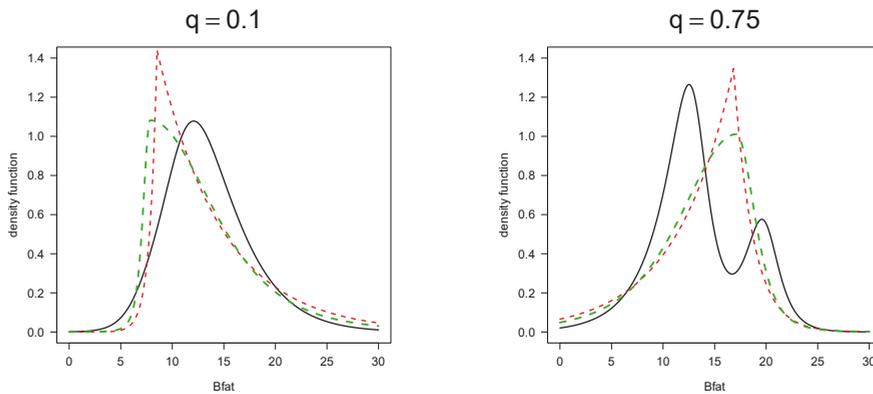


Figure 6. Distribution for 0.1 and 0.75 quantiles of body fat percentage considering body mass index and lean body mass equal to 22.96 and 64.87, respectively. Curves in black, red, and green represent the density functions estimated by the GSC, SKL, and SKT models, respectively.

5. Final Comments

This paper proposes the GSC distribution, which is flexible in its modes and contains the SC distribution as a special case. We implemented ML estimation in a quantile regression, obtaining significant results. In the applications to real data it performs very well, better than potential rival models. Some further characteristics of the GSC distribution are:

- The GSC distribution contains the SC and hyperbolic secant models as special cases.
- The GSC distribution presents great flexibility in its modes, as can be observed in Figure 1.
- The proposed model has a closed-form expression for its cdf.
- In the two applications, we show that the GSC model fits better than the other models.

Author Contributions: All of the authors contributed significantly to this research article.

Funding: The research of Yolanda M. Gómez was supported by proyecto DIUDA programa de inserción No. 22367 of the Universidad de Atacama. The research of Héctor W. Gómez was supported by Grant SEMILLERO UA-2019 (Chile).

Acknowledgments: The authors would like to thank the editor and the anonymous referees for their comments and suggestions, which significantly improved our manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
2. Ma, Y.; Genton, M.G. Flexible class of skew-symmetric distributions. *Scand. J. Stat.* **2004**, *31*, 459–468. [[CrossRef](#)]
3. Kim, H.J. On a class of two-piece skew-normal distributions. *Statistics* **2005**, *39*, 537–553. [[CrossRef](#)]
4. Lin, T.I.; Lee, J.C.; Hsieh, W.J. Robust mixture models using the skew-t distribution. *Stat. Comput.* **2007**, *17*, 81–92. [[CrossRef](#)]
5. Lin, T.I.; Lee, J.C.; Yen, S.Y. Finite mixture modeling using the skew-normal distribution. *Stat. Sin.* **2007**, *17*, 909–927.
6. Elal-Olivero, D.; Gómez, H.W.; Quintana, F.A. Bayesian Modeling using a class of Bimodal skew-Elliptical distributions. *J. Stat. Plan. Infer.* **2009**, *139*, 1484–1492. [[CrossRef](#)]
7. Arnold, B.C.; Gómez, H.W.; Salinas, H.S. On multiple constraint skewed models. *Statistics* **2009**, *43*, 279–293. [[CrossRef](#)]

8. Arnold, B.C.; Gómez, H.W.; Salinas, H.S. A doubly skewed normal distribution. *Statistics* **2015**, *49*, 842–858. [[CrossRef](#)]
9. Venegas, O.; Salinas, H.S.; Gallardo, D.I.; Bolfarine, H.; Gómez, H.W. Bimodality based on the generalized skew-normal distribution. *J. Stat. Comput. Simul.* **2018**, *88*, 156–181. [[CrossRef](#)]
10. McLachlan, G.J.; Peel, D. *Finite Mixture Models*; Wiley Interscience: New York, NY, USA, 2000.
11. Marin, J.M.; Mengersen, K.; Robert, C. Bayesian modeling and inference on mixtures of distributions. *Handbook Stat.* **2005**, *25*, 459–503.
12. Cobb, L.; Koppstein, P.; Chen, N.H. Estimation and moment recursion relations for multimodal distributions of the exponential families. *J. Amer. Stat. Assoc.* **1983**, *78*, 124–130. [[CrossRef](#)]
13. Fisher, R.A. On the Mathematical Foundations of Theoretical Statistics. *Philos. Trans. R. Soc. London Ser. A* **1922**, *222*, 309–368. [[CrossRef](#)]
14. Rao, K.S.; Narayana, J.L.; Sastry, V.P. A bimodal distribution. *Bull. Calcutta Math. Soc.* **1988**, *80*, 238–240.
15. Famoye, F.; Lee, C.; Eugene, N. Beta-normal distribution: Bimodality properties and application. *J. Modern Appl. Stat. Method* **2004**, *3*, 85–103. [[CrossRef](#)]
16. Everitt, B.S.; Hand, D.J. *Finite Mixture Distributions*; Chapman & Hall: London, UK, 1981.
17. Chatterjee, S.; Handcock, M.S.; Simonoff, J.S. *A Casebook for a First Course in Statistics and Data Analysis*; John Wiley & Sons: New York, NY, USA, 1995.
18. Weisberg, S. *Applied Linear Regression*; John Wiley & Sons: New York, NY, USA, 2005.
19. Bansal, B.; Gokhale, M.R.; Bhattacharya, A.; Arora, B.M. InAs/InP quantum dots with bimodal size distribution: Two evolution pathways. *J. Appl. Phys.* **2007**, *101*, 1–6. [[CrossRef](#)]
20. Zografos, K.; Balakrishnan, N. On families of beta- and generalized gamma-generated distributions and associated inference. *Stat. Methodol.* **2009**, *6*, 344–362. [[CrossRef](#)]
21. Talacko, J. Perks' distributions and their role in the theory of Wiener's stochastic variables. *Trabajos de Estadística* **1956**, *7*, 159–174. [[CrossRef](#)]
22. Team, R.C. R: A Language and Environment for Statistical Computing. Available online: <http://www.R-project.org> (accessed on 26 May 2019).
23. Cooray, K. Exponentiated Sinh Cauchy distribution with applications. *Commun. Stat. Theor. Method* **2013**, *42*, 3838–3852. [[CrossRef](#)]
24. Martínez-Flórez, G.; Salinas, H.S.; Bolfarine, H. Bimodal regression model. *Colombian J. Stat.* **2017**, *40*, 65–83.
25. Galarza, C.E.; Lachos, V.H.; Barbosa, C.; Castro, L.M. Robust quantile regression using a generalized class of skewed distributions. *Stat* **2017**, *6*, 113–130.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Article

Univariate and Bivariate Models Related to the Generalized Epsilon–Skew–Cauchy Distribution

Barry C. Arnold ¹, Héctor W. Gómez ², Héctor Varela ^{2,*} and Ignacio Vidal ³¹ Statistics Department, University of California Riverside, Riverside, CA 92521, USA; barry.arnold@ucr.edu² Departamento de Matemática, Universidad de Antofagasta, Antofagasta 1240000, Chile; hector.gomez@uantof.cl³ Instituto de Matemática y Física, Universidad de Talca, Casilla 747, Talca, Chile; ivaldal@inst-mat.otalca.cl

* Correspondence: hector.varela@uantof.cl

Received: 8 May 2019; Accepted: 10 June 2019; Published: 14 June 2019

Abstract: In this paper, we consider a stochastic representation of the epsilon-skew-Cauchy distribution, viewed as a member of the family of skewed distributions discussed in Arellano-Valle et al. (2005). The stochastic representation facilitates derivation of distributional properties of the model. In addition, we introduce symmetric and asymmetric extensions of the Cauchy distribution, together with an extension of the epsilon-skew-Cauchy distribution. Multivariate versions of these distributions can be envisioned. Bivariate examples are discussed in some detail.

Keywords: Epsilon-skew-Normal; Epsilon-skew-Cauchy; bivariate densities; generalized Cauchy distributions

1. Introduction

Mudholkar and Hutson (2000) [1] studied an asymmetric normal distribution that they called the epsilon-skew-normal $\{ESN(\varepsilon); |\varepsilon| < 1\}$, with asymmetry or skewness parameter ε . When the parameter ε assumes the value 0, the distribution simplifies to become a standard normal distribution. The family thus consists of a parameterized set of usually asymmetric distributions that includes the symmetric standard normal density as a special case. Specifically, we say that $X \sim ESN(\varepsilon)$ if its density is of the form:

$$g(x; \varepsilon) = \phi\left(\frac{x}{1 - \operatorname{sgn}(x)\varepsilon}\right)$$

where $x \in \mathbb{R}$, ϕ is the standard normal density and $\operatorname{sgn}(\cdot)$ is the sign function.

Arellano-Valle et al. (2005) [2] discuss extension of this model, together with associated inference procedures. They consider a class of Epsilon-skew-symmetric distributions associated with a particular symmetric density $f(\cdot)$ that is indexed by an asymmetry parameter ε with densities given by

$$h(x; \varepsilon) = f\left(\frac{x}{1 - \operatorname{sgn}(x)\varepsilon}\right) \quad (1)$$

where $x \in \mathbb{R}$ and $|\varepsilon| < 1$.

If X has density of the form (1), then we say that X is an epsilon-skew-symmetric random variable and we write $X \sim ESf(\varepsilon)$. Arellano-Valle et al. (2005) [2] extend this family to the model epsilon-skew-exponential-power, a model that has major and minor asymmetry and kurtosis that the ESN model. On the other hand Gómez et al. (2007) [3] study the Fisher information matrix for epsilon-skew-t model, which was used before in the study a financial series by Hansen (1994) [4]; see also Gómez et al. (2008) [5]. Note that if in (1) we set $f(t) = 1/(\pi(1+t^2))$, we obtain the epsilon-skew-Cauchy model.

We will write $X \sim N(0, 1)$ to indicate that X has a standard normal distribution, and we will write $Y \sim HN(0, 1)$ to indicate that Y has a standard half-normal distribution, i.e., that $Y = |X|$ where $X \sim N(0, 1)$.

The distribution of the ratio X/Y of two random variables is of interest in problems in biological and physical sciences, econometrics, and ranking and selection. It is well known that if $X \sim N(0, 1)$, $Y_1 \sim N(0, 1)$ and $Y_2 \sim HN(0, 1)$ are independent, then the random variables X/Y_1 and X/Y_2 both have Cauchy distributions; see Johnson et al. (1994, [6] Chapter 16). Behboodian, et al. (2006) [7] and Huang and Chen (2007) [8] study the distribution of such quotients when the component random variables are skew-normal (of the form studied in Azzalini (1985) [9]). The principle objective of the present paper is to study the behavior of such quotients when the component random variables have epsilon-skew-normal distributions.

The paper is organized in the following manner. In Section 2, we describe a representation of the epsilon-skew-Cauchy model. In Section 3, we consider the distribution of the ratio of two independent random variables, one of which has an $ESN(\epsilon)$ distribution and the other a standard normal distribution. In addition, an extension of the epsilon-skew-Cauchy (ESC) distribution is introduced. Bivariate versions of these distributions are discussed in Section 4. Extensions to higher dimensions can be readily envisioned, but are not discussed here. In Section 5, some of the bivariate distributions introduced in this paper are considered as possible models for a particular real-world data set.

2. Representation of the ESC (Epsilon-Skew-Cauchy) Model

Proposition 1. *If $X \sim ESN(\epsilon)$ and $Y \sim HN(0, 1)$ are independent random variables, then $Z_1 \stackrel{d}{=} \frac{X}{Y}$ has an epsilon-skew-Cauchy distribution with asymmetry parameter ϵ and density given by*

$$f_{Z_1}(z; \epsilon) = \frac{1}{\pi \left[1 + \left(\frac{z}{1 - \text{sgn}(z)\epsilon} \right)^2 \right]},$$

where $z \in \mathbb{R}$, $|\epsilon| < 1$ and we write $Z_1 \sim ESC(\epsilon)$.

Proof. With the transformation $Z_1 = X/Y$ and $W = Y$, whose Jacobian $|J| = w$, we obtain

$$f_{Z_1, W}(z, w) = \frac{w}{\pi} \exp \left\{ -\frac{1}{2} \left[\frac{z^2}{(1 - \text{sgn}(z)\epsilon)^2} + 1 \right] w^2 \right\},$$

where $z \in \mathbb{R}, w > 0$.

It follows directly that

$$f_{Z_1}(z; \epsilon) = \int_0^\infty f_{Z_1, W}(z, w) dw = \frac{1}{\pi \left[1 + \left(\frac{z}{1 - \text{sgn}(z)\epsilon} \right)^2 \right]}.$$

□

Figure 1 depicts the behavior of the ESC density for a variety of parameter values.

