

Article

Properties and Applications of a New Family of Skew Distributions

Emilio Gómez-Déniz ¹, Barry C. Arnold ², José M. Sarabia ³ and Héctor W. Gómez ^{4,*}

- ¹ Department of Quantitative Methods in Economics and TiDES Institute, University of Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canarias, Spain; emilio.gomez-deniz@ulpgc.es
- ² Statistics Department, University of California Riverside, Riverside, CA 92504, USA; barry.arnold@ucr.edu
- ³ Department of Quantitative Methods, CUNEF University, 28040 Madrid, Spain; josemaria.sarabia@cunef.edu
- ⁴ Departamento de Matemática, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta 1240000, Chile
- * Correspondence: hector.gomez@uantof.cl; Tel.: +56-55-2637-278

Abstract: We introduce two families of continuous distribution functions with not-necessarily symmetric densities, which contain a parent distribution as a special case. The two families proposed depend on two parameters and are presented as an alternative to the skew normal distribution and other proposals in the statistical literature. The density functions of these new families are given by a closed expression which allows us to easily compute probabilities, moments and related quantities. The second family can exhibit bimodality and its standardized fourth central moment (kurtosis) can be lower than that of the Azzalini skew normal distribution. Since the second proposed family can be bimodal we fit two well-known data set with this feature as applications. We concentrate attention on the case in which the normal distribution is the parent distribution but some consideration is given to other parent distributions, such as the logistic distribution.

Keywords: logistic distribution; normal distribution; skew normal distribution; symmetric distribution



Citation: Gómez-Déniz, E.; Arnold, B.C.; Sarabia, J.M.; Gómez, H.W. Properties and Applications of a New Family of Skew Distributions. *Mathematics* **2021**, *9*, 87. <https://doi.org/10.3390/math9010087>

Received: 16 November 2020

Accepted: 29 December 2020

Published: 3 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

There are many situations in which empirical data show slight or marked asymmetry. This is frequently the case, for example, with actuarial and financial data which, in addition to this feature, have heavy tails reflecting the existence of extreme values. The se features mean that the data can not be adequately modeled by the Gauss (or normal) distribution. Furthermore, bimodal distributions appear naturally in many different scenarios. For example, in certain disease patterns, as well as in certain cancer incidence curves. Behind the bimodality (and multimodality as well) of some cancer incidence curves, and their study, clinicians can improve their understanding of cancer, the development process as well as the potential characteristics that identify cancer and that separate a particular type of cancer of all other types. This occurs, for example, in cases where there are two peaks of occurrence per age. The se cancers include Kaposi's sarcoma and Hodgkin's lymphoma. The latter type of cancer has two peaks of occurrence: in young people adults and middle-aged adults. On the other hand, the normal skew distribution appears naturally in stochastic frontier analysis when a normal distribution is assumed to represent the noise or idiosyncratic component and a half-normal distribution to represent the inefficiency term, in the event that the researcher imposes inefficient behavior on all firms in the sample of interest. See, for instance [1]. Recently, [2] introduces (using a finite mixture model) the zero inefficiency stochastic frontier model which can accommodate the presence of both efficient and inefficient firms in the sample by appearing various bimodal scenarios. The refore, it seems plausible to try to obtain families of distributions that incorporate bias to the normal distribution but that at the same time are more versatile in the sense of being able to adapt to the bimodal scenario that appears in different situations.

Although there are various mechanisms to obtain skewed distributions from an initial that is not skewed, (Two well-known procedures that allow to generalize an initial probability distribution, symmetric or not, are those provided in the works of [3,4], among others). Our attention here will focus on the mechanism for this purpose introduced by [5], which enjoys an undoubted popularity and has been the subject of research in numerous works. Let g and G be, respectively, the probability density function (pdf) and the cumulative distribution function (cdf) of a symmetric distribution. A random variate Z is said to have a skew distribution if its pdf is given by

$$f_Z(z) = 2g(z)G(\lambda z), \quad -\infty < z < \infty, \lambda \in \mathbb{R}. \quad (1)$$

This family of distributions has been widely studied as an extension of the normal distribution by means of a shape parameter, λ , which accounts for the skewness. In this case g and G are replaced in Equation (1) by the pdf and cdf of the standard normal distribution and the resulting distribution is called the skew normal distribution. It should be pointed out that the function g does not have to be precisely the derivative of the cdf G to ensure that the pdf given in Equation (1) is a genuine pdf, although this case has not been studied in depth in the statistical literature. Following the notation provided in Reference [6] we denote the family of distributions given by $g(z) = 2\phi(z)\Phi(\lambda z)$, where ϕ and Φ are the pdf and cdf of the standard normal distribution, respectively, by $SN = \{SN(\lambda) : \lambda \in \mathbb{R}\}$. Furthermore, when a random variable follows a skew normal distribution with location parameter $-\infty < \mu < \infty$ and scale parameter $\sigma > 0$ we will write $SN(\lambda, \mu, \sigma)$.

In this article, a new generalization of the family of skew distributions given in Equation (1) is proposed, which also includes the skew family of distributions of Azzalini as a particular case; that is, the expression (1). The methodology used is based on the combination of Azzalini's proposal and a result provided by [7] which led us to add a new parameter to the family (1). Later, from this new family a second family, very similar to the first, is introduced. This new family of distributions can exhibit bimodality and its standardized fourth central moment (kurtosis) can be lower than the kurtosis of the Azzalini skew normal distribution (and can be positive or negative).

In recent decades, starting from Azzalini's proposal, several generalizations and extensions of the skew-normal distribution have been introduced (see for example [8]). For multivariate extensions, see References [9–11], among others. The methods applied in the present paper can be considered as extensions and alternatives to the well-known skew-normal distribution (see [5,12]), whose properties (see [12,13]), and corresponding estimation [14] have been widely discussed. Other ways of obtaining skewed normal distributions have also been introduced, such as the one proposed by Reference [15], the Balakrishnan skew-normal density in Reference [16], the proposed model of Reference [17] and the generalized normal distribution in References [18–20], among others. For an exhaustive and comprehensive study of the skew-normal distribution, see the recent book by Reference [21].

The class of probability models proposed in the present paper can also be considered as alternatives or as approximations to the usual collective risk models in actuarial settings (see [22,23] among others). Data sets in these settings are typically skewed and the generalized models of the present paper expected to provide better fits than the standard models. In collective risk settings the right tail of the distribution is of considerable interest since the likelihood of large claims is of concern. In addition, the total claim distribution is of interest. Normal approximations are frequently resorted to when dealing with these variables. The use of more flexible generalized normal models can be expected to yield improvement.

The organization of this paper is as follows. The main result from which we constructed the two proposed families is shown in Section 2. Due to the importance that the normal distribution plays in numerous problems of applied statistics we dedicate a complete section, Section 3, to the study of this distribution. For this purpose, the pdf, which appears in closed form, is shown for the two families. We also give expressions for the mean, variance and the third and fourth standardized cumulant, to compare

with their equivalents corresponding to the classical (Azzalini) skew normal distribution. In Section 4, the parameter estimation problem is studied. In order to obtain numerical solutions to the maximum likelihood (ML) estimation problem, suitable software has to be used. Multivariate extensions are described briefly in Section 5. Some examples and applications are described in Section 6. Finally, some conclusions are drawn and promising fields for further research are proposed in the last Section.

2. Main Results

We recall (see [24]) that if X and Y are independent and indentially distributed random variables with a finite fractional moment and if for all real λ , $\Pr(X + \lambda Y > 0) = 1/2$, then they are symmetric. Also, the following Theorem, which appears in [7], is required for the main result which will appear later.

Theorem 1. (see [7]) *Let G be the cdf of an arbitrary distribution that is symmetric about zero. The n,*

$$\int_{-a}^a G(z) dz = a, \quad \forall a \in \mathbf{R}^+. \tag{2}$$

Expression (2) in Reference [25] also establishes this assertion.

2.1. The First Family of Skew Distributions

The following result presents the key contribution of this work, consisting of proposing, given a family of symmetrical distributions, a more general family not necessarily symmetric that includes as a particular case the first family.

