

Article

# On Properties of the Bimodal Skew-Normal Distribution and an Application

David Elal-Olivero <sup>1</sup>, Juan F. Olivares-Pacheco <sup>1</sup>, Osvaldo Venegas <sup>2,\*</sup>, Heleno Bolfarine <sup>3</sup> and Héctor W. Gómez <sup>4</sup>

<sup>1</sup> Departamento de Matemáticas, Facultad de Ingeniería, Universidad de Atacama, Copiapó 1530000, Chile; david.elal@uda.cl (D.E.-O.); jfolivar@uda.cl (J.F.O.-P.)

<sup>2</sup> Departamento de Ciencias Matemáticas y Físicas, Facultad de Ingeniería, Universidad Católica de Temuco, Temuco 4780000, Chile

<sup>3</sup> Instituto de Matemática e Estatística (IME), Universidade de São Paulo, São Paulo 05508-090, Brazil; hbolfar@ime.ups.br

<sup>4</sup> Departamento de Matemática, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta 1240000, Chile; hector.gomez@uantof.cl

\* Correspondence: ovenegas@uct.cl; Tel.: +56-45-220-5410

Received: 10 March 2020; Accepted: 29 April 2020; Published: 2 May 2020



**Abstract:** The main object of this paper is to develop an alternative construction for the bimodal skew-normal distribution. The construction is based upon a study of the mixture of skew-normal distributions. We study some basic properties of this family, its stochastic representations and expressions for its moments. Parameters are estimated using the maximum likelihood estimation method. A simulation study is carried out to observe the performance of the maximum likelihood estimators. Finally, we compare the efficiency of the new distribution with other distributions in the literature using a real data set. The study shows that the proposed approach presents satisfactory results.

**Keywords:** bimodal; simulation; skew-normal distribution; stochastic representation

## 1. Introduction

The skew-normal distribution was introduced by Azzalini [1], typically denoted as  $\{SN(\lambda), \lambda \in \mathbb{R}\}$ , with asymmetry parameter  $\lambda$ , so that  $SN(0)$  becomes the standard normal distribution. Hence,  $Y \sim SN(\lambda)$  has a density function given by

$$f(y|\lambda) = 2\phi(y)\Phi(\lambda x), \quad y \in \mathbb{R} \quad (1)$$

with  $\lambda \in \mathbb{R}$  and where  $\phi$  and  $\Phi$  denote, respectively, the density and distribution functions of the  $N(0, 1)$  distribution. Some well known facts about this distribution are

$$E(Y) = \sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{1+\lambda^2}}; \quad Var(Y) = 1 - \frac{2\lambda^2}{\pi(1+\lambda^2)} \quad (2)$$

$$\sqrt{\beta_1} = \frac{1}{2}(4-\pi) \left( \frac{E^2(X)}{Var(Y)} \right)^{3/2}; \quad \beta_2 = 2(\pi-3) \left( \frac{E^2(Y)}{Var(Y)} \right)^2 - 3 \quad (3)$$

where  $\sqrt{\beta_1}$  and  $\beta_2$  are asymmetry and kurtosis coefficients, respectively. From (3) it is well known that

$$-0.9953 \leq \sqrt{\beta_1} \leq 0.9953 \quad (4)$$

$$0 \leq \beta_2 \leq 0.8692. \quad (5)$$

Henze [2] developed the stochastic representation for the skew-normal density and computed its odd moments. The stochastic representation and some more general representations for skew models are also discussed in Azzalini [3]. Arnold et al. [4] make use of the above results to develop truncations of the normal model. Pewsey [5] studied inference problems faced by the skew-normal model with some general results, especially consequences of the singularity of the Fisher information matrix (FIM) in the vicinity of symmetry. Gupta and Chen [6] developed a goodness of fit test. Reliability studies for the skew-normal model were developed by Gupta and Brown [7]. Univariate extensions to the skew-normal model are studied in Arellano-Valle et al. [8], Azzalini [9], Gómez et al. ([10,11]), Venegas et al. [12], etc.

When it is necessary to model data with more than one mode, mixtures of distributions are always used. The importance of such studies rests on the fact that these models present computational difficulties due to identifiability problems. One major motivation of this paper is to develop models that can be seen as alternative parametric models to replace the use of mixtures of distributions, as these present estimation problems from either classical or Bayesian points of view ([13,14]). Bimodal distributions generated from skew distributions can be found in Azzalini and Capitanio [15], Ma and Genton [16], Arellano-Valle et al. [17], Kim [18], Lin et al. ([19,20]), Elal-Olivero et al. [21], Arnold et al. [22], Arellano-Valle et al. [23], Elal-Olivero [24], Gómez et al. [25], Arnold et al. [26], Braga et al. [27], Venegas et al. [28], Shah et al. [29], Gómez-Déniz et al. [30], Esmaeili et al. [31], Imani and Ghoreishi [32], Maleki et al. [33], etc. The main object of this paper is to study the properties of the bimodal skew-normal model introduced by Elal-Olivero et al. [21]. In particular, we derive results related to stochastic representation of the distribution and density function; this makes it simple to derive distributional moments and inferences by maximum likelihood (ML) estimation, among other quantities. The paper is organized as follows. Section 2 develops a bimodal normal distribution, its basic properties, representation, moments and moment generating function. Section 3 develops a skew-normal bimodal distribution, its basic properties, stochastic representation, moments and moment generating function. In Section 4, we perform a small scale simulation study of the ML estimators for parameters. A real data application is discussed in Section 5, which illustrates the usefulness of the proposed model. Conclusions and future work are presented in Section 6.