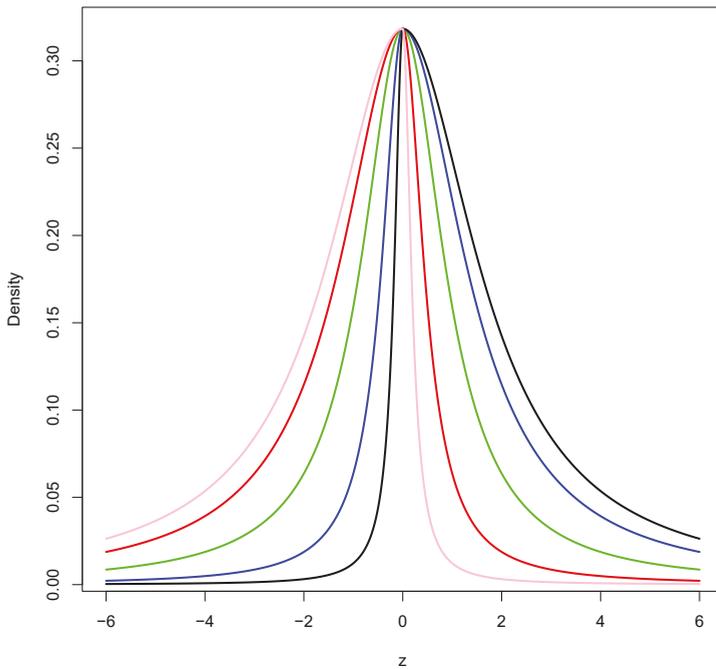


Figure 1. Examples of the ESC(ε) density for : $\varepsilon = 0$ (green line), $\varepsilon = -0.5$ (blue line), $\varepsilon = -0.8$ (black line), $\varepsilon = 0.5$ (red line) and $\varepsilon = 0.8$ (pink line).

3. Generalized Cauchy Distribution

Proposition 2. If $X \sim ESN(\varepsilon_1)$ and $Y \sim ESN(\varepsilon_2)$ are independent random variables, then $Z \stackrel{d}{=} \frac{X}{Y}$ has what we call a generalized epsilon-skew-Cauchy (GESC) distribution with parameters ε_1 and ε_2 and with density given by

$$f_Z(z; \varepsilon_1, \varepsilon_2) = \frac{1}{2\pi \left[\left(\frac{1}{1+\varepsilon_2} \right)^2 + \left(\frac{z}{1+\text{sgn}(z)\varepsilon_1} \right)^2 \right]} + \frac{1}{2\pi \left[\left(\frac{1}{1-\varepsilon_2} \right)^2 + \left(\frac{z}{1-\text{sgn}(z)\varepsilon_1} \right)^2 \right]},$$

where $z \in \mathbb{R}$, $|\varepsilon_1| < 1$, $|\varepsilon_2| < 1$ and we write $Z \sim GESC(\varepsilon_1, \varepsilon_2)$.

Proof. Arguing as in Proposition 1, we have that

$$f_Z(z; \varepsilon_1, \varepsilon_2) = \int_{-\infty}^{\infty} \frac{|w|}{2\pi} \exp \left\{ -\frac{1}{2} \left[\frac{z^2}{(1-\text{sgn}(zw)\varepsilon_1)^2} + \frac{1}{(1-\text{sgn}(w)\varepsilon_2)^2} \right] w^2 \right\} dw.$$

By considering separately the cases in which $z \geq 0$ and $z < 0$, result follows directly. \square

From Proposition 2 two special cases are obtained directly,

1. If $\varepsilon_2 = 0$ a generalized Cauchy (GC) distribution is obtained. In this case we write $Z \sim GC(\varepsilon)$, and its pdf is given by

$$f_Z(z; \varepsilon) = \frac{1}{2\pi \left[1 + \left(\frac{z}{1+\text{sgn}(z)\varepsilon} \right)^2 \right]} + \frac{1}{2\pi \left[1 + \left(\frac{z}{1-\text{sgn}(z)\varepsilon} \right)^2 \right]},$$

where $z \in \mathbb{R}$, $0 < \varepsilon < 1$ and $Z \stackrel{d}{=} X/Y$ where $X \sim ESN(\varepsilon)$ and $Y \sim N(0, 1)$ are independent random variables.

2. If $\varepsilon_1 = \varepsilon_2 = \varepsilon$ an epsilon-skew-Cauchy distribution is obtained. In this case we write $Z \sim ESC2(\varepsilon)$, and its pdf is given by

$$f_Z(z; \varepsilon) = \frac{1}{2\pi \left[\left(\frac{1}{1+\varepsilon} \right)^2 + \left(\frac{z}{1+\operatorname{sgn}(z)\varepsilon} \right)^2 \right]} + \frac{1}{2\pi \left[\left(\frac{1}{1-\varepsilon} \right)^2 + \left(\frac{z}{1-\operatorname{sgn}(z)\varepsilon} \right)^2 \right]},$$

where $z \in \mathbb{R}$, $|\varepsilon| < 1$ and $Z \stackrel{d}{=} X/Y$ where $X \sim ESN(\varepsilon)$ and $Y \sim ESN(\varepsilon)$ are independent random variables.

4. General Bivariate Mudholkar-Hutson Distribution

Define Z_1, Z_2, Z_3 to be i.i.d. standard normal random variables. For $i = 1, 2, 3$ define

$$U_i = \begin{cases} -\alpha_i & \text{with probability } \gamma_i \\ \beta_i & \text{with probability } 1 - \gamma_i \end{cases}$$

where $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3$ are positive numbers and $0 < \gamma_1, \gamma_2, \gamma_3 < 1$. So, the parameters α_i and β_i indicate the propensity in which the discrete random variable takes negative and positive values, respectively. The parameters γ_i control how often negative and positive values are taken by U_i .

In addition assume that all six random variables Z_1, Z_2, Z_3, U_1, U_2 and U_3 are independent, and define

$$(X, Y) = \left(\frac{U_1 |Z_1|}{U_3 |Z_3|}, \frac{U_2 |Z_2|}{U_3 |Z_3|} \right). \tag{2}$$

The model (2) is highly flexible since it allows for different behavior in each of the four quadrants of the plane. From (2) it may be recognized that only fractional moments of X and Y exist.

Note that if we define $(W_1, W_2) = \left(\frac{|Z_1|}{|Z_3|}, \frac{|Z_2|}{|Z_3|} \right)$, it is readily verified that

$$f_{W_1, W_2}(w_1, w_2) = \frac{2}{\pi} \left(1 + w_1^2 + w_2^2 \right)^{-3/2} I(w_1 > 0, w_2 > 0), \tag{3}$$

in which case we say that (W_1, W_2) has a standard bivariate half-Cauchy distribution.

Using (3) and conditioning on U_1, U_2 and U_3 we obtain the density of (X, Y) in the albeit complicated form:

$$\begin{aligned}
 f(x, y) = & \frac{2}{\pi} \left\{ \gamma_1 \gamma_2 \gamma_3 \frac{\alpha_3^2}{\alpha_1 \alpha_2} \left[1 + \left(\frac{\alpha_3}{\alpha_1} x \right)^2 + \left(\frac{\alpha_3}{\alpha_2} y \right)^2 \right]^{-3/2} I(x > 0, y > 0) \right\} \\
 & + \frac{2}{\pi} \left\{ \gamma_1 \gamma_2 (1 - \gamma_3) \frac{\beta_3^2}{\alpha_1 \alpha_2} \left[1 + \left(\frac{\beta_3}{\alpha_1} x \right)^2 + \left(\frac{\beta_3}{\alpha_2} y \right)^2 \right]^{-3/2} I(x < 0, y < 0) \right\} \\
 & + \frac{2}{\pi} \left\{ \gamma_1 (1 - \gamma_2) \gamma_3 \frac{\alpha_3^2}{\alpha_1 \beta_2} \left[1 + \left(\frac{\alpha_3}{\alpha_1} x \right)^2 + \left(\frac{\alpha_3}{\beta_2} y \right)^2 \right]^{-3/2} I(x > 0, y < 0) \right\} \\
 & + \frac{2}{\pi} \left\{ \gamma_1 (1 - \gamma_2) (1 - \gamma_3) \frac{\beta_3^2}{\alpha_1 \beta_2} \left[1 + \left(\frac{\beta_3}{\alpha_1} x \right)^2 + \left(\frac{\beta_3}{\beta_2} y \right)^2 \right]^{-3/2} I(x < 0, y > 0) \right\} \\
 & + \frac{2}{\pi} \left\{ (1 - \gamma_1) \gamma_2 \gamma_3 \frac{\alpha_3^2}{\beta_1 \alpha_2} \left[1 + \left(\frac{\alpha_3}{\beta_1} x \right)^2 + \left(\frac{\alpha_3}{\alpha_2} y \right)^2 \right]^{-3/2} I(x < 0, y > 0) \right\} \\
 & + \frac{2}{\pi} \left\{ (1 - \gamma_1) \gamma_2 (1 - \gamma_3) \frac{\beta_3^2}{\beta_1 \alpha_2} \left[1 + \left(\frac{\beta_3}{\beta_1} x \right)^2 + \left(\frac{\beta_3}{\alpha_2} y \right)^2 \right]^{-3/2} I(x > 0, y < 0) \right\} \\
 & + \frac{2}{\pi} \left\{ (1 - \gamma_1) (1 - \gamma_2) \gamma_3 \frac{\alpha_3^2}{\beta_1 \beta_2} \left[1 + \left(\frac{\alpha_3}{\beta_1} x \right)^2 + \left(\frac{\alpha_3}{\beta_2} y \right)^2 \right]^{-3/2} I(x < 0, y < 0) \right\} \\
 & + \frac{2}{\pi} \left\{ (1 - \gamma_1) (1 - \gamma_2) (1 - \gamma_3) \frac{\beta_3^2}{\beta_1 \beta_2} \left[1 + \left(\frac{\beta_3}{\beta_1} x \right)^2 + \left(\frac{\beta_3}{\beta_2} y \right)^2 \right]^{-3/2} I(x > 0, y > 0) \right\}.
 \end{aligned} \tag{4}$$

Some special cases which might be considered include the following:

1. Mudholkar and Hutson type. For this we set: $\alpha_i = 1 + \varepsilon_i, \beta_i = 1 - \varepsilon_i$ and $\gamma_i = (1 + \varepsilon_i) / 2$ for $i = 1, 2, 3$.

In this case the density (4) simplifies somewhat to become:

$$\begin{aligned}
 f(x, y) = & \frac{1}{4\pi} \left\{ (1 + \varepsilon_3)^3 \left[1 + \left(\frac{1 + \varepsilon_3}{1 + \varepsilon_1} x \right)^2 + \left(\frac{1 + \varepsilon_3}{1 + \varepsilon_2} y \right)^2 \right]^{-3/2} I(x > 0, y > 0) \right\} \\
 & + \frac{1}{4\pi} \left\{ (1 - \varepsilon_3)^3 \left[1 + \left(\frac{1 - \varepsilon_3}{1 + \varepsilon_1} x \right)^2 + \left(\frac{1 - \varepsilon_3}{1 + \varepsilon_2} y \right)^2 \right]^{-3/2} I(x < 0, y < 0) \right\} \\
 & + \frac{1}{4\pi} \left\{ (1 + \varepsilon_3)^3 \left[1 + \left(\frac{1 + \varepsilon_3}{1 + \varepsilon_1} x \right)^2 + \left(\frac{1 + \varepsilon_3}{1 - \varepsilon_2} y \right)^2 \right]^{-3/2} I(x > 0, y < 0) \right\} \\
 & + \frac{1}{4\pi} \left\{ (1 - \varepsilon_3)^3 \left[1 + \left(\frac{1 - \varepsilon_3}{1 + \varepsilon_1} x \right)^2 + \left(\frac{1 - \varepsilon_3}{1 - \varepsilon_2} y \right)^2 \right]^{-3/2} I(x < 0, y > 0) \right\} \\
 & + \frac{1}{4\pi} \left\{ (1 + \varepsilon_3)^3 \left[1 + \left(\frac{1 + \varepsilon_3}{1 - \varepsilon_1} x \right)^2 + \left(\frac{1 + \varepsilon_3}{1 + \varepsilon_2} y \right)^2 \right]^{-3/2} I(x < 0, y > 0) \right\} \\
 & + \frac{1}{4\pi} \left\{ (1 - \varepsilon_3)^3 \left[1 + \left(\frac{1 - \varepsilon_3}{1 - \varepsilon_1} x \right)^2 + \left(\frac{1 - \varepsilon_3}{1 + \varepsilon_2} y \right)^2 \right]^{-3/2} I(x > 0, y < 0) \right\} \\
 & + \frac{1}{4\pi} \left\{ (1 + \varepsilon_3)^3 \left[1 + \left(\frac{1 + \varepsilon_3}{1 - \varepsilon_1} x \right)^2 + \left(\frac{1 + \varepsilon_3}{1 - \varepsilon_2} y \right)^2 \right]^{-3/2} I(x < 0, y < 0) \right\} \\
 & + \frac{1}{4\pi} \left\{ (1 - \varepsilon_3)^3 \left[1 + \left(\frac{1 - \varepsilon_3}{1 - \varepsilon_1} x \right)^2 + \left(\frac{1 - \varepsilon_3}{1 - \varepsilon_2} y \right)^2 \right]^{-3/2} I(x > 0, y > 0) \right\}.
 \end{aligned} \tag{5}$$

A further specialization of the density (5) can be considered, as follows.

2. Homogenous Mudholkar and Hutson type. For this we set: $\alpha_i = 1 + \varepsilon$, $\beta_i = 1 - \varepsilon$ and $\gamma_i = (1 + \varepsilon) / 2$ for $i = 1, 2, 3$.

