REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000-000
https://doi.org/00.00000/revstat.v00i0.000

Bimodal and Multimodal Extensions of the Normal and Skew Normal Distributions

Authors: Emilio Gómez–Déniz
– Department of Quantitative Methods in Economics and TiDES Institute,
University of Las Palmas de Gran Canaria, Spain
emilio.gomez-deniz@ulpgc.es

Enrique Calderín-Ojeda ✉
– Department of Economics,
University of Melbourne, Australia
enrique.calderin@unimelb.edu.au

José M. Sarabia
– Department of Quantitative Methods,
CUNEF Universidad, Spain
josemaria.sarabia@cunef.edu

Received: Month 0000 Revised: Month 0000 Accepted: Month 0000

Abstract:

- A transformation of a density function is introduced to derive two families of continuous densities, the first symmetric and the second not-necessarily symmetric, exhibiting both unimodality and bimodality. Their respective density functions are provided in closed form, allowing us to simply obtain moments and related quantities. We focus on the case where the normal distribution is considered, although it can be applied to other models, such as the logistic and Cauchy distributions. This transformation is also extended to derive a family of asymmetric unimodal and bimodal distributions via Azzalini's scheme. An example related to environmental science illustrate these models' practical performance.

Keywords:

- *Multimodality; Old faithful geyser data; Skewness; Unimodality; Univariate distribution.*

AMS Subject Classification:

- 62P05, 97M30, 91B30.

✉ Corresponding author

1. INTRODUCTION

We use an old theorem proven over ninety years ago to obtain bimodal and multimodal extensions of the normal distribution and the skew-normal distribution. One can almost certainly say that the normal distribution constitutes the queen of the comprehensive family of the continuous probability distributions. Since the end of the 19th century, numerous researchers, such as the distinguished F. Y. Edgeworth, and also Chas. H. Kummel, Arthur L. Bowley, Morgan W. Crofton, among many others derived modifications of the normal law to discuss situations where the empirical data presented some asymmetry that the normal distribution could not explain. A review of the normal distribution and some of its modifications can be found in [27].

Bimodal distributions arise in nature in many different scenarios. Perhaps, one of the most relevant phenomena that can be explained with distributions is the disease patterns. For example, the incidence of some types of cancers by age displays a major mode for young adults and minor mode for older adults see [2]. In addition, the occurrence of bimodality has also implications in geoscience see [22]. 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 and skew-normal densities that may be unimodal or bimodal. This new family of distributions that arises from an old Theorem provided by [33] comprises flexible parametric families of continuous distributions that are useful in statistical practice.

In the last years, different techniques to extend the normal family have been deemed in the statistical literature. The skew-normal distribution in [6] see also [7], the Balakrishnan skew-normal density in [31] more details in [35], the generalization proposed by [5], the Sinh-arcsinh family introduced by [24], the generalized normal one in [16], [18] and [19], and the recently proposed models provided by [36] and [34], among others. Some other works related to the normal and skew normal densities are [3], [4] and [17]. For a comprehensive review of the skew normal families the reader is referred to [8].

The density function introduced here resembles some important properties satisfied by the normal distribution. The first family is symmetric with positive real support. The second family is asymmetric and defined on the positive real numbers. In general, both families show bimodality. An overview of this work that will undoubtedly help the reader to understand better the elements that are not so essential is illustrated in the flowchart displayed in Figure 4.

The rest of this paper is structured as follows. In Section 2 we derive the methodology based on the use of a result provided in [33] to derive the new family of distribution. Here, expressions for the mean, variance, and other features for the general model are also provided. Next, we also examine the special case of considering the classical normal distribution as the parent distribution. Then, to break the symmetry of the latter case, we introduce the skew-normal distribution as the baseline model. In Section 3, the parameter estimation problem is discussed. Some illustrative examples related to environmental issues, in particular in geoscience, are analyzed in Section 4. Finally, closing comments and modifications of the models proposed are shown in the last Section.

2. The proposed model

This section gives the main results of this paper, from which we derive the two families of probability density functions that will be described later. The first family is introduced in the second theorem of this section. Although any distribution with support on the real line can be used as a candidate of this new distribution, the normal case is the one we are examining in this section. It can be simply shown after a change of variable that this model is connected to the generalized inverse Gaussian distribution. This probabilistic family is symmetric and has two modal values that are equidistant with respect to the axis of symmetry. The second family presents the advantage of having an asymmetric density function. We begin with the following Theorem found in [33] that is required for the main result of this work.

Theorem 2.1 (Slobin, 1927)**.** Let the function $\omega(x) = x - 1/x$, $x \neq 0$. Then, if the function $m(x)$ is a function integrable on $\mathbb{R} = (-\infty, \infty)$ and if the function $m(\omega(x))$ is also integrable in $\mathbb{R} = (-\infty, \infty)$, we have that

$$\int_{-\infty}^{\infty} m\left(\omega(x)\right)\, \mathrm{d}x = \int_{-\infty}^{\infty} m(x)\, \mathrm{d}x. \tag{2.1}$$

Following the same arguments that the ones provided in the proof of the above Theorem given in [33], it is simple to observe that (2.1) is also valid for $\omega_\alpha(x) = x - \alpha/x$, being $\alpha \geq 0$. The following result provides an alternative and more simple proof than the one given in [33] for this case. Previously we need the following Lemma, which is provided in [10].

Lemma 2.1 (Behboodian, 1978)**.** Let X be a symmetric random variable, and let $y = h(x)$ be an odd real-valued function. Then, the random variable $Y = h(X)$ is also symmetric.

As a result of this Lemma, if X is a symmetric random variable then the random variable $Y = \omega_\alpha(X)$ is also symmetric. In the next result we derive an expression for the density function of $Y = \omega_\alpha(X)$.

Theorem 2.2. Let $f(x)$ be a probability density function (pdf hereafter) symmetric about 0 and consider the function $f(\omega_\alpha(x))$, with $\omega_\alpha(x) = x - \alpha/x$, being $\alpha \geq 0$. Then, if $\mathrm{d}f(\omega_\alpha(x))/(\mathrm{d}\alpha)$ is also a symmetric function we have that $\int_{-\infty}^{\infty} f(\omega_\alpha(x))\, \mathrm{d}x = 1$.

Proof: Since $f(x)$ is symmetrical and $\omega_\alpha(x)$ is an odd function, using Lemma 2.1 we have that $f(\omega_\alpha(x))$ is also symmetrical. Now, consider the function $\nu(\alpha) = \int_{-\infty}^{\infty} f(\omega_\alpha(x))\,\mathrm{d}x$ for which we have that

$$\nu'(\alpha) = \frac{\mathrm{d}}{\mathrm{d}\alpha}\nu(\alpha) = -\int_{-\infty}^{\infty} \frac{1}{x}\frac{\mathrm{d}}{\mathrm{d}\alpha} f(\omega_\alpha(x))\,\mathrm{d}x = 0,$$

because $\mathrm{d}f(\omega_\alpha(x))/(\mathrm{d}\alpha)$ is symmetrical (by assumption). Therefore, $\nu(\alpha)$ is constant and since $\nu(0) = 1$ we have the result. □

Based on the use of Theorem 2.2 we can build a family of pdf's by taking

$$g_\alpha(x) = \begin{cases} f(\omega_\alpha(x)), & x \neq 0, \\ f(0), & x = 0, \end{cases} \tag{2.2}$$

where $\alpha \geq 0$. Note that this is a two piece-wise pdf.