## 2. Preliminaries

### 2.1. Bimodal Normal Distribution

**Definition 1.** If the random variable  $X$  has density function

$$f(x) = x^2\phi(x), \quad x \in \mathbb{R} \quad (6)$$

where  $\phi$  is the  $N(0,1)$  density, we say that  $X$  follows a bimodal normal distribution (see Elal-Olivero [24]) which is denoted by  $X \sim BN$ .

**Remark 1.** If  $X \sim BN$  with  $f(x)$  its density, then  $f(x)$  is bimodal. This can be verified by noticing that  $f'(x) = x^2(-x\phi(x)) + 2x\phi(x) = x\phi(x)(2 - x^2)$ , and it follows that it reaches its minimum value at  $x = 0$  and maximum value at  $x = -\sqrt{2}$  and  $x = \sqrt{2}$ . Notice that the maximum value is the same; this fact will play an important role in defining a more flexible model by adding an extra parameter which will control the height at the modes.

For the sake of completeness, some important results derived in Elal-Olivero [24] are presented below.

1. If  $X \sim BN$  and  $F_X(t)$  and  $M_X(t)$  are the corresponding distribution and the moment generating functions then

- (a)  $F_X(t) = \Phi(t) - t\phi(t)$
- (b)  $M_X(t) = (1 + t^2)\exp(t^2/2)$
- (c)  $X^2 \sim \chi^2_{(3)}$

2. Let  $X$  and  $U$  be independent random variables with  $X \sim \chi^2_{(3)}$  and  $U$  such that  $P[U = 1] = P[U = -1] = \frac{1}{2}$ . If  $W = \sqrt{X}U$  then  $W \sim BN$ .

2.2. Bimodal Normal Distribution with Shape Parameter

**Definition 2.** If random variable  $X$  has density given by

$$f(x|\alpha) = \left(\frac{1 + \alpha x^2}{1 + \alpha}\right) \phi(x), \quad x \in \mathbb{R}, \alpha \geq 0 \tag{7}$$

where  $\phi$  is the density of the  $N(0, 1)$  distribution, we say that  $X$  is distributed according to the bimodal normal distribution with parameter  $\alpha$  which we denote by  $X \sim BN(\alpha)$ .

**Remark 2.** Some basic properties are shown in the following. Under the assumption that  $X \sim BN(\alpha)$  and that  $f(x|\alpha)$  is the density function of  $X$ , then

- (a)  $f(x|\alpha = 0) = \phi(x)$
- (b)  $f(x|\alpha)$  is symmetric around zero  $\forall \alpha \geq 0$
- (c) As  $\alpha \rightarrow +\infty$ , then  $f(x|\alpha) \rightarrow BN$ .
- (d)  $f(x|\alpha)$  is bimodal  $\forall \alpha > \frac{1}{2}$  where  $x_0 = \sqrt{\frac{2\alpha-1}{\alpha}}$  and  $x_1 = -\sqrt{\frac{2\alpha-1}{\alpha}}$  are the points where the density reaches its maximum.
- (e)  $f(x|\alpha)$  is unimodal  $\forall \alpha \leq \frac{1}{2}$ .

**Proposition 1.** If  $W \sim BN(\alpha)$  then the cumulative distribution and the moment generating function (MGF), denoted by  $F_W(t)$  and  $M_W(t)$  respectively, are given by:

- 1.  $F_W(t) = \Phi(t) - \frac{\alpha t}{1+\alpha} \phi(t)$ ;
- 2.  $M_W(t) = \left(1 + \frac{\alpha t^2}{1+\alpha}\right) \exp(t^2/2)$ .

**Proof.** Follows directly from the corresponding definitions.  $\square$

The next result presents a stochastic representation for the  $BN(\alpha)$  distribution.

**Proposition 2.** Let  $X$  and  $Y$  be independent random variables such that  $X \sim BN$  and  $Y \sim N(0, 1)$ . If  $W = \sqrt{\frac{\alpha}{1+\alpha}} X + \sqrt{\frac{1}{1+\alpha}} Y$  then  $W \sim BN(\alpha)$ .