This homogeneity results in a little simplification of (4), thus:

$$\begin{aligned}
 f(x, y; \varepsilon) = & \frac{1}{4\pi} \left\{ [(1 + \varepsilon)^3 + (1 - \varepsilon)^3] [1 + x^2 + y^2]^{-3/2} I(x > 0, y > 0) \right\} \\
 & + \frac{1}{4\pi} \left\{ (1 - \varepsilon)^3 \left[1 + \left(\frac{1 - \varepsilon}{1 + \varepsilon} x \right)^2 + \left(\frac{1 - \varepsilon}{1 + \varepsilon} y \right)^2 \right]^{-3/2} I(x < 0, y < 0) \right\} \\
 & + \frac{1}{4\pi} \left\{ (1 + \varepsilon)^3 \left[1 + x^2 + \left(\frac{1 + \varepsilon}{1 - \varepsilon} y \right)^2 \right]^{-3/2} I(x > 0, y < 0) \right\} \\
 & + \frac{1}{4\pi} \left\{ (1 - \varepsilon)^3 \left[1 + \left(\frac{1 - \varepsilon}{1 + \varepsilon} x \right)^2 + y^2 \right]^{-3/2} I(x < 0, y > 0) \right\} \\
 & + \frac{1}{4\pi} \left\{ (1 + \varepsilon)^3 \left[1 + \left(\frac{1 + \varepsilon}{1 - \varepsilon} x \right)^2 + y^2 \right]^{-3/2} I(x < 0, y > 0) \right\} \\
 & + \frac{1}{4\pi} \left\{ (1 - \varepsilon)^3 \left[1 + x^2 + \left(\frac{1 - \varepsilon}{1 + \varepsilon} y \right)^2 \right]^{-3/2} I(x > 0, y < 0) \right\} \\
 & + \frac{1}{4\pi} \left\{ (1 + \varepsilon)^3 \left[1 + \left(\frac{1 + \varepsilon}{1 - \varepsilon} x \right)^2 + \left(\frac{1 + \varepsilon}{1 - \varepsilon} y \right)^2 \right]^{-3/2} I(x < 0, y < 0) \right\}.
 \end{aligned} \tag{6}$$

It is easy to see that the parameter ε is not identifiable in (6) because $f(x, y; \varepsilon) = f(x, y; -\varepsilon)$. An adjustment to ensure identifiability involves introducing the constraint $\varepsilon \geq 0$.

3. Equal weights. In this case we assume that $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3$ are positive numbers and $\gamma_i = 1/2$ for $i = 1, 2, 3$.

$$\begin{aligned}
 f(x, y) = & \frac{1}{4\pi} \left\{ \frac{\alpha_3^2}{\alpha_1 \alpha_2} \left[1 + \left(\frac{\alpha_3}{\alpha_1} x \right)^2 + \left(\frac{\alpha_3}{\alpha_2} y \right)^2 \right]^{-3/2} + \frac{\beta_3^2}{\beta_1 \beta_2} \left[1 + \left(\frac{\beta_3}{\beta_1} x \right)^2 + \left(\frac{\beta_3}{\beta_2} y \right)^2 \right]^{-3/2} \right\} \\
 & \times I(x > 0, y > 0) \\
 & + \frac{1}{4\pi} \left\{ \frac{\beta_3^2}{\alpha_1 \alpha_2} \left[1 + \left(\frac{\beta_3}{\alpha_1} x \right)^2 + \left(\frac{\beta_3}{\alpha_2} y \right)^2 \right]^{-3/2} + \frac{\alpha_3^2}{\beta_1 \beta_2} \left[1 + \left(\frac{\alpha_3}{\beta_1} x \right)^2 + \left(\frac{\alpha_3}{\beta_2} y \right)^2 \right]^{-3/2} \right\} \\
 & \times I(x < 0, y < 0) \\
 & + \frac{1}{4\pi} \left\{ \frac{\alpha_3^2}{\alpha_1 \beta_2} \left[1 + \left(\frac{\alpha_3}{\alpha_1} x \right)^2 + \left(\frac{\alpha_3}{\beta_2} y \right)^2 \right]^{-3/2} + \frac{\beta_3^2}{\beta_1 \alpha_2} \left[1 + \left(\frac{\beta_3}{\beta_1} x \right)^2 + \left(\frac{\beta_3}{\alpha_2} y \right)^2 \right]^{-3/2} \right\} \\
 & \times I(x > 0, y < 0) \\
 & + \frac{1}{4\pi} \left\{ \frac{\alpha_3^2}{\beta_1 \alpha_2} \left[1 + \left(\frac{\alpha_3}{\beta_1} x \right)^2 + \left(\frac{\alpha_3}{\alpha_2} y \right)^2 \right]^{-3/2} + \frac{\beta_3^2}{\alpha_1 \beta_2} \left[1 + \left(\frac{\beta_3}{\alpha_1} x \right)^2 + \left(\frac{\beta_3}{\beta_2} y \right)^2 \right]^{-3/2} \right\} \\
 & \times I(x < 0, y > 0).
 \end{aligned} \tag{7}$$

The pdf (7) is not identifiable because the values of α_i can be interchanged with those of β_i and $f(x, y)$ does not change. Moreover, multiplying all of the α 's and β 's by a constant does not change $f(x, y)$. So, one way to get identifiability in the model (7) is to set $\alpha_i = \beta_i$ ($i = 1, 2, 3$) and α_3 equal to 1. In that case, Equation (7) takes the form

$$f(x, y) = \frac{1}{2\pi} \frac{1}{\alpha_1 \alpha_2} \left[1 + \left(\frac{x}{\alpha_1} \right)^2 + \left(\frac{y}{\alpha_2} \right)^2 \right]^{-3/2}.$$

However, this is then recognizable as being simply a scaled version of the standard bivariate Cauchy density (compare with Equation (3)).

We now consider the marginal densities for the random variable (X, Y) defined by (2). From (2) we have $X = \frac{U_1|Z_1|}{U_3|Z_3|}$ and $Y = \frac{U_2|Z_2|}{U_3|Z_3|}$ and the density of $W_1 = \frac{|Z_1|}{|Z_3|} = \left| \frac{Z_1}{Z_3} \right|$ is a standard half-Cauchy density, i.e., $f_{W_1}(w_1) = \frac{2}{\pi} (1 + w_1^2)^{-1}$.

Consequently, the density of X is of the form

$$f_X(x) = \frac{2}{\pi} \gamma_1 \gamma_3 \frac{\alpha_3}{\alpha_1} \left[1 + \left(\frac{\alpha_3}{\alpha_1} x \right)^2 \right]^{-1} I(x > 0) + \frac{2}{\pi} \gamma_1 (1 - \gamma_3) \frac{\beta_3}{\alpha_1} \left[1 + \left(\frac{\beta_3}{\alpha_1} x \right)^2 \right]^{-1} I(x < 0) + \frac{2}{\pi} (1 - \gamma_1) \left\{ \gamma_3 \frac{\alpha_3}{\beta_1} \left[1 + \left(\frac{\alpha_3}{\beta_1} x \right)^2 \right]^{-1} I(x < 0) + (1 - \gamma_3) \frac{\beta_3}{\beta_1} \left[1 + \left(\frac{\beta_3}{\beta_1} x \right)^2 \right]^{-1} I(x > 0) \right\},$$

which is a mixture of half-Cauchy densities. The density of Y is then of the same form with α_1, β_1 replaced by α_2, β_2 .

To get a more direct bivariate version of the original Mudholkar-Hutson density in which the α 's, β 's and γ 's were related to each other, we go back to the original representation, i.e., Equation (2), but now we set $\alpha_1 = (1 + \varepsilon_1)$, $\beta_1 = (1 - \varepsilon_1)$, $\gamma_1 = (1 + \varepsilon_1) / 2$, $\alpha_2 = (1 + \varepsilon_2)$, $\beta_2 = (1 - \varepsilon_2)$, $\gamma_2 = (1 + \varepsilon_2) / 2$ and $\alpha_3 = \beta_3 = \gamma_3 = 1$. So the density is of the form

$$f(x, y) = \frac{1}{2\pi} \left[1 + \left(\frac{x}{1 + \varepsilon_1} \right)^2 + \left(\frac{y}{1 + \varepsilon_2} \right)^2 \right]^{-3/2} I(x > 0, y > 0) + \frac{1}{2\pi} \left[1 + \left(\frac{x}{1 + \varepsilon_1} \right)^2 + \left(\frac{y}{1 - \varepsilon_2} \right)^2 \right]^{-3/2} I(x > 0, y < 0) + \frac{1}{2\pi} \left[1 + \left(\frac{x}{1 - \varepsilon_1} \right)^2 + \left(\frac{y}{1 + \varepsilon_2} \right)^2 \right]^{-3/2} I(x < 0, y > 0) + \frac{1}{2\pi} \left[1 + \left(\frac{x}{1 - \varepsilon_1} \right)^2 + \left(\frac{y}{1 - \varepsilon_2} \right)^2 \right]^{-3/2} I(x < 0, y < 0)$$

with marginal density

$$f_X(x) = \frac{1}{\pi} \left[1 + \left(\frac{x}{1 + \varepsilon_1} \right)^2 \right]^{-1} I(x > 0) + \frac{1}{\pi} \left[1 + \left(\frac{x}{1 - \varepsilon_1} \right)^2 \right]^{-1} I(x < 0).$$

A more general version of the density with $\gamma_3 = 1$, is of the form: (without loss of generality set $\alpha_3 = 1$)

$$f(x, y) = \frac{2}{\pi} \left\{ \gamma_1 \gamma_2 \frac{1}{\alpha_1 \alpha_2} \left[1 + \left(\frac{x}{\alpha_1} \right)^2 + \left(\frac{y}{\alpha_2} \right)^2 \right]^{-3/2} I(x > 0, y > 0) \right\} + \frac{2}{\pi} \left\{ \gamma_1 (1 - \gamma_2) \frac{1}{\alpha_1 \beta_2} \left[1 + \left(\frac{x}{\alpha_1} \right)^2 + \left(\frac{y}{\beta_2} \right)^2 \right]^{-3/2} I(x > 0, y < 0) \right\} + \frac{2}{\pi} \left\{ (1 - \gamma_1) \gamma_2 \frac{1}{\beta_1 \alpha_2} \left[1 + \left(\frac{x}{\beta_1} \right)^2 + \left(\frac{y}{\alpha_2} \right)^2 \right]^{-3/2} I(x < 0, y > 0) \right\} + \frac{2}{\pi} \left\{ (1 - \gamma_1) (1 - \gamma_2) \frac{1}{\beta_1 \beta_2} \left[1 + \left(\frac{x}{\beta_1} \right)^2 + \left(\frac{y}{\beta_2} \right)^2 \right]^{3/2} I(x < 0, y < 0) \right\}$$

with corresponding marginal

$$f_X(x) = \frac{2}{\pi} \gamma_1 \frac{1}{\alpha_1} \left[1 + \left(\frac{x}{\alpha_1} \right)^2 \right]^{-1} I(x > 0) + \frac{2}{\pi} (1 - \gamma_1) \frac{1}{\beta_1} \left[1 + \left(\frac{x}{\beta_1} \right)^2 \right]^{-1} I(x < 0).$$

Needless to say we can consider analogous models in which, instead of assuming that $(W_1, W_2) = \left(\frac{|Z_1|}{|Z_3|}, \frac{|Z_2|}{|Z_3|}\right)$ has a density of the form (3) we assume that it has another bivariate density with support $\mathbb{R}^+ \times \mathbb{R}^+$, e.g., bivariate normal restricted to $\mathbb{R}^+ \times \mathbb{R}^+$, or bivariate Pareto, etc.

Another general bivariate MH model which includes the bivariate skew-Cauchy distribution given by (2) is the bivariate skew t model. This model can be obtained replacing $|Z_3|$ by $V_3 \sim \chi^2_\nu$ in Equation (2). Thus, if we assume that all six random variables $Z_1, Z_2 \stackrel{iid}{\sim} N(0, 1), V_3 \sim \chi^2_\nu; U_1, U_2$ and U_3 are independent, and define

$$(X, Y) = \left(\frac{U_1 |Z_1|}{U_3 \sqrt{V_3/\nu}}, \frac{U_2 |Z_2|}{U_3 \sqrt{V_3/\nu}} \right), \tag{8}$$

then, because $(W_1, W_2) = \left(\frac{|Z_1|}{\sqrt{V_3/\nu}}, \frac{|Z_2|}{\sqrt{V_3/\nu}}\right)$ has density

$$f_{W_1, W_2}(w_1, w_2) = \frac{2}{\pi} \left(1 + \frac{w_1^2 + w_2^2}{\nu} \right)^{-(\nu+2)/2} I(w_1 > 0, w_2 > 0),$$