Theorem 2. *Let X and Y two random variables with symmetric cdf's $G_X(x)$ and $G_Y(y)$ and pdf's $g_X(x)$ and $g_Y(y)$, respectively. The n,*

$$f_Y(y; \lambda, \alpha) = \frac{g_Y(y)}{\alpha} \int_{-\alpha}^{\alpha} G_X(z + \lambda y) dz \tag{3}$$

represents a genuine pdf for $\alpha \in \mathbf{R} - \{0\}$ and $\lambda \in \mathbf{R}$.

Proof. Without loss of generality, assume that X and Y are independent random variables. Taking into account the fact that $X - \lambda Y$ is symmetric and using the result provided in Theorem 1, we get

$$\begin{aligned} \alpha &= \int_{-\alpha}^{\alpha} \Pr(X - \lambda Y < z) dz \\ &= \int_{-\infty}^{\infty} \left[\int_{-\alpha}^{\alpha} \Pr(X < z + \lambda y | Y = y) dz \right] g_Y(y) dy \\ &= \int_{-\infty}^{\infty} \left(\int_{-\alpha}^{\alpha} \Pr(X < z + \lambda y) dz \right) g_Y(y) dy \\ &= \int_{-\infty}^{\infty} g_Y(y) \left(\int_{-\alpha}^{\alpha} G_X(z + \lambda y) dz \right) dy. \end{aligned}$$

Hence the result. \square

Expression (3) can instead be viewed in the following form related to an infinite mixture construction. Let $G_X(\xi + \lambda y)$, $\lambda \in \mathbf{R}$, $\xi \in \mathbf{R}$ and $y \in \mathbf{R}$, be the cdf of a symmetric distribution with support in the real line. Suppose now that ξ is random and follows a uniform distribution in the interval $[-\alpha, \alpha]$, then

$$H_X(y; \lambda, \alpha) = \frac{1}{2\alpha} \int_{-\alpha}^{\alpha} G(\xi + \lambda y) d\xi,$$

is also a genuine cdf symmetric around zero. That is, $H_X(y) = 1 - H_X(-y)$. Now, Equation (3) is derived by taking into account the fact that $2g_Y(y)H_X(y)$ is a genuine pdf. Another elegant way to see that Equation (3) defines a genuine pdf is to consider the argument given in Lemma 1 in [15]. That is, let $S(\lambda, \alpha) = \int_{-\infty}^{\infty} f_Y(y; \lambda, \alpha) dy$. Now, we have that $S(0, \alpha) = \int_{-\infty}^{\infty} g_Y(y) dy = 1$ and since

$$\frac{\partial}{\partial \lambda} S(\lambda, \alpha) = \int_{-\infty}^{\infty} \frac{y g_Y(y)}{\alpha} [G_X(\alpha + \lambda y) - G_X(-\alpha + \lambda y)] dy = 0$$

we have that $S(\lambda, \alpha) = 1$.

The results in the following proposition are readily verified and consequently are stated without proof.

Proposition 1. For the density (3) the following results hold.

- (i) $f_Y(0; \lambda, \alpha) = g_Y(0)$.
- (ii) $f_Y(y; \lambda, \alpha) = f_Y(y; \lambda, -\alpha)$ for $\lambda \in \mathbf{R}, \alpha \in \mathbf{R} - \{0\}$.
- (iii) $f_Y(y; 0, \alpha) = g_Y(y)$ for $\alpha \in \mathbf{R} - \{0\}$.
- (iv) $\lim_{\alpha \rightarrow 0} f_Y(y; \lambda, \alpha) = 2g_Y(y)G_X(\lambda y)$ for $\lambda \in \mathbf{R}$.
- (v) $G_Y(y; \lambda, \alpha) = 1 - G_Y(-y; -\lambda, \alpha)$.
- (vi) If Y has the pdf (3) then $-Y$ has the same distribution but with the parameter λ replaced by $-\lambda$.
- (vii) $f_Y(y; \lambda, \alpha) + f_Y(y; -\lambda, \alpha) = 2g_Y(y)$.
- (viii) Let $\alpha = \lambda$ in (3) and consider the two random variates Z_1 and Z_2 following the pdf (3) with parameters $\lambda_1 \in \mathbf{R}$ and $\lambda_2 \in \mathbf{R}$, respectively, then, if $\lambda_1 < \lambda_2, Z_1 <_{st} Z_2$. That is, Z_1 is stochastically smaller than Z_2 .

Because of Result (ii) in Proposition 1, an identifiability problem will arise in (3) if we allow α to assume both positive and negative values. To avoid this problem we can and will restrict α to be non-negative when discussing inference for this model. Observe that (iv) establishes that when $\alpha \rightarrow 0$ we get as a special case the well studied skew family of distributions appearing in References [5,12].

In most cases the density function (3) does not have a simple, closed form but it can be computed numerically. However, closed-form expressions for the pdf can be obtained for some specific choices of well-known distributions. For example, if we utilize the pdf and cdf of the logistic distribution with location parameter $\mu = 0$ and scale parameter $s = 1$ in (3), i.e., take $g_Y(y) = \exp(-y)/(1 + \exp(-y))^2$ and $G_X(x) = (1 + \exp(-x))^{-1}$, respectively, then (3) becomes

$$f_Y(y; \lambda, \alpha) = \frac{\exp(y)}{\alpha(1 + \exp(yx))^2} \log \left[\frac{1 + \exp(\alpha + \lambda y)}{1 + \exp(-\alpha + \lambda y)} \right].$$

On the other hand, combining the pdf of the standard normal distribution with the cdf of the logistic distribution we have, after applying (3) the new pdf

$$f_Y(y) = \frac{\phi(y)}{\alpha} \log \left[\frac{1 + \exp(\alpha + \lambda y)}{1 + \exp(-\alpha + \lambda y)} \right].$$

If in Equation (3), we use a normal pdf and a normal cdf, i.e., take $g_Y(y) = \phi(y)$ and $G_X(x) = \Phi(x)$, where $\phi(\cdot)$ and $\Phi(\cdot)$ represent the pdf and cdf of the standard normal distribution, respectively, then (3) becomes:

$$f_Y(y; \lambda, \alpha) = \frac{\phi(y)}{\alpha} [2\alpha + \phi(\alpha + \lambda y) - \phi(\alpha - \lambda y) + (\lambda y - \alpha)\Phi(\alpha - \lambda y) - (\lambda y + \alpha)\Phi(-\alpha - \lambda y)], \tag{4}$$

which does not appear to be a very attractive analytical expression. However it is not intractable and does represent a flexible two parameter family of densities. Figure 1 shows some graphs of this distribution in comparison with the skew normal distribution with the same mean. Since the mean and variance do not have a closed form for this distribution, we have chosen to calculate the mean numerically and match the mean of the SN distribution with parameters (λ, μ, σ) to obtain the value of the skew parameter λ . In all cases, except for the first graph, both distributions have the same mean and approximately the same variance.) and with parameters λ, μ and σ . Similar to the latter, the distribution is unimodal. Furthermore, the degree of skewness increases when λ grows. Positive skewness corresponds to the case $\lambda > 0$. As can be seen, the new distribution can take the same shape as the normal skew distribution, even having one less parameter when the highest probability mass percentage is around the ordinate axis. Otherwise, shape and scale parameters will have to be incorporated, as will be proposed later. The refore, taking into account the Ockham’s razor principle, it would seem logical, if one had to choose between both models, to opt for the new modeling proposed in this work.

