Journal Pre-proof

Bimodal normal distribution: Extensions and applications

E. Gómez-Déniz, J.M. Sarabia, E. Calderín-Ojeda

PII: S0377-0427(20)30583-5
DOI: https://doi.org/10.1016/j.cam.2020.113292
Reference: CAM 113292

To appear in: *Journal of Computational and Applied Mathematics*

Received date : 20 April 2020
Revised date : 4 October 2020

Please cite this article as: E. Gómez-Déniz, J.M. Sarabia and E. Calderín-Ojeda, Bimodal normal distribution: Extensions and applications, *Journal of Computational and Applied Mathematics* (2020), doi: https://doi.org/10.1016/j.cam.2020.113292.

Bimodal Normal Distribution: Extensions and Applications

[a]E. Gómez–Déniz, [b]J.M. Sarabia and [c]E. Calderín–Ojeda
[a]*Department of Quantitative Methods and Institute of Tourism and Sustainable Economic Development (TIDES). University of Las Palmas de Gran Canaria, Spain.*
[b]*CUNEF University, Department of Quantitative Methods, Leonardo Prieto Castro 2, 28040 Madrid, Spain*
[c] *Centre for Actuarial Studies, Department of Economics, The University of Melbourne, Australia*

Abstract

In this paper, a new family of continuous random variables with non-necessarily symmetric densities is introduced. Its density function can incorporate unimodality and bimodality features. Special attention is paid to the normal distribution which is included as a particular case. Its density function is given in closed-form which allows to easily compute probabilities, moments and other related measures such as skewness and kurtosis coefficients. Also, a stochastic representation of the family that enables us to generate random variates of this model is also presented. This new family of distributions is applied to explain the incidence of Hodgkin's disease by age. Other applications include the implications of bimodality in geoscience. Finally, the multivariate counterpart of this distribution is briefly discussed.

Key words: Bimodal Distribution; Folded Normal Distribution; Hyperbolic Function; Normal Distribution

Mathematics Subject Classification (2020): 62F03, 62A05.

Acknowledgements

The authors thank the Associate Editor and an anonymous referee for their constructive comments and suggestions, which have greatly helped us to improve the paper. EGD was partially funded by grant ECO2017-85577-P (Ministerio de Economía, Industria y Competitividad. Agencia Estatal de Investigación)) JMS was partially funded by grant PID2019-105986GB-C22 (Ministerio de Ciencia e Innovación).

Bimodal Normal Distribution: Extensions and Applications

Abstract

In this paper, a new family of continuous random variables with non-necessarily symmetric densities is introduced. Its density function can incorporate unimodality and bimodality features. Special attention is paid to the normal distribution which is included as a particular case. Its density function is given in closed-form which allows to easily compute probabilities, moments and other related measures such as skewness and kurtosis coefficients. Also, a stochastic representation of the family that enables us to generate random variates of this model is also presented. This new family of distributions is applied to explain the incidence of Hodgkin's disease by age. Other applications include the implications of bimodality in geoscience. Finally, the multivariate counterpart of this distribution is briefly discussed.

Key words: Bimodal Distribution; Folded Normal Distribution; Hyperbolic Function; Normal Distribution

Mathematics Subject Classification (2020): 62F03, 62A05.

1 Introduction

Bimodal distributions arise naturally in many different scenarios. Perhaps, as seen above, one of the most relevant phenomena that can be explained through these distributions is the disease patterns. By understanding the reason behind the multimodality of some cancer incidence curves, the practitioners may enhance their knowledge of the cancer developmental process and the potential features that identify cancer and that separate a particular type of cancer from all other types of cancer. Therefore, the importance of properly identifying cancer occurrences by age is vital to improve the tumor diagnosis. There exists some type of cancers where it is observed the existence of two spikes in occurrences. In this regard, there exists some type of cancers that have two peaks in occurrences, by age. Examples of recognized bimodal cancers include Kaposi's sarcoma and Hodgkin lymphoma. The latter type of cancer has two peaks in occurrence: in young adults and middle-aged adults. Hodgkin's lymphoma (see MacMahon (1966)) is uncommon cancer that develops in the lymphatic system, which is a network of vessels and glands spread throughout the organism.

The occurrence of bimodality has also implications in biogeoscience (see Hirota et al., 2011). Finding appropriate probabilistic models that can explain bivariate datasets is an issue of vital importance. In this work, we propose an extension of the normal distribution that may be either unimodal or bimodal. This new family of distributions arises from the folded normal distribution suggested by Leone et al. (1961). The latter model that generalizes the half-normal distribution has probability density function (pdf) given by the expression,

$$f(z) = \frac{1}{\sigma}\sqrt{\frac{2}{\pi}} \exp\left\{-\frac{1}{2\sigma^2}\left[z^2+(\sigma\theta)^2\right]\right\} \cosh\left(\frac{\theta z}{\sigma}\right), \quad z > 0.$$

It is worth mentioning that the folded normal distribution has been scarcely discussed in the literature, see for example Johnson (1962) and the more recent paper written by Gui et al. (2013). In the last decades, different techniques to extend the normal family have been deemed in the statistical literature. The skew-normal distribution in Azzalini (1985) (see also Azzalini, 1986), the beta-normal distribution suggested by Eugene et al. (2002) (see also Eugene, 2001, Famoye et al., 2004 and Rêgo et al., 2012), the Balakrishnan skew-normal density in Sharafi and Behboodian (2008) (more details in Teimouri and Nadarajah, 2016), the generalization proposed by Arnold and Beaver (2002), the Sinh-arcsinh family introduced by Jones and Pewsey (2009) and the generalized normal one in García et al. (2010), among others. For a comprehensive review of the skew normal families the reader is referred to Azzalini (2013).

In order to make the paper self-contained, we introduce here some functions that will be used throughout the paper. The basic hyperbolic sine and cosine functions are defined as,

$$\sinh(z) = \frac{1}{2}\left[\exp(z) - \exp(-z)\right], \quad \cosh(z) = \frac{1}{2}\left[\exp(z) + \exp(-z)\right]. \qquad (1)$$

From these two expressions the hyperbolic secant and tangent functions given by $\operatorname{sech}(z) = 1/\cosh(z)$ and $\tanh(z) = \sinh(z)/\cosh(z)$ are simply derived. Additionally, we will use $\phi(z)$ and $\Phi(z)$ to denote the standard normal ($N(0,1)$) density and distribution functions respectively and $\phi_{\mu,\sigma}(z)$ the normal density with mean μ and standard deviation σ.

The rest of this paper is structured as follows. In Section 2, the mechanism to derive the new family of distributions from which the proposed model is derived. Here, expressions for the mean, variance and the third and fourth standardized cumulative are also provided. In Section 3, the parameter estimation problem is discussed. Numerical applications are given in Section 4. The multivariate version of this model is presented in Section 5. Finally, conclusions and further comments are shown in the last section.