The following proposition displays some essential properties related to this distribution.

Proposition 2.1. The pdf given in (2.2) satisfies the following properties:

(i) $g_\alpha(x)$ is symmetric about zero. That is, $g_\alpha(x) = g_\alpha(-x)$ for all $x \in \mathbb{R}$. In fact, the random variable $Z = -X$ follows the same distribution that X.

(ii) $g_0(x) = f(x)$.

(iii) $g_\alpha(0) = f(0)$ for all $\alpha \geq 0$.

(iv) $\mathbb{E}(X^{2\kappa+1}) = 0$, $\kappa \in \{0, 1, \dots\}$. That is, all odd raw moments are zero.

(v) The random variables $Y = \omega_\alpha(X)$ and $Z = g_\alpha(X)$ are uncorrelated and therefore $cov(Y, Z) = 0$, provided that all the first and second moments of Y and Z exist.

Proof: Properties $(i) - (iv)$ are direct. To show (v), observe that $\omega_\alpha(x)$ is an odd function, $g_\alpha(x)$ is an even real-valued (measurable) function and the random variable $T = Y\,Z$ satisfies that $T(-x) = \omega_\alpha(-x)g_\alpha(-x) = -\omega_\alpha(x)g_\alpha(x) = -T(x)$, therefore is an odd function. Thus, $\mathrm{cov}(Y, Z) = \mathbb{E}(Y\,Z) - \mathbb{E}(Y)\mathbb{E}(Z) = 0$, because $\mathbb{E}(Y) = 0$ (due to Lemma 2.1, Y is symmetrical) and $\mathbb{E}(Y\,Z) = 0$ ($T = Y\,Z$ is an odd function). For more deails see [10]. □

2.1. The normal case

Natural choices for $f(x)$ to be plugged into (2.2) 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. $f(x) = \phi(x)$, being $\phi(x)$ the pdf of the standard normal distribution. Then, it is simple to see that

$$g_\alpha(x) = \begin{cases} \phi(\omega_\alpha(x)), & x \neq 0, \\ \phi(0), & x = 0 \end{cases} \tag{2.3}$$

is a genuine pdf for $\alpha \geq 0$. Note that the special case $\alpha = 0$ represents the standard normal distribution. Simple algebra provides that the distribution is symmetric about zero and has mean and variance given by 0 and $1 + \alpha$, respectively. The distribution is always bimodal, with two modes in $x = -\sqrt{\alpha}$ and $x = \sqrt{\alpha}$. To see this, observe that

$$g'_\alpha(x) = -g_\alpha(x)\left(x - \frac{\alpha}{x}\right)\left(1 + \frac{\alpha}{x^2}\right) = 0$$

for $x = \pm\sqrt{\alpha}$. Now, it is simple to see that $g''_\alpha(\pm\sqrt{\alpha}) < 0$. The antimode is obviously $x = 0$. Henceforward, we will write $X \sim \text{BN}(\alpha)$ when the random variable X follows the pdf given in (2.3), denoting that is a bimodal generalization of the normal distribution.

The entropy does not depend on α and is equivalent to the one of the standard normal distribution. Observe that $\lim_{x\to 0^+} g_\alpha(x) = \lim_{x\to 0^-} g_\alpha(x) = \phi(0)$ and thus the pdf defined in (2.3) is a continuous function.

Figure 1 displays the graphs of the pdf given in (2.3) for selected values of parameter $\alpha \geq 0$. The α parameter, the only parameter of the distribution, clearly indicates two fundamental things: first, if it takes the value zero, we are in the case of the standard normal distribution; second, a value other than zero provides a distribution with two modes that are equidistant with respect to the axis of symmetry. The distance between the modes increases with the value of α.

2.2. Connection with others distributions

The following result connects the proposed distribution with the generalized inverse Gaussian distribution. Recall that a continuous variable $Z > 0$ follows a generalized inverse Gaussian distribution see [25] and [23, Chapter 15] with parameters $a > 0$, $b > 0$ and $r \in \mathbb{R}$ if its pdf is given by

$$f(z) = \frac{(a/b)^{r/2}}{2K_r(\sqrt{ab})} z^{r-1} \exp\left[-\frac{1}{2}\left(az + \frac{b}{z}\right)\right], \quad z > 0, \tag{2.4}$$

where $K_\nu(s)$ gives the modified Bessel function of the second kind. Furthermore, if Z follows a generalized inverse Gaussian distribution, then $1/Z$ follows a reciprocal generalized inverse Gaussian distribution. Additionally, simple computation provides that the random variable $1/X^2$ follows a reciprocal generalized inverse Gaussian distribution.

Proposition 2.2. Let $X \sim \text{BN}(\alpha)$ with the pdf given in (2.3). Then, the random variable $V = X^2$ follows a generalized inverse Gaussian distribution with parameters $a = 1$, $b = \alpha^2$ and $r = 1/2$.

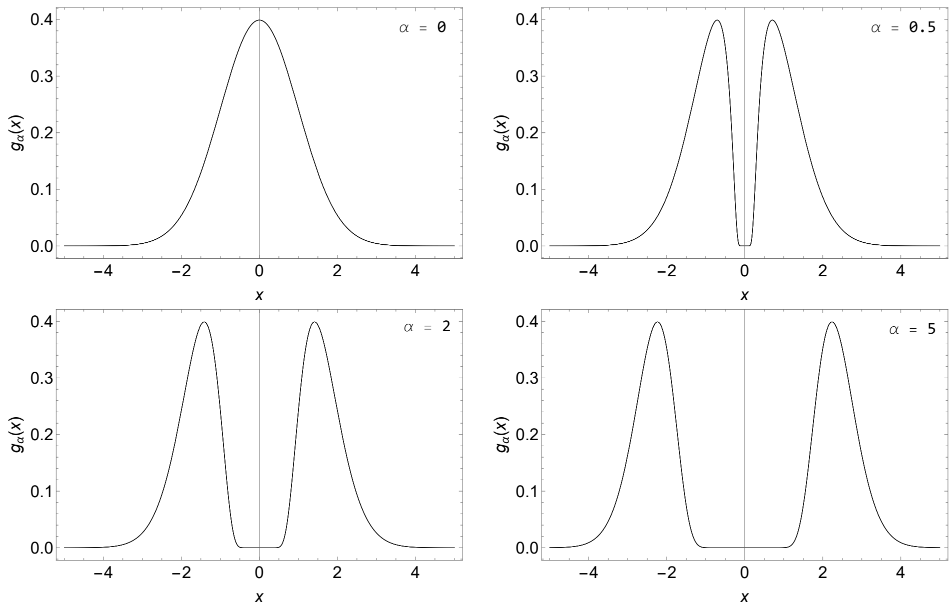

Figure 1: Plots of the pdf $g_\alpha(x)$ for selected values of the parameter α

Proof: Since $\mathrm{d}x = 1/(2\sqrt{v})\mathrm{d}v$ we have that

$$\begin{aligned} g_\alpha(v) &= \frac{1}{2\sqrt{2v\pi}} \exp\left[-\frac{1}{2}\left(\sqrt{v} - \frac{\alpha}{\sqrt{v}}\right)^2\right] \\ &= \frac{v^{-1/2}\exp(\alpha)}{2\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(v + \frac{\alpha^2}{v}\right)\right]. \end{aligned} \tag{2.5}$$