**Proof.** Using moments definition,

$$\begin{aligned} M_W(t) &= E[\exp(tW)] \\ &= E\left[\exp\left\{t\left(\sqrt{\frac{\alpha}{1+\alpha}}X + \sqrt{\frac{1}{1+\alpha}}Y\right)\right\}\right] = M_X\left(\sqrt{\frac{\alpha}{1+\alpha}}t\right) M_Y\left(\sqrt{\frac{1}{1+\alpha}}t\right) \\ &= \left(1 + \frac{\alpha t^2}{1+\alpha}\right) \exp\left\{\frac{\alpha t^2}{2(1+\alpha)}\right\} \exp\left\{\frac{t^2}{2(1+\alpha)}\right\} = \left(1 + \frac{\alpha t^2}{1+\alpha}\right) \exp(t^2/2), \end{aligned}$$

which agrees with the MGF of  $BS(\alpha)$  derived above.  $\square$

**Proposition 3.** Let  $W \sim BN(\alpha)$ . Then for  $k = 1, 2, \dots$  the moments are given by:  $\mu_{2k-1} = E[W^{2k-1}] = 0$  and  $\mu_{2k} = E[W^{2k}] = \left(\frac{1+\alpha(2k+1)}{1+\alpha}\right) \prod_{j=1}^k (2j-1)$ .

**Proof.** Follows directly from moment definition.  $\square$

Hence, the kurtosis coefficient is given by  $\beta_2 = \frac{3(1+5\alpha)(1+\alpha)}{(1+3\alpha)^2} - 3$ . Recall that the kurtosis coefficient is given by

$$\beta_2 = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{(\mu_2 - \mu_1^2)^2} - 3,$$

given that the even moments are zero and  $\mu_2 = \frac{1+3\alpha}{1+\alpha}$  and  $\mu_4 = \frac{3+15\alpha}{1+\alpha}$ , which proves the result.

**Remark 3.** Considering that the  $BN(\alpha)$  model is unimodal for  $\alpha \in [0, 0.5]$ , Table 1 shows the variation for the kurtosis coefficient as parameter  $\alpha$  ranges the interval.

**Table 1.** Kurtosis values for  $BN(\alpha)$  model.

$\alpha$	Kurtosis
$[0, 0.5]$	$[-0.48, 0]$

### 3. Bimodal Asymmetric Distribution

We start by dealing first with the extension of the ordinary normal bimodal distribution.

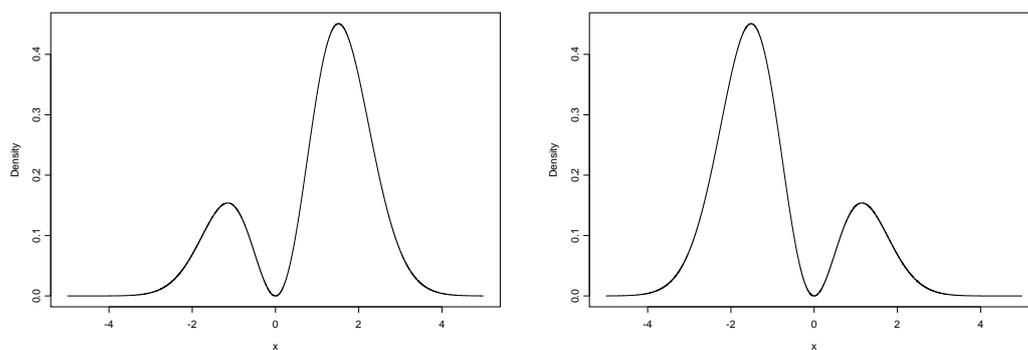
#### 3.1. One-Parameter Bimodal Skew-Normal Distribution

**Definition 3.** If the random variable  $X$  is distributed according to the density function

$$f(x|\lambda) = 2x^2\phi(x)\Phi(\lambda x), \quad x \in \mathbb{R}, \lambda \in \mathbb{R} \tag{8}$$

then we say that it is distributed according to the bimodal skew normal distribution with parameter  $\lambda$  which we denote by  $X \sim BSN(\lambda)$ .

Figure 1 depicts examples of the bimodal skew-normal (BSN) distribution given in Equation (8) for different values of parameter  $\lambda$ .



**Figure 1.** Density  $BSN(\lambda)$  with parameter  $\lambda = 0.5$  (left) and  $\lambda = -0.5$  (right).

**Proposition 4.** Let  $X \sim BSN(\lambda)$ , then  $X^2 \sim \chi^2_{(3)}$ .

**Proof.** Let  $Y = X^2$ , then, for  $y \geq 0$ , it follows that

$$F_Y(y) = P[Y \leq y] = P[X^2 \leq y] = P[-\sqrt{y} \leq X \leq \sqrt{y}] = F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

Differentiating this last expression we conclude the proof by showing that

$$f_Y(y) = \frac{1}{2\sqrt{y}}f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}}f_X(-\sqrt{y}) = \sqrt{y}\phi(\sqrt{y})[\Phi(\lambda\sqrt{y}) + \Phi(-\lambda\sqrt{y})] = \sqrt{y}\phi(\sqrt{y}), \quad y > 0.$$

□

**Remark 4.** The density function  $f(x|\lambda) = 2x^2\phi(x)\Phi(\lambda x)$ ,  $x \in \mathbb{R}$  behaves like the density function of the bimodal normal model  $g(x) = x^2\phi(x)$  by the perturbation function  $h(x) = 2\Phi(x)$ . We note that the heights at the modes for a bimodal symmetric distribution are the same. However, for an asymmetric distribution the heights at the modes are not the same as shown in the next proposition.