2 The proposed model

Let be $G(x)$ an absolutely continuous distribution function such that $G'(x) = g(x)$ is symmetric about zero. Then, the following generalization of $G(x)$ is proposed,

$$F_{\theta,\lambda}(x) = \frac{1}{2}\left[\exp(-\lambda\theta)G(x-\theta) + \exp(\lambda\theta)G(x+\theta)\right]\operatorname{sech}(\lambda\theta), \tag{2}$$

for $-\infty < x < \infty$, $\theta \in \mathbb{R}$ and $\lambda \in \mathbb{R}$. Obviously, $G(x) = F_{0,\lambda}(x)$. Natural choices for $G(x)$ to be plugged into (2) are the Cauchy distribution, the Student's t distribution, and the normal distribution that will be the one considered in the rest of this work, i.e. $G(x) = \Phi(x)$.

For this particular member of the family, the resulting pdf is

$$f_{\theta,\lambda}(x) = \sqrt{2\pi}\operatorname{sech}(\lambda\,\theta)\,\phi(\theta)\,\phi(x)\,T_{\theta,\lambda}(x), \quad -\infty < x < \infty, \tag{3}$$

where $T_{\theta,\lambda}(x) = \cosh(\theta(x-\lambda))$, $\theta \in \mathbb{R}$ and $\lambda \in \mathbb{R}$. Using (1) and the moment generating function of $\phi(x)$ it is straightforward to observe that (3) represents a genuine probability density function in $(-\infty, \infty)$. Obviously (3) reduces to the normal distribution, $N(0,1)$, when $\theta = 0$. Note that apart from the hyperbolic functions, no other special functions appear in the density (3). Hereafter, we will denote $X \sim GN(\theta, \lambda)$ for a random variate X that follows (3).

Both parameters θ and λ control the unique mode or the two modes (and the corresponding antimode) of the distribution that can be numerically obtained by solving the equation

$$\theta\,\tanh(\theta(x-\lambda)) - x = 0,$$

where parameters λ and θ play the role of the shape and location parameters. When $\lambda = 0$, (3) defines a family of symmetric distributions yielding heavier tails than the normal distribution. In this case, the distribution is unimodal when $|\theta| < 1$ and bimodal in the rest of the domain. Furthermore, the degree of skewness increases when λ grows. Positive skewness corresponds to the case $\lambda > 0$. This is confirmed by Figure 1 below where some plots of the pdf (3) for special cases of parameters are shown.

Simple calculations provide the moment generating function of the $GN(\theta, \lambda)$ distribution which is given by

$$M_X(t) = \operatorname{sech}(\theta\lambda)\cosh(\theta(t-\lambda))\exp\left(\frac{t^2}{2}\right). \tag{4}$$

The latter expression allows us to identify completely this model since according to Chiogna (1998) the *knowledge of the moments of a distribution is equivalent to a knowledge of the distribution function, in the sense that it should be possible to exhibit all the properties of the distribution in terms of the moments.* For the distribution introduced in this paper, odd and even non-central moments, are provided in closed-form. The odd moments are given by

$$\begin{aligned} E(X^r) &= \frac{1}{2}\sqrt{\frac{2^r}{\pi}}\left[1+\exp(2\theta\lambda)\right]\exp\left[-\frac{\theta}{2}(\theta+2\lambda)\right] \\ &\quad \times\Gamma\left(\frac{1+r}{2}\right)\,{}_1F_1\left(\frac{1+r}{2},\frac{1}{2},\frac{\theta^2}{2}\right), \quad r=1,3,\ldots, \end{aligned}$$

whereas the even moments result

$$\begin{aligned} E(X^r) &= \frac{\theta}{2}\sqrt{\frac{2^{r+1}}{\pi}}\left[1-\exp(2\theta\lambda)\right]\exp\left[-\frac{\theta}{2}(\theta+2\lambda)\right] \\ &\quad \times\Gamma\left(1+\frac{r}{2}\right)\,{}_1F_1\left(1+\frac{r}{2},\frac{3}{2},\frac{\theta^2}{2}\right), \quad r=2,4,\ldots, \end{aligned}$$

where ${}_1F_1$ is the Kummer confluent hypergeometric function.

In particular, the mean and variance of this distribution are given by,

$$\begin{aligned} E(X) &= -\theta\tanh(\theta\lambda), \\ var(X) &= 1+\left(\theta\operatorname{sech}(\theta\lambda)\right)^2. \end{aligned}$$

Although they are not derived in this work, the incomplete moments can also be obtained in closed-form. The third (skewness) and fourth (kurtosis)

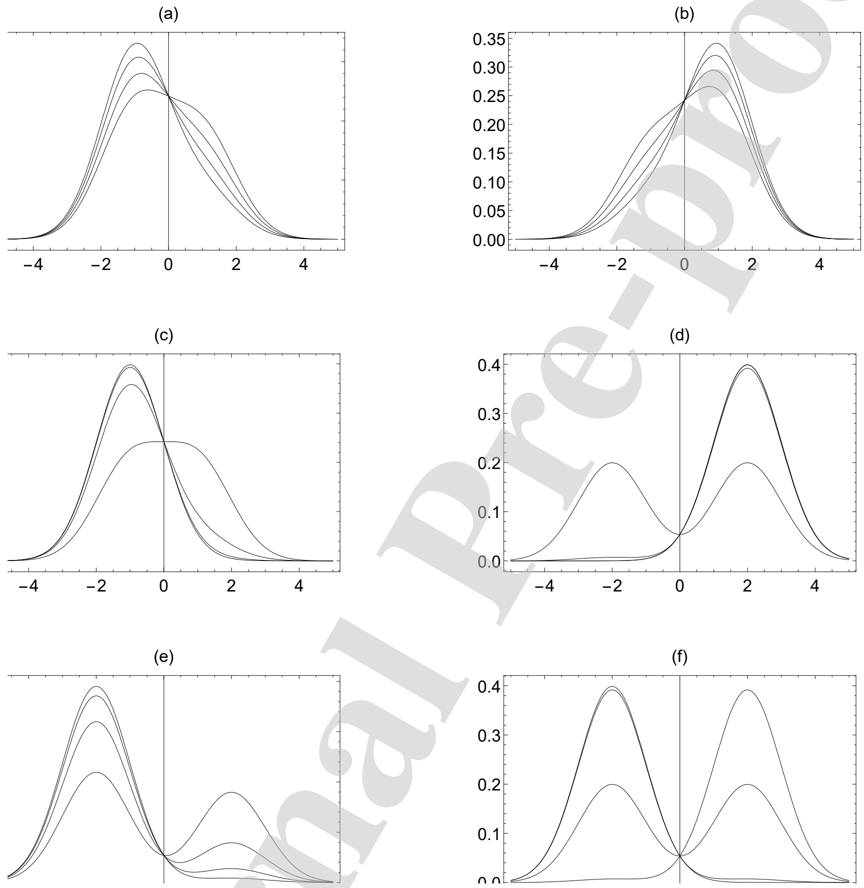

Figure 1: Plot of the probability density function (3) for selected values of the parameters. From left to right, panel (a): $\theta = 1$, $\lambda = 0.7, 0.5, 0.3, 0.1$; panel (b): $\theta = 1$, $\lambda = -0.2, -0.4, -0.6, -0.8$; panel (c): $\theta = 1$, $\lambda = 3, 2, 1, 0$; panel (d): $\theta = 2$, $\lambda = 0, -1, -2, -3$; panel (e): $\theta = 2$, $\lambda = 0.95, 0.65, 0.35, 0.05$; panel (f): $\theta = 2$, $\lambda = 2, 1, 0, -1$.

standardized cumulants are given by,

$$\gamma_1 = -\frac{2\theta^2 \mathrm{sech}^2(\theta\lambda)E(X)}{(var(X))^{3/2}},$$
$$\gamma_2 = \frac{8\theta^4(\cosh(2\theta\lambda)-2)}{(1+2\theta^2\cosh(2\theta\lambda))^2}.$$

The cumulative distribution function (cdf) is given by,

$$F(x) = \frac{1}{2}\left[\exp(-\theta\lambda)\Phi(x-\theta)+\exp(\theta\lambda)\Phi(x+\theta)\right]\mathrm{sech}(\theta\lambda).$$

The following proposition, which is stated without proof, displays some important properties of this distribution.