Now, having into account that $K_{1/2}(\alpha) = \exp(-\alpha)\sqrt{\pi/(2\alpha)}$, the result follows by comparing (2.5) with (2.4). □

Proposition 2.3. Let $X \sim \mathrm{BN}(\alpha)$ with the pdf given in (2.3). Then, it is verified that $\mathbb{E}(X^\kappa) = 0$ if κ (positive or negative) is odd while the even moments (positive or negative) are given by

$$\mathbb{E}(X^{2\kappa}) = \sqrt{\frac{2\alpha^{1+2\kappa}}{\pi}} \exp(\alpha) K_{\kappa+\frac{1}{2}}(\alpha), \quad \kappa \in \{0, 1, \dots\}. \tag{2.6}$$

Proof: Since the distribution given in (2.3) is symmetrical, then all odd-order moments are equal to zero. To see that (2.6) is true, then it is simple to see that the distribution is symmetrical since we have that

$$\mathbb{E}(X^\kappa) = 2\int_0^\infty \phi(\omega_\alpha(x)\,\mathrm{d}x$$

and by making the change of variable $u = x^2$ we get

$$\mathbb{E}(X^\kappa) = \frac{2\exp(\alpha)}{\sqrt{2\pi}} \int_0^\infty u^{(\kappa-1)/2} \exp\left[-\frac{1}{2}\left(u + \frac{\alpha^2}{u}\right)\right] \mathrm{d}u \tag{2.7}$$

from which the result follows immediately by arranging parameters in (2.7) and identifying it with the pdf of the generalized inverse Gaussian distribution given in (2.4). $\square$

In particular, if $\kappa = 1$ we get the second row moment of the distribution, which coincides with the variance, given by $\mathrm{var}(X) = 1 + \alpha$. Furthermore, if $\kappa = -1$ by using (2.6) we have that

$$\mathbb{E}\left(\frac{1}{X^2}\right) = \frac{1}{\alpha}, \quad \alpha \neq 0. \tag{2.8}$$

and

$$\mathbb{E}\left[\left(X - \frac{\alpha}{X}\right)^{2\kappa}\right] = (2\kappa - 1)!!, \tag{2.9}$$

where $n!! = n(n-2)(n-4)\cdots 2 \cdot 1$ represents the double factorial.

Note that property given in (2.9) is shared with the standard normal distribution. Using the series representation of the exponential function, we derive the moment generating function of the distribution, which is given by

$$\mathrm{M}_X(t) = \mathbb{E}[\exp(tX)] = \sum_{j=0}^{\infty} \frac{t^{2j}}{(2j)!} \sqrt{\frac{2\alpha^{1+2j}}{\pi}} \exp(\alpha) K_{j+\frac{1}{2}}(\alpha).$$

Proposition 2.4. The cumulative distribution function (cdf henceforward), $G_\alpha(x) = \Pr(X \leq x)$, for a continuous random variable following the pdf given in (2.3) is

$$G_\alpha(x) = \frac{1}{2}\left[\Phi(\omega_\alpha(x)) + \Phi(\tau_\alpha(x)) \exp(2\alpha)\right], \quad x < 0, \tag{2.10}$$

$$G_\alpha(x) = 1 - \frac{1}{2}\left[\bar{\Phi}(\omega_\alpha(x)) + \bar{\Phi}(\tau_\alpha(x)) \exp(2\alpha)\right], \quad x > 0, \tag{2.11}$$

and $G_\alpha(0) = 1/2$, where $\tau_\alpha(x) = x + \alpha/x$ and $\bar{\Phi}(z) = 1 - \Phi(z)$ is the survival function of the standard normal distribution.

Proof: The proof is obtained in the following way. Let $G_\alpha(-x) = \Pr(X \leq -x)$. Thus,

$$G_\alpha(-x) = \int_{-\infty}^{-x} \phi(\omega_\alpha(t)\, \mathrm{d}t = \int_x^\infty \phi(\omega_\alpha(t)\, \mathrm{d}t,$$

which can be written, after the change of variable $Y = X^2$, as

$$G_\alpha(-x) = \int_x^\infty \frac{\exp(\alpha)}{\sqrt{2y\pi}} \exp\left[-\frac{1}{2}\left(y + \frac{\alpha^2}{y}\right)\right] \mathrm{d}y.$$

Now, by using the cdf of the generalized inverse Gaussian distribution provided in [26] we get, after simple algebra (2.10). Expression (2.11) is obtained in a similar way. $\square$

A random variate X from the random variable with pdf given by (2.3) is derived as follows:

- Generate a random number u from the standard uniform distribution, $\mathrm{U}(0,1)$.
- Generate random variate v from the generalized inverse Gaussian distribution with parameters $a = 1$, $b = \alpha^2$ and $r = 1/2$.
- If $u < 0.5$ then $x = -\sqrt{v}$; otherwise $x = \sqrt{v}$.

2.3. Extensions

The major disadvantage of the family of distributions given in (2.3) lies in its symmetry and also in the fact that the two modes are equidistant with respect to the axis of symmetry. Since $f(\omega_\alpha(x)$ is a symmetric pdf, by using the representation provided by [6], we can consider the more flexible family of pdf's given by

$$g_{\alpha,\lambda}(x) = \begin{cases} 2\Phi(\lambda\, x)\phi(\omega_\alpha(x)), & x \neq 0, \\ \phi(0), & x = 0, \end{cases} \tag{2.12}$$

where $\alpha \geq 0$ and $\lambda \in \mathbb{R}$.

In practice $\Phi(\lambda x)$ can be replaced by $\Phi(\lambda\, m(x))$ for any odd function $m(\cdot)$ in order to ensure that (2.13) represents a proper density function. In particular, we can take $m(x) = \omega_\beta(x)$, $\beta \in \mathbb{R}$, to build the family of pdf's given by

$$g_{\alpha,\beta,\lambda}(x) = \begin{cases} 2\Phi(\lambda\, \omega_\beta(x))\phi(\omega_\alpha(x)), & x \neq 0, \\ \phi(0), & x = 0, \end{cases} \tag{2.13}$$

where $\alpha \geq 0$, $\beta \in \mathbb{R}$ and $\lambda \in \mathbb{R}$. See for instance [8]. Observe that when $\alpha = \beta = 0$ the pdf given in (2.13) reduces to the skew normal density provided in [6]. See also, [7] and [9], among others. [6], [7], [14], [21] and [20], among other papers, provide many properties of the skew normal density. The standard normal distribution is obtained for $\alpha = \lambda = 0$. A probabilistic representation of this family of distribution can be obtained in a similar fashion as the one provided in [7] and [21] see also [8].