the density of (X, Y) is

$$\begin{aligned} f(x, y) &= \int f(x, y | u_1, u_2, u_3) f(u_1, u_2, u_3) du_1 du_2 du_3 \\ &= \int \frac{u_3^2}{u_1 u_2} f_{W_1, W_2} \left(\frac{u_3 x}{u_1}, \frac{u_3 y}{u_2} \mid u_1, u_2, u_3 \right) f(u_1, u_2, u_3) du_1 du_2 du_3 \\ &= \int \frac{2}{\pi} \frac{u_3^2}{u_1 u_2} \left[1 + \frac{1}{\nu} \left(\frac{u_3 x}{u_1} \right)^2 + \frac{1}{\nu} \left(\frac{u_3 y}{u_2} \right)^2 \right]^{-(\nu+2)/2} f(u_1, u_2, u_3) du_1 du_2 du_3 \\ &= \frac{2}{\pi} \left\{ \gamma_1 \gamma_2 \gamma_3 \frac{\alpha_3^2}{\alpha_1 \alpha_2} \left[1 + \frac{1}{\nu} \left(\frac{\alpha_3 x}{\alpha_1} \right)^2 + \frac{1}{\nu} \left(\frac{\alpha_3 y}{\alpha_2} \right)^2 \right]^{-(\nu+2)/2} I(x > 0, y > 0) \right\} \\ &\quad + \frac{2}{\pi} \left\{ \gamma_1 \gamma_2 (1 - \gamma_3) \frac{\beta_3^2}{\alpha_1 \alpha_2} \left[1 + \frac{1}{\nu} \left(\frac{\beta_3 x}{\alpha_1} \right)^2 + \frac{1}{\nu} \left(\frac{\beta_3 y}{\alpha_2} \right)^2 \right]^{-(\nu+2)/2} I(x < 0, y < 0) \right\} \\ &\quad + \frac{2}{\pi} \left\{ \gamma_1 (1 - \gamma_2) \gamma_3 \frac{\alpha_3^2}{\alpha_1 \beta_2} \left[1 + \frac{1}{\nu} \left(\frac{\alpha_3 x}{\alpha_1} \right)^2 + \frac{1}{\nu} \left(\frac{\alpha_3 y}{\beta_2} \right)^2 \right]^{-(\nu+2)/2} I(x > 0, y < 0) \right\} \\ &\quad + \frac{2}{\pi} \left\{ \gamma_1 (1 - \gamma_2) (1 - \gamma_3) \frac{\beta_3^2}{\alpha_1 \beta_2} \left[1 + \frac{1}{\nu} \left(\frac{\beta_3 x}{\alpha_1} \right)^2 + \frac{1}{\nu} \left(\frac{\beta_3 y}{\beta_2} \right)^2 \right]^{-(\nu+2)/2} I(x < 0, y > 0) \right\} \\ &\quad + \frac{2}{\pi} \left\{ (1 - \gamma_1) \gamma_2 \gamma_3 \frac{\alpha_3^2}{\beta_1 \alpha_2} \left[1 + \frac{1}{\nu} \left(\frac{\alpha_3 x}{\beta_1} \right)^2 + \frac{1}{\nu} \left(\frac{\alpha_3 y}{\alpha_2} \right)^2 \right]^{-(\nu+2)/2} I(x < 0, y > 0) \right\} \\ &\quad + \frac{2}{\pi} \left\{ (1 - \gamma_1) \gamma_2 (1 - \gamma_3) \frac{\beta_3^2}{\beta_1 \alpha_2} \left[1 + \frac{1}{\nu} \left(\frac{\beta_3 x}{\beta_1} \right)^2 + \frac{1}{\nu} \left(\frac{\beta_3 y}{\alpha_2} \right)^2 \right]^{-(\nu+2)/2} I(x > 0, y < 0) \right\} \\ &\quad + \frac{2}{\pi} \left\{ (1 - \gamma_1) (1 - \gamma_2) \gamma_3 \frac{\alpha_3^2}{\beta_1 \beta_2} \left[1 + \frac{1}{\nu} \left(\frac{\alpha_3 x}{\beta_1} \right)^2 + \frac{1}{\nu} \left(\frac{\alpha_3 y}{\beta_2} \right)^2 \right]^{-(\nu+2)/2} I(x < 0, y < 0) \right\} \\ &\quad + \frac{2}{\pi} \left\{ (1 - \gamma_1) (1 - \gamma_2) (1 - \gamma_3) \frac{\beta_3^2}{\beta_1 \beta_2} \left[1 + \frac{1}{\nu} \left(\frac{\beta_3 x}{\beta_1} \right)^2 + \frac{1}{\nu} \left(\frac{\beta_3 y}{\beta_2} \right)^2 \right]^{-(\nu+2)/2} I(x > 0, y > 0) \right\}. \end{aligned} \tag{9}$$

The pair of variables in (8) allows one to model a wider variety of paired data sets than that given in (2) because it also can model light and heavy “tails” in a different way for each quadrant of the coordinate axis. Additionally, it is possible to compute the r -order moments.

Proposition 3. *The expected value of (X, Y) in (8) is given by*

$$\mathbb{E}(X, Y) = \frac{\sqrt{\nu}}{\sqrt{\pi}} (\beta_1 (1 - \gamma_1) - \alpha_1 \gamma_1, \beta_2 (1 - \gamma_2) - \alpha_2 \gamma_2) \left(\frac{1 - \gamma_3}{\beta_3} - \frac{\gamma_3}{\alpha_3} \right) \frac{\Gamma(\nu/2 - 1/2)}{\Gamma(\nu/2)},$$

provided that $\nu > 1$.

Proof. For $i = 1, 2$ it is immediate that $\mathbb{E}(U_i) \mathbb{E}(|Z_i|) = (\beta_i(1 - \gamma_i) - \alpha_i\gamma_i) \sqrt{2}/\sqrt{\pi}$ and $\mathbb{E}(U_3^{-1}) = (1 - \gamma_3) / \beta_3 - \gamma_3 / \alpha_3$.

On the other hand, since $V_3 \sim \chi_v^2$ then

$$\mathbb{E}\left(\frac{\sqrt{v}}{\sqrt{V_3}}\right) = \int_0^\infty \frac{\sqrt{v}}{\sqrt{x}} \frac{x^{v/2-1}}{2^{v/2}\Gamma(v/2)} e^{-x/2} dx \stackrel{v \geq 1}{=} \frac{\sqrt{v}\Gamma(v/2 - 1/2)}{\sqrt{2}\Gamma(v/2)}.$$

Therefore, from (8) we have

$$\mathbb{E}(X, Y) = (\mathbb{E}(U_1) \mathbb{E}(|Z_1|), \mathbb{E}(U_2) \mathbb{E}(|Z_2|)) \mathbb{E}\left(\frac{1}{U_3}\right) \mathbb{E}\left(\frac{\sqrt{v}}{\sqrt{V_3}}\right)$$

and the result is obtained straightforward. \square

Following the proof of the previous proposition it is possible to obtain the r -order moments provided that $v > r$.

In applications, it will usually be appropriate to augment these models by the introduction of location, scale and rotation parameters, i.e., to consider

$$\left(\tilde{X}, \tilde{Y}\right) = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2} (X, Y),$$

where $\boldsymbol{\mu} \in (-\infty, \infty)^2$ and $\boldsymbol{\Sigma}$ is positive definite.

5. Application

The data that we will use were collected by the Australian Institute of Sport and reported by Cook and Weisberg (1994) [10]. The data set consists of values of several variables measured on $n = 202$ Australian athletes. Specifically, we shall consider the pair of variables (Ht,Wt) which are the height (cm) and the weight (Kg) measured for each athlete.

We fitted the bivariate Mudholkar-Hutson distributions for five different cases. In addition to the general case which is given by the pdf $f(x, y; \alpha_i, \beta_i, \gamma_i, \nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, based on (9), where $i = 1, 2, 3$, and $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ is the location parameter and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

is the symmetric positive definite scale matrix, we also consider four special cases:

1. When the pdf is given by Equation (4). That is taking $\nu \rightarrow \infty$ in the bivariate skew t MH model.
2. When the pdf is given by

$$f(x, y; \varepsilon_i, \mu_j, \Sigma_{11}, \Sigma_{22}, \Sigma_{12}) = f\left(x, y; 1 + \varepsilon_i, 1 - \varepsilon_i, \frac{1 + \varepsilon_i}{2}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right),$$

which is the bivariate MH distribution specified using the special case (1.), where $|\varepsilon_i| < 1$ for all $i = 1, 2, 3$.

3. When the pdf is given by

$$f(x, y; \varepsilon, \mu_j, \Sigma_{11}, \Sigma_{22}, \Sigma_{12}) = f\left(x, y; 1 + \varepsilon, 1 - \varepsilon, \frac{1 + \varepsilon}{2}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right),$$

which is the bivariate MH distribution specified using the special case (2.), where $|\varepsilon| < 1$.

4. When the pdf is given by

$$f(x, y; \mu_j, \Sigma_{11}, \Sigma_{22}, \Sigma_{12}) = f\left(x, y; 1, 1, \frac{1}{2}, \mu, \Sigma\right),$$

which is the bivariate Cauchy distribution (as in (3)).

All fits were done by maximizing the likelihood through numerical methods which combine algorithms based on the Hessian matrix and the Simulated Annealing algorithm. Standard errors of the estimations were computed based on 1000 bootstrap data samples.

To compare model fits, we used the Akaike criterion (see Akaike, 1974 [11]), namely

$$AIC = -2 \ln [L(\hat{\theta})] + 2k,$$

where k is the dimension of θ which is the vector of parameters of the model being considered.

Table 1 displays the results of the fits for 100 women. In Table 1, we see the results of fitting the five competing models. They are the general bivariate skew $-t$ MH model with pdf given in Equation (9), General bivariate MH with pdf given in Equation (4), Bivariate MH (1.) with pdf given in Equation (5), Bivariate MH (2.) with pdf given in Equation (6) and Bivariate Cauchy. It shows the maximum likelihood estimates (mle's) of the five models. The last column shows the estimated standard errors (se) of the estimates. Non identifiability of the bivariate skew $-t$ MH model is evidenced by the huge value of the estimated standard error of the estimate of ν . However, the AIC criterion indicates that data are better fitted by the general bivariate skew $-t$ MH (9) model. Figure 2 shows the contour lines of the fitted pdf.

Table 1. Bivariate Mudholkar-Hutson fits for women.

Model	AIC	$(\alpha_i, \beta_i, \gamma_i, \mu, \Sigma)$ Estimates	Bootstrap Standard Errors of the Estimates
Bivariate skew t MH	1354.5	$(\alpha_1, \alpha_2, \alpha_3) = (13.48, 16.62, 1.25)$ $(\beta_1, \beta_2, \beta_3) = (10.65, 36.47, 3.00)$ $(\gamma_1, \gamma_2, \gamma_3) = (1.00, 0.97, 0.94)$ $\nu = 10.93$ $(\mu_1, \mu_2) = (161.97, 52.50)$ $(\Sigma_{11}, \Sigma_{22}, \Sigma_{12}) = (1.82, 1.67, 1.58)$	(10.02, 29.96, 8.05) (1.53, 17.79, 19.75) (0.006, 0.02, 0.03) 168.56 (2.18, 2.22) (17.80, 14.36, 14.33)
General bivariate MH	1413.5	$(\alpha_1, \alpha_2, \alpha_3) = (13.26, 16.04, 1.05)$ $(\beta_1, \beta_2, \beta_3) = (10.64, 36.50, 4.99)$ $(\gamma_1, \gamma_2, \gamma_3) = (1.00, 0.98, 0.94)$ $(\mu_1, \mu_2) = (161.69, 50.90)$ $(\Sigma_{11}, \Sigma_{22}, \Sigma_{12}) = (1.13, 1.17, 1.09)$	(2.70, 4.30, 0.90) (0.92, 3.85, 3.64) (0.003, 0.02, 0.02) (1.79, 2.39) (1.57, 1.59, 1.47)
Bivariate MH (1.)	1429.4	$(\varepsilon_1, \varepsilon_2, \varepsilon_3) = (0.85, 0.83, -1.00)$ $(\mu_1, \mu_2) = (186.30, 83.80)$ $(\Sigma_{11}, \Sigma_{22}, \Sigma_{12}) = (169.37, 334.65, 224.88)$	(0.07, 0.08, 0.003) (1.45, 1.47) (33.30, 51.34, 36.83)
Bivariate MH (2.)	1453.8	$\varepsilon = 0.50$ $(\mu_1, \mu_2) = (171.60, 62.30)$ $(\Sigma_{11}, \Sigma_{22}, \Sigma_{12}) = (32.49, 61.40, 33.96)$	0.15 (1.86, 2.80) (8.59, 16.41, 10.87)
Bivariate Cauchy	1665.5	$(\mu_1, \mu_2) = (177.21, 69.08)$ $(\Sigma_{11}, \Sigma_{22}, \Sigma_{12}) = (77.84, 124.86, 41.53)$	(1.64, 2.01) (2.23, 1.39, 8.32)

Table 2 displays the results of the fits for 102 men. In Table 2 we again compare five competing models. The maximum likelihood estimates (mle's) of the five models, the corresponding Akaike criterion values and estimated standard errors of the estimates are displayed in the table. Again, the AIC indicates that the data are better fitted by the general bivariate skew $-t$ MH (9) model. Figure 3 shows the contour lines of the fitted pdf. Non identifiability of the general bivariate skew $-t$ MH (9)

and General bivariate MH (4) models is shown by the large values of the estimated standard errors for some estimates.

Table 2. Bivariate Mudholkar-Hutson fits for men.