Proposition 1 *The $GN(\theta,\lambda)$ distribution satisfies the following properties:*

(*i*) $f_{\theta,\lambda}(0) = \phi(\theta)$.

(*ii*) $f_{\theta,\lambda}(x) = f_{-\theta,\lambda}(x)$.

(*iii*) $f_{0,\lambda}(x) = f_{0,0}(x) = \phi(x)$.

(*iv*) *If X follows the pdf (3) then $-X$ follows the same distribution but with the parameter λ replaced by $-\lambda$. Thus, we have that $F_{\theta,\lambda}(x) = 1 - F_{\theta,-\lambda}(-x)$.*

(*v*) $f_{\theta,\lambda}(x) + f_{\theta,-\lambda}(x) = f_{\theta,0}(x)$.

(*vi*) *Let $\theta = \lambda$ in (3) and consider the two random variates Z_1 and Z_2 following the pdf (3) with parameter $\lambda_1 < \lambda_2$, $\lambda_1 < (>)\lambda_2$, respectively, then $Z_1 <_{st} (>_{st})Z_2$. That is, Z_1 is stochastically smaller (larger) than Z_2.*

Since $f_{\theta,\lambda}(0) > 0$ then $E(X^{-1})$ will be infinite. Furthermore, higher negative moments does not exist. See for example Piegorsch and Casella (1985).

Figure 2 displays the hazard rate function, $r_{\theta,\lambda}(x) = f_{\theta,\lambda}(x)/(1-F_{\theta,\lambda}(x))$, of the $GN(\theta,\lambda)$ distribution for different values of the parameters θ and λ. It is observable that this function is monotonically increasing when $|\theta| < 1$ regardless of the values of the parameter λ (see panel (a)). When the parameter θ increases (panels (b) and (c)) and $\lambda > 0$, i.e. the density is right skewed, the hazard rate function has a local maximum (then the hazard rate is initially increasing and later decreasing). Moreover, if $\lambda_1 > \lambda_2$ then $r_{\theta,\lambda_1}(x) > r_{\theta,\lambda_2}(x)$. In addition, when $\lambda < 0$, the failure rate is monotonically increasing.

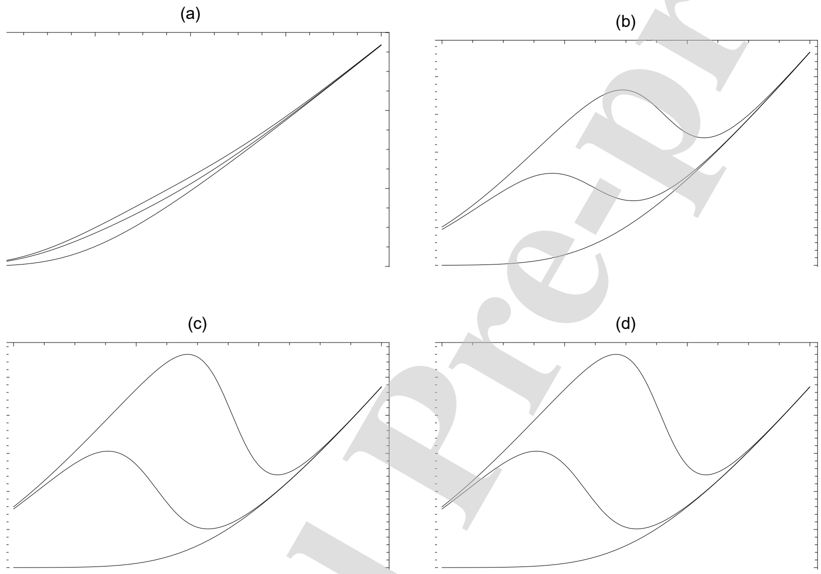

Figure 2: Plot of the hazard rate function for selected values of the parameters. From left to right and below to above, panel (a): $\theta = 0.5$, $\lambda = -5, 1, 2$; panel (b): $\theta = 1$, $\lambda = -2, 1, 2$; panel (c): $\theta = 1.5$, $\lambda = -5, 1, 2$; panel (d): $\theta = 2$, $\lambda = -2, 1, 2$.

2.1 Stochastic representation

The pdf (3) can also be written in the following way,

$$f_{\theta,\lambda}(x) = p f_{N(-\theta,1)}(x) + (1-p) f_{N(\theta,1)}(x), \quad x \in \mathbb{R}, \tag{5}$$

where

$$p = \operatorname{sech}(\lambda\theta)\exp(-\lambda\theta)/2, \tag{6}$$

and $f_{N(\mu,1)}(x)$ represents the pdf of a normally distributed random variable with mean μ and standard deviation one. Note that $0 < p < 1$. Then, if $X \sim GN(\theta, \lambda)$, it admits the stochastic representation,

$$X := \begin{cases} N(-\theta, 1) & \text{w.p.} \quad p, \\ N(\theta, 1) & \text{w.p.} \quad 1-p. \end{cases}$$

This stochastic representation allows us to simulate random variates from the model (3) in a straightforward manner.

2.2 Transformations

The family of distributions given in (3) can be generalized to incorporate a location and scale parameters using a linear transformation. Obviously, a more flexible model is obtained. For that reason, let us consider the location-scale generalization of the proposed distribution defined by the following change of variable $Y = \mu + \sigma X$, where $X \sim \mathrm{GN}(\theta, \lambda)$ given in (3), where $\mu \in \mathbb{R}$ and $\sigma > 0$. Its pdf is given by

$$f(x) = \sqrt{2\pi}\,\mathrm{sech}\psi\,\phi(\theta)\phi_{\mu,\sigma}(x)\cosh\left(T(x)\right), \quad -\infty < x < \infty, \tag{7}$$

where $T(x) \equiv T_{\theta,\lambda,\sigma}(x) = \frac{\theta}{\sigma}(x-\lambda)$, $\psi \equiv \psi(\theta, \lambda, \mu, \sigma) = \theta(\mu-\lambda)/\sigma$, $\theta \in \mathbb{R}$, $\sigma > 0$, $\mu \in \mathbb{R}$ and $\lambda \in \mathbb{R}$.

Recall that the hyperbolic tangent function can also be written as $\tanh(z) = 1-2(1+\exp(2z))^{-1}$ and the hyperbolic secant function as $\mathrm{sech}(z) = 2(\exp(z)+\exp(-z))^{-1}$.