**Proposition 5.** Let  $f(x|\lambda) = 2x^2\phi(x)\Phi(\lambda x)$ ,  $x \in \mathbb{R}$ . Moreover, let  $x_0$  and  $x_1$  ( $x_0 < x_1$ ) be the points at which the function  $f(x|\lambda)$  reaches its the maximum value. Then,

- (a) If  $\lambda \geq 0$  then  $f(x_0|\lambda) \leq f(x_1|\lambda)$
- (b) If  $\lambda \leq 0$  then  $f(x_0|\lambda) \geq f(x_1|\lambda)$ .

**Proof.** If  $\lambda = 0$  (symmetric case) then  $f(x_0|\lambda) = f(x_1|\lambda)$  with  $x_0 = -\sqrt{2}$  and  $x_1 = \sqrt{2}$ . Suppose now that  $\lambda > 0$ ; it should then be required that  $x_0 < 0$  and  $x_1 > 0$ .

Hence,  $\forall x \geq 0$ ,  $f(x_1|\lambda) > f(x|\lambda)$  and, in particular,  $f(x_1|\lambda) > f(\sqrt{2}|\lambda)$ . On the other hand, for the symmetric case, it is known that  $x_0^2\phi(x_0) \leq (\sqrt{2})^2\phi(\sqrt{2})$  and given that  $\Phi(\lambda x_0) < \frac{1}{2} < \Phi(\lambda\sqrt{2})$ , it follows that

$$f(x_0|\lambda) < f(\sqrt{2}|\lambda) \leq f(x_1|\lambda).$$

The case  $\lambda \leq 0$  is proved similarly. □

**Proposition 6.** Let  $X \sim BSN(\lambda)$ ,  $Y \sim SN(\lambda)$  and  $F_X, F_Y$  be the distribution functions of the random variables  $X, Y$ , respectively, and  $f_Y$  the density function of  $Y$ . Therefore,

$$F_X(t) = F_Y(t) - \left[ t f_Y(t) + \frac{\lambda b}{1 + \lambda^2} \phi(t\sqrt{1 + \lambda^2}) \right]$$

with  $b = \sqrt{\frac{2}{\pi}}$ .

**Proof.** Let

$$F_X(t) = \int_{-\infty}^t 2x^2\phi(x)\Phi(\lambda x)dx,$$

the distribution function of the random variable  $X \sim BSN(\lambda)$ . Integrating by parts, with  $u = x\Phi(\lambda x)$  and  $dv = x\phi(x)$ , it follows that  $du = (\lambda x\phi(\lambda x) + \Phi(\lambda x))dx$  and  $v = -\phi(x)$ , so that

$$\begin{aligned} F_X(t) &= 2 \left( -t\phi(t)\Phi(\lambda t) + \frac{1}{2}F_Y(t) \right) + 2 \int_{-\infty}^t \lambda x\phi(x)\phi(\lambda x)dx \\ &= -2t\phi(t)\Phi(\lambda t) + F_Y(t) - \left( \frac{\lambda b}{1 + \lambda^2} \right) \phi(t\sqrt{1 + \lambda^2}) \\ &= F_Y(t) - \left[ t f_Y(t) + \frac{\lambda b}{1 + \lambda^2} \phi(t\sqrt{1 + \lambda^2}) \right]. \end{aligned}$$

□

**Proposition 7.** Let  $X \sim \text{BSN}(\lambda)$  and  $Y \sim \text{SN}(\lambda)$  with MGF  $M_X$ , and  $M_Y$ , respectively. Then,

$$M_X(t) = (1 + t^2)M_Y(t) + bt\delta(2 - \delta^2)\exp\left(\frac{t^2(1 - \delta^2)}{2}\right)$$

with  $b = \sqrt{\frac{2}{\pi}}$  and  $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$ .

**Proof.** We have

$$M_X(t) = E[\exp(tX)] = \int_{-\infty}^{\infty} \exp(tx)2x^2\phi(x)\Phi(\lambda x)dx$$

Integrating by parts, with  $u = 2\exp(tx)x\Phi(\lambda x)$  and  $dv = x\phi(x)$ , it follows that  $du = 2\exp(tx)[\lambda x\phi(\lambda x) + \Phi(\lambda x) + tx\Phi(\lambda x)]dx$  and  $v = -\phi(x)$  so that

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} 2\exp(tx)\phi(x)[\lambda x\phi(\lambda x) + \Phi(\lambda x) + tx\Phi(\lambda x)]dx \\ &= (1 + t^2)M_Y(t) + \left(\frac{b\lambda t}{\sqrt{(1 + \lambda^2)^3}} + \frac{b\lambda t}{\sqrt{1 + \lambda^2}}\right)\exp\left(\frac{t^2}{2(1 + \lambda^2)}\right) \\ &= (1 + t^2)M_Y(t) + bt\delta(2 - \delta^2)\exp\left(\frac{t^2(1 - \delta^2)}{2}\right). \end{aligned}$$