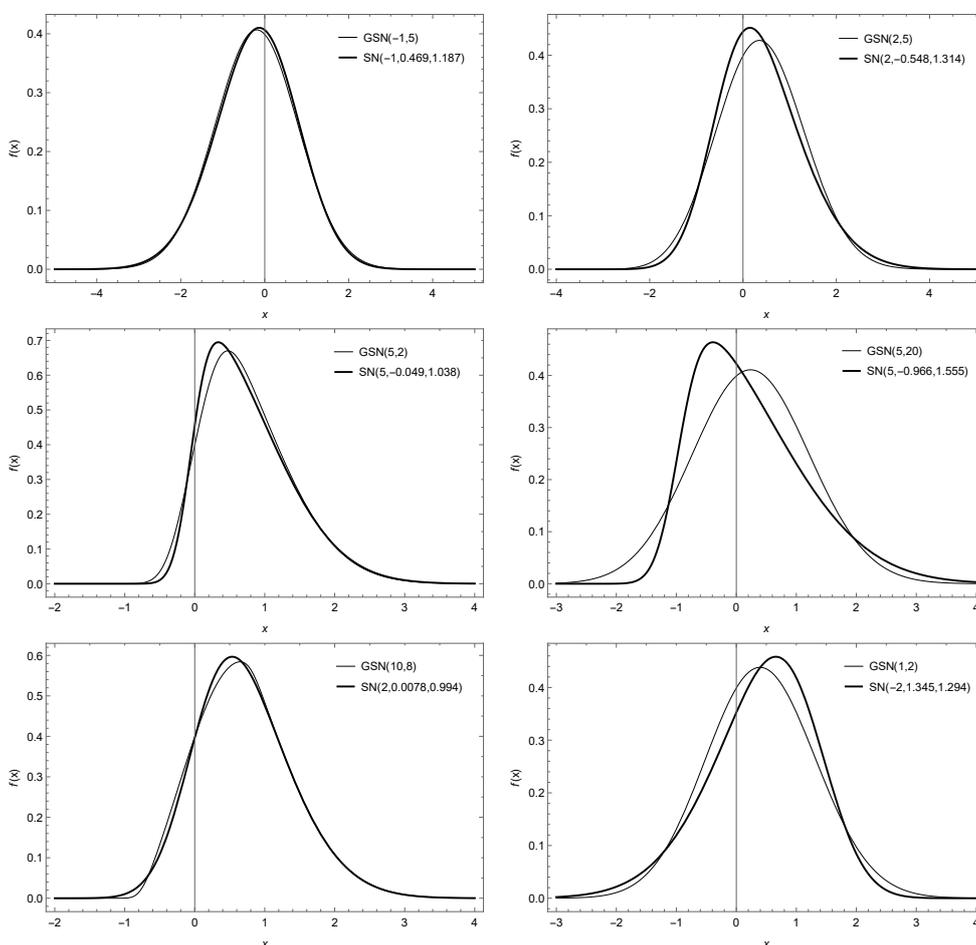


Figure 1. Plot of the pdf (4) (thick line) denoted as $GSN(\lambda, \alpha)$ and the $SN(\lambda, \mu, \sigma)$ (thin line) for selected values of the parameters α and λ .

Other possibilities for which we can get simple expressions for the corresponding pdf by applying (3) include the hyperbolic secant distributions (see [26,27], among others) and the ArcSin distributions (see [28]).

On some occasions, it is convenient expressed (3) as

$$f_Y(y; \lambda, \alpha) = \frac{g_Y(y)}{\alpha} \int_{\lambda y - \alpha}^{\lambda y + \alpha} G_X(u) du, \tag{5}$$

which is obtained after the change of variable $u = z + \lambda y$ in (3). In fact the family of skew distributions given in [5] can alternatively be obtained from (5) by applying the first mean value Theorem to the integral appearing on (5) in the interval $[\lambda y - \alpha, \lambda y + \alpha]$. To see this take $c = \lambda y$ in the well known formula $\int_b^a f(x) dx = f(c)(b - a)$, $c \in [a, b]$, to obtain $f_Y(y; \lambda, \alpha) = 2g_Y(y)G_X(\lambda y)$.

2.2. The Second Family of Skew Distributions

Now we present the second family of skew distributions proposed in this work. This is derived from Equation (5) as follows.

Theorem 3. *If $g_Y(y)$ be a density function that is symmetric about zero and G_X a cdf also symmetric about zero, then,*

$$f_Y(y; \lambda, \alpha) = g_Y(y)[G_X(\lambda y + \alpha) + G_X(\lambda y - \alpha)] \tag{6}$$

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

Proof. From Equation (3) it follows that

$$\begin{aligned} \alpha &= \int_{-\infty}^{\infty} g_Y(y) \left(\int_{-\alpha}^{\alpha} G_X(z + \lambda y) dz \right) dy \\ &= \int_{-\infty}^{\infty} g_Y(y) \left[\int_{-\alpha}^0 G_X(z + \lambda y) dz + \int_0^{\alpha} G_X(z + \lambda y) dz \right] dy. \end{aligned} \tag{7}$$

If we now take the derivative with respect to α on both sides in (7) and apply the Fundamental Theorem of calculus we get

$$\int_{-\infty}^{\infty} g_Y(y)[G_X(\lambda y + \alpha) + G_X(\lambda y - \alpha)] dy = 1.$$

Thus Equation (6) is a genuine pdf. \square

It can be verified that the pdf's given in (6) also satisfy the properties listed in Proposition 1. Furthermore, it is straightforward to see that $f_Y(y; \lambda, 0) = 2g_Y(y)G_X(y)$, $f_Y(y; \lambda, \infty) = g_Y(y)$ and $f_Y(y; 0, \alpha) = g_Y(y)$.

Note that the two proposed families, (3) and (6), are different and the only density belonging to both families is the basic density $g_Y(y)$. A major difference between the two proposed families is that in the first family α is not permitted to take the value zero while that is permitted in the second family. Both families are very similar and differ markedly only in small regions of the support of the distributions. To see this, note that applying the trapezoid rule to (3) we get

$$f_Y(y; \lambda, \alpha) \approx g_Y(y)[G_X(\lambda y + \alpha) + G_X(\lambda y - \alpha)],$$

which coincides with (5).

Because $G_X(\alpha) = 1 - G_X(-\alpha)$, from (6) we get that when $\lambda = 0$, $f_Y(y) = g_Y(y)$ while for $\alpha = 0$ we get the skew family of distributions proposed in Reference [5]. Again, it can be verified that $f_Y(y; \lambda, \alpha) = f_Y(y; \lambda, -\alpha)$ for the family given in Equation (6). Thus the same identifiability problem arises for the model (6) as did in the model (3) if we allow α to assume both positive and negative values. To avoid this problem here too we can and will restrict α to be non-negative when discussing inference for the model. If desired, a more general class than the one proposed in (6) is one corresponding to finite mixtures of densities of the form (6) as follows

$$f_Y(y) = g_Y(y) \sum_{i=1}^m \delta_i [G_X(\lambda_i y + \alpha_i) + G_X(\lambda_i y - \alpha_i)], \tag{8}$$

where the δ_i 's are positive and $\sum_{i=1}^m \delta_i = 1$.

At this point, we give a more general result than the one provided in [5,29,30] for a symmetric random variable.

Proposition 2. *If U be a random variable that is symmetric about zero with cdf given by $G(u)$, pdf $g(u)$ and if λ and α are two real numbers, then,*

$$E[G(\lambda U + \alpha)] = 1 - E[G(\lambda U - \alpha)]. \tag{9}$$

Proof. This is obtained in a straightforward way directly from the result given in Theorem 3 by integrating on both sides of the equality over the support $(-\infty, \infty)$. \square

When U follows a standard normal distribution (9) reduces to

$$E[\Phi(\lambda U + \alpha)] = 1 - E[\Phi(\lambda U - \alpha)] = \Phi\left(\frac{\alpha}{\sqrt{1 + \lambda^2}}\right),$$

a result which is well known in the statistical literature (see [5,29,30]).

3. The Normal Distribution Case

Of greater interest, because it is expressed in a simpler formulation, is the pdf obtained from Equation (6) when g and G are replaced by the pdf and cdf of the standard normal distribution, respectively. This is given by

$$f_X(x) = \phi(x)[\Phi(\lambda x + \alpha) + \Phi(\lambda x - \alpha)]. \tag{10}$$

In the following discussion, when a random variate X has its pdf given by Equation (10) it will be denoted by $X \sim \text{GSN}(\lambda, \alpha)$. Figure 2 show some graphs of this pdf and the corresponding skew normal pdf with the same mean and variance. (In this case, the skew parameter λ of the SN distribution has been set so that, once equal to the mean and variance of the new distribution, the values of μ and σ were obtained numerically so that both distributions should have the same mean and the same variance. It can be seen that the new model is very versatile and that the value of the parameters provide a distribution which can exhibit unimodality and bimodality. Again, as with Figure 1, the distribution can take approximately the same shape as the normal skew distribution with fewer parameter).

We next provide the moment generating function of the family given in Equation (10).

Proposition 3. *The moment generating function of a random variable X having its pdf given by Equation (10) is of the form*

$$M_X(t) = [\Phi(\delta(\lambda t + \alpha)) + \Phi(\delta(\lambda t - \alpha))] \exp\left(\frac{t^2}{2}\right), \tag{11}$$

where $\delta = 1/\sqrt{1 + \lambda^2}$.