3 Statistical inference

The method of moments can be used to estimate the parameters. Obviously, since moments are not be obtained in closed-form, numerical methods are required to solve the resulting equations. For that reason, this method of estimation is no longer considered in this work. In the rest of this Section, we will focus on the maximum likelihood method. Let us consider a random sample of n observations $\tilde{x} = (x_1, \ldots, x_n)$ from the distribution (7) and let $\Theta = (\theta, \lambda, \mu, \sigma)$ be the vector of parameters to be estimated. The log-likelihood function for Θ is proportional to

$$\begin{aligned}\ell(\tilde{x};\Theta) \quad \propto \quad & -n \log\cosh\psi - n\log\sigma - \frac{n\theta^2}{2} - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2} \\ & + \sum_{i=1}^{n}\log\cosh T(x_i),\end{aligned}$$

from which the normal equations are obtained. These equations are given by,

$$\begin{aligned}\frac{\partial \ell(\tilde{x};\Theta)}{\partial\theta} &= n\left[\mu\tanh\psi + \theta\,\sigma\right] - \sum_{i=1}^{n}(x_i-\lambda)\tanh T(x_i) = 0, \\ \frac{\partial \ell(\tilde{x};\Theta)}{\partial\lambda} &= n\tanh\psi - \sum_{i=1}^{n}\tanh T(x_i) = 0, \\ \frac{\partial \ell(\tilde{x};\Theta)}{\partial\mu} &= -\bar{x} + \mu + \theta\,\sigma\tanh\psi = 0, \\ \frac{\partial \ell(\tilde{x};\Theta)}{\partial\sigma} &= n\,\sigma\,(\theta\,\mu\tanh\psi - \sigma) + \sum_{i=1}^{n}(x_i-\mu)^2 \\ & \quad -\theta\,\sigma\sum_{i=1}^{n}(x_i-\lambda)\tanh T(x_i) = 0.\end{aligned}$$

From these equations, the entries of the Fisher's information matrix can be simply derived by differentiating these equations with respect to the four parameters, multiplying by negative one and taking expectation. They are displayed in the Appendix. Recall that the Fisher's information matrix of the skew-normal distribution proposed by Azzalini (1985) is singular for the skew parameter and, consequently, the maximum likelihood estimate of this parameter can be infinite with a positive probability.

3.1 Conjugate distribution

It is well-known that if $Y \sim N(\mu, \sigma^2)$ and μ follows also a normal distribution, $N(a, \tau^2)$, then the posterior distribution of μ given the sample information $\tilde{y} = (y_1, \ldots, y_n)$ is $N((a\sigma^2 + n\bar{y}\tau^2)/(\sigma^2 + n\tau^2), (\sigma^2 + \tau^2)/(\sigma^2 + n\tau^2))$. An analogue result is sustained when the prior distribution is provided by (7) always that $\theta = \sigma > 0$. That is, let $Y \sim N(a, \tau^2)$, $a \in \mathbb{R}$, $\tau > 0$ and $a \sim GN(\theta, \mu, \lambda)$, then the posterior distribution of a given the sample information is also a generalized normal distribution $GN(\theta^*, \lambda^*, \mu^*)$ with ${\theta^*}^2 = \theta^2\tau^2/(\tau^2 + n\theta^2)$, $\lambda_2^* = \lambda_2$ and $\mu^* = (\mu\tau^2 + n\bar{y}\theta^2)/(\tau^2 + n\theta^2)$.

4 Numerical Illustrations

4.1 Bimodality in Cancer Incidence

The incidence of some type of cancers by age displays a major mode for young adult and minor mode for older adults (see Anderson et al., 2006). The contour of the curve of cancer incident by age, for the different types of cancer, is a compelling puzzle. By understanding the reason behind the multimodality of some cancer incidence curves, the practitioners may enhance their knowledge of the developmental process of cancer and the potential features that identify cancer and that separate a particular type of cancer from all other types of cancer. Therefore, the importance of properly identifying cancer occurrences by age is vital to improve the tumor diagnosis

In this regard, there exists some type of cancers that have two peaks in occurrences, by age. Examples of recognized bimodal cancers include Kaposi's sarcoma and Hodgkin lymphoma. The latter type of cancer has two peaks in occurrence: in young adults and middle-aged adults. Hodgkin lymphoma is an uncommon type of cancer that develops in the lymphatic system, which is a network of vessels and glands spread throughout the organism. According to the *National Health Service* (`https://www.nhs.uk/conditions/hodgkin-lymphoma/`), *"Hodgkin lymphoma can develop at any age, but it mostly affects young adults in their early 20s and older adults over the age of 70."* We have downloaded Hodgkin's disease incidence by age data in England from the aforementioned website that correspond to the ICD-9 code 2010-2019 from the period 1971-1994 and to the ICD-10 code (C810, C811, C812, C813, C817 and, C819) from the period 1995-2016. The datasets were collected in nine different regions across England (North East, North West, Yorkshire and The Humber, East Midlands, West Midlands, East, London, South East and, Sound West). The first dataset includes 29,187 (17,380 males and 11,807 females) diagnosed cases whereas the second one contained 30,906 (17,581 males and 13,325 females). In Figure 3, we have plotted the empirical incidence by age in the period 1971-1994 (left panel) and 1995-2016 (right panel).

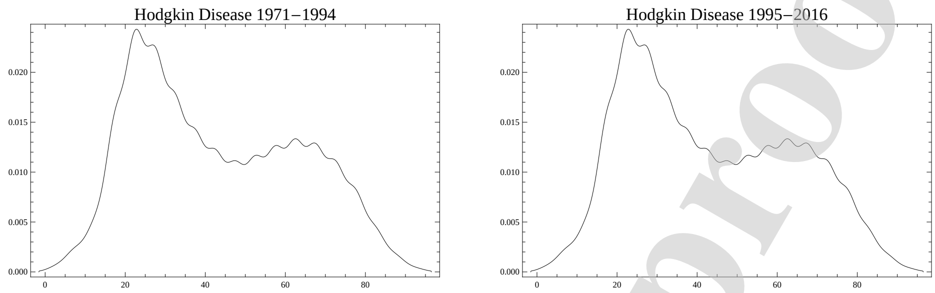

Figure 3: Observed incidence of Hodgkin disease in England by age in the period 1971-1994 (left) and 1995-2016 (right).

The bimodality of these datasets can be measured by the maximum difference, across sample points through the Hartigan's dip test Hartigan and Hartigan (1985). This test is available in the "diptest" package developed for R environment. In this test, we test the null hypothesis the distribution is unimodal against the alternative hypothesis that the distribution is non-unimodal, i.e. at least bimodal. For incidence by age in the period 1971-1994 dataset, there is empirical evidence to reject the null hypothesis. In this case, we have obtained a test statistic value of 0.0556 with a p-value less than 0.0001. This value was calculated via Monte Carlo simulation based on 5000 replications. Similarly, for the period 1995-2016, the null hypothesis is also rejected, the value of the test statistic is 0.0511 with a p-value less than 0.0001. Therefore, it seems reasonable to use a bivariate parametric model to explain these cancer datasets. We have fitted the four-parameter distribution provided by expression (7) to the two datasets presented in the first Section. Below in Figure 4, we have plotted again the empirical incidence by age of this cancer (solid line) in the period 1971-1994 (left panel) and 1995-2016 (right panel). The location-scale generalization of the proposed distribution (dashed line) has been superimposed to the observed data. It is observable that the model (7) can reproduce the two modes of the age incidence of this cancer for both periods.