To see that (2.13) represents a genuine pdf, we proceed in a similar way as we did in Theorem 2.2. In this case, we have to add that $\Phi(\cdot)$ is a bounded function with a derivative being a symmetric density function about zero. The family (2.13) contains the normal, the skew normal density and others for $\lambda \neq 0$. Furthermore, density (2.3) also appears by mixture (see the discussion of M. Cuadras about the work of [5]). To see this, note that if λ follows a symmetric distribution $\pi(\lambda)$, with $-\infty < \lambda < \infty$, then

$$\int_{-\infty}^{\infty} 2\Phi(\lambda\,\omega_\beta(x))\phi(\omega_\alpha(x))\pi(\lambda)\,\mathrm{d}\lambda = \phi(\omega_\alpha(x)).$$

Hereafter, we will write $X \sim \mathrm{GSN}(\alpha, \beta, \lambda)$ to denote that the pdf of the random variable X follows the pdf given in (2.13).

Generation of random variates from (2.13) is now easy via the following representation of the distribution. Let $X \sim \mathrm{BN}(\alpha)$ and $Z = X\,S_X$ where, conditionally on $X = x \neq 0$, we have

$$S_X = \begin{cases} +1 & \text{with probability } \ \Phi(\lambda\,\omega_\beta(x)), \\ -1 & \text{with probability } \ 1 - \Phi(\lambda\,\omega_\beta(x)). \end{cases}$$

Therefore, a random variate z from the random variable with density function given by (2.13) is derived as follows:

- Generate a random number u from the standard uniform distribution, $\mathrm{U}(0, 1)$.
- Generate random variate x from the distribution with pdf (2.3).
- Compute $\Phi(\lambda\,\omega_\beta(x))$.
- If $u < \Phi(\lambda\,\omega_\beta(x))$ then $z = x$; otherwise $z = -x$.

Then, the random variable Z has the density function given in (2.13). Figure 2 displays some plots of the pdf (2.13) for special values of the parameters.

It is straightforward to verify that properties (2.8) and (2.9) are satisfied also for the distribution (2.13). Some additional results of (2.13) are given below.

Proposition 2.5. The following results are verified:

(i) If $X \sim g_{\alpha,\beta,\lambda}(x)$ then the random variable $Z = -X \sim g_{\alpha,\beta,-\lambda}(z)$. That is, $g_{\alpha,\beta,\lambda}(-x) = g_{\alpha,\beta,-\lambda}(x)$ for all x.

(ii) For all $x \in \mathbb{R}$, the cdf $G_{\alpha,\beta,\lambda}(x) = \Pr(X \leq x)$, verifies:
$G_{\alpha,\beta,\lambda}(x) = G_{\alpha,\beta,-\lambda}(-x)$.

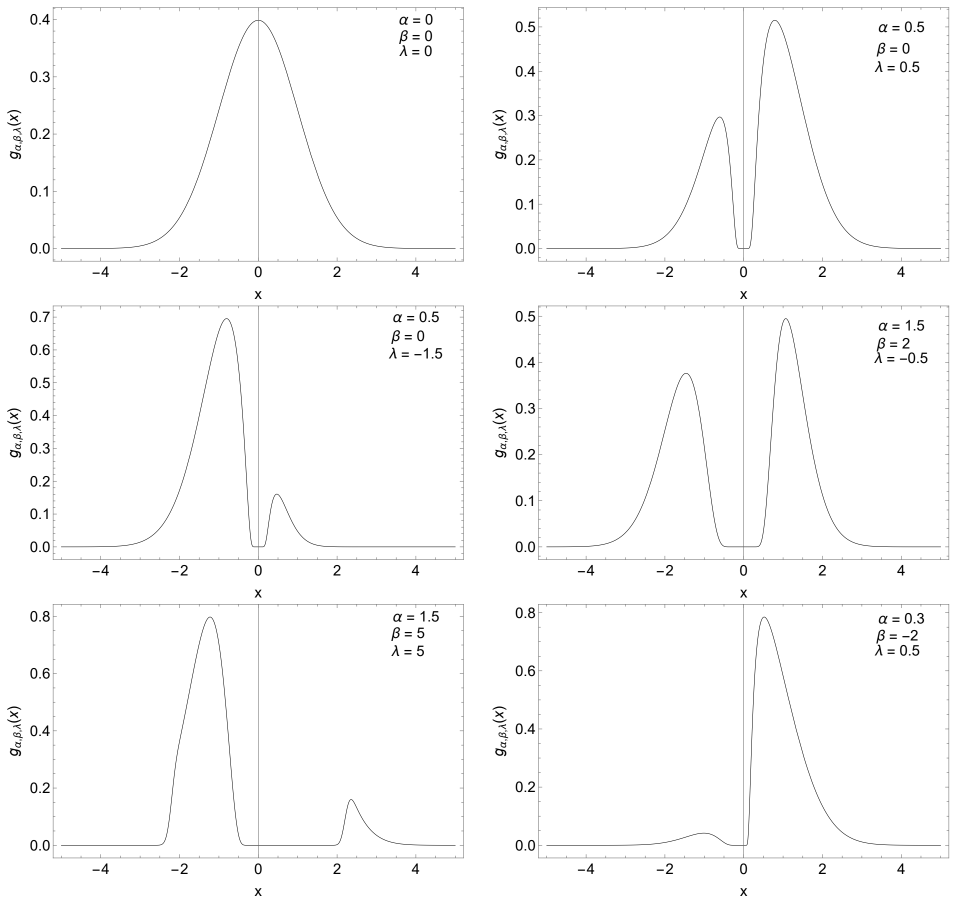

Figure 2: Plots of the pdf (2.13) for selected values of the parameters α, β and λ

Proof: To see (i), observe that given $Z = -X$ we have that $|dz| = |dx|$. Now the result follows having into account that $\lambda\,\omega_\beta(-z) = \lambda(-z+\beta/z) = -\lambda(z-\beta/z) = -\lambda\,\omega_\beta(z)$ and $\phi(\omega_\alpha(-x)) = \phi(\omega_\alpha(x))$. Finally, (ii) follows from (i). $\square$

Proposition 2.6. As $\lambda \to \infty$ and $\beta \to 0$ the pdf given in (2.13) tends to $g_\alpha(x) = 2\phi(\omega_\alpha(x))$, i.e. a generalized half-normal density.

Proof: It is derived as a result of writing (2.13) as

$$g_{\alpha,\beta,\lambda}(x) = 2\left(\int_{-\infty}^{\lambda\omega_\beta(x)} \phi(t)\,\mathrm{d}t\right)\phi(\omega_\alpha(x)),$$

and taking $\lambda \to \infty$. $\square$

For $\lambda \to \infty$ and $\alpha \to 0^+$ the classical half-normal density is obtained.