Proof. It is straightforward following the same argument as the one used in Reference [5] in order to get the moment generating function of the skew-normal distribution. \square

Moments can then be readily obtained by differentiation of Equation (11). In particular, the mean and variance are given by

$$E(X) = \lambda b \delta \exp\left[-\frac{1}{2}(\alpha \delta)^2\right], \tag{12}$$

$$\text{var}(X) = 1 - (\lambda b \delta)^2 \exp\left[-(\alpha \delta)^2\right], \tag{13}$$

where $b = \sqrt{2/\pi}$. Table 1 shows the mean and variance of the proposed distribution and the corresponding ones for the skew normal distribution for some special cases of parameters. Similar to case of the skew normal distribution, it can be verified that $E(X^2) = 1$. Another important property that $GSN(\lambda, \alpha)$ shares with the skew normal distribution is that if $Z \sim GSN(\lambda, \alpha)$ then $Z^2 \sim \chi_1^2$ for all values of λ and α . In complete moments can also be studied following the work of Reference [31]. The third (skewness) and fourth (kurtosis) standardised cumulants are given by,

$$\gamma_1 = \frac{E(X)}{[var(X)]^{3/2}}(\delta\lambda)^2 [2b^2 \exp(-(\alpha\delta)^2) + (\alpha\delta)^2 - 1],$$

$$\gamma_2 = \frac{\pi b^2(3b^2\pi - 4\delta^4 E^2(X)\kappa_1) - 12E^4(X)}{(1 - var(X))^2\kappa_2} - 3,$$

where

$$\kappa_1 = 3 + 2(2 + \alpha^2)\lambda^2 + \lambda^4,$$

$$\kappa_2 = \frac{\pi \exp[(\alpha\delta)^2] - 2(\lambda\delta)^2}{(\lambda\delta)^4},$$

and $E(X)$ and $var(X)$ are given by Equations (12) and (13), respectively. Comparisons of these values with the standard skew normal distribution are shown in Table 2. As can be seen, the standardized fourth central moment (kurtosis) can be lower for the GSN distribution than it is for Azzalini’s skew normal distribution.

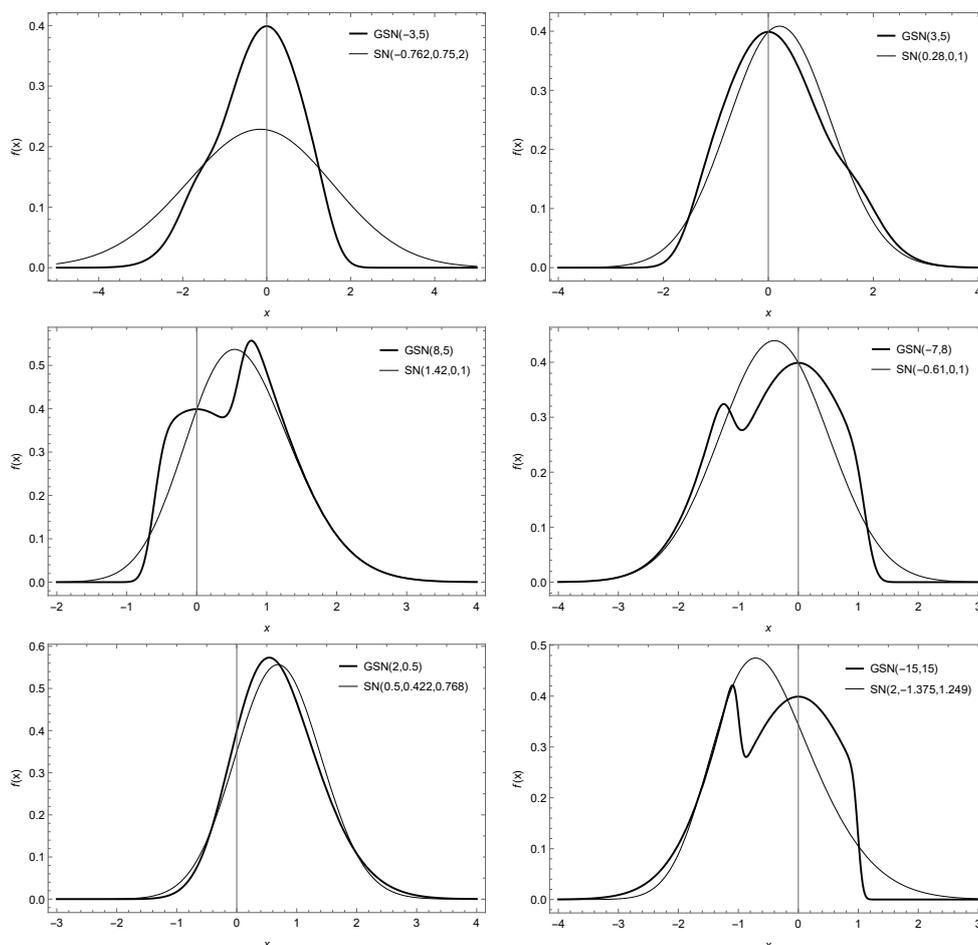


Figure 2. Plot of the pdf in (10) for selected values of parameters and comparison with the skew normal one.

Table 1. Mean and variance of the GSN and the SN variates for some parameter values.

λ	0	1	2	3	4	5
Mean SN(λ)	0.000	0.564	0.713	0.756	0.774	0.782
Mean GSN($\lambda, 0.1$)	0.000	0.562	0.712	0.756	0.773	0.782
Variance SN(λ)	1.000	0.681	0.490	0.427	0.400	0.387
Variance GSN($\lambda, 0.1$)	1.000	0.683	0.491	0.427	0.401	0.388
Mean SN(λ)	0.000	0.564	0.713	0.756	0.774	0.782
Mean GSN($\lambda, 1$)	0.000	0.439	0.645	0.720	0.751	0.767
Variance SN(λ)	1.000	0.681	0.490	0.427	0.400	0.387
Variance GSN($\lambda, 1$)	1.000	0.806	0.583	0.481	0.435	0.410
Mean SN(λ)	0.000	0.564	0.713	0.756	0.774	0.782
Mean GSN($\lambda, 5$)	0.000	0.001	0.058	0.216	0.371	0.483
Variance SN(λ)	1.000	0.681	0.490	0.427	0.400	0.387
Variance GSN($\lambda, 5$)	1.000	0.999	0.996	0.952	0.862	0.765

Table 2. Third and fourth standardized cumulants of the GSN and the SN variates for some parameter values.

λ	0	1	2	3	4	5
$\gamma_1, SN(\lambda)$	0.000	0.136	0.453	0.667	0.784	0.850
$\gamma_1, \text{Mean GSN}(\lambda, 0.1)$	0.000	0.135	0.451	0.664	0.782	0.849
$\gamma_2, SN(\lambda)$	0.000	0.061	0.305	0.509	0.632	0.705
$\gamma_2, \text{GSN}(\lambda, 0.1)$	0.000	0.060	0.302	0.507	0.630	0.703
$\gamma_1, SN(\lambda)$	0.000	0.136	0.453	0.667	0.784	0.850
$\gamma_1, \text{GSN}(\lambda, 1)$	0.000	0.082	0.281	0.488	0.639	0.738
$\gamma_2, SN(\lambda)$	0.000	0.061	0.305	0.509	0.632	0.705
$\gamma_2, \text{GSN}(\lambda, 1)$	0.000	-0.046	0.071	0.289	0.458	0.571
$\gamma_1, SN(\lambda)$	0.000	0.136	0.453	0.667	0.784	0.850
$\gamma_1, \text{GSN}(\lambda, 5)$	0.000	0.006	0.188	0.336	0.332	0.311
$\gamma_2, SN(\lambda)$	0.000	0.061	0.305	0.509	0.632	0.705
$\gamma_2, \text{GSN}(\lambda, 5)$	0.000	-0.000	-0.044	-0.294	-0.480	-0.501

Let $\Phi_{\lambda,\alpha}(z)$ denote the cdf of $Z \sim \text{GSN}(\lambda, \alpha)$.