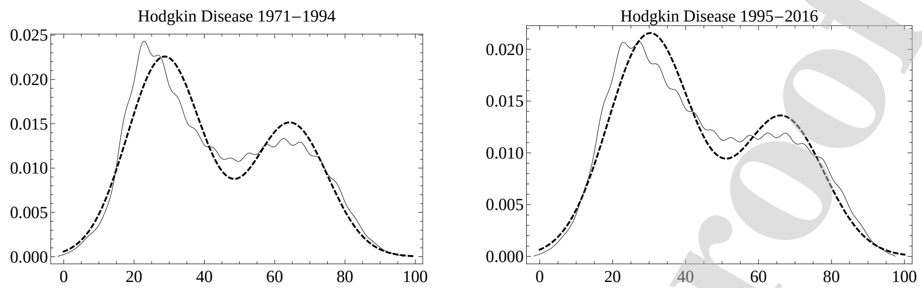

Figure 4: Observed (solid) and fitted (dashed) incidence of Hodgkin disease in England by age in the period 1971-1994 (left) and 1995-2016 (right).

4.2 Other applications

The performance of the location-scale generalization of the generalized normal distribution is tested by using two other datasets. Firstly, we consider data from the Old Faithful Geyser in Yellowstone National Park. The data consists of 299 pairs of measurements referring to the time interval between the starts of successive eruptions and duration of subsequent eruptions in minutes. See Azzalini and Bowman (1990) for more details. Applications of Hartigan's dip test conclude that both datasets are non-unimodal (i.e. bimodal) with a value of the test statistics are 0.0390 for the eruption time and 0.1025 for the waiting time between eruptions respectively. The associated p-values are 0.0030 and less than 0.0001 respectively. This is graphically confirmed in Figure 5, where we have depicted the distribution of the empirical eruption time (left panel) and waiting time (right panel) by using a solid line. Again, the four-parameter generalized distribution (dashed line) captures the bimodality feature of the distribution.

Our third illustrative example consists of an analysis of the breaking strengths of 63 glass fibres of a length of 1.5 cm. This set of data appears in Jones and Pewsey (2009). Unlike the previous examples, the Hartigan's dip test concludes that there is no empirical evidence to reject the null hypothesis that this dataset is unimodal. Here a value of 0.0311 was obtained for the test statistic and with a p-value of 0.9622. However, visual inspection of the observed data (see Figure 6) reveals that there exists a small bump in the observed data (solid line) that the model (7) (dashed line) can describe.

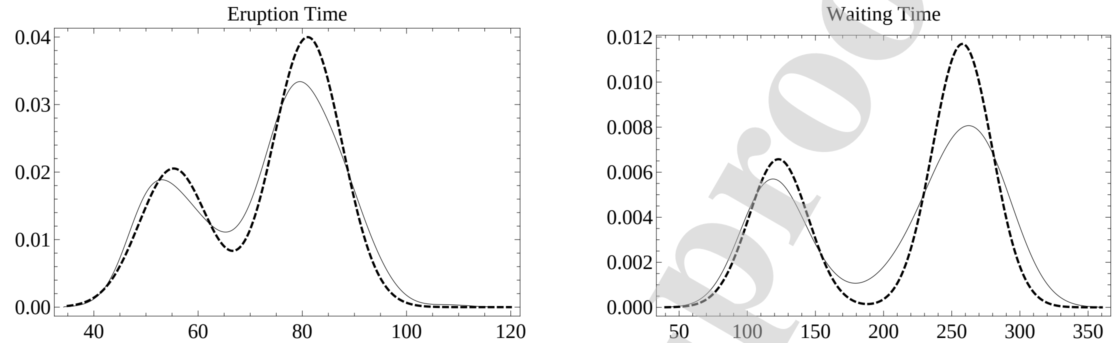

Figure 5: Time interval between the starts of successive eruptions and duration of subsequent eruptions in minutes of Old Faithful Geyser. See Azzalini and Bowman (1990) Smith and Naylor (1987).

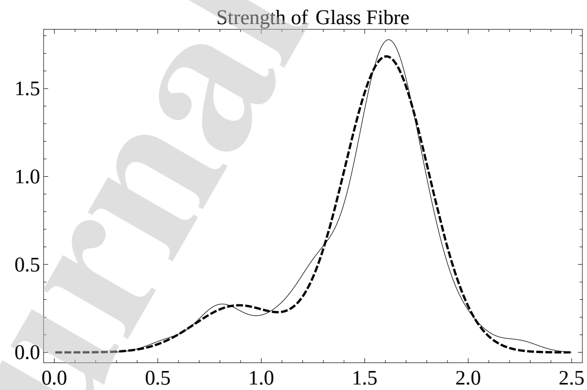

Figure 6: Observed (solid) and fitted (dashed) breaking strengths of 63 glass fibres of a length of 1.5 cm. See Jones and Pewsey (2009).

4.3 Summary of results

Parameter estimates for the four-parameter generalized normal $GN(\theta, \lambda, \mu, \sigma)$ and standard errors for the five aforementioned datasets are illustrated in Table 1. This Table also included the estimation of the parameters when the skew-normal distribution with parameters (λ, μ, σ) and a finite mixture of two skew-normal distributions with the same parameters μ and σ and different skew parameters, say $\lambda_1 \in \mathbb{R}$ and $\lambda_2 \in \mathbb{R}$ are considered. The mixture parameter is given by $p \in (0, 1)$. Observe that for the parameter θ consistently satisfies that $|\theta| > 1$ to reflect the fact that the bimodality is captured by the model. Furthermore, the positive skewness of these datasets is explained by the positive values of the estimates of the parameter λ.

Table 1: Parameter estimates and standard errors (in brackets) for the different datasets considered.