If $X \sim \mathrm{GSN}(\alpha, \beta, \lambda)$ then its distribution function

$$G_{\alpha,\beta,\lambda}(x) = 2\int_{-\infty}^{x}\int_{-\infty}^{\lambda\omega_\beta(s)} \phi(t)\phi(\omega_\alpha(s))\,\mathrm{d}t\,\mathrm{d}s \tag{2.14}$$

can be represented as the cdf of a bivariate normal distribution. To see this take $\delta = \lambda/\sqrt{1+\lambda^2}$ and consider the change of variable

$$t = \frac{\eta + \delta\,\omega_\beta(s)}{\sqrt{1-\delta^2}}.$$

Then, some algebra provides that (2.14) can be rewritten as

$$G_{\alpha,\beta,\lambda}(x) = \frac{2}{\sqrt{1-\delta^2}}\int_{-\infty}^{x}\left(\int_{-\infty}^{0}\phi\left(\frac{\eta+\delta\,\omega_\beta(s)}{\sqrt{1-\delta^2}}\right)\mathrm{d}\eta\right)\phi(\omega_\alpha(s))\,\mathrm{d}s.$$

Unfortunately, we have not been able to find either the generating moment function or the ordinary moments of the distribution given in (2.13). Finally, by taking logarithm in (2.13), it is simple to verify that this pdf can have two modes which are the solutions of the equation

$$\lambda\left(1+\frac{\beta}{x^2}\right)\phi(\lambda\omega_\beta(x)) - \left(1+\frac{\alpha}{x^2}\right)\Phi(\lambda\omega_\beta(x))\,\omega_\alpha(x) = 0.$$

As most of the multimodal datasets considered in practice are defined on the positive real values, it is convenient to reparametrized the distribution given by (2.3) via a linear transformation, i.e. $Y = \mu + \sigma X$, where $X \sim g_\alpha(x)$, where $\alpha \geq 0$, $\mu \in \mathbb{R}$ and $\sigma > 0$ given in (2.3) to obtain a more general family of densities. Its pdf is given by

$$g_{\alpha,\mu,\sigma}(x) = \begin{cases} \phi(\omega_\alpha\left(\frac{x-\mu}{\sigma}\right)), & x \neq \mu, \\ \phi_{\mu,\sigma}(\mu), & x = \mu. \end{cases} \tag{2.15}$$

For the sake of simplicity, we will consider the value $\mu = 0$ when estimating the parameters of the distribution, in that case the distribution coincides with (2.3). A value $x = 0$ is better identifiable in an empirical data source than another value that is unlikely to be an integer. For the case that $\mu = 0$, the parameter can be estimated by using a similar procedure as the one used in the composite models, see [13].

2.4. Extensions

A variant of the approach used to derived (2.13) can be simply implemented as follows

$$g_{\alpha_1,\alpha_2,\beta_1,\beta_2,\lambda}(x) = 2\Phi(\lambda\,\omega_{\beta_1,\beta_2}(x))\phi(\omega_{\alpha_1,\alpha_2}(x)) \tag{2.16}$$

for $x \neq 0$, $x \neq \sqrt{\beta_i}$, $x \neq \sqrt{\alpha_i}$, while $g_{\alpha_1,\alpha_2,\beta_1,\beta_2,\lambda}(0) = \phi(0)$, $g_{\alpha_1,\alpha_2,\beta_1,\beta_2,\lambda}(\sqrt{\alpha_i}) = \phi(\sqrt{\alpha_i})$, $g_{\alpha_1,\alpha_2,\beta_1,\beta_2,\lambda}(\sqrt{\beta_i}) = \phi(\sqrt{\beta_i})$, where $\beta_i \in \mathbb{R}$, $\alpha_i \geq 0$ $(i = 1, 2)$ and

$$\begin{aligned} \omega_{\alpha_1,\alpha_2}(x) &= x - \alpha_2 - \frac{\alpha_1}{x - \frac{\alpha_2}{x}}, \\ \omega_{\beta_1,\beta_2}(x) &= x - \beta_2 - \frac{\beta_1}{x - \frac{\beta_2}{x}}. \end{aligned}$$

This modified family of distributions would allow us to obtain densities with more than two modal values. The extension of this distribution to generate multimodality is immediate. For the particular case (2.16), two graphs of the pdf have been plotted in Figure 3.

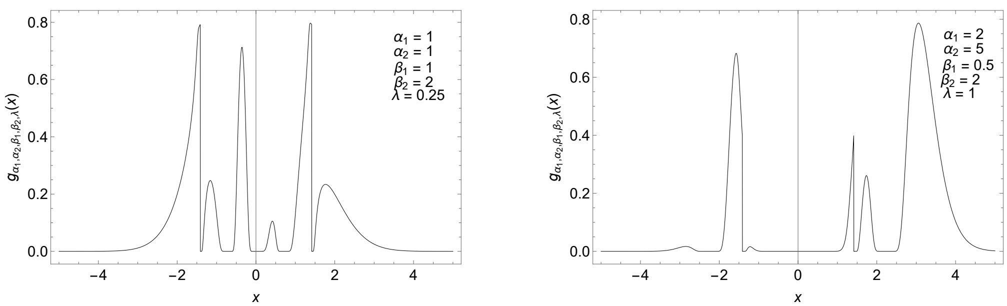

Figure 3: Plot of the probability density function (2.16) for selected values of the parameters α_i, β_i $(i = 1, 2)$ and λ

This new multimodal family of probability distributions can be utilized to explain the size of the claims in cyber risk. In this regard, some multimodal and asymmetric distribution can be effortlessly applied to capture the multimodality and extremely skewed feature of the severity of the cyber breaches.

2.5. Summary of the proposed methodology

Before continuing with the usual elements of distribution theory, such as statistical inference and applications, it is essential to summarize the methodology we have carried out in this work in a diagram. Figure 4 shows a flowchart outlining the methods developed in this article. This diagram can help the reader observe the work's general perspective and allow, if desired, to ignore those elements that could be of lesser interest.