Proposition 4. If $Z \sim \text{GSN}(\lambda, \alpha)$, then its cdf is given by

$$\begin{aligned} \Phi_{\lambda,\alpha}(z) = & \Phi(z) + T\left(z, \frac{\alpha}{z} - \lambda\right) - T\left(z, \frac{\alpha}{z} + \lambda\right) \\ & + T\left(\alpha\delta, \frac{y}{\alpha\delta^2} - \lambda\right) - T\left(\alpha\delta, \frac{z}{\alpha\delta^2} + \lambda\right), \quad z \neq 0, \alpha \neq 0, \end{aligned}$$

where $T(x, a)$ is the Owen's function see [32] given by

$$T(x, a) = \frac{1}{2\pi} \int_0^a \frac{1}{1+t^2} \exp\left[-x^2(1+t^2)/2\right] dt, \quad a \in \mathbf{R}.$$

Proof. The proof is direct by applying result B.21 in Reference [21]. □

Proposition 5. If $Z \sim \text{GSN}(\lambda, \alpha)$, then it follows that:

- (i) $\Phi_{\lambda,0}(z) = \Phi(z) + T\left(z, \frac{\alpha}{z} - \lambda\right) - T\left(z, \frac{\alpha}{z} + \lambda\right).$
- (ii) $\Phi_{\lambda,\alpha}(0) = \frac{1}{2} - 2T(\alpha\delta, \lambda).$
- (iii) $\Phi_{\lambda,\alpha}(z) = \Phi_{-\lambda,\alpha}(-z).$

Proof. The proof is also direct this time by applying the result B.23 in [21]. □

To end this Section we provide a result related to probability transformations which generalises a result appearing in Reference [5] and provided also in Reference [31].

Proposition 6. *Let W and Z be independent random variables with distribution $N(0, 1)$ and $GSN(\lambda, \alpha)$, respectively. The n , the random variable $Y = (hW + kZ) / \sqrt{h^2 + k^2}$, where $h, k \in \mathbb{R}$, follows a $GSN(\tilde{\lambda}, \tilde{\alpha})$, where $\tilde{\lambda} = \delta k \lambda / \sqrt{h^2 + k^2}$ and $\tilde{\alpha} = \delta \alpha$.*

Proof. It can be proved following the same argument as that used in Lemma 1 in Reference [31]. □

4. Estimation

The family of distributions $GSN(\lambda, \alpha)$ can be generalized by means of a linear transformation in order to introduce a location and a scale parameter adding more flexibility to the model (10). We thus will consider the location-scale generalization of the skew-normal distribution defined as the distribution of $Y = \mu + \sigma X$, where $X \sim GSN(\lambda, \alpha)$ given in Equation (10), where $\mu \in \mathbb{R}$ and $\sigma > 0$. Its pdf is given by

$$f_Y(y) = \frac{1}{\sigma} \phi_Y\left(\frac{y - \mu}{\sigma}\right) \left[\Phi_Y\left(\frac{\lambda(y - \mu)}{\sigma} + \alpha\right) + \Phi_Y\left(\frac{\lambda(y - \mu)}{\sigma} - \alpha\right) \right]. \tag{14}$$

When $\lambda = \alpha = 0$ this pdf reduces to the $N(\mu, \sigma)$ and when $\alpha = 0$ to the $SN(\mu, \sigma, \lambda)$. For a sample $\underline{y} = \{y_1, \dots, y_n\}$ we can estimate the four parameters, $\Theta = (\lambda, \alpha, \mu, \sigma)$, of the model given in Equation (14) by a direct search for the maximum of the log-likelihood surface given by

$$\ell(\underline{y}; \Theta) \propto -n \log \sigma - \sum_{i=1}^n \frac{(y_i - \mu)^2}{2\sigma^2} + \sum_{i=1}^n \log \left[\Phi_Y\left(\frac{\lambda(y_i - \mu)}{\sigma} + \alpha\right) + \Phi_Y\left(\frac{\lambda(y_i - \mu)}{\sigma} - \alpha\right) \right]. \tag{15}$$

From Equation (15) we get the normal equations which are given by:

$$\begin{aligned} \frac{\partial \ell(\underline{y}; \Theta)}{\partial \lambda} &= \sum_{i=1}^n \frac{y_i - \mu}{\sigma} \mathcal{I}(r_{1i}, r_{2i}) = 0, \\ \frac{\partial \ell(\underline{y}; \Theta)}{\partial \alpha} &= \sum_{i=1}^n \mathcal{H}(r_{1i}, r_{2i}) = 0, \\ \frac{\partial \ell(\underline{y}; \Theta)}{\partial \mu} &= \frac{n(\bar{y} - \mu)}{\sigma^2} - \frac{\lambda}{\sigma} \sum_{i=1}^n \mathcal{I}(r_{1i}, r_{2i}) = 0, \\ \frac{\partial \ell(\underline{y}; \Theta)}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (y_i - \mu)^2 - \frac{\lambda}{\sigma^2} \sum_{i=1}^n (y_i - \mu) \mathcal{I}(r_{1i}, r_{2i}) = 0, \end{aligned}$$

where

$$\mathcal{H}(r_{1i}, r_{2i}) = \frac{\phi(r_{1i}) - \phi(r_{2i})}{\Phi(r_{1i}) + \Phi(r_{2i})}, \quad \mathcal{I}(r_{1i}, r_{2i}) = \frac{\phi(r_{1i}) + \phi(r_{2i})}{\Phi(r_{1i}) + \Phi(r_{2i})}$$

and

$$r_{1i} = \frac{\lambda(y_i - \mu)}{\sigma} + \alpha, \quad r_{2i} = \frac{\lambda(y_i - \mu)}{\sigma} - \alpha.$$

Since it is not possible to obtain closed expressions for the maximum likelihood estimators, algorithms based on numerical methods, such as Newton-Raphson or Broyden-Fletcher-Goldfarb-Sanno (BFGS), among others, will have to be used. It is recommended to use different seed points as initial values to ensure that the solution obtained constitutes a global maximum of the logarithm-likelihood function.

The standard errors of the estimators can be obtained by inverting the Hessian matrix. Both Mathematica and WinRats have at least two methods for this purpose. The first is to use Cholesky factors (this package is available on the web upon request). The second, faster method, involves by finite differentiation. Furthermore, WinRats package also offers the possibility to get the maximum of the log-likelihood directly giving us the elements of the Fisher information matrix. In fact, for the examples considered later these two packages were used to get the maximum likelihood estimators in a fast way.

5. Multivariate Versions

Multivariate extensions of the univariate distributions arise in an easy way as we can see in the next result.

Theorem 4. Let X and \underline{Y} be two random variables where $X \sim N(0, 1)$ and $\underline{Y} \sim N^{(m)}(\underline{0}, \Sigma)$. The n,

$$f(\underline{y}; \underline{\lambda}, \alpha) = \frac{f_{\underline{Y}}(\underline{y})}{\alpha} \int_{-\alpha}^{\alpha} F_X(z + \underline{\lambda}^T \underline{y}) dz \tag{16}$$

represents a genuine pdf for $\alpha \in \mathbb{R} - \{0\}$ and $\underline{\lambda} \in \mathbb{R}^m$.

Proof. Without loss of generality, assume that X and \underline{Y} are independent random variables. Taking into account the fact that $X - \underline{\lambda}^T \underline{Y}$ is symmetric and using the result provided in Theorem 1, we get

$$\begin{aligned} \alpha &= \int_{-\alpha}^{\alpha} \Pr(X - \underline{\lambda}^T \underline{Y} < z) dz \\ &= \int_{\mathbb{R}^m} \left[\int_{-\alpha}^{\alpha} \Pr(X < z + \underline{\lambda}^T \underline{y} | \underline{Y} = \underline{y}) dz \right] f_{\underline{Y}}(\underline{y}) d\underline{y} \\ &= \int_{\mathbb{R}^m} \left[\int_{-\alpha}^{\alpha} \Pr(X < z + \underline{\lambda}^T \underline{y}) dz \right] f_{\underline{Y}}(\underline{y}) d\underline{y} \\ &= \int_{\mathbb{R}^m} f_{\underline{Y}}(\underline{y}) \left(\int_{-\alpha}^{\alpha} F_X(z + \underline{\lambda}^T \underline{y}) dz \right) d\underline{y}. \end{aligned}$$

Hence the result. \square

Remark 1. The only important required property of the distribution of \underline{Y} is that, for any $\underline{\lambda}$, the random variable $\underline{\lambda}^T \underline{Y}$ is symmetric about zero. The only required property for the distribution of X is that it be symmetric about zero. This is true for the above Theorem and the next.