		Hodgkin's disease 1971-1994	Hodgkin's disease 1995-2016	Old Faithful Eruption Time	Old Faithful Waiting Time	Breaking stress of glass fiber
GN	$\widehat{\theta}$	1.687	1.590	1.953	3.090	1.765
		(0.011)	(0.011)	(0.112)	(0.148)	(0.276)
	$\widehat{\lambda}$	47.813	50.001	67.041	188.330	1.136
		(0.114)	(0.132)	(0.568)	(1.496)	(0.066)
	$\widehat{\mu}$	46.549	48.326	68.168	190.364	1.243
		(0.090)	(0.100)	(0.484)	(1.419)	(0.054)
	$\widehat{\sigma}$	10.607	11.416	6.596	21.838	0.205
		0.053	(0.057)	(0.299)	(0.947)	(0.021)
SN	$\widehat{\lambda}$	-1.9E-4	-3.2E-4	5.4E-4	-0.020	-2.679
		(0.872)	(1.0E-6)	(4.9E-5)	(0.213)	(0.802)
	$\widehat{\mu}$	43.005	45.180	72.308	3.479	1.850
		(14.309)	(1.0 E-6)	(5.4E-5)	(0.189)	(0.049)
	$\widehat{\sigma}$	20.500	20.029	13.867	1.146	0.470
		(0.094)	(2.1E-4)	(5.8E-6)	(0.050)	(0.055)
MSN	$\widehat{p}$	0.305	0.305	0.464	0.996	0.721
		(0.008)	(0.008)	(0.062)	(0.004)	(0.248)
	$\widehat{\lambda}_1$	-5.823	-5.810	11.689	38.373	-11.992
		(0.466)	(0.469)	(7.742)	(21.387)	(19.563)
	$\widehat{\lambda}_2$	0.688	0.690810	-0.851	-5.988	-1.220
		(0.021)	(0.021)	(0.248)	(10.712)	(1.097)
	$\widehat{\mu}$	40.097	40.021	72.309	1.651	1.837
		(0.183)	(0.128)	(0.767)	(0.033)	(0.047)
	$\widehat{\sigma}$	20.704	20.606	13.867	2.141	0.461
		(0.087)	(0.086)	(0.566)	(0.091)	(0.050)

Comparisons of this distribution with the three-parameter skewed normal distribution (SN) and the finite mixture of two skew-normal distributions with the same location and scale parameters μ and σ and different skew parameters, say $\lambda_1 \in \mathbb{R}$ and $\lambda_2 \in \mathbb{R}$ (MSN) are displayed in Table 2 for the five datasets considered. We have used the following measures of model selection, the negative of the maximum of the likelihood function ($-\ell_{\max}$), Akaike's information criterion (AIC), Bayesian information criterion (BIC) and consistent Akaike's information criterion (CAIC). A lower value of these measures of model selection is desirable. It is observable that the generalized-normal (GN) distribution is preferable to the SN and MSN models for the five datasets considered. We have also fitted a finite mixture of two skew-normal distributions with the different location parameters μ_1 and μ_2, different scale parameters σ_1 and σ_2 and different skew parameters λ_1 and λ_2. The following values for the consistent Akaike's information criterion (CAIC): 252333.850 (Hodgkin's disease:1971-1994), 269746.656 (Hodgkin's disease:1995-2016), 2355.511 (eruption time), 623.015 (waiting time) and 56.302 (breaking stress of glass fiber). It can be seen that the GN distribution provides a better fit than the seven-parameter mixture for the eruption time and breaking stress of glass fiber datasets in terms of this measure of model selection.

Table 2: Different measures of model selection for the generalized normal (GN) and skewed-normal (SN) distributions for the different datasets considered.

		Hodgkin's disease		Old Faithful		Breaking stress
		1971-1994	1995-2016	Eruption Time	Waiting Time	of glass fiber
GN	$-\ell_{max}$	126430.807	1352605.502	1161.709	305.538	10.334
	AIC	252869.615	270529.004	2331.419	619.076	28.669
	BIC	252902.741	270562.359	2346.220	627.649	37.241
	CAIC	252906.741	270566.359	2350.220	631.649	41.241
SN	$-\ell_{max}$	129571.764	137996.198	1210.488	465.005	13.957
	AIC	259149.528	275998.396	2426.977	936.010	33.914
	BIC	259174.372	276023.412	2438.078	947.111	40.344
	CAIC	259177.372	276026.412	2441.078	950.111	43.344
MSN	$-\ell_{max}$	128876.013	137325.559	1192.415	457.297	12.361
	AIC	257762.027	274661.118	2394.830	924.594	34.722
	BIC	257803.434	274702.811	2413.330	943.096	45.437
	CAIC	257808.434	274707.811	2418.330	948.096	50.437

5 Multivariate version

A multivariate version of (3) arises naturally using the representation (5). We define the vector $(X_1, \ldots, X_m)^\top$ with joint density

$$
\begin{aligned}
f(\boldsymbol{x}) &= \frac{p}{(2\pi)^{m/2}|\Sigma_1|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\mu_{-\theta})^\top \Sigma_1^{-1}(\boldsymbol{x}-\mu_{-\theta})\right\} \\
&\quad + \frac{1-p}{(2\pi)^{m/2}|\Sigma_2|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\mu_{\theta})^\top \Sigma_2^{-1}(\boldsymbol{x}-\mu_{\theta})\right\}, \qquad (8)
\end{aligned}
$$

where p is defined as in (6), $\mu_{-\theta} = (-\theta, \dots, -\theta)^\top$, $\mu_\theta = (\theta, \dots, \theta)^\top$ are the mean vectors, and both covariance matrices Σ_i, $i = 1, 2$ have unit variances. Then, by construction, the marginal distributions in (8) are dependent and identically distributed $X_i \sim GN(\theta, \lambda)$, $i = 1, 2, \dots, m$.

Figure 7 and 8 below shows the three-dimensional density plots and the contour plots respectively for the bivariate case and different values of the location-shape parameter θ, shape parameter λ, and the correlation coefficients associated to Σ_1 and Σ_2, ρ_1 and ρ_2 respectively. The situation $\theta = 0$, $\rho_1 = 0$ and $\rho_2 = 0$ corresponds to the case that X and Y are jointly normal and uncorrelated, i.e. independent. Note that this bivariate model can be either unimodal or bimodal.

We have implemented the method of maximum likelihood estimation for the bivariate case. For that reason, the Old Faithful geyser dataset that includes 272 pairs of measurements, about the time interval between the starts of successive eruptions and the duration of the subsequent eruption is considered. We have considered four different versions of our model: the first model includes the extra location parameters μ_1 and μ_2 and scale parameters σ_1 and σ_2; the second one location parameters μ_1 and μ_2 and scale parameters σ_1, σ_2, $\tilde{\sigma}_1$ and $\tilde{\sigma}_2$; the third one contains the location parameters μ_1, μ_2, $\tilde{\mu}_1$ and $\tilde{\mu}_2$ and scale parameters σ_1 and σ_2 and finally the fourth model has location parameters μ_1, μ_2, $\tilde{\mu}_1$ and $\tilde{\mu}_2$ and scale parameters σ_1, σ_2, $\tilde{\sigma}_1$ and $\tilde{\sigma}_2$. The result is shown in Table 3. It can be seen that the model with 12 parameters provides the best fit to this dataset in terms of the four measures of model validation. By comparing these results in terms of the BIC with the mixture of two-component skew-normal bivariate distribution provided in Prates et al. (2013) (see also Jain et al., 2013), model 3 and model 4 have better performance than the FMNOR distribution. On the other hand, FMSN, FMST and FMSCN provide a slightly better fit for this dataset.