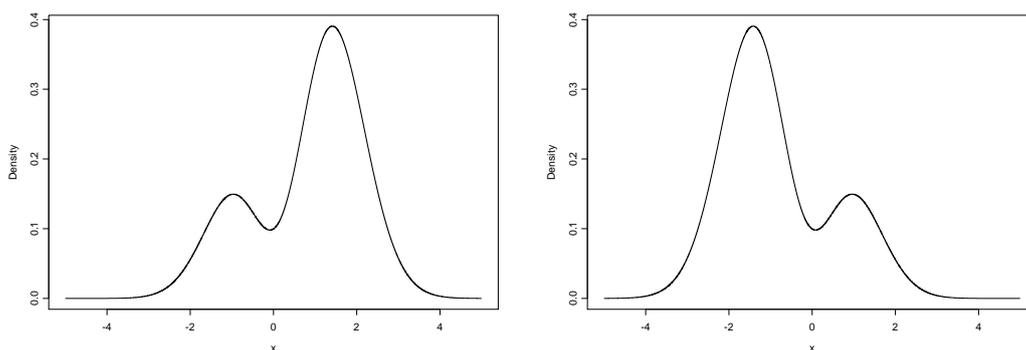
□

### 3.2. Two-Parameter Bimodal Skew-Normal Distribution

**Definition 4.** If the density function of a random variable  $X$  is such that

$$f(x|\lambda, \alpha) = 2\frac{1 + \alpha x^2}{1 + \alpha}\phi(x)\Phi(\lambda x), \quad x \in \mathbb{R}, \lambda \in \mathbb{R}, \alpha \geq 0 \tag{9}$$

then we say that  $X$  is distributed as the bimodal skew-normal distribution (see Elal-Oliviero et al. [21]) with parameters  $\lambda$  and  $\alpha$ , which we denote by  $X \sim \text{BSN}(\lambda, \alpha)$ . Figure 2 depicts examples of the BSN distribution given in Equation (9) for different values of parameters  $\lambda$  and  $\alpha$ .



**Figure 2.** Density function for  $\text{BSN}(\lambda, \alpha)$  with  $\lambda = 0.5$  and  $\alpha = 3$  (left);  $\lambda = -0.5$  and  $\alpha = 3$  (right).

**Remark 5.** It is already known that changes in parameter  $\lambda$  lead to changes in values (heights) of the density of the  $\text{BSN}(\lambda)$  model, that is, in  $f(x|\lambda) = 2x^2\phi(x)\Phi(\lambda x)$ . By incorporating the extra parameter  $\alpha$  in the density  $f(x|\lambda, \alpha) = 2\left(\frac{1 + \alpha x^2}{1 + \alpha}\right)\phi(x)\Phi(\lambda x)$ , corresponding to the  $\text{BSN}(\lambda, \alpha)$  model, as  $\alpha$  ranges in the interval  $[0, \infty]$  the density changes from unimodal to bimodal, and vice versa, with great flexibility.

As we mentioned in the introduction, the BSN model was introduced by Elal-Oliviero et al. [21] and used for the Bayesian inference. Now we observe that we have constructed it based on a mixture

of two distributions; below we study some of its properties, carry out a simulation study to see the behavior of the ML estimators and present an application to a real data set.

**Proposition 8.** Let  $X \sim BSN(\lambda, \alpha)$ , so that the following properties hold:

- (a) If  $\lambda = 0$  then  $X \sim BN(\alpha)$
- (b) If  $\lambda \rightarrow \infty$  then  $X \sim HalfBN(\alpha)$
- (c) If  $\alpha = 0$  then  $X \sim SN(\lambda)$
- (d) If  $\alpha \rightarrow \infty$  then  $X \sim BSN(\lambda)$
- (e) Let  $W \sim BSN(\lambda, \alpha)$ ,  $X \sim \chi^2_{(1)}$ ,  $Y \sim \chi^2_{(3)}$  and  $U = W^2$ . Then,

$$f_U(w) = \left(\frac{1}{1+\alpha}\right) f_X(w) + \left(\frac{\alpha}{1+\alpha}\right) f_Y(w).$$

**Proof.** For  $w \geq 0$ , it follows that

$$F_U(w) = P[U \leq w] = P[W^2 \leq w] = P[-\sqrt{w} \leq W \leq \sqrt{w}] = F_W(\sqrt{w}) - F_W(-\sqrt{w}),$$

which upon differentiation leads to:

$$\begin{aligned} f_U(w) &= \frac{1}{2\sqrt{w}} f_W(\sqrt{w}) + \frac{1}{2\sqrt{w}} f_W(-\sqrt{w}) \\ &= \frac{1}{\sqrt{w}} \left(\frac{1+\alpha w}{1+\alpha}\right) \phi(\sqrt{w}) [\Phi(\lambda\sqrt{w}) + \Phi(-\lambda\sqrt{w})] \\ &= \left(\frac{1}{1+\alpha}\right) f_X(w) + \left(\frac{\alpha}{1+\alpha}\right) f_Y(w). \end{aligned}$$

□

- (f) Let  $Y \sim SN(\lambda)$ ,  $X \sim BSN(\lambda)$  and  $W \sim BSN(\lambda, \alpha)$ . Then

$$f_W(x) = \left(\frac{1}{1+\alpha}\right) f_Y(x) + \left(\frac{\alpha}{1+\alpha}\right) f_X(x).$$