Theorem 5. Let X and \underline{Y} be two random variables where $X \sim N(0, 1)$ and $\underline{Y} \sim N^{(m)}(\underline{0}, \Sigma)$. The n,

$$f(\underline{y}; \underline{\lambda}, \alpha) = f_{\underline{Y}}(\underline{y}) \left[F_X(\underline{\lambda}^T \underline{y} + \alpha) + F_X(\underline{\lambda}^T \underline{y} - \alpha) \right] \tag{17}$$

is a valid pdf for $\alpha \in \mathbb{R}$ and $\underline{\lambda} \in \mathbb{R}^m$.

Proof. From (16) it follows that

$$\begin{aligned} \alpha &= \int_{\mathbb{R}^m} f_{\underline{Y}}(\underline{y}) \left(\int_{-\alpha}^{\alpha} F_X(z + \underline{\lambda}^T \underline{y}) dz \right) d\underline{y} \\ &= \int_{\mathbb{R}^m} f_{\underline{Y}}(\underline{y}) \left[\int_{-\alpha}^0 F_X(z + \underline{\lambda}^T \underline{y}) dz + \int_0^{\alpha} F_X(z + \underline{\lambda}^T \underline{y}) dz \right] d\underline{y}. \end{aligned} \tag{18}$$

If we now take the derivative with respect to α on both sides in Equation (18) and apply the Fundamental Theorem of calculus we get

$$\int_{\mathbb{R}^m} f_Y(\underline{y}) \left[F_X(\underline{\lambda}^T \underline{y} + \alpha) + F_X(\underline{\lambda}^T \underline{y} - \alpha) \right] d\underline{y} = 1.$$

Thus (17) is a genuine pdf. \square

Note that if we set $\alpha = 0$ in (17), we obtain

$$f(\underline{y}; \underline{\lambda}) = 2f_Y(\underline{y})F_X(\underline{\lambda}^T \underline{y}),$$

which was one of the first multivariate skew-normal models to appear in the literature. See [10,11], for instance.

Figure 3 shows the density for bivariate generalized skew-normal (BGSN) model for some combinations of the parameters. Perusal of Figure 3, will confirm that the density of the BGSN model exhibits a more interesting array of possible shapes than do many of its competitors. The flexibility of this model can be expected to be useful in fitting the model to a variety of different data sets.

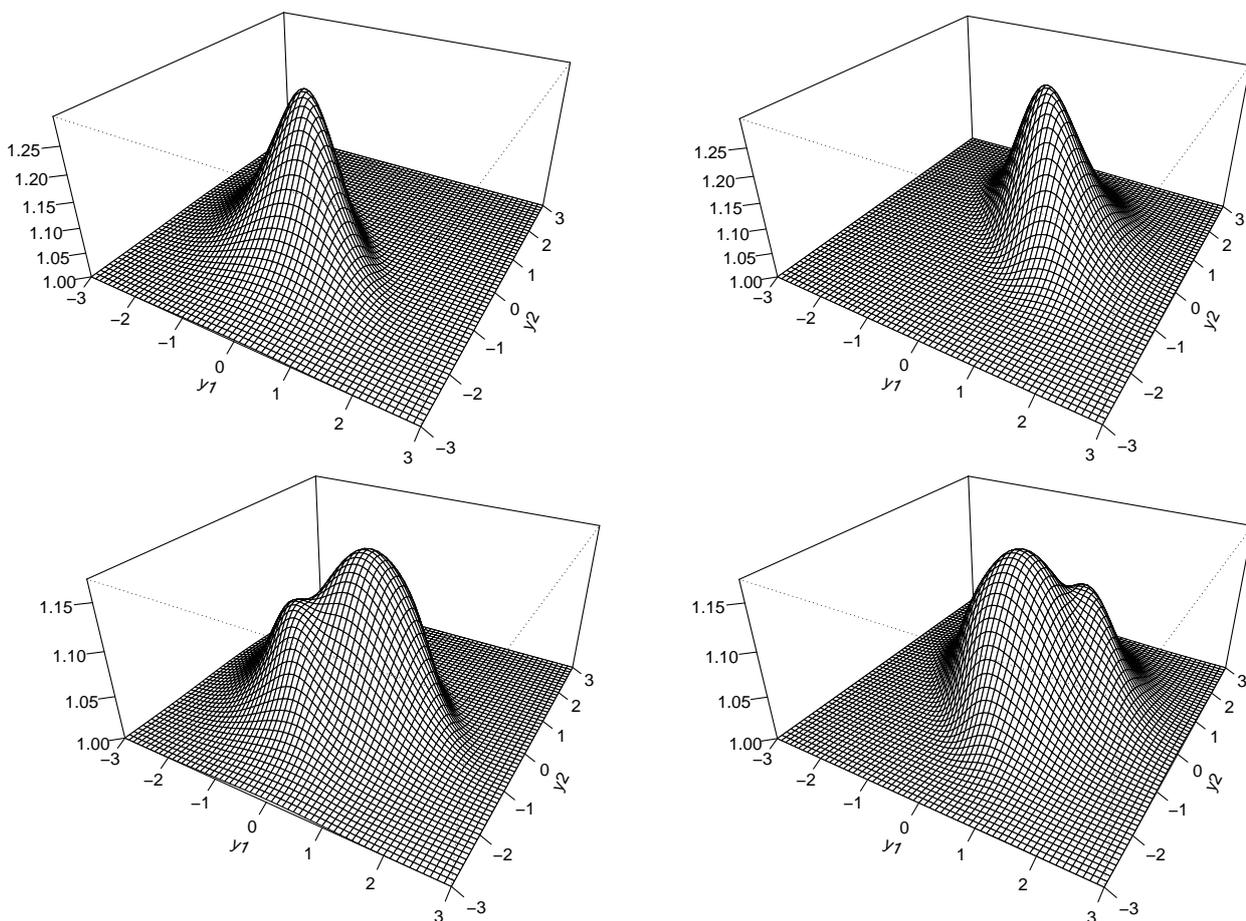


Figure 3. Density for BGSN for: $\lambda_1 = \lambda_2 = -3$ and $\alpha = 1$ (upper left panel); $\lambda_1 = \lambda_2 = 3$ and $\alpha = 1$ (upper right panel); $\lambda_1 = \lambda_2 = -3$ and $\alpha = 5$ (lower left panel); $\lambda_1 = \lambda_2 = 3$ and $\alpha = 5$ (lower right panel).

6. Numerical Illustrations

In this section, three examples for the GSN model given in Equation (14) are carried out and the results are compared with the flexible epsilon skew-normal (FESN) model introduced by Reference [33] in the first example, with the mixture of two normals (MN)

model in the second example and flexible skew-normal (FSN) model introduced by Reference [34] in the third example. The three densities respectively are given by:

$$\begin{aligned}
 1. \quad f(y; \mu, \sigma, \lambda, \varepsilon) &= \begin{cases} \frac{1}{2\sigma c_\lambda} \phi\left(\frac{y-\mu}{\sigma(1+\varepsilon)} - \lambda\right) & \text{if } y < \mu \\ \frac{1}{2\sigma c_\lambda} \phi\left(\frac{y-\mu}{\sigma(1-\varepsilon)} + \lambda\right) & \text{if } y \geq \mu \end{cases} \\
 2. \quad f(y; \mu, \mu_1, \sigma, \sigma_1, p) &= \frac{p}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) + \frac{1-p}{\sigma_1} \phi\left(\frac{y-\mu_1}{\sigma_1}\right) \\
 3. \quad f(y; \mu, \sigma, \lambda, \alpha) &= \frac{2}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) \Phi\left(\lambda\left(\frac{y-\mu}{\sigma}\right) + \alpha\left(\frac{y-\mu}{\sigma}\right)^3\right)
 \end{aligned}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and distribution functions of the standard normal distribution, $c_\lambda = 1 - \Phi(\lambda)$, $\mu, \mu_1, \lambda, \alpha \in \mathbb{R}$, $\sigma, \sigma_1 \in \mathbb{R}^+$, $-1 < \varepsilon < 1$ and $0 \leq p \leq 1$.