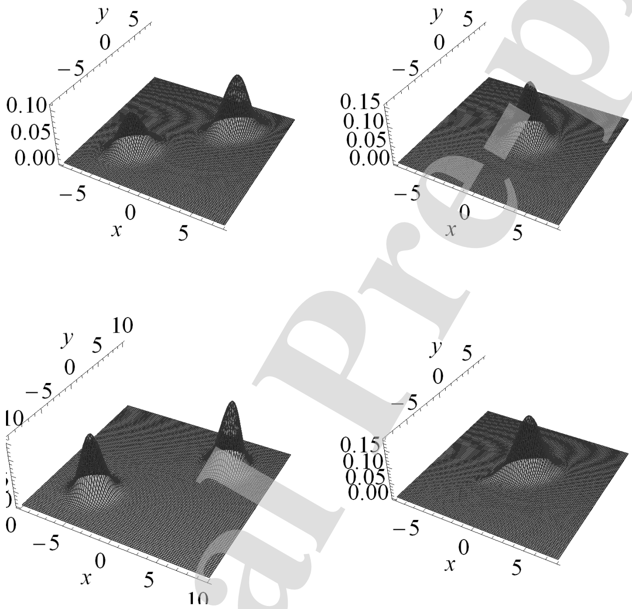

Figure 7: Three-dimensional density plots for the bivariate case of (8) and different values of θ, λ, ρ_1 and ρ_2. From top to bottom and left to right: $\theta = 3.5$, $\lambda = 0.05$, $\rho_1 = 0.4$, $\rho_2 = 0.5$; $\theta = 0.75$, $\lambda = 2$, $\rho_1 = 0.075$, $\rho_2 = 0.05$; $\theta = 5$, $\lambda = 0.015$, $\rho_1 = 0.15$, $\rho_2 = 0.015$ and $\theta = 0.75$, $\lambda = 2$, $\rho_1 = 0.9$, $\rho_2 = 0.5$.

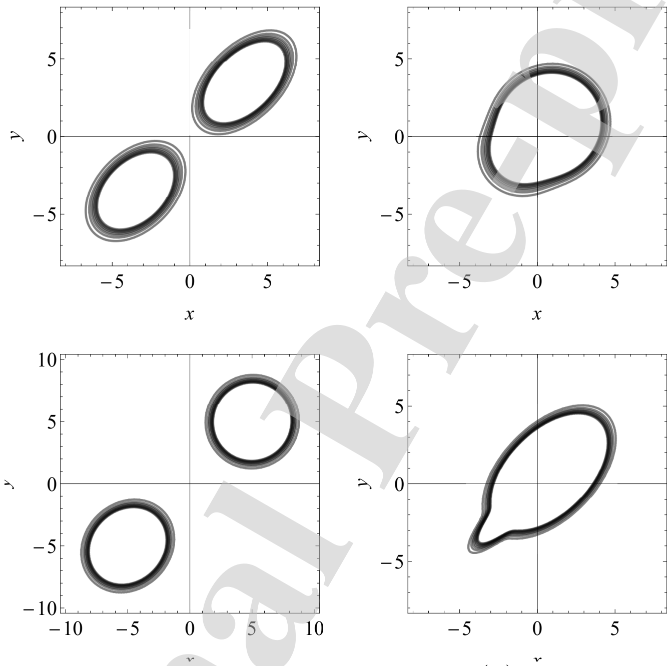

Figure 8: Contour plots for the bivariate case of (8) and different values of θ, λ, ρ_1 and ρ_2. From top to bottom and left to right: $\theta = 3.5$, $\lambda = 0.05$, $\rho_1 = 0.4$, $\rho_2 = 0.5$; $\theta = 0.75$, $\lambda = 2$, $\rho_1 = 0.075$, $\rho_2 = 0.05$; $\theta = 5$, $\lambda = 0.015$, $\rho_1 = 0.15$, $\rho_2 = 0.015$ and $\theta = 0.75$, $\lambda = 2$, $\rho_1 = 0.9$, $\rho_2 = 0.5$.

Table 3: Parameter estimates and standard errors (in brackets) for the different bivariate models considered.

Estimate (S.E.)	Model 1	Model 2	Model 3	Model 4
$\widehat{\theta}$	6.8707	1.0089	0.2207	0.7907
	(0.1368)	(0.0531)	(1.3211)	(8.6092)
$\widehat{\lambda}$	9.2711	-0.3288	-1.3061	-0.3751
	(0.6211)	(0.0685)	(7.8184)	(4.0838)
$\widehat{\mu}_1$	-22.3299	3.2356	4.0761	3.4990
	(2.7274)	(0.0672)	(1.3215)	(8.6093)
$\widehat{\mu}_2$	-4.3391	77.7687	79.8243	79.1775
	(0.1903)	(0.7111)	(1.3971)	(8.6213)
$\widehat{\tilde{\mu}}_1$			2.2680	2.8271
			(1.3217)	(8.6092)
$\widehat{\tilde{\mu}}_2$			54.8258	55.2649
			(1.4549)	(8.6295)
$\widehat{\sigma}_1$	13.5688	0.4411	0.3642	0.4123
	(0.5781)	(0.0418)	(0.0158)	(0.0229)
$\widehat{\sigma}_2$	13.1392	6.4008	5.9573	6.0025
	(0.0488)	(0.6924)	(0.2610)	(0.3277)
$\widehat{\tilde{\sigma}}_1$		0.3156		0.2631
		(0.0819)		(0.0201)
$\widehat{\tilde{\sigma}}_2$		23.1677		5.8042
		(1.8057)		(0.4180)
$\widehat{\rho}_1$	0.8336	0.4464	0.3295	0.3800
	(0.0033)	(0.1185)	(0.0657)	(0.0667)
$\widehat{\rho}_2$	0.9008	0.6781	0.4039	0.2855
	(0.0117)	(0.1774)	(0.1059)	(0.0949)
$-\ell_{max}$	-1289.7887	-1266.8696	-1139.9866	-1130.2640
AIC	2595.5775	2553.7393	2299.9733	2284.5280
BIC	2624.4239	2589.7973	2336.0313	2327.7977
CAIC	2632.4239	2599.7973	2346.0313	2339.7977

6 Final comments

In this paper, a new family of continuous random variables with non-necessarily symmetric densities has been introduced that is useful to explain the incidence by age of the Hodgkin's disease. This parametric family of distributions can include unimodality and bimodality features. The distribution, which generalizes the normal distribution, depends on four parameters and it improves the performance of the normal distributions and other related models available in the statistical literature. Some basic properties of the new distribution are examined, including moments, connections with other distributions and transformations. The maximum likelihood estimators were derived straightforwardly. The versatility of this model is confirmed through other applications in the field of geoscience. Of course, this new family of distributions might be employed in other areas. In this regard, we believe that this family could be useful to derive new stochastic frontier models (see Aigner et al., 1977, Battese and Coelli, 1988 or Gómez-Déniz and Pérez-Rodríguez, 2015).

References

Aigner, D., Lovell, C., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function model. *Journal of Econometrics*, 12:21–37.

Anderson, W., Pfeiffer, R., Dores, G., and Sherman, M. (2006). Bimodal breast cancer incidence patterns provide support for a dualistic model of mammary carcinogenesis. *Journal of Clinical Oncology*, 24(18):595–595.

Arnold, B. and Beaver, R. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting (with discussion). *Test*, 11(1):1–54.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12:171–178.

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica*, 46:199–208.

Azzalini, A. (2013). *The Skew-Normal and Related Families*. Cambridge University Press, Cambridge.

Azzalini, A. and Bowman, A. (1990). A look at some data on the old faithful geyser. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 39(3):357–365.