Model	AIC	$(\alpha_i, \beta_i, \gamma_i, \mu, \Sigma)$ Estimates	Bootstrap Standard Errors of the Estimates
Bivariate skew <i>t</i> MH	1396.9	$(\alpha_1, \alpha_2, \alpha_3) = (13.43, 20.16, 2.62)$ $(\beta_1, \beta_2, \beta_3) = (35.02, 11.84, 6.01)$ $(\gamma_1, \gamma_2, \gamma_3) = (0.94, 0.96, 0.88)$ $v = 9.99$ $(\mu_1, \mu_2) = (177.92, 70.30)$ $(\Sigma_{11}, \Sigma_{22}, \Sigma_{12}) = (4.77, 4.71, 4.04)$	(14.66, 5.34, 4.12) (8.42, 1.82, 7.83) (0.02, 0.02, 0.04) 64.18 (0.65, 1.22) (1.89, 6.12, 1.83)
General bivariate MH	1441.8	$(\alpha_1, \alpha_2, \alpha_3) = (52.54, 52.17, 14.58)$ $(\beta_1, \beta_2, \beta_3) = (29.83, 62.76, 31.76)$ $(\gamma_1, \gamma_2, \gamma_3) = (0.96, 0.98, 0.96)$ $(\mu_1, \mu_2) = (174.60, 66.6)$ $(\Sigma_{11}, \Sigma_{22}, \Sigma_{12}) = (9.32, 17.40, 11.78)$	(0.89, 1.28, 0.43) (1.52, 1.39, 0.59) (0.01, 0.01, 0.02) (1.13, 1.16) (0.76, 0.50, 0.55)
Bivariate MH (1.)	1456.2	$(\varepsilon_1, \varepsilon_2, \varepsilon_3) = (-0.87, -0.91, -1.0)$ $(\mu_1, \mu_2) = (172.70, 61.00)$ $(\Sigma_{11}, \Sigma_{22}, \Sigma_{12}) = (195.26, 445.26, 279.93)$	(0.06, 0.05, 0.00) (1.39, 1.94) (32.88, 62.62, 38.52)
Bivariate MH (2.)	1512.6	$\varepsilon = 0.17$ $(\mu_1, \mu_2) = (185.10, 80.54)$ $(\Sigma_{11}, \Sigma_{22}, \Sigma_{12}) = (36.15, 70.27, 38.53)$	0.21 (2.73, 3.54) (14.66, 26.93, 19.09)
Bivariate Cauchy	1775.2	$(\mu_1, \mu_2) = (184.45, 79.36)$ $(\Sigma_{11}, \Sigma_{22}, \Sigma_{12}) = (65.57, 155.19, 59.31)$	(3.01, 3.43) (3.62, 1.54, 8.29)

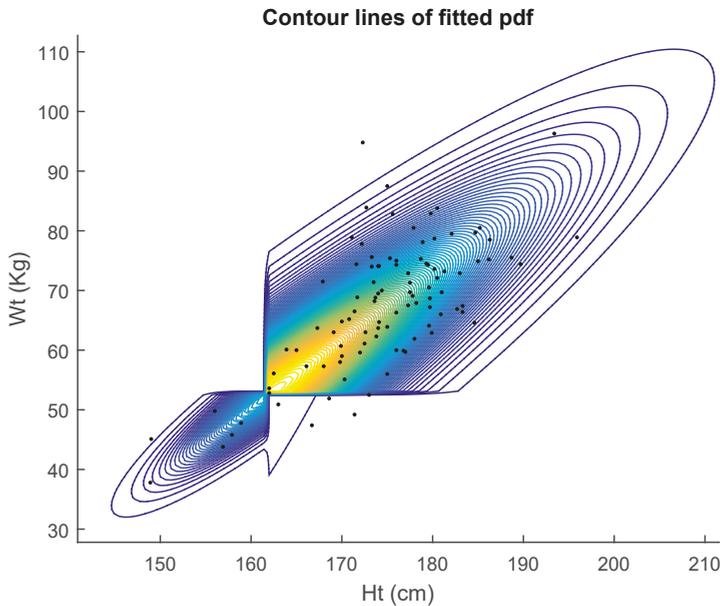


Figure 2. Female Australian athletes data: scatter plot (Ht, Wt) and fitted General bivariate MH.

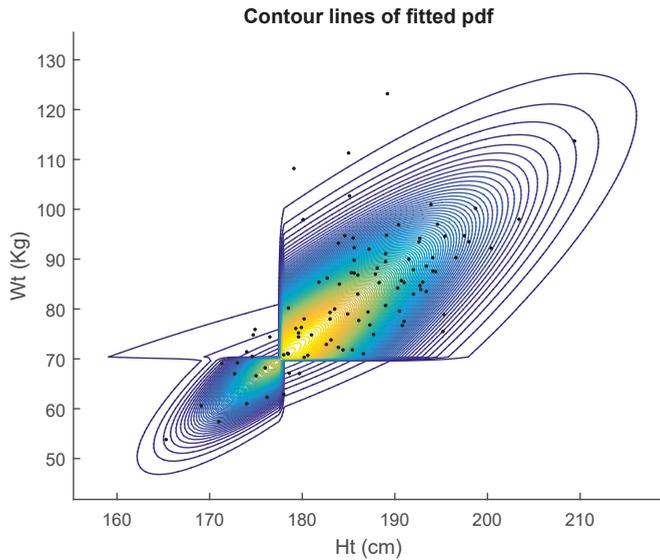


Figure 3. Male Australian athletes data: scatter plot (Ht, Wt) and fitted General bivariate MH.

Table 3 displays the results of the fits for the full set of $n = 202$ Australian athletes, regardless of gender. Table 3 shows the maximum likelihood estimates (mle’s) of the five models together with the corresponding Akaike criterion values and estimated standard errors of the estimates. The AIC here also indicates that the data are better fitted by the general bivariate skew- t MH model (9). Figure 4 shows the contour lines of the fitted pdf. Thus, in all three cases, males, females and combined, the best fitting model was the general bivariate skew- t MH.

Table 3. Bivariate Mudholkar-Hutson fits.

Model	AIC	$(\alpha_i, \beta_{ir}, \gamma_{ir}, \mu, \Sigma)$ Estimates	Bootstrap Standard Errors of the Estimates
Bivariate skew t MH	2803.3	$(\alpha_1, \alpha_2, \alpha_3) = (14.15, 17.35, 1.09)$ $(\beta_1, \beta_2, \beta_3) = (49.35, 67.05, 1.62)$ $(\gamma_1, \gamma_2, \gamma_3) = (0.98, 0.96, 0.91)$ $\nu = 10.93$ $(\mu_1, \mu_2) = (171.09, 61.00)$ $(\Sigma_{11}, \Sigma_{22}, \Sigma_{12}) = (1.01, 1.32, 1.04)$	$(8.18, 10.35, 3.57)$ $(5.22, 5.57, 6.86)$ $(0.01, 0.01, 0.02)$ 9.42 $(0.24, 0.76)$ $(5.74, 7.75, 6.07)$
General bivariate MH	2901.4	$(\alpha_1, \alpha_2, \alpha_3) = (15.42, 17.67, 2.25)$ $(\beta_1, \beta_2, \beta_3) = (49.28, 66.53, 3.36)$ $(\gamma_1, \gamma_2, \gamma_3) = (0.98, 0.96, 0.88)$ $(\mu_1, \mu_2) = (170.54, 60.95)$ $(\Sigma_{11}, \Sigma_{22}, \Sigma_{12}) = (2.16, 3.02, 2.35)$	$(10.74, 12.92, 2.61)$ $(5.00, 5.97, 5.10)$ $(0.01, 0.02, 0.02)$ $(0.54, 1.02)$ $(1.55, 1.84, 1.32)$
Bivariate MH (1.)	2984.0	$(\epsilon_1, \epsilon_2, \epsilon_3) = (1.00, 0.39, -0.13)$ $(\mu_1, \mu_2) = (181.06, 74.48)$ $(\Sigma_{11}, \Sigma_{22}, \Sigma_{12}) = (11.41, 50.90, 18.73)$	$(0.01, 0.07, 0.06)$ $(0.76, 0.84)$ $(1.59, 5.00, 2.17)$
Bivariate MH (2.)	3022.2	$\epsilon = 0.70$ $(\mu_1, \mu_2) = (171.37, 61.32)$ $(\Sigma_{11}, \Sigma_{22}, \Sigma_{12}) = (48.62, 103.69, 64.24)$	0.10 $(2.42, 3.07)$ $(53.89, 89.24, 68.07)$
Bivariate Cauchy	3536.9	$(\mu_1, \mu_2) = (178.84, 71.81)$ $(\Sigma_{11}, \Sigma_{22}, \Sigma_{12}) = (106.75, 199.84, 83.64)$	$(0.83, 1.55)$ $(2.16, 1.16, 5.53)$

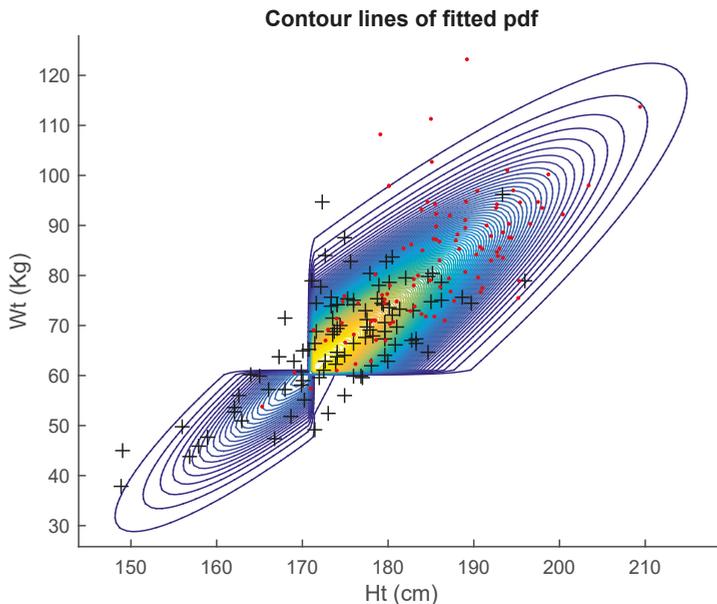


Figure 4. Australian athletes data: scatter plot (Ht, Wt) and fitted General bivariate MH. Red point for Men and sign + for women.

For the three real cases analyzed the general bivariate skew-t MH model (9) was indicated as the best fitted. That means that it seems worth considering the more general model to explain the variability of these data sets.

6. Concluding Remarks

The Mudholkar–Hutson skewing mechanism admits flexible extensions in both the univariate and multivariate cases. Stochastic representations of such extended models typically have corresponding likelihood functions that are somewhat complicated (for example Equations (4) and (9)). This is partly compensated for by the ease of simulation for such models using the representation in terms of latent variables (the U_i 's and Z_i 's in (2) or (8)). The data set analyzed in Section 5 illustrates the potential advantage of considering the extended bivariate Mudholkar–Hutson model, since the basic bivariate M–H model does not provide an acceptable fit to the “athletes” data. Needless to say, applying Occam’s razor, it would always be desirable to consider the hierarchy of the five nested bivariate models that were considered in Section 5, in order to determine whether one of the simpler models might be adequate to describe a particular data set. Indeed there may be other special cases of the model (4), intermediate between (4) and (5) say, that might profitably be considered for some data sets. However, it should be remarked that selection of such sub-models for consideration should be done prior to inspecting the data. The addition of the GBMH model to the data analysts tool-kit should provide desirable flexibility for modeling data sets which exhibit behavior somewhat akin to, but not perfectly adapted to, more standard bivariate Cauchy models.

Author Contributions: All the authors contributed significantly to this research article.

Funding: This research received no external funding.

Acknowledgments: The research of H. Varela and H.W. Gómez was supported by Grant SEMILLERO UA-2019 (Chile).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mudholkar, G.S.; Hutson, A.B. The epsilon-skew-normal distribution for analyzing near-normal data. *J. Stat. Plan. Inference* **2000**, *83*, 291–309. [[CrossRef](#)]
2. Arellano-Valle, R.B.; Gómez, H.W.; Quintana, F.A. Statistical inference for a general class of asymmetric distributions. *J. Stat. Plan. Inference* **2005**, *128*, 427–443. [[CrossRef](#)]
3. Gómez, H.W.; Torres, F.J.; Bolfarine, H. Large-Sample Inference for the Epsilon-Skew-t Distribution. *Commun. Stat. Theory Methods* **2007**, *36*, 73–81. [[CrossRef](#)]
4. Hansen, B.E. Autoregressive conditional density estimation. *Int. Econ. Rev.* **1994**, *35*, 705–730. [[CrossRef](#)]
5. Gómez, H.W.; Torres, F.J.; Bolfarine, H. Letter to the Editor. *Commun. Stat. Theory Methods* **2008**, *37*, 986. [[CrossRef](#)]
6. Johnson, N.; Kotz, S.L.; Balakrishnan, N. *Continuous Univariate Distributions*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 1994; Volume 1.
7. Behboodian, J.; Jamalizadeh, A.; Balakrishnan, N. A new class of skew-Cauchy distributions. *Stat. Probab. Lett.* **2006**, *76*, 1488–1493. [[CrossRef](#)]
8. Huang, W.J.; Chen, Y.H. Generalized skew-Cauchy distribution. *Stat. Probab. Lett.* **2007**, *77*, 1137–1147. [[CrossRef](#)]
9. Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
10. Cook, R.D.; Weisberg, S. *An Introduction to Regression Graphics*; John Wiley & Sons: New York, NY, USA, 1994.
11. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Article

A Note on Ordering Probability Distributions by Skewness

V. J. García ^{1,†}, M. Martel-Escobar ^{2,†} and F. J. Vázquez-Polo ^{2,*,†}

¹ Department of Statistics and Operations Research, University of Cádiz, 11002 Cádiz, Spain; victoriano.garcia@uca.es

² Department of Quantitative Methods & TiDES Institute, University of Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Las Palmas, Spain; maria.martel@ulpgc.es or francisco.vazquezpolo@ulpgc.es

* Correspondence: francisco.vazquezpolo@ulpgc.es; Tel.: +34-928-451-800

† These authors contributed equally to this work.

Received: 15 June 2018; Accepted: 10 July 2018; Published: 16 July 2018

Abstract: This paper describes a complementary tool for fitting probabilistic distributions in data analysis. First, we examine the well known bivariate index of skewness and the aggregate skewness function, and then introduce orderings of the skewness of probability distributions. Using an example, we highlight the advantages of this approach and then present results for these orderings in common uniparametric families of continuous distributions, showing that the orderings are well suited to the intuitive conception of skewness and, moreover, that the skewness can be controlled via the parameter values.

Keywords: positive and negative skewness; ordering; fitting distributions

MSC: 62E10; 62P99

1. Introduction

Detailed knowledge of the characteristics of probability models is desirable (if not essential) if data are to be modeled properly. In studying these properties, many authors have considered orderings within probability distribution families, according to diverse measuring criteria. The usual approach taken by researchers in this field is to evaluate or measure one or more theoretical characteristics of a given distribution and to study the effect produced by the value of its parameters on this measurement. In actuarial science, stochastic orders are widely used in order to make risk comparisons [1].

Some parametric distributions can be ordered according to the evaluation made of a given property, merely by comparing some of its parameters. Although most related orders are actually preorders, each one presents interesting applications. Many studies have been conducted in this area, and the following are particularly significant: Lehmann (1955) [2], which is of seminal importance; Arnold (1987) [3], who compared random variables according to stochastic ordering in a particular Lorenz order; Shaked and Shanthikumar (2006) [1], on stochastic orders; Nanda and Shaked (2001) [4], on reversed hazard rate orders; Ramos-Romero and Sordo-Díaz (2001) [5], on the likelihood ratio order; and Gupta and Aziz (2010) [6], on convex orders.