We use these three models, since they have been used in the applied statistics literature to explain bimodal empirical data. We chose the MN model since it is a classic model that is used to model bimodal data sets, we chose the FESN model since it is one of the first bimodal extensions of the family of epsilon-skew-simetric distributions and We chose the FSN model since it is one of the first bimodal extensions of the family of skew-simetric distributions.

6.1. Example 1

The data in this example is a set of fiber levels for 315 patients and is available online at http://Lib.stat.cmu.edu/datasets/Plasma_Retinol and contains values for 14 variables for each patient. For our analysis we will use only the variable called Fiber (Grams of fiber consumed per day). Low levels of this variable may be associated with higher risk of certain types of cancer. Descriptive statistics for the data set are displayed in Table 3. In the table b_1 and b_2 denote sample skewness and kurtosis coefficients. Note that the data exhibits a high level of kurtosis.

Table 3. Fiber: Descriptive statistics.

n	\bar{y}	s^2	b_1	b_2
315	12.789	28.411	1.147	5.425

The estimated values of the parameters for the two models are shown in Table 4 together with the standard errors (SE) in parentheses. The table also includes the maximum of the log-likelihood function (ℓ_{\max}), the Akaike information criterion (AIC) and the consistent Akaike information criterion (CAIC), proposed in References [35,36] respectively. A model with a lower AIC or CAIC is preferred to a model with a higher value.

Table 4. Parameter estimates (SE) for FESN and GSN models.

Parameter	FESN	GSN
μ	7.176 (0.405)	6.714 (0.329)
σ	4.396 (0.446)	8.076 (0.406)
λ	-0.288 (0.235)	9.692 (2.510)
α	-	2.486 (0.764)
ε	-0.695 (0.048)	-
ℓ_{\max}	-949.458	-945.575
AIC	1906.916	1899.150
CAIC	1925.926	1918.160

Graphs of histogram of the data and fitted densities are given in Figure 4. As it can be seen, the GSN distribution is the better of the two models with regard to reflecting the

nature of the empirical data. All computations here were done using Mathematica v.11.0 and WinRATS v.7.0. These codes are available from the authors upon request.

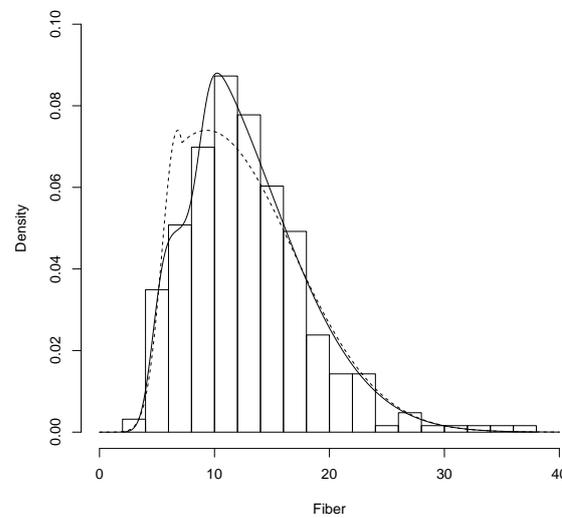


Figure 4. FESN distribution (dashed line) and GSN distribution (solid line) for the Fiber data.

6.2. Example 2

We consider the variables *M-Sweet* available in the database creaminess of cream cheese which can be found at <http://www.models.kvl.dk/research/data/Cream/index.asp>.

Table 5 shows summary statistics for the M-Sweet dataset.

Table 5. M-Sweet: Descriptive statistics.

n	\bar{y}	s^2	b_1	b_2
240	3.276	1.964	0.882	5.049

In Table 6, presents parameter estimates (SE) for both, the MN and GSN models. It can be seen that the log-likelihood for the GSN model is higher than the for the MN model. The AIC and CAIC criterion are used again to compare the estimated models, it can be seen that the GSN model presents the best fit (smallest AIC and CAIC values).

Table 6. Parameter estimates (SE) for MN and GSN models.

Parameter	MN	GSN
μ	2.115 (0.115)	2.577 (0.126)
σ	0.498 (0.104)	1.564 (0.091)
λ	-	3.780 (0.852)
α	-	3.685 (0.944)
μ_1	3.727 (0.176)	-
σ_1	1.376 (0.084)	-
p	0.280 (0.077)	-
AIC	828.214	826.585
CAIC	850.617	844.507

Finally, the histogram of the data and the fitted densities are shown in Figure 5.

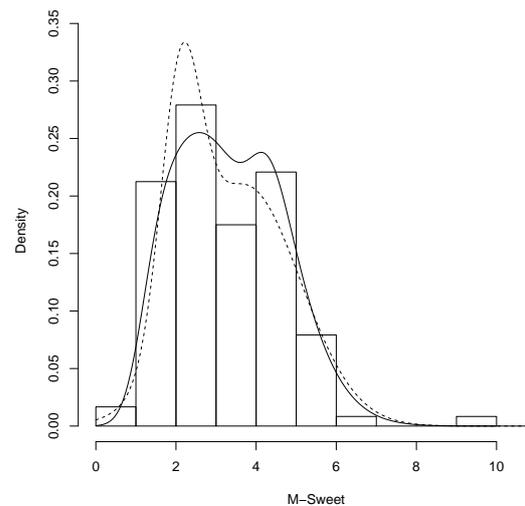


Figure 5. MN distribution (dashed line) and GSN distribution (solid line) for the M-Sweet data.

6.3. Example 3

Finally, data corresponding to the age and frequency of cancer cataloged as Kaposi’s sarcoma have been taken. This is a type of cancer that can form masses in the skin, lymph nodes, or other organs without distinguishing the sub-types. The data has been collected from the website of the Office for National Statistics (ONS, section Health statistics) and it can be seen in Table A1 in the Appendix A. It can be observed that there is a higher incidence in individuals with an age around 25 years as well as for those with an age around 60 years. The se records were taken during the years 1995–2016 and correspond to different regions of the United Kingdom. Table 7 shows summary statistics for the Kaposi’s sarcoma dataset.

Table 7. Kaposi’s sarcoma: Descriptive statistics.

n	\bar{y}	s^2	b_1	b_2
29131	45.396	416.487	0.313	1.936

The two fitted models are represented in Figure 6 and the correspondent estimated values can be seen in Table 8. The GSN model presents better fit for the data, since the AIC and CAIC values are smaller.

Table 8. Parameter estimates (SE) for FSN and GSN models.

Parameter	FSN	GSN
μ	17.896 (0.119)	37.039 (0.139)
σ	34.245 (0.171)	22.052 (0.105)
λ	6.016 (0.118)	4.898 (0.118)
α	−1.007 (0.090)	5.525 (0.138)
AIC	255,356.2	253,832.6
CAIC	255,393.3	253,869.7

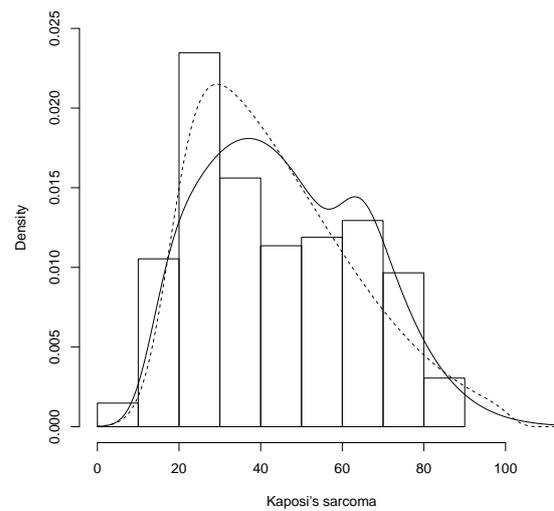


Figure 6. FSN distribution (dashed line) and GSN distribution (solid line) for the Kaposi’s sarcoma data.

7. Conclusions

We have proposed two families of skew distributions which can be considered as alternatives to the well-known skew normal distribution for fitting skewed data.