**Proof.**

$$f_W(x) = \left(\frac{1}{1+\alpha}\right) 2\phi(x)\Phi(\lambda x) + \left(\frac{\alpha}{1+\alpha}\right) 2x^2\phi(x)\Phi(\lambda x) = \left(\frac{1}{1+\alpha}\right) f_Y(x) + \left(\frac{\alpha}{1+\alpha}\right) f_X(x).$$

We note that the density for the  $BSN(\lambda, \alpha)$  model can be seen as a mixture between the  $SN(\lambda)$  and  $BSN(\lambda)$  models. □

**Proposition 9.** Let  $W \sim BSN(\lambda, \alpha)$ ,  $Y \sim SN(\lambda)$  and  $F_W$  and  $F_Y$  be the distribution function for the random variables  $W$  and  $Y$  respectively and  $f_Y(\cdot)$  the density function  $Y$ . Then,

$$F_W(t) = F_Y(t) - \frac{\alpha}{1+\alpha} \left[ t f_Y(t) + \frac{\lambda b}{1+\lambda^2} \phi(t\sqrt{1+\lambda^2}) \right]$$

**Proof.** Let  $F_X$  be the distribution function of the random variable  $X$ , where  $X \sim BSN(\lambda)$ . Then, the result follows by noticing that

$$\begin{aligned} F_W(t) &= \left(\frac{1}{1+\alpha}\right) \int_{-\infty}^t 2\phi(x)\Phi(\lambda x)dx + \left(\frac{\alpha}{1+\alpha}\right) \int_{-\infty}^t 2x^2\phi(x)\Phi(\lambda x)dx \\ &= \left(\frac{1}{1+\alpha}\right) F_Y(t) + \left(\frac{\alpha}{1+\alpha}\right) F_X(t) \\ &= \left(\frac{1}{1+\alpha}\right) F_Y(t) + \left(\frac{\alpha}{1+\alpha}\right) \left[ F_Y(t) - \left( tf_Y(t) + \frac{\lambda b}{1+\lambda^2} \phi\left(t\sqrt{1+\lambda^2}\right) \right) \right] \\ &= F_Y(t) - \frac{\alpha}{1+\alpha} \left[ tf_Y(t) + \frac{\lambda b}{1+\lambda^2} \phi\left(t\sqrt{1+\lambda^2}\right) \right]. \end{aligned}$$

□

**Proposition 10.** Let  $W \sim BSN(\lambda, \alpha)$  and  $Y \sim SN(\lambda)$ , and let,  $M_W$  and  $M_Y$  be the moment generating functions for the random variables  $W$  and  $Y$ , respectively; then,

$$M_W(t) = M_Y(t) + \frac{\alpha}{1+\alpha} \left[ t^2 M_Y(t) + \delta b t (2 - \delta^2) \exp\left(\frac{t^2(1 - \delta^2)}{2}\right) \right],$$

with  $M_Y(t) = E[\exp(tY)] = 2\Phi(\delta t) \exp\left(\frac{t^2}{2}\right)$ ,  $b = \sqrt{\frac{2}{\pi}}$  and  $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$ .

**Proof.** The result follows by noticing that

$$\begin{aligned} M_W(t) &= \left(\frac{1}{1+\alpha}\right) \int_{-\infty}^t \exp(tx) 2\phi(x)\Phi(\lambda x)dx + \left(\frac{\alpha}{1+\alpha}\right) \int_{-\infty}^t \exp(tx) 2x^2\phi(x)\Phi(\lambda x)dx \\ &= \left(\frac{1}{1+\alpha}\right) M_Y(t) + \left(\frac{\alpha}{1+\alpha}\right) M_X(t) \\ &= \left(\frac{1}{1+\alpha}\right) M_Y(x) + \left(\frac{\alpha}{1+\alpha}\right) \left[ (1+t^2)M_Y(t) + bt\delta(2 - \delta^2) \exp\left(\frac{t^2(1 - \delta^2)}{2}\right) \right] \\ &= M_Y(t) + \frac{\alpha}{1+\alpha} \left[ t^2 M_Y(t) + \delta b t (2 - \delta^2) \exp\left(\frac{t^2(1 - \delta^2)}{2}\right) \right]. \end{aligned}$$

□

**Remark 6.** The properties for the  $BSN(\lambda, \alpha)$  model presented next follow from general properties of the  $f(z) = 2f_0(z)G(w(z))$  model presented in Azzalini [3] where  $f_0$  is a symmetric (around zero) density function,  $G$  is a unidimensional (symmetric) distribution function such that  $G'$  exists and  $w(\cdot)$  is an even function.

1. A stochastic representation for the  $BSN(\lambda, \alpha)$  model. Let  $X \sim BN(\alpha)$  and  $Y \sim N(0, 1)$  be independent random variables and define

$$W = \begin{cases} X & \text{if } Y < \lambda X \\ -X & \text{if } Y \geq \lambda X. \end{cases}$$

Then  $W \sim BSN(\alpha, \lambda)$ .

2. Invariance perturbation property.

If  $W \sim BSN(\lambda, \alpha)$  and  $Z \sim BN(\alpha)$  then  $|X| \stackrel{d}{=} |Z|$  are identically distributed.