In this paper, we study the relationship between the skewness of some parametric distributions and the value of one of their parameters. The first question to be addressed is that of measuring the skewness. In this respect, Oja (1981) [7] introduced a set of axioms to be verified by any measurement of skewness considered. These axioms were established for indexes of skewness with one main constraint: that the skewness of a distribution should be evaluated by a single real number. This point is discussed below.

Many authors have proposed and obtained different descriptive elements to measure skewness (see, for instance, [8–13]). Ref [10] suggested a measurement of skewness corresponding to the (unique) mode, M , given by the following index:

$$\gamma_M(F) = 1 - 2F(M). \quad (1)$$

Ref [10] applied this index to ordering the gamma, log-logistic, lognormal and Weibull families of distributions by their skewness, taking into account the feasible values of their respective parameters. Index (1), which is proven to satisfy those axioms derived from Oja (1981) [7], is also recommended in [14] as a (very) good index of skewness. However, notice that (1) only compares the probability weight on the left side of a central point (the mode) with the value $1/2$, but it does not account for how the weights are distributed to each side of the centre.

García et al. (2015) [15] introduced some further elements to be incorporated into the list of skewness measurements of a probability distribution. According to these authors, given a unimodal probability distribution $F(x)$, its skewness is considered to be a local function of a given distance, z , from the mode, M . For such a distance, and given the interval $[M - z, M + z]$, the aggregate skewness function, $\nu_F(z)$, compares the probability weight of F at either side of the interval:

$$\nu_F(z) = \Pr(X > M + z) - \Pr(X < M - z), \quad (2)$$

where $z \geq 0$. Thus, the (maximum) right skewness of the distribution F and its (minimum) left skewness are respectively given by

$$S^+(F) = \max_{z \geq 0} \nu_F(z), \quad S^-(F) = \min_{z \geq 0} \nu_F(z). \quad (3)$$

The distances, z_p and z_n , where these extreme values are achieved, are termed the critical distances to the mode. As the skewness function is bounded inside the interval $[-1, 1]$ and $\nu_F(\infty) = 0$, the bivariate index $(S^-(F), S^+(F))$ belongs to $[-1, 0] \times [0, 1]$. A given distribution function F such that $\nu_F(z) \geq 0$ for all $z \geq 0$ is said to be only skewed to the right; and if $\nu_F(z) \leq 0$ for all $z \geq 0$, it is said to be only skewed to the left.

The relationship $F <_c G$ (F c -precedes G) means that $G^{-1}[F(x)]$ is a convex function. For a continuous distribution F , the bivariate measurement of skewness $(S^-(F), S^+(F))$ verifies the following properties, where $aF + b$ and $-F$ mean the distributions of the corresponding transformation of a random variable that is F -distributed:

1. $(S^-(F), S^+(F)) = \mathbf{0}$, for any symmetric distribution F .
2. $(S^-(aF + b), S^+(aF + b)) = (S^-(F), S^+(F))$, for all $a > 0, -\infty < b < \infty$.
3. $(S^-(-F), S^+(-F)) = (-S^+(F), -S^-(F))$.
4. If $F <_c G$, then $(S^-(F), S^+(F)) \leq (S^-(G), S^+(G))$, understood as vector dominance.

These properties can be considered as a vectorial interpretation of the axioms given by Oja (1981) [7].

As it is easily proven that $\nu_F(0) = \gamma_M(F)$, we can establish that (2) and (3) give considerably clearer and more complete information than (1) about the skewness of any distribution function.

Most families of continuous distributions are only skewed to the right (or only to the left), while doubles-sign skewness is abundant within the discrete families, as shown in [15]. Nevertheless, the joint use of the function (2) and the bivariate index (3) makes it possible to improve the ordering of the skewness-based distribution discussed in [10], as can be seen in the following example.

Example 1. Assume the following random variable $X \in [-2, \infty)$ with PDF given by:

$$f(x) = \begin{cases} \frac{1}{4}x + \frac{1}{2}, & -2 \leq x < 0; \\ \frac{1}{2} \exp(-x), & x \geq 0. \end{cases}$$

Assume also the PDF, $g(y)$ of $Y = -X$. Then, $\gamma_M(F) = 0 = \gamma_M(G)$. That is, according to coefficient $\gamma_M(\cdot)$, both distributions have the same null skewness, although they do not even have a symmetric support set. However, using expression (2), we find that

$$v_F(z) = \begin{cases} \frac{1}{2} \exp(-z) - \frac{1}{8} (z-2)^2, & 0 \leq z < 2; \\ \frac{1}{2} \exp(-z), & z \geq 2, \end{cases}$$

and $v_G(z) = -v_F(z)$, for all $z \geq 0$. These functions are plotted in Figure 1, where it can be seen that $v_F(z) \geq v_G(z)$ for all $z \geq 0$, $S^+(F) = -S^-(G) \neq 0$ and $S^-(F) = 0 = S^+(G)$. Clearly, the information about skewness obtained from the aggregate skewness function $v(z)$ and the indices $S^+(\cdot)$ and $S^-(\cdot)$ is considerably more comprehensive than that obtained from $\gamma_M(\cdot)$.

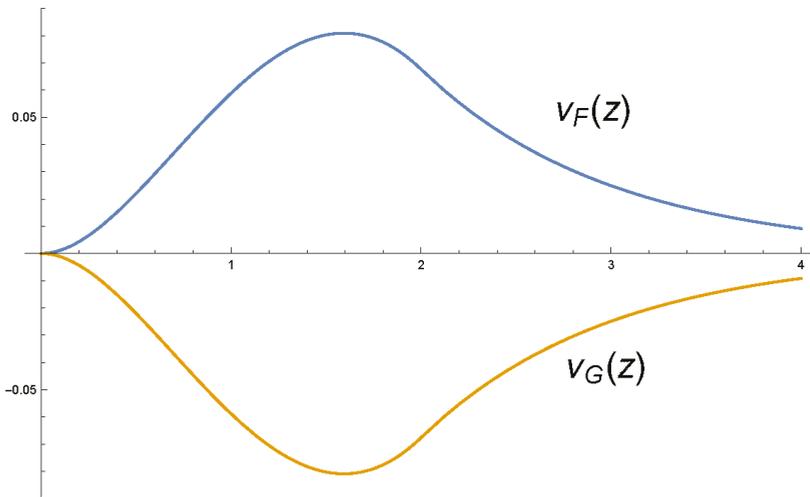


Figure 1. Skewness functions $v_F(z)$ and $v_G(z)$ in Example 1.

Outline

In applied statistical analysis, it is useful to have a large catalogue of plausible distributions with which to fit the data. According to García et al. (2015) [15], common measures of skewness can be complemented with a bivariate index of positive–negative skewness, and the authors show that the mode is the relevant central value to study both right and left skewness. In this paper, we extend the tool-box approach to fit data from probability distributions, introducing two orderings that are deduced from the skewness measures given in [15]. The first of those orderings is based on the positive part of the bivariate index of skewness, which in many instances coincides with the well known $\gamma_M(F)$. Nevertheless, the differences can be highly significant, as in the previous example. The second, more

noteworthy, order is based on the skewness function $v_F(z)$ and meets the first of the conditions, but not the reciprocal.

There are two reasons for ordering a family of distributions according to a given measurement of skewness. Firstly, as a property of the distribution, this ordering allows us to control its skewness by the appropriate selection of the parameter. When this is done (and the parameter is readily determined), the theoretical results have immediate applications in the data-fitting process. Secondly, when a given family of distributions is conceived as being more or less skewed according to the value of a parameter, and a measurement of skewness ratifies the ordering, it may be concluded that the functioning of this measurement provides a reasonably good fit with an intuitive conception of skewness.

The rest of this paper is organized as follows. In Section 2, we study the aggregate skewness function and the resultant skewness-based ordering of the gamma, log-logistic, lognormal, Weibull and asymmetric Laplace families of continuous probability distributions. In Section 3, we study the ordering of two of the most well-known distributions commonly used in PERT methods: the beta and the asymmetric triangular distributions. Finally, conclusions are presented in Section 4.

2. Families of Uniparametric Distributions Ordered by Skewness

Let F and G be unimodal distribution functions, with no centre or scale parameters, and modes M_F and M_G , respectively. We compare their respective skewness by two different criteria.

Definition 1. If

$$v_F(z) - v_G(z) \geq 0, \quad \forall z \geq 0, \quad (4)$$

then we say that F has equal or more aggregate skewness to the right at any point than G . We denote this by $F \geq_v G$.

Definition 2. If F and G are both skewed only to the right, we say that F has equal or more maximum aggregate skewness to the right than G when

$$S^+(F) - S^+(G) \geq 0, \quad (5)$$

and we denote this by $F \geq_+ G$.

With these definitions, it immediately follows that:

Proposition 1. If $F \geq_v G$, then $F \geq_+ G$.

The reverse implication is not true in general.

Proof. The proof follows immediately from the definitions given in (4) and (5). \square

In the next section, we consider some well known uniparametric families of continuous distributions, with no centre or scale parameters but depending on a skewness parameter, and examine whether they are ordered by aggregate skewness, or by maximum aggregate skewness. The gamma family is a very broad one, which includes many other well known distributions as particular cases. A study of the log-logistic, lognormal, Weibull and asymmetric Laplace families, one by one and in turn, when not included inside the previous one, will produce widely varying results.

2.1. Uniparametric Gamma Distributions

Let X be a uniparametric gamma distributed random variable, $G(\alpha)$. That is, its CDF $G(x; \alpha)$ is given by

$$G(x; \alpha) = \frac{1}{\Gamma(1 + \alpha)} \int_0^x t^\alpha e^{-t} dt, \quad (6)$$

for $x > 0$, where $-1 < \alpha < \infty$, and the mode is given by $M = \max \{\alpha, 0\}$. Then, for $-1 < \alpha \leq 0$, the density function decreases on x along the positive real line and we obtain that

$$v_G(z; \alpha) = 1 - G(z; \alpha) = \frac{1}{\Gamma(1 + \alpha)} \int_z^\infty t^\alpha e^{-t} dt.$$

In these cases, $v_G(z; \alpha)$ is a decreasing function on z , and $S^+(G(\alpha_1)) = v_G(0; \alpha_1) = 1$.

Proposition 2. Let $G(\alpha_1)$ and $G(\alpha_2)$ be gamma distributions with CDF as in (6). Then:

1. If $-1 < \alpha_1 < \alpha_2 < 0$, then $G(\alpha_2) \geq_v G(\alpha_1)$.
2. If $0 < \alpha_1 < \alpha_2$, then $G(\alpha_1) \geq_+ G(\alpha_2)$.

Proof. Part 1. We can write

$$v_G(z; \alpha_1) - v_G(z; \alpha_2) = G(z; \alpha_2) - G(z; \alpha_1).$$

By denoting $\alpha_2 = \alpha_1 + \varepsilon$, $\varepsilon > 0$, and then considering $u(z) = \frac{d}{dz} [v_G(z; \alpha_1) - v_G(z; \alpha_2)]$, we obtain

$$u(z) = \frac{z^{\alpha_2} e^{-z}}{\Gamma(1 + \alpha_2)} - \frac{z^{\alpha_1} e^{-z}}{\Gamma(1 + \alpha_1)} = z^{\alpha_1} e^{-z} \left[\frac{z^\varepsilon \Gamma(1 + \alpha_1) - \Gamma(1 + \alpha_2)}{\Gamma(1 + \alpha_2) \Gamma(1 + \alpha_1)} \right].$$

Therefore, $u(z) = 0$ when

$$z = z_0 = \left[\frac{\Gamma(1 + \alpha_2)}{\Gamma(1 + \alpha_1)} \right]^{1/\varepsilon},$$

$u(z)$ is negative for $0 < z < z_0$, and positive for $z > z_0$. Also, $v_G(0^+; \alpha_1) - v_G(0^+; \alpha_2) = 0$, $v_G(\infty; \alpha_1) - v_G(\infty; \alpha_2) = 0$. Then,

$$v_G(z_0; \alpha_1) - v_G(z_0; \alpha_2) = \frac{1}{\Gamma(1 + \alpha_1) \Gamma(1 + \alpha_2)} \int_0^{z_0} t^{\alpha_1} e^{-t} [\Gamma(1 + \alpha_1) t^\varepsilon - \Gamma(1 + \alpha_2)] dt,$$

is the integral of a negative function, so it is negative, and the proof is complete.