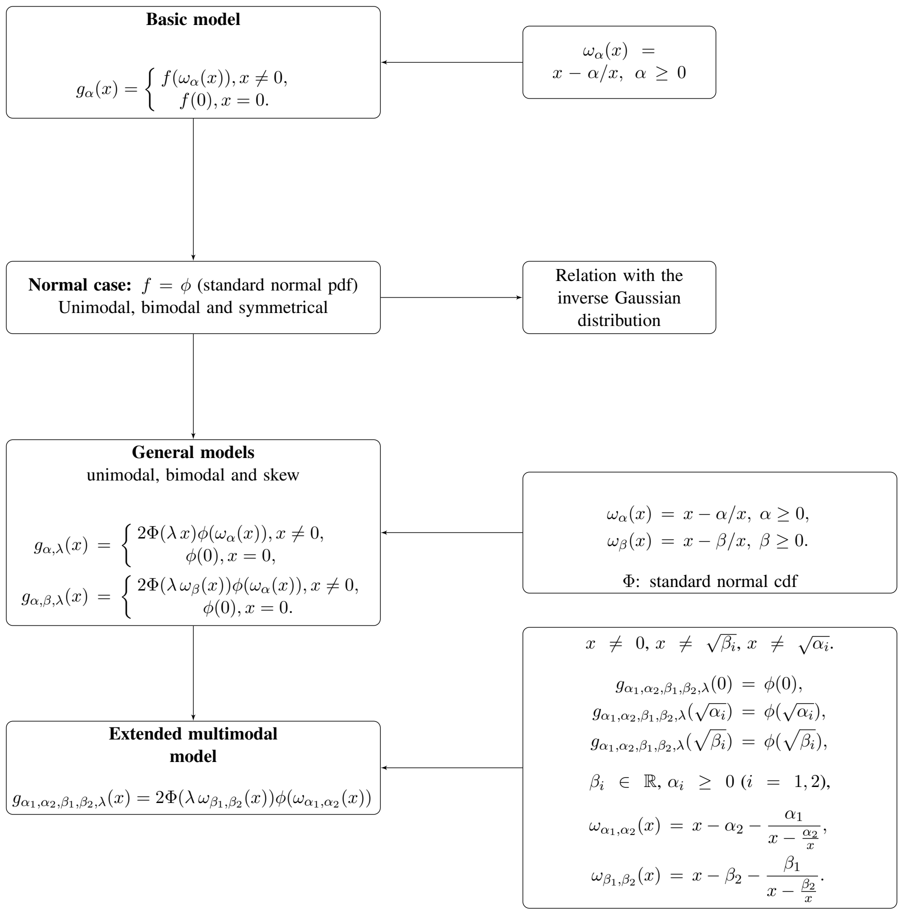

Figure 4: Flowchart showing the methodology proposed in this paper

3. Statistical inference

Let us consider a random sample of n observations $\boldsymbol{x} = (x_1, \ldots, x_n)$, in which there are n_0 observations that are zeros and n_1 non-zero observations; $n_0 + n_1 = n$. Now by using the pdf (2.3), the log-likelihood function is proportional to $\ell(\alpha; \boldsymbol{x}) \propto -1/2 \sum_{i \in \{1,\ldots,n_1\}} (\omega_\alpha(x_i))^2$. By equating the first derivative with respect to α to zero, we get the maximum likelihood estimator of the parameter α is given by $\widehat{\alpha} = n_1 \left\{ \sum_{i \in \{1\ldots,n_1\}} x_i^{-2} \right\}^{-1}$, $x_i \neq 0$. Now, by computing the second derivative of the log-likelihood function and its expectation, the corresponding standard error, that can be obtained from the Fisher's information entry, is $(n/\widehat{\alpha})^{-1/2}$. To obtain this result, it is necessary the expectation of $1/X^2$ with respect to the random variable with pdf (2.3) which is given by $1/\alpha$.

Let us now examine the pdf (2.15) with $\mu = 0$. In this case, the log-likelihood function is proportional to

$$\ell(\alpha, \sigma; \boldsymbol{x}) \propto -n \log \sigma - \frac{1}{2} \sum_{i \in \{1,\ldots,n_1\}} (\omega_\alpha(x_i/\sigma))^2, \tag{3.1}$$

where n_1 is the number of non-zero observations in the sample. From (3.1) we derive the normal equations given by

$$\frac{n_1}{\sigma} - \alpha\sigma \sum_{i\in\{1,\ldots,n_1\}} \left(\frac{1}{x_i}\right)^2 = 0, \tag{3.2}$$

$$\frac{n}{\sigma} - \sigma \sum_{i\in\{1,\ldots,n_1\}} \left[\left(\frac{x_i}{\sigma^2}\right)^2 - \left(\frac{\alpha}{x_i}\right)^2\right] = 0. \tag{3.3}$$

After simple algebra, equations (3.2)-(3.3) provides the maximum likelihood estimators of the parameters which are given by

$$\begin{aligned}
\widehat{\alpha} &= \frac{nn_1}{\left(\sum_{i\in\{1,\ldots,n_1\}} x_i^{-2}\right)\left(\sum_{i\in\{1,\ldots,n_1\}} x_i^2\right) - n_1^2},\\
\widehat{\sigma} &= \left\{\frac{1}{n}\left[\sum_{i\in\{1,\ldots,n_1\}} x_i^2 - n_1^2\left(\sum_{i\in\{1,\ldots,n_1\}} x_i^{-2}\right)^{-1}\right]\right\}^{1/2}.
\end{aligned}$$

The second partial derivatives are provided by

$$\begin{aligned}
\frac{\partial\ell(\alpha,\sigma;\boldsymbol{x})}{\partial\alpha^2} &= -\sigma^2 \sum_{i\in\{1,\ldots,n_1\}} \left(\frac{1}{x_i}\right)^2,\\
\frac{\partial\ell(\alpha,\sigma;\boldsymbol{x})}{\partial\alpha\partial\sigma} &= -2\alpha\sigma \sum_{i\in\{1,\ldots,n_1\}} \left(\frac{1}{x_i}\right)^2,\\
\frac{\partial\ell(\alpha,\sigma;\boldsymbol{x})}{\partial\sigma^2} &= \frac{n}{\sigma^2} - \sum_{i\in\{1,\ldots,n_1\}} \left[\frac{3x_i^2}{\sigma^4} + \left(\frac{\alpha}{x_i}\right)^2\right].
\end{aligned}$$

Now, taking into account that $\mathbb{E}(X^2) = \sigma^2(1+\alpha)$ and $\mathbb{E}(1/X_i^2) = 1/(\alpha\,\sigma^2)$, it is a simple exercise to note that the Fisher's information matrix is

$$\mathcal{I}(\widehat{\alpha},\widehat{\sigma}) = \begin{bmatrix} n_1/\widehat{\alpha} & 2n_1/\widehat{\sigma} \\ 2n_1/\widehat{\sigma} & (2n(2\widehat{\alpha}+1)-n_1)/\widehat{\sigma}^2 \end{bmatrix}.$$

Finally, when the pdf (2.13) is considered, the log-likelihood function is proportional to

$$\ell(\boldsymbol{\theta};\boldsymbol{x}) \propto -n\log\sigma + \sum_{i\in\{1,\ldots,n_1\}} \log\Phi\left(\lambda\,\omega_\beta(x_i/\sigma)\right) - \frac{1}{2}\sum_{i\in\{1,\ldots,n_1\}} \left(\omega_\alpha(x_i/\sigma)\right)^2, \tag{3.4}$$

where $\boldsymbol{\theta} = (\alpha,\beta,\lambda,\sigma)$ is the vector of parameters to be estimated.