Future research could address the following issue. We can ask whether the methodology proposed here can be applied to the generalized skew-normal distribution proposed by Reference [9] to obtain a more flexible distribution. For the case with standard normal components, one might consider the following model which is an average of two Arnold and Beaver densities.

$$f_X(x) = \frac{1}{2}\phi(x) \left[\frac{\Phi(\lambda x + \alpha_1)}{\Phi\left(\frac{\alpha_1}{\sqrt{1+\lambda^2}}\right)} + \frac{\Phi(\lambda x - \alpha_2)}{\Phi\left(\frac{-\alpha_2}{\sqrt{1+\lambda^2}}\right)} \right],$$

where $\lambda, \alpha_1, \alpha_2 \in \mathbb{R}$. Note that this model is not obtainable by methods analogous to those used to develop the model (10). However, it is a simple more flexible extension of the Arnold-Beaver model. But, once we recognize it as a mixture with equal weights, it is reasonable to add more flexibility by considering unequal weights, as follows

$$f_X(x) = \phi(x) \left[\gamma \frac{\Phi(\lambda x + \alpha_1)}{\Phi\left(\frac{\alpha_1}{\sqrt{1+\lambda^2}}\right)} + (1 - \gamma) \frac{\Phi(\lambda x - \alpha_2)}{\Phi\left(\frac{-\alpha_2}{\sqrt{1+\lambda^2}}\right)} \right],$$

where $\gamma \in [0, 1]$.

Author Contributions: Conceptualization, E.G.-D., B.C.A. and J.M.S.; Formal analysis, E.G.-D., B.C.A., J.M.S. and H.W.G.; Investigation, E.G.-D., B.C.A., J.M.S. and H.W.G.; Methodology, E.G.-D. and J.M.S.; Software, E.G.-D. and H.W.G.; Supervision, E.G.-D. and B.C.A.; Validation, E.G.-D. and H.W.G. All of the authors contributed significantly to this research article. All authors have read and agreed to the published version of the manuscript.

Funding: The authors thank to the Ministerio de Economía y Competitividad, Spain (project ECO2017-85577-P, EGD and Ministerio de Ciencia e Innovación, Spain (PID2019-105986GB-C22), JMS for partial support of this work. The research of H.W. Gómez was supported by PUENTE-UA project, Chile.

Acknowledgments: The authors thank the Editor and three anonymous referees for their constructive comments and suggestions, which have greatly helped them to improve the paper.

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

Appendix A

Table A1. Data corresponding to Kaposi's sarcoma.

Age	Number
1	1
5	89
10	342
15	718
20	2352
25	3593
30	3243
35	2533
40	2015
45	1747
50	1562
55	1662
60	1801
65	1915
70	1855
75	1611
80	1203
85	642
90	247

References

- Aigner, D.J.; Lovell, C.A.K.; Schmidt, P. Formulation and estimation of stochastic frontier production functions. *J. Econ.* **1977**, *6*, 21–37. [\[CrossRef\]](#)
- Kumbhakar, S.C.; Parmeter, C.F.; Tsionas, E.G. A zero inefficiency stochastic frontier model. *J. Econ.* **2013**, *172*, 66–76. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Jones, M.C. Families of distributions arising from distributions of order statistics. *Test* **2007**, *13*, 1–43. [\[CrossRef\]](#)
- Azzalini, A. A class of distributions which includes the normal ones. *Scan. J. Stat.* **1985**, *12*, 171–178.
- Henze, N. A probabilistic representation of the Skew-Normal distribution. *Scan. J. Stat.* **1986**, *4*, 271–275.
- Silver, E.A.; Costa, D. A property of symmetric distributions and a related order statistic result. *Am. Stat.* **1997**, *51*, 32–33.
- Gupta, R.C.; Gupta, R.D. Generalized skew normal model. *Test* **2004**, *12*, 501–524. [\[CrossRef\]](#)
- Arnold, B.C.; Beaver, R.J. Skewed multivariate models related to hidden truncation and/or selective reporting (with discussion). *Test* **2002**, *11*, 1–54. [\[CrossRef\]](#)
- Azzalini, A.; Capitanio, A. Statistical applications of the multivariate skew-normal distribution. *J. R. Stat. Soc. Ser. B* **1999**, *61*, 579–602. [\[CrossRef\]](#)
- Azzalini, A.; Valle, A. The multivariate skew-normal distribution. *Biometrika* **1996**, *83*, 715–726. [\[CrossRef\]](#)
- Azzalini, A. Further results on a class of distributions which includes the normal ones. *Statistica* **1986**, *46*, 199–208.
- Azzalini, A.; Chiogna, M. Some results on the stress-strength model for skew-normal variates. *METRON* **2004**, *LXII*, 315–326.
- Gupta, R.D.; Gupta, R.C. Analyzing skewed data by power normal model. *Test* **2008**, *17*, 197–210. [\[CrossRef\]](#)
- Gómez, H.; Venegas, O.; Bolfarine, H. Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics* **2007**, *18*, 395–407. [\[CrossRef\]](#)
- Sharafi, M.; Behboodan, J. The Balakrishnan skew-normal density. *Stat. Paper.* **2008**, *49*, 769–778. [\[CrossRef\]](#)
- Jones, M.; Pewsey, A. Sinh-arcsinh distributions. *Biometrika* **2009**, *96*, 761–780. [\[CrossRef\]](#)
- Gómez-Déniz, E.; Iriarte, Y.A.; Calderín-Ojeda, E.; Gómez, H.W. Modified Power-Symmetric Distribution. *Symmetry* **2019**, *11*, 1410. [\[CrossRef\]](#)
- Alavi, S.M.R.; Tarhani, M. On a Skew Bimodal Normal-Normal distribution fitted to the Old-Faithful geyser data. *Commun. Stat. Theor. Method.* **2017**, *46*, 7301–7312. [\[CrossRef\]](#)
- García, V.; Gómez-Déniz, E.; Vázquez-Polo, F. A new skew generalization of the normal distribution: Properties and applications. *Comput. Stat. Data Anal.* **2010**, *54*, 2021–2034. [\[CrossRef\]](#)
- Azzalini, A. *The Skew Normal and Related Families*; Cambridge University Press: Cambridge, UK, 2014.
- Dickson, D. *Insurance Risk and Ruin*; Cambridge University Press: Cambridge, UK, 2005.
- Rolski, T.; Schmidli, H.; Schmidt, V.; Teugels, J. *Stochastic Processes for Insurance and Finance*; John Wiley and Sons: Hoboken, NJ, USA, 1999.

24. Burdick, D. A note on symmetric random variables. *Annal. Math. Stat.* **1972**, *43*, 2039–2040. [[CrossRef](#)]
25. Jones, M. Generating distributions by transformation of scale. *Stat. Sinica* **2014**, *24*, 749–771. [[CrossRef](#)]
26. Ding, P. Three Occurrences of the Hyperbolic-Secant Distribution. *Am. Stat.* **2014**, *68*, 32–35. [[CrossRef](#)]
27. Johnson, N.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*; Wiley: New York, NY, USA, 1995.
28. Arnold, B.C.; Groeneveld, R.A. Some Properties of the Arcsine Distribution. *J. Am. Stat. Assoc.* **1980**, *75*, 173–175. [[CrossRef](#)]
29. Ellison, B. Two theorems for inferences about the normal distribution with applications in acceptance sampling. *J. Am. Stat. Assoc.* **1964**, *59*, 89–95. [[CrossRef](#)]
30. Zacks, S. *Parametric Statistical Inference*; Pergamon Press: Oxford, UK, 1981.
31. Chiogna, M. Some results on the scalar skew-normal distribution. *J. Ital. Statist. Soc.* **1998**, *1*, 1–13. [[CrossRef](#)]
32. Owen, D. Tables for computing bivariate normal probabilities. *Ann. Math. Stat.* **1956**, *27*, 1075–1090. [[CrossRef](#)]
33. Arellano-Valle, R.B.; Cortés, M.A.; Gómez, H.W. An Extension of the Epsilon-Skew-Normal Distribution. *Communicat. Stat. Theor. Method.* **2010**, *39*, 912–922. [[CrossRef](#)]
34. Yanyuan, M.A.; Genton, M.G. Flexible Class of Skew-Symmetric Distributions. *Scand. J. Stat.* **2004**, *31*, 459–468.
35. Bozdogan, H. The general theory and its analytical extension. *Psychometrika* **1987**, *52*, 345–370. [[CrossRef](#)]
36. Akaike, H. A new look at the statistical model. *IEEE Transact. Automat. Control.* **1974**, *19*, 716–723. [[CrossRef](#)]