Part 2. For $0 < \alpha < \infty$, we have that

$$v_G(z; \alpha) = \begin{cases} \frac{1}{\Gamma(1+\alpha)} \left(\int_{z+\alpha}^\infty t^\alpha e^{-t} dt - \int_0^{\alpha-z} t^\alpha e^{-t} dt \right), & 0 \leq z < \alpha, \\ \frac{1}{\Gamma(1+\alpha)} \int_{z+\alpha}^\infty t^\alpha e^{-t} dt, & z \geq \alpha. \end{cases}$$

and,

$$\frac{dv_G}{dz} = \begin{cases} \frac{1}{\Gamma(1+\alpha)} \left[(\alpha - z)^\alpha e^{-(\alpha-z)} - (\alpha + z)^\alpha e^{-(\alpha+z)} \right], & 0 \leq z < \alpha, \\ \frac{-1}{\Gamma(1+\alpha)} z^\alpha e^{-z}, & z \geq \alpha. \end{cases}$$

Then, clearly we have that $dv_G/dz < 0$ for all $z \geq \alpha$. For $0 \leq z < \alpha$, if we denote

$$w(z) = (\alpha - z)^\alpha e^z - (\alpha + z)^\alpha e^{-z},$$

then the sign of dv_G/dz is the sign of $w(x)$. As $w(0) = 0$, $w(\alpha) = -(2\alpha)^\alpha e^{-\alpha} < 0$, and

$$\frac{dw}{dz} = z(z + \alpha)^{\alpha-1} e^{-z} - z(\alpha - z)^{\alpha-1} e^z \leq 0,$$

we conclude that v_G is a decreasing function on $z \geq 0$ and $S^+(G(\alpha)) = v_G(0; \alpha)$.

$$S^+(G(\alpha)) = \frac{\Gamma(1 + \alpha, \alpha)}{\Gamma(1 + \alpha)},$$

where $\Gamma(1 + \alpha, \alpha)$ is the incomplete Gamma function, and then $S^+(G(\alpha))$ is a decreasing function on α , when $\alpha \rightarrow \infty$. Nevertheless, a simple plotting of the functionals $v_G(z; \alpha_i)$ for any $0 < \alpha_1 < \alpha_2$ shows that both functionals cross each other and that they are not ordered by " \geq_v ". Thus, the proof is completed. \square

2.2. Log-Logistic Distributions

The CDF of a uniparametric log-logistic distributed random variable X is given by

$$F_{LL}(x; \theta) = \left(1 + x^{-\theta}\right)^{-1}, \tag{7}$$

for $x > 0$, with $\theta > 0$. The mode of these distributions depends on θ . If $0 < \theta \leq 1$, then $M = 0$, and

$$v_{LL}(z; \theta) = \frac{1}{1 + z^\theta},$$

and $S^+(F_{LL}(\theta)) = 1$. The functionals $v_{LL}(z; \theta)$ for different values of θ inside the rank cross each other at $z = 1$, and these distributions are ordered neither by skewness function nor by skewness indexes. Nevertheless, for $\theta > 1$, the mode is

$$0 < M = \left(\frac{\theta - 1}{\theta + 1}\right)^{1/\theta} < 1. \tag{8}$$

Notice that M is an increasing function of θ when $\theta > 1$, because

$$\frac{dM}{d\theta} = M \cdot \left[\frac{2}{(\theta^2 - 1)\theta} - \frac{1}{\theta^2} \ln \frac{\theta - 1}{\theta + 1} \right] > 0. \tag{9}$$

When $\theta > 1$, it is also known from Arnold and Groeneveld (1995) that

$$v_{LL}(0; \theta) = \frac{1}{\theta}.$$

As $v_{LL}(z; \theta)$ is a decreasing function, it is then stated that $1 < \theta_1 < \theta_2$ implies $F_{LL}(\theta_1) \geq_+ F_{LL}(\theta_2)$. Furthermore, the skewness functions are ordered, as we prove below.

Proposition 3. Let be $F_{LL}(\theta_1)$ and $F_{LL}(\theta_2)$ log-logistic distributions with CDF as in (7), where $1 < \theta_1 < \theta_2$. Then,

$$F_{LL}(\theta_1) \geq_v F_{LL}(\theta_2). \tag{10}$$

Proof. Let $\theta > 1$. Then,

$$v_{LL}(z; \theta) = \begin{cases} \frac{1 - (M^2 - z^2)^\theta}{1 + (M + z)^\theta + (M - z)^\theta + (M^2 - z^2)^\theta}, & 0 \leq z \leq M, \\ \frac{1}{1 + (M + z)^\theta}, & z > M, \end{cases}$$

If we consider $1 < \theta_1 < \theta_2$, such that the respective modes verify $0 < M_1 < M_2 < 1$, we can then denote

$$a = (M_1 - z)^{\theta_1} < b = (M_1 + z)^{\theta_1},$$

$$c = (M_2 - z)^{\theta_2} < d = (M_2 + z)^{\theta_2}.$$

and consider the function h given by

$$h(\theta) = (M \pm z)^\theta, \quad 0 \leq z \leq M,$$

with M as in (8). Then,

$$\frac{dh}{d\theta} = \frac{(M \pm z)^{\theta-1}}{\theta(\theta+1)^2} \left(\frac{2\theta}{M^{\theta-1}} + \theta(1+\theta)^2(M \pm z) \ln(M \pm z) + (1+\theta)^2 M \ln \frac{\theta+1}{\theta-1} \right),$$

For $z < M$, this implies that $a < c, b < d$. With this notation, we can write $v_{LL}(z; \theta_1) - v_{LL}(z; \theta_2)$ as follows.

Firstly, for $0 \leq z \leq M_1$,

$$v_{LL}(z; \theta_1) - v_{LL}(z; \theta_2) = \frac{1-ab}{(1+a)(1+b)} - \frac{1-cd}{(1+c)(1+d)}$$

$$= \frac{(c-a) + (d-b) + ac(d-b) + bd(c-a) + 2(cd-ab)}{(a+1)(b+1)(c+1)(d+1)} > 0.$$

Secondly, for $M_1 < z \leq M_2$, we only need to compare $d - b$, because

$$v_{LL}(z; \theta_1) - v_{LL}(z; \theta_2) = \frac{1}{1+b} - \frac{1-cd}{(1+c)(1+d)}$$

$$= \frac{c + (d-b) + 2cd + bcd}{(1+b)(1+c)(1+d)} > 0.$$

Finally, when $z > M_2$,

$$v_{LL}(z; \theta_1) - v_{LL}(z; \theta_2) = \frac{1}{1+b} - \frac{1}{1+d} = \frac{d-b}{(1+b)(1+d)} > 0.$$

Hence, the proof is completed. \square

2.3. Lognormal Variance Distributions

$$LN(x; \sigma) = \Phi\left(\frac{\ln x}{\sigma}\right), \tag{11}$$

for $x, \sigma > 0$, where $\Phi(\cdot)$ is the standard normal distribution function. The mode is given by $M_\sigma = \exp(-\sigma^2)$ and

$$v_{LN}(z; \sigma) = 1 - \Phi\left(\frac{\ln[z + \exp(-\sigma^2)]}{\sigma}\right) - \Phi\left(\frac{\ln[\exp(-\sigma^2) - z]}{\sigma}\right).$$

Proposition 4. Let $LN(\sigma_1)$ and $LN(\sigma_2)$ be lognormal distributions with CDF as in (11), where $0 < \sigma_1 < \sigma_2$. Then,

$$LN(\sigma_2) \geq_v LN(\sigma_1).$$

Proof. For $0 < \sigma_1 < \sigma_2$, the corresponding modes are $M_1 > M_2$, and

$$\Phi\left(\frac{\ln(z + M_1)}{\sigma_1}\right) > \Phi\left(\frac{\ln(z + M_2)}{\sigma_2}\right),$$

$$\Phi\left(\frac{\ln(-z + M_1)}{\sigma_1}\right) > \Phi\left(\frac{\ln(-z + M_2)}{\sigma_2}\right),$$

because Φ is a strictly increasing function. Thus, we obtain that $v_{LN}(z; \sigma_1) > v_{LN}(z; \sigma_2)$ for all $z > 0$ and the proof is completed. \square

2.4. Uniparametric Weibull Distributions

Consider the uniparametric Weibull distributions family given by the CDF

$$W(x; c) = 1 - \exp(-x^c), \quad x > 0, c > 0. \tag{12}$$

The mode is known to be at 0, for $c \leq 1$ (as a limit, when $c < 1$) and at

$$0 < M_c = \left(\frac{c-1}{c}\right)^{1/c} < 1,$$

for $c > 1$. The expression for v_W is given by

$$v_W(z; c) = \begin{cases} \exp[-(M_c + z)^c] + \exp[-(M_c - z)^c] - 1, & 0 < z < M_c \\ \exp[-(M_c + z)^c], & z \geq M_c \end{cases}$$

On the one hand, when $c < 1$, note that $v_{W(c)}(1) = e^{-1}$, so all these functions intersect at this point. Graphically, it can be seen that there is no ordering by “ \geq_v ”, and also that $S^+(W(c)) = 1$, when $c < 1$. On the other hand, for $1 \leq c_1 < c_2$, the following result is obtained.

Proposition 5. Let $W(c_1)$ and $W(c_2)$ be Weibull distributions with $1 \leq c_1 < c_2$ and CDF as in (12). Then,

$$W(c_1) \geq_v W(c_2).$$

Proof. For $1 \leq c_1 < c_2$, the corresponding modes are $M_1 < M_2$. Then, for $0 < z < M_1$,

$$v_W(z; c_1) - v_W(z; c_2) = \{ \exp[-(M_1 + z)^{c_1}] - \exp[-(M_2 + z)^{c_2}] \}$$

$$+ \{ \exp[-(M_1 - z)^{c_1}] - \exp[-(M_2 - z)^{c_2}] \} > 0,$$

because each part of the expression inside brackets $\{\cdot\}$ is positive. If we take $M_1 \leq z < M_2$, then

$$v_W(z; c_1) - v_W(z; c_2) = \{ \exp[-(M_1 + z)^{c_1}] - \exp[-(M_2 + z)^{c_2}] \}$$

$$+ \{ 1 - \exp[-(M_2 - z)^{c_2}] \} > 0,$$

for a similar reason. Finally, if we take $z > M_2$, then

$$v_W(z; c_1) - v_W(z; c_2) = \exp[-(M_1 + z)^{c_1}] - \exp[-(M_2 + z)^{c_2}] > 0,$$

and the proof is completed. \square

2.5. Asymmetric Laplace Distributions

The asymmetric Laplace distribution has been introduced in the literature by different ways ([16,17]). In this paper we will use Kozubowski and Podgórski (2002) [18] (later refined in [19]) to refer it. This distribution is obtained by using the scheme introduced by Fernández and Steel (1998) [20] to produce skewness on a symmetric distribution. In this way, the pdf of a skewed or asymmetric Laplace distribution can be written in the form

$$f(x; \mu, \sigma, \kappa) = \begin{cases} \frac{\sqrt{2}}{\sigma} \frac{\kappa}{1+\kappa^2} \exp\left[-\frac{\sqrt{2}}{\kappa\sigma}(\mu - x)\right], & x < \mu, \\ \frac{\sqrt{2}}{\sigma} \frac{\kappa}{1+\kappa^2} \exp\left[-\frac{\kappa\sqrt{2}}{\sigma}(x - \mu)\right], & x \geq \mu, \end{cases}$$

where $\sigma, \kappa > 0$, and $-\infty < \mu < \infty$. Then, we assign values (0, 1) to the centre and scale parameters (μ and σ , respectively) in order to study the aggregate skewness function, and the extreme right and left skewness indices then depend only on the skewness parameter $\kappa > 0$. Thus, it is easily proven that:

1. The aggregate skewness function of an AL (κ) distribution can be written as

$$v_{AL}(z; \kappa) = \frac{1}{1 + \kappa^2} \left[\exp(-\sqrt{2}\kappa z) - \kappa^2 \exp\left(-\frac{\sqrt{2}}{\kappa}z\right) \right].$$

2. $v_{AL}(z; \kappa)$ is an increasing negative function of z when $\kappa > 1$, and it is a decreasing positive function of z when $0 < \kappa < 1$. $v_{AL,1}(z; 1) = 0$, for all $z \geq 0$. That is, any AL distribution is skewed only to the right or to the left, depending on κ . In any case, the function verifies $\lim_{z \rightarrow \infty} v_{AL}(z; \kappa) = 0$ but, when $\kappa \neq 1$, the function never reaches that limit value. To prove these results, it is sufficient to note that

$$\frac{dv_{AL}(z; \kappa)}{dz} = \frac{\sqrt{2}\kappa}{\kappa^2 + 1} \left[\exp\left(-\frac{\sqrt{2}}{\kappa}z\right) - \exp(-\sqrt{2}\kappa z) \right].$$

3. At $z = 0$, the skewness function takes the following value:

$$v_{AL}(0; \kappa) = \frac{1 - \kappa^2}{1 + \kappa^2}.$$

Then, $v_{AL}(0; \kappa)$ is the value for $S^+(F_{AL}(\kappa))$ or $S^-(F_{AL}(\kappa))$, depending on its sign.

4. $v_{AL}(z; \kappa)$ is a strictly decreasing function on κ . This is easily shown by means of

$$\frac{dv_{AL}(z; \kappa)}{d\kappa} = -\frac{\sqrt{2}z\kappa^2 + 2\kappa + \sqrt{2}z}{(\kappa^2 + 1)^2} \left[\exp\left(-\sqrt{2}\frac{z}{\kappa}\right) + \exp(-\sqrt{2}\kappa z) \right] < 0,$$

for all $z > 0$, and all $\kappa > 0$.

As a conclusion, we can enunciate the following Proposition, whose proof is straightforward and hence omitted.

Proposition 6. Assume $0 < \kappa_1 < \kappa_2 < \infty$, and let $F_{AL}(\kappa_1)$ and $F_{AL}(\kappa_2)$ be the respective asymmetric Laplace distributions. Then:

1. $F_{AL}(\kappa_1) \geq_v F_{AL}(\kappa_2)$.
2. If $0 < \kappa_1 < 1$, then $F_{AL}(\kappa_1)$ is skewed only to the right.
3. If $\kappa_1 > 1$, then $F_{AL}(\kappa_1)$ is skewed only to the left.

3. The Beta and the AST Distributions

The methods for Project Management and Review Technique (PERT) are well known and widely applied when the needed activities for a given project must be ordered according to precedence in

time. Some of these methods require modelling the time length of each activity as a random variable, following an expert’s opinion. The beta and the asymmetric triangular distributions are commonly used by engineers to describe these time lengths. In any case, the indications of the experts can be related to a maximum and a minimum values and a mode, often completed with further considerations about the shape and skewness of the PDF of the time random variable. Then, a deep study of the skewness of both families of probability distributions would be welcome to improve the model fit.