Battese, G. and Coelli, T. (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics*, 38:387–399.

Chiogna, M. (1998). Some results on the scalar skew-normal distribution. *J. Ital. Statist. Soc.*, 1:1–13.

Eugene, N. (2001). *A Generalized Normal Distribution: Properties, Estimation and Applications*. PhD thesis, Central Michigan University, Mount Pleasant, Michigan. Unpublished Thesis.

Eugene, N., Lee, C., and Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics-Theory and Methods*, 31(4):497–512.

Famoye, F., Lee, C., and Eugene, N. (2004). Beta-normal distribution: Bimodality properties and application. *Journal of Modern Applied Statistical Methods*, 3(1):85–103.

García, V., Gómez-Déniz, E., and Vázquez-Polo, F. (2010). A new skew generalization of the normal distribution: properties and applications. *Computational Statistics and Data Analysis*, 54:2021–2034.

Gómez-Déniz, E. and Pérez-Rodríguez, J. (2015). Closed-form solution for a bivariate distribution in stochastic frontier models with dependent errors. *Journal of Productivity Analysis*, 43:215–223.

Gui, W., Chen, P.-H., and Wu, H. (2013). A folded normal slash distribution and its applications to non-negative measurements. *Journal of Data Science*, 11:231–247.

Hartigan, J. and Hartigan, P. (1985). The dip test of unimodality. *The Annals of Statistics*, 13(1):70–84.

Hirota, M., Holmgren, M., Nes, E. V., and Scheffer, M. (2011). Globa lresilience of tropical forest and savanna to critical transitions. *Science*, 334:232–235.

Jain, S., Lavine, M., Radivojac, P., and Trosset, M. (2013). Identifiability of twocomponent skew normal mixtures with one known component. *Scandinavian Journal of Statistics*, 46(4):955–986.

Johnson, N. L. (1962). The folded normal distribution: Accuracy of estimation by maximum likelihood. *Technometrics*, 4(2):249–256.

Jones, M. and Pewsey, A. (2009). Sinh-arcsinh distributions. *Biometrika*, 96(4):761–780.

Leone, F., Nelson, L., and Nottingham, R. (1961). The folded normal distribution. *Technometrics*, 3:543–550.

MacMahon, B. (1966). Epidemiology of hodgkin's disease. *Cancer Research*, 26:1189–1200.

Piegorsch, W. and Casella, G. (1985). The existence of the first negative moment. *The American Statistician*, 39(1):60–62.

Prates, M. O., Barbosa, C., and Lachos, V. (2013). mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions. *Journal of Statistical Software*, 54(12):1–20.

Rêgo, L., Cintra, R., and Cordeiro, G. (2012). On some properties of the beta normal distribution. *Communications in StatisticsTheory and Methods*, 41(20):3722–3738.

Sharafi, M. and Behboodian, J. (2008). The Balakrishnan skew-normal density. *Statistical Papers*, 49:769–778.

Smith, R. L. and Naylor, J. C. (1987). Comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 36(3):358–369.

Teimouri, M. and Nadarajah, S. (2016). On simulating Balakrishnan skew-normal variates. *Computational Statistics and Data Analysis*, 45(11):3310–3322.

Appendix

Here we provide the entries of the Fisher's information matrix.

$$
\begin{aligned}
E\left(-\frac{\partial^2 \ell(\tilde{x};\Theta)}{\partial\theta^2}\right) &= \frac{n\mu^2}{\sigma^2}\mathrm{sech}^2\psi + n - \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\lambda)^2\mathrm{sech}^2 T(x_i),\\
E\left(-\frac{\partial^2 \ell(\tilde{x};\Theta)}{\partial\theta\partial\lambda}\right) &= -\frac{n}{\sigma^2}\left[\theta\mu\mathrm{sech}^2\psi + \sigma\tanh\psi\right.\\
&\quad \left.-\frac{1}{n}\sum_{i=1}^{n}\theta(x_i-\mu_2)\mathrm{sech}^2 T(x_i) + \sigma\tanh T(x_i)\right],\\
E\left(-\frac{\partial^2 \ell(\tilde{x};\Theta)}{\partial\theta\partial\mu}\right) &= \frac{n}{\sigma^2}\left(\theta\mu\mathrm{sech}^2\psi + \sigma\tanh\psi\right),\\
E\left(-\frac{\partial^2 \ell(\tilde{x};\Theta)}{\partial\theta\partial\sigma}\right) &= -\frac{n\mu}{2\sigma^3}\mathrm{sech}^2\psi\left[2\theta\mu + \sigma\sinh(2\psi)\right.\\
&\quad \left.+\frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\lambda)\mathrm{sech}^2 T(x_i) + \sigma\tanh T(x_i)\right],\\
E\left(-\frac{\partial^2 \ell(\tilde{x};\Theta)}{\partial\lambda^2}\right) &= -\frac{n\theta^2}{\sigma^2}\left[1-\mathrm{sech}^2\psi - \frac{1}{n}\sum_{i=1}^{n}\tanh^2 T(x_i)\right],\\
E\left(-\frac{\partial^2 \ell(\tilde{x};\Theta)}{\partial\lambda\partial\mu}\right) &= -\frac{n\theta^2}{\sigma^2}\mathrm{sech}^2\psi,\\
E\left(-\frac{\partial^2 \ell(\tilde{x};\Theta)}{\partial\lambda\partial\sigma}\right) &= \frac{n\theta}{\sigma^3}\left(\mu\,\theta\,\mathrm{sech}^2\psi + \sigma\tanh\psi\right)\\
&\quad +\frac{\theta}{\sigma^3}\left[\theta(\lambda-x_i)\mathrm{sech}^2 T(x_i) - \sigma\tanh T(x_i)\right],\\
E\left(-\frac{\partial^2 \ell(\tilde{x};\Theta)}{\partial\mu^2}\right) &= \frac{n}{\sigma^2}\left(1+\theta^2\mathrm{sech}^2\psi\right),\\
E\left(-\frac{\partial^2 \ell(\tilde{x};\Theta)}{\partial\mu\partial\sigma}\right) &= \frac{n\theta}{\sigma^3}\left(\sigma\tanh\psi - \theta\mu\mathrm{sech}^2\psi\right),\\
E\left(-\frac{\partial^2 \ell(\tilde{x};\Theta)}{\partial\sigma^2}\right) &= \frac{n}{\sigma^4}\left[2\mu\,\sigma\,\theta\,\tanh\psi + \theta^2\mu^2\mathrm{sech}^2\psi + 3\mu^2 - \theta^2\mu_2^2 - \sigma^2\right.\\
&\quad -2(3\mu-\theta^2\mu_2)\left(\mu+\sigma\theta\tanh\psi\right)\\
&\quad +\frac{1}{n}\sum_{i=1}^{n}x_i^2(3-\theta^2) + \theta\left((x_i-\lambda)\tanh T(x_i)\right)\\
&\quad \left.\times\left(-2\sigma + \theta(x_i-\lambda)\tanh T(x_i)\right)\right],
\end{aligned}
$$

From the inverse of the matrix evaluated at the corresponding maximum likelihood estimates, asymptotic standard errors can be derived in the usual way.