**Remark 7.** The moments of the  $BSN(\lambda, \alpha)$  model can be computed easily, separating even from odd moments.

1. Considering anti-perturbation invariance, the even moments of  $X$  and  $Z$  are the same, so that  $\mu_{2k} = E[X^{2k}] = E[Z^{2k}] = \left(\frac{1+\alpha(2k+1)}{1+\alpha}\right) \prod_{j=1}^k (2j - 1)$  with  $k = 1, 2, 3, \dots$

2. For the odd moments, considering that  $Y \sim SN(\lambda)$ , it follows that:

$$\mu_{2k-1} = E[X^{2k-1}] = \frac{1}{1+\alpha} E[Y^{2k-1}] + \frac{\alpha}{1+\alpha} E[Y^{2k+1}]$$

with  $k = 1, 2, 3, \dots$

where the odd moments  $E[Y^{2k+1}]$  can be computed using derivations in Henze [2], leading to:

$$E[Y^{2k+1}] = \frac{b\delta(2k+1)!}{(2+2\lambda^2)^k} \sum_{\nu=0}^k \frac{\nu!(2\lambda)^{2\nu}}{(2\nu+1)!(k-\nu)!}$$

with  $k = 0, 1, 2, 3, \dots$

If we denote by  $\sqrt{\beta_1}$  and  $\beta_2$  the asymmetry and kurtosis coefficients, respectively, then:

$$\sqrt{\beta_1} = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{(\mu_2 - \mu_1^2)^{\frac{3}{2}}} \quad \text{and} \quad \beta_2 = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{(\mu_2 - \mu_1^2)^2},$$

where  $\mu_k = E(X^k)$ , with  $k = 1, 2, 3, 4$  are as given by

$$\mu_1 = \frac{b\delta}{(1+\alpha)(1+\lambda^2)} [(1+2\alpha)\lambda^2 + 3\alpha + 1], \quad \mu_2 = \frac{1+3\alpha}{1+\alpha},$$

$$\mu_3 = \frac{b\delta}{(1+\alpha)(1+\lambda^2)^2} [(2+8\alpha)\lambda^4 + (5+20\alpha)\lambda^2 + 15\alpha + 3] \quad \text{and} \quad \mu_4 = \frac{3+15\alpha}{1+\alpha}.$$

Representing the asymmetry for specific values of  $\alpha$  and  $\lambda$  by  $\sqrt{\beta_1} = A(\alpha, \lambda)$ , then the following relation holds:  $A(\alpha, \lambda) = -A(\alpha, -\lambda)$ .

Tables 2 and 3 show asymmetry and kurtosis values for different values of  $\alpha$  and  $\lambda$ .

**Table 2.** Possible values of the standardized skewness coefficient.

$\alpha$	$\lambda$									
	0	0.5	1	1.5	2	5	10	50	100	$\infty$
0	0.0000	0.0239	0.1369	0.3003	0.4538	0.8510	0.9556	0.9936	0.9949	0.9953
0.5	0.0000	-0.1648	-0.0617	0.1411	0.3069	0.6282	0.6935	0.7157	0.7164	0.7167
1	0.0000	-0.3161	-0.2517	-0.0229	0.1624	0.4797	0.5355	0.5537	0.5543	0.5545
1.5	0.0000	-0.4119	-0.3789	-0.1275	0.0779	0.4045	0.4559	0.4722	0.4727	0.4728
2	0.0000	-0.4771	-0.4689	-0.1983	0.0258	0.3642	0.4128	0.4277	0.4282	0.4283
3	0.0000	-0.5600	-0.5878	-0.2869	-0.0311	0.3299	0.3748	0.3879	0.3883	0.3884
5	0.0000	-0.6444	-0.7149	-0.3740	-0.0733	0.3221	0.3624	0.3732	0.3735	0.3736
10	0.0000	-0.7223	-0.8390	-0.4497	-0.0900	0.3510	0.3853	0.3929	0.3931	0.3932
50	0.0000	-0.7969	-0.9648	-0.5145	-0.0743	0.4287	0.4539	0.4570	0.4571	0.4571
100	0.0000	-0.8072	-0.9827	-0.5225	-0.0688	0.4446	0.4681	0.4705	0.4705	0.4705
$\infty$	0.0000	-0.8177	-0.0012	-0.5305	-0.0621	0.4625	0.4842	0.4857	0.4857	0.4857