On the one hand, the asymmetric standard triangular distribution (ASTD), free of center and scale parameters, depends on only one parameter $0 \leq \theta \leq 1$, and has the pdf:

$$f(x|\theta) = \begin{cases} 2x\theta^{-1}, & 0 \leq x \leq \theta, \\ 2(1-x)(1-\theta)^{-1}, & \theta \leq x \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

There is a large body of literature that shows the use of the ASTD in PERT methods (see [21] and [19] and cites therein). Note that cases $\theta = 0, 1$ are members of the beta family of distributions.

For $0 < \theta < 1$, the ASTD(θ) CDF can be written as follows:

$$F(x|\theta) = \begin{cases} x^2\theta^{-1}, & 0 \leq x \leq \theta, \\ (2x - x^2 - \theta)(1-\theta)^{-1}, & \theta \leq x \leq 1. \end{cases}$$

As the mode is found to be at $x = \theta$, its skewness function is found to be

$$v_{ASTD}(z; \theta) = (1 - 2\theta) - \frac{(1 - 2\theta)}{\theta(1 - \theta)}z^2,$$

for $0 \leq z \leq \min\{\theta, 1 - \theta\}$. In the case $\theta = 0.5$, the skewness function is null. Then, for $0 < \theta < 0.5$ and $\theta < z \leq 1 - \theta$,

$$v_{ASTD}(z; \theta) = \frac{(z - 1 + \theta)^2}{1 - \theta}.$$

In the case $0.5 < \theta < 1$, for $1 - \theta < z \leq \theta$,

$$v_{ASTD}(z; \theta) = -\frac{(\theta - z)^2}{\theta},$$

and it is easily found that

$$v_{ASTD}(z; \theta) = -v_{ASTD}(z; 1 - \theta), \tag{13}$$

for $0 \leq z < \infty$.

Some algebra allows to prove that, being $0 < \theta_1 < \theta_2 < 1$,

1. $ASTD(\theta_1) \geq_v ASTD(\theta_2)$.
2. If $0 < \theta_i < 0.5$, then $S_{ASTD}^+(\theta_i) = v_{ASTD}(0; \theta_i) = 1 - 2\theta_i$, and $S_{ASTD}^-(\theta_i) = 0$.
3. If $0.5 < \theta_i < 1$, then $S_{ASTD}^+(\theta_i) = 0$ and $S_{ASTD}^-(\theta_i) = v_{ASTD}(0; \theta_i) = 1 - 2\theta_i$.

Therefore, the skewness of the ASTD distributions is completely controlled by the parameter θ .

On the other hand, the pdf of a beta distribution is given by

$$f_B(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 \leq x \leq 1,$$

where $\alpha, \beta > 0$, and $B(\cdot, \cdot)$ is the beta function. Given that its CDF $F(x; \alpha, \beta)$ verifies that $F(x; \alpha, \beta) = 1 - F(x; \beta, \alpha)$ and the sign of its skewness depends only on the condition $\beta \geq \alpha$ or $\beta \leq \alpha$, we can study only the case $\beta > \alpha$.

We are interested on the cases $\alpha, \beta > 1$, where there is an unique mode M ,

$$M = \frac{\alpha - 1}{\alpha + \beta - 2} \doteq b(\alpha, \beta).$$

Hence, we only consider cases where $1 < \alpha < \beta$, where there exists a right skewness; the cases $1 < \beta < \alpha$, with left skewness, can be immediately deducted by taking the parameters in reverse.

Notice that $\Pr(X > M + z) > 0$ requires $0 \leq z \leq b(\beta, \alpha)$, and that $\Pr(X < M - z) > 0$ requires $0 \leq z \leq b(\alpha, \beta)$. Then,

$$v_B(z; \alpha, \beta) = \begin{cases} 1 - I_{M+z}(\alpha, \beta) - I_{M-z}(\alpha, \beta) > 0, & 0 \leq z \leq b(\alpha, \beta) \\ 1 - I_{M+z}(\alpha, \beta) > 0, & b(\alpha, \beta) < z \leq b(\beta, \alpha) \\ 0, & z > b(\beta, \alpha), \end{cases} \tag{14}$$

where,

$$I_z(\alpha, \beta) = \int_0^z \frac{t^{\alpha-1} (1-t)^{\beta-1}}{B(\alpha, \beta)} dt$$

is the well known Beta Regularized function.

Firstly, observe that

$$v_B(0; \alpha, \beta) = 1 - \frac{2}{B(\alpha, \beta)} \int_0^M x^{\alpha-1} (1-x)^{\beta-1} dx,$$

and

$$\frac{1}{B(\alpha, \beta)} \int_0^M x^{\alpha-1} (1-x)^{\beta-1} dx < \frac{1}{B(\alpha, \beta)} \int_0^m x^{\alpha-1} (1-x)^{\beta-1} dx \approx \frac{1}{2},$$

where

$$m = \frac{\alpha - \frac{1}{3}}{\alpha + \beta - \frac{2}{3}}$$

is the approximate median of the distribution.

Secondly, if $0 \leq z \leq b(\alpha, \beta) < b(\beta, \alpha)$, then

$$B(\alpha, \beta) \cdot \frac{dv_B(z; \alpha, \beta)}{dz} = -[b(\alpha, \beta) + z]^{\alpha-1} [b(\beta, \alpha) - z]^{\beta-1} - [b(\alpha, \beta) - z]^{\alpha-1} [b(\beta, \alpha) + z]^{\beta-1},$$

which is negative within the rank of z . For $b(\alpha, \beta) < z \leq b(\beta, \alpha)$,

$$B(\alpha, \beta) \cdot \frac{dv_B(z; \alpha, \beta)}{dz} = -[b(\alpha, \beta) + z]^{\alpha-1} [b(\beta, \alpha) - z]^{\beta-1} < 0.$$

Hence, for $0 \leq z \leq b(\beta, \alpha)$, $v_B(z; \alpha, \beta)$ is a strictly decreasing continuous function with $v_B(0; \alpha, \beta) > 0$ and $v_B(b(\beta, \alpha); \alpha, \beta) = 0$.

Now we focus on the family of Beta distributions with given mode, M . That is, we consider the subfamily of Beta distributions:

$$B\left(\alpha + 1, 1 + \frac{1 - M}{M} \alpha\right),$$

with $\alpha > 0$. Then, with the aid of a proper software (we have used Wolfram Mathematica 10), one can obtain the derivative

$$\frac{\partial}{\partial \alpha} v_B\left(z; \alpha + 1, 1 + \frac{1 - M}{M} \alpha\right),$$

and maximize this function, in two cases:

First case, the constrains are $\alpha \geq 1$, $0 < m < 1/2$, $0 \leq z \leq b(\alpha + 1, 1 + (1 - M)\alpha/M)$. The maximum value of the function is 0, and it is achieved when $M = 0.5$, $\alpha \simeq 3.54147$, $z \simeq 0.309936$.

Second case, the constrain are $\alpha \geq 1$, $0 < m < 1/2$, $b(\alpha + 1, 1 + (1 - M)\alpha/M) < z \leq b(1 + (1 - M)\alpha/M, \alpha + 1)$. The maximum value of the function is -5.07056×10^{-6} , and it is achieved when $M = 0.123564$, $\alpha \simeq 1.62726$, $z \simeq 0.632457$.

With these results, we can conclude that $\nu_B(z; \alpha + 1, 1 + (1 - M)\alpha/M)$ decreases with the feasible values of α . That way, the subsets of Beta distributions with fixed mode are ordered on skewness (see Figure 2). As the parameter values increase, these Beta distributions become less skewed.

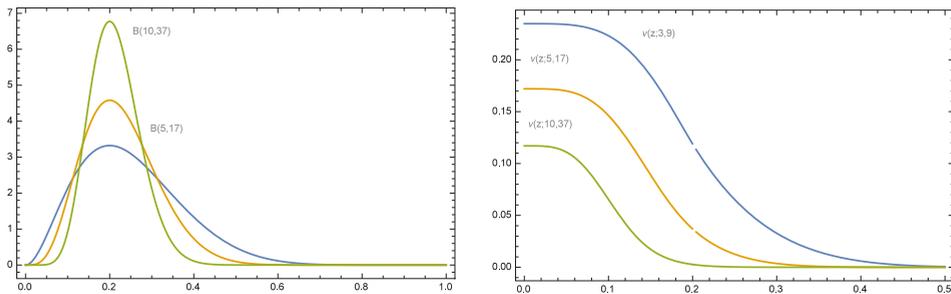


Figure 2. Beta distributions with common given mode $M = 0.2$ (left panel) for $\alpha = 2, 4$ and 9 and their skewness functions ν_B (right panel).

4. Conclusions

In this paper two main objectives are achieved: on the one hand, the given examples show that the skewness function orders the mesh in good accordance with the intuitive conception of skewness. Moreover, these examples show that the skewness of a distribution obtained from certain parametric families can be controlled by reference to their parameters.

As we show, the function $\nu_F(z)$ facilitates the description of a random variable by means of a probability distribution, by making any skewness in the model easily observable and should be undertaken to examine the use of these properties in data fitting.

In practice, much can be learned from this model, but there remains the risk that it may be wrongly specified in real applications. Thus, in practice we must be willing to assume that the underlying distribution has a unique mode and belongs to a uniparametric family of distributions.

In many practical situations, the maximum skewness index coincides with the well known $\gamma_M(F)$, but this second index only takes into account the difference of probability weights at each side of the mode, while the first takes a value from the point where this difference is maximum. Moreover, the aggregate skewness function gives more accurate information about how the probability weight is distributed along both sides of the mode. Accordingly, the condition $F \geq_\nu G$ provides highly valuable information.

Author Contributions: All authors have contributed equally to this paper.

Funding: This research received no external funding.

Acknowledgments: This research was partially funded by MINECO (Spain) grant number EC02017-85577-P. The authors are grateful for helpful suggestions made by two reviewers.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shaked, M.; Shanthikumar, J.G. Stochastic Orders. In *Springer Series in Statistics 43*; Springer: New York, NY, USA, 2007.
2. Lehmann, E.L. Ordered families of distributions. *Ann. Math. Statist.* **1955**, *26*, 399–419. [[CrossRef](#)]

3. Arnold, B.C. Majorization and the Lorenz Order: A brief introduction. In *Lecture Notes in Statistics* 43; Springer: New York, NY, USA, 1987.
4. Nanda, A.K.; Shaked, M. The hazard rate and reverse hazard rate orders, with applications to order statistics. *Ann. Inst. Statist. Math.* **2001**, *53*, 853–864. [[CrossRef](#)]
5. Ramos-Romero, H.M.; Sordo-Díaz, M.A. The proportional likelihood ratio order and applications. *Questiio* **2001**, *25*, 211–223.
6. Gupta, A.K.; Aziz, M.A.S. Convex Ordering of Random Variables and its Applications in Econometrics and Actuarial Science. *Eur. J. Pure Appl. Math.* **2010**, *3*, 79–85.
7. Oja, H. On location, scale, skewness and kurtosis of univariate distributions. *Scand. J. Stat.* **1981**, *8*, 154–168.
8. Van Zwet, W.R. Mean, Median, Mode II. *Stat. Neerl.* **1979**, *33*, 1–5. [[CrossRef](#)]
9. MacGillivray, H.L. Skewness and asymmetry: measures and orderings. *Ann. Stat.* **1986**, *14*, 994–1011. [[CrossRef](#)]
10. Arnold, B.C.; Groeneveld, R.A. Measuring Skewness with respect to the Mode. *Am. Stat.* **1995**, *49*, 34–38.
11. Sato, M. Some remarks on the mean, median, mode and skewness. *Aust. J. Stat.* **1997**, *39*, 219–224. [[CrossRef](#)]
12. Von Hippel, P.T. Mean, Median and Skew: Correcting a Textbook Rule. *J. Stat. Educ.* **2005**, *13*. [[CrossRef](#)]
13. Das, S.; Mandal, P.K.; Ghosh, D. On Homogeneous Skewness of Unimodal Distributions. *Indian J. Stat.* **2009**, *71-B*, 187–205.
14. Rubio, F.J.; Steel, M. On the Marshall–Olkin transformation as a skewing mechanism. *Comput. Stat. Data Anal.* **2012**, *56*, 2251–2257. [[CrossRef](#)]
15. García, V.J.; Martel-Escobar, M.; Vázquez-Polo, F.J. Complementary information for skewness measures. *Stat. Neerl.* **2015**, *69*, 442–459. [[CrossRef](#)]
16. Mc Gill, W.J. Random fluctuations of response rate. *Psychometrika* **1962**, *27*, 3–17. [[CrossRef](#)]
17. Holla, M.S.; Bhattacharya, S.K. On a compound Gaussian distribution. *Ann. Instit. Stat. Math.* **1968**, *20*, 331–336. [[CrossRef](#)]
18. Kozubowski, T.J.; Podgórski, K. Maximum likelihood estimation of asymmetric Laplace parameters. *Ann. Inst. Stat. Math.* **2002**, *54*, 816–826.
19. Kotz, S.; van Dorp, J.R. Uneven two-sided power distribution: with applications in econometric models. *Stat. Methods Appl.* **2004**, *13*, 285–313. [[CrossRef](#)]
20. Fernández, C.; Steel, M.F.J. On Bayesian Modeling of Fat Tails and Skewness. *J. Am. Stat. Assoc.* **1998**, *93*, 359–371.
21. Johnson, D. The triangular distribution as a proxy for the beta distribution in risk analysis. *The Statistician* **1997**, *46*, 387–398. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

Symmetry Editorial Office
E-mail: symmetry@mdpi.com
www.mdpi.com/journal/symmetry



MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland

Tel: +41 61 683 77 34
Fax: +41 61 302 89 18

www.mdpi.com



ISBN 978-3-03936-647-7