In practice, although both normal equations and Fisher's information matrix can be obtained after tedious algebra, the estimates and the entries of this matrix can be achieved by directly maximizing the log-likelihood function given in (3.4). Moreover, this procedure can be extended, as it is seen in the numerical illustrations, for the case where a location parameter μ is included. Recall that the Fisher's information matrix of the skew-normal distribution proposed by [6] is singular for the skew parameter and, consequently, the maximum likelihood estimate of this parameter can be infinite with a positive probability. With respect to the singularity of the Fisher information matrix of the generalized skew normal (GSN) distribution with pdf (2.13), we could use the Theorem 3 in [28] to derive a reparametrization of (2.13) and provide a solution to the singularity problem for (α, β, λ) as in [36]. In order to show the asymptotic behaviour of the maximum likelihood estimator, we carry out the following simulation experiment where the algorithm illustrated in the previous section is used, a complete simulation analysis for the GSN distribution with density function (2.13) is carried out by generating $N := 1000$ samples of sizes $n := 50,\ 100,\ 200$ for different values of the parameters α, β and λ. The value of these parameters have been chosen for the sake of simplicity in estimation. For each parameter, the analysis computes the following measures:

- Average bias (AB) of the simulated estimates:
  $\mathrm{AB}(\Lambda^*) = \dfrac{1}{N} \displaystyle\sum_{j \in \{1,\ldots,N\}} (\Lambda_j^* - \Lambda);$
- Mean square error (MSE) of the simulated estimates:
  $\mathrm{MSE}(\Lambda^*) = \dfrac{1}{N} \displaystyle\sum_{j \in \{1,\ldots,N\}} (\Lambda_j^* - \Lambda)^2;$

where Λ_j^* represents the maximum likelihood estimate of each parameter in the jth sample and Λ is the true value of the parameter. Table 1 shows the average bias and mean square errors of the parameter estimates for different values of α, β and λ for different values of n. In the first row of this table, the case of the skew parameter $\lambda = 0$ is considered, i.e. symmetric case. As expected, the mean square error decreases when n increases. Also, the average bias is positive and decreases with n. It is also noted that the MSE increases with the value of the parameter α. However, the mean square errors for the parameters β and λ seem to be influenced by the value considered for the parameter α. In general, the MSE's decrease with the sample size satisfying that $\lim\limits_{n \to \infty} \mathrm{MSE}(\Lambda^*) = 0$, and therefore, the estimates are consistent in mean square error. It implies that the estimate gets closer and closer to the parameter's true value as data accumulates. Also, for large values of n, the maximum likelihood estimators are normally distributed with the mean equals to the true value of the parameter and variance equal to the reciprocal of the information function evaluated at the mean.

Table 1: Average bias (AB) and mean square error (MSE) of the maximum likelihood estimates for different values of the parameters of the GSN distribution for different samples sizes n with simulation size $N := 1000$

n		$\alpha = 0.25$	$\beta = 0.5$	$\lambda = 0$	$\alpha = 1$	$\beta = 1$	$\lambda = 0$
50	AB	0.0015	–	–	0.0160	–	–
	MSE	0.0003	–	–	0.0224	–	–
100	AB	0.0013	–	–	0.0138	–	–
	MSE	0.0002	–	–	0.0108	–	–
200	AB	0.0000	–	–	0.0021	–	–
	MSE	0.0001	–	–	0.0049	–	–
n		$\alpha = 0.25$	$\beta = 0.5$	$\lambda = 0.5$	$\alpha = 1$	$\beta = 1$	$\lambda = 1$
50	AB	0.0008	0.0922	0.0552	0.0230	0.0057	0.0606
	MSE	0.0003	0.1302	0.1640	0.0211	0.0445	0.0756
100	AB	0.0001	0.1028	0.0472	0.0169	0.0072	0.0386
	MSE	0.0002	0.0795	0.1068	0.0105	0.0209	0.0419
200	AB	0.0001	0.0848	0.0336	0.0032	0.0075	0.0212
	MSE	0.0001	0.0606	0.0767	0.0054	0.0105	0.0209
n		$\alpha = 0.5$	$\beta = 0.25$	$\lambda = 0.25$	$\alpha = 0.75$	$\beta = 1.5$	$\lambda = 1.2$
50	AB	0.0040	-0.0224	-0.0024	0.0135	0.0322	0.0544
	MSE	0.0025	0.0470	0.0725	0.0092	0.0524	0.0890
100	AB	0.0030	-0.0181	0.0122	0.0054	0.0158	0.0320
	MSE	0.0012	0.0437	0.0407	0.0046	0.0284	0.0581
200	AB	0.0022	-0.0176	0.0234	-0.0003	0.0078	0.0276
	MSE	0.0007	0.0363	0.0219	0.0021	0.0143	0.0391

4. Numerical illustrations

In this section, some numerical applications of the GSN distribution given in (2.13) are carried out. The results are compared with those ones of the skew-normal distribution with parameters $\mu \in \mathbb{R}$, $\sigma > 0$ and $\lambda \in \mathbb{R}$, i.e. SN(μ, σ, λ).

The example considered uses the well-known old faithful geyser (Yellowstone Park, Wyoming, USA) data set. This data set consists of 299 measurements of the numerical eruption time in minutes and the waiting time to the next eruption (also in minutes). This popular dataset has been examined extensively in the literature. See, for example, [32], [9] and [15], among others. It is already known that these two datasets show bimodality. There are different versions of these datasets in the statistical literature. The one examined here is taken from the **R** package MASS available in the website

`https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/faithful.html`

Descriptive statistics of these two datasets are shown in Table 2.

Table 2: Descriptive statistics of the two variables considered in the Old Faithful dataset

	Time eruption	Time waiting
Mean	3.461	72.314
Variance	1.313	192.296
$\min$	0.833	43.000
$\max$	5.450	108.000

The estimated values of the parameters for the two models are shown in Table 3 together with the standard errors (in brackets). This Table also includes the value of the maximum log-likelihood function ($\ell_{\max}$), the Akaike's information criterion (AIC) see [1] and the consistent Akaike's information criteria (CAIC), proposed by [11]. The last measure of model selection was chosen to overcome the tendency of the AIC to overestimate the complexity of the underlying model since it lacks certain properties of asymptotic consistency as it does not directly depend on the sample size. Then, to calculate the CAIC, a correction factor based on the sample size is used to compensate for the overestimating nature of AIC. The CAIC is defined as twice $\ell_{\max}$ plus $k\,(1 + \log(n))$, where k is the number of free parameters and n refers to the sample size. Note that a model with a lower AIC and CAIC values is preferred to one with a higher value. It is observable that the GSN distribution has a better performance than the skew normal (SN).

Table 3: Parameters estimates, standard errors (in brackets), maximum of the log-likelihood function ($\ell_{\max}$), AIC and CAIC values for the two variables considered in the old faithful geyser dataset

	Time eruption		Time waiting	
	SN	GSN	SN	GSN
$\widehat{\lambda}$	10.310	0.676	–7.975	0.247
	(3.851)	(0.116)	(1.512)	(0.078)
$\widehat{\alpha}$	–	0.468	–	0.551
	–	(0.058)	–	(0.062)
$\widehat{\beta}$	–	0.227	–	-0.216
	–	(0.096)	–	(0.334)
$\widehat{\mu}$	48.454	65.185	4.897	3.135
	(0.944)	(0.258)	(0.049)	(0.009)
$\widehat{\sigma}$	27.597	13.088	1.837	0.956
	(1.393)	(0.557)	(0.084)	(0.038)
$\ell_{\max}$	-1231.57	-1116.427	-425.737	-399.229
AIC	2469.13	2242.85	857.474	808.458
CAIC	2483.24	2266.36	871.575	831.960

Graphs of the empirical smooth kernel density and theoretical distribution model (GSN) are shown in Figure 5. This former density function was derived by using the in-built function `SmoothKernelDistribution` in `Mathematica`® `v.12.0`. We used an smoothing Gaussian kernel and automatically computed bandwidth parameter. As it can be seen, the GSN is able to capture the bimodal nature of the empirical data although there is an underestimation produced by the adjustment of the proposed distributions. Maximization techniques were completed using `Mathematica`® `v.12.0` and corroborated with `WinRATS v.7.0` (the codes are available upon request) and the computer used was a Intel(R) Core(TM) i7-4790 CPU @ 3.60GHz with 16,0 GB RAM and a processor based on x64 getting acceptable time of processing. Details about these two software can be found in [30] and [12], among others. The routines employed were standard, including among others the `FindMaximum` to compute the maximum likelihood estimates and the `Experimental`CreateNumericalFunction` to obtain the Hessian matrix.

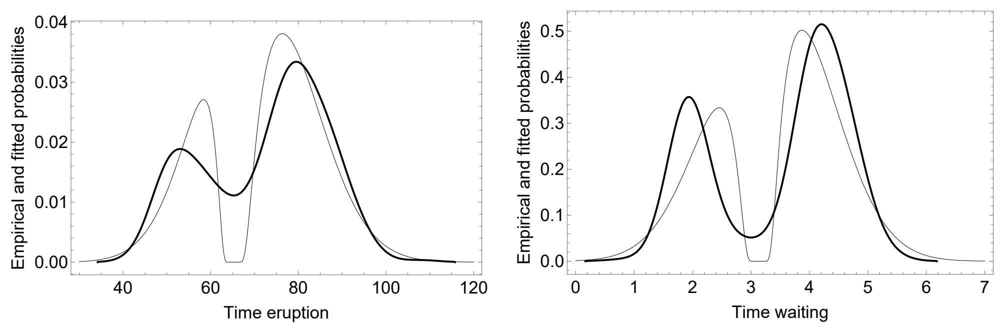

Figure 5: Smooth kernel density estimate of the empirical data (thick line) and the GSN (thin line) for the old faithful data set

5. Conclusions, limitations and future research

In this work, we have studied two families of distributions with support on the real line, the first symmetric and the second not necessarily symmetric. Both families can present more than one mode and include the normal distribution as a special case. In addition, the second one includes, as a particular case, the skew normal distribution. The model has been applied to environmental data, and it can also be used in other scenarios where bimodality is present.

One of the limitations of the distribution proposed in this work is based on the fact that the value that the first distribution takes at zero (at μ for the second model) is fixed, what make these models inflexible. This is an issue that that undoubtedly deserves to be deeply studied to guarantee a more versatile and flexible proposal than the ones presented in this work.

It should also be noted that the extension shown in the Subsection 2.4 requires a separate analysis outside this work's scope. This indeed constitutes a promising probabilistic family that allows to model multimodal data.

ACKNOWLEDGMENTS

EGD and ECO work were partially funded by grant PID2021-127989OB-I00 (Ministerio de Economía y Competitividad, Spain. Also ECO thanks Ministerio de Universidades and Next Generation EU for funding the Program "Ayudas María Zambrano". The authors thank the Associate Editor and three anonymous referees for their constructive comments and suggestions, which have greatly helped them to improve the paper.

REFERENCES

[1] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

[2] 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.

[3] Arellano-Valle, R., Gómez, H., and Quintana, F. (2004). A new class of skew-normal distributions. *Communications in Statistics: Theory and Methods*, 33(7):1465–1480.

[4] Arellano-Valle, R., Gómez, H., and Quintana, F. (2005). Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference*, 128(2):427–443.

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

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

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

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

[9] 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.

[10] Behboodian, J. (1978). Uncorrelated dependent random variables. *Mathematics Magazine*, 51(5):303–304.

[11] Bozdogan, H. (1987). The general theory and its analytical extension. *Psychometrika*, 52(3):345–370.

[12] Brooks, C. (2009). *RATS Handbook to Accompany Introductory Econometrics for Finance*. Cambridge: Cambridge University Press.

[13] Calderín-Ojeda, E. (2015). On the Composite Weibull-Burr Model to describe claim data. *Communications in Statistics: Case Studies, Data Analysis and Applications*, 1,(1):59–69.

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

[15] Dekking, F., Kraaikamp, C., Lopuhaa, H., and Meester, L. (2005). *A Modern Introduction to Probability and Statistics Understanding Why and How.* Springer-Verlag. London.

[16] 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.

[17] Gómez, H., Venegas, O., and Bolfarine, H. (2007). Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics*, 18:395–407.

[18] Gómez-Déniz, E., Arnold, B.C., Sarabia, J.M. and Gómez, W.H. (2021). Properties and applications of a new family of skew distributions. *Mathematics*, 9, 87, 1–18.

[19] Gómez-Déniz, E., Sarabia, J.M. and Calderín-Ojeda, E. (2021). Bimodal normal distribution: extensions and applications. *Journal of Computational and Applied Mathematics* (in press).

[20] Gupta, A., Nguyen, T., and Sanqui, J. (2004). Characterization of the skew-normal distribution. *Ann. Inst. Statist. Math.*, 56(2):351–360.

[21] Henze, N. (1986). A probabilistic representation of the skew-normal distribution. *Scandinavian Journal of Statistics*, 13(4):271–275.

[22] 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.

[23] Johnson, N., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions.* Wiley, New York.

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

[25] Jørgensen, B. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution. Lecture Notes in Statistics.* Springer–Verlag, New York.

[26] Malinovskii, V.K. (2017). Generalized inverse Gaussian distributions and the time of first level crossing. *arXiv:1708.08671v1.*

[27] Patel, J. and Campbell, B. (1984). *Handbook of the Normal Distribution.* Marcel Dekker, Inc. New York and Basel.

[28] Rotnitzky, A., Cox, D.R., Bottai, M., et al. (2000). Likelihood-based inference with singular information matrix. *Bernoulli*, 6, 243–284.

[29] Ribeiro-Reis, L.D., Cordeiro, G.M. and de Santana e Silvas, J.J. (2022). The Mc-Donald Chen distribution: A new bimodal distribution with properties and applications. *Chilean Journal of Statistics*, 13(1), 91–111.

[30] Ruskeepaa, H. (2009). *Mathematica Navigator: Mathematics, Statistics, and Graphics*. New York: Academic Press.

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

[32] Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis.* London, UK: Chapman and Hall.

[33] Slobin, H. (1927). A theorem on improper definite integrals. *The American Mathematical Monthly*, 34(5):265–266.

[34] Sulewski, P. (2022). New members of the Johnson family of probability distributions: properties and application. *REVSTAT* (accepted)

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

[36] Venegas, O., Salinas, H.S., Gallardo, D.I., Bolfarine, H. and Gómez, H.W. (2018). Bimodality based on the generalized skew-normal distribution. *Journal of Statistical Computation and Simulation*, 88(1), 156–181.