**Table 3.** Possible values of the standardized kurtosis coefficient.

$\alpha$	$\lambda$									
	0	$\pm 0.5$	$\pm 1$	$\pm 1.5$	$\pm 2$	$\pm 5$	$\pm 10$	$\pm 50$	$\pm 100$	$\pm \infty$
0	3.0000	3.0060	3.0617	3.1759	3.3051	3.7053	3.8232	3.8673	3.8687	3.8692
0.5	2.5200	2.7015	2.8741	2.9361	2.9760	3.0847	3.1132	3.1233	3.1237	3.1238
1	2.2500	2.6043	2.9357	2.9645	2.9361	2.9016	2.9014	2.9017	2.9018	2.9018
1.5	2.1074	2.5869	3.0642	3.0787	2.9987	2.8720	2.8579	2.8538	2.8537	2.8537
2	2.0204	2.5920	3.1946	3.2032	3.0804	2.8836	2.8612	2.8548	2.8546	2.8545
3	1.9200	2.6169	3.4159	3.4233	3.2319	2.9301	2.8984	2.8897	2.8894	2.8893
5	1.8281	2.6620	3.7139	3.7304	3.4436	3.0049	2.9658	2.9560	2.9557	2.9556
10	1.7513	2.7204	4.0663	4.1060	3.6930	3.0833	3.0411	3.0322	3.0319	3.0319
50	1.6843	2.7904	4.4864	4.5692	3.9787	3.1393	3.1012	3.0967	3.0966	3.0966
100	1.6755	2.8011	4.5514	4.6424	3.0212	3.1393	3.1064	3.1029	3.1029	3.1028
$\infty$	1.6667	2.8123	4.6199	4.7198	3.0653	3.1453	3.1105	3.1082	3.1082	3.1082

The parameter  $\lambda$  produces asymmetry in the symmetric  $BSN(\alpha)$  model and, in particular when  $\alpha > \frac{1}{2}$ , the asymmetry is reflected in the change of height of the modes of the symmetric bimodal model. The following proposition shows the identifiability of the model.

**Proposition 11.** *The  $BSN(\lambda, \alpha)$  model is identifiable.*

**Proof.** Let  $\lambda = \lambda_0$  fixed. We will prove that if  $\alpha_1 \neq \alpha_2$  then  $f(x, \alpha_1, \lambda_0) \neq f(x, \alpha_2, \lambda_0)$ , for all  $x \in \mathbb{R}$ . Let us assume, for a contradiction, that  $f(x, \alpha_1, \lambda_0) = f(x, \alpha_2, \lambda_0)$ , for all  $x \in \mathbb{R}$  and  $\lambda_0$  fixed. Thus, after some algebraic procedures, we have that  $(\alpha_1 - \alpha_2)(x^2 - 1) = 0$ , for all  $x \in \mathbb{R}$ . In consequence,  $\alpha_1 = \alpha_2$  is a contradiction of our assumption.

To prove that  $f$  is an injective function with respect to parameter  $\lambda$ , we use an analogous procedure, which concludes the proof.  $\square$

Below we show some properties involving conditional distributions and their relations with the model introduced in this paper.

**Proposition 12.** *If  $U|V = \lambda \sim BSN(\lambda, \alpha)$  and  $V \sim BN(\alpha)$ . Then  $U \sim BN(\alpha)$ .*

**Proof.** The proof follows by noticing that

$$\begin{aligned}
 f_U(u) &= \int_{-\infty}^{\infty} 2 \frac{1 + \alpha u^2}{1 + \alpha} \phi(u) \Phi(\lambda u) \left( \frac{1 + \alpha \lambda^2}{1 + \alpha} \right) \phi(\lambda) d\lambda \\
 &= \frac{1 + \alpha u^2}{1 + \alpha} \phi(u) \int_{-\infty}^{\infty} 2 \left( \frac{1 + \alpha \lambda^2}{1 + \alpha} \right) \phi(\lambda) \Phi(\lambda u) d\lambda = \left( \frac{1 + \alpha u^2}{1 + \alpha} \right) \phi(u).
 \end{aligned}$$

$\square$

**Corollary 1.** *If  $U|V = \lambda \sim SN(\lambda)$  and  $V \sim BN(\alpha)$ , then  $U \sim N(0, 1)$ .*

**Proposition 13.** *Let  $U|V = \lambda \sim BSN(\lambda, \alpha)$  and  $V \sim BN(\alpha)$ . Then  $V|U = u \sim BSN(u, \alpha)$ .*

**Proof.**

$$f_{V|U}(v) = \frac{2 \left( \frac{1 + \alpha u^2}{1 + \alpha} \right) \phi(u) \Phi(vu) \left( \frac{1 + \alpha v^2}{1 + \alpha} \right) \phi(v)}{\left( \frac{1 + \alpha u^2}{1 + \alpha} \right) \phi(u)} = 2 \left( \frac{1 + \alpha v^2}{1 + \alpha} \right) \phi(v) \Phi(vu).$$

$\square$

### 3.3. Location-Scale Extension

In practical scenarios, it is common to work with the location-scale transformation  $X = \mu + \sigma Z$ , where  $\mu \in \mathbb{R}$ ,  $\sigma > 0$  and  $Z \sim \text{BSN}(\alpha, \lambda)$ , with  $\alpha \geq 0$  and  $\lambda \in \mathbb{R}$ . Therefore, the density function for the random variable  $X$ , denoted as  $X \sim \text{BSN}(\mu, \sigma, \alpha, \lambda)$  is given by

$$f_X(x; \mu, \sigma, \alpha, \lambda) = 2 \frac{\sigma^2 + \alpha(x - \mu)^2}{\sigma^3(1 + \alpha)} \phi\left(\frac{x - \mu}{\sigma}\right) \Phi\left(\lambda \frac{x - \mu}{\sigma}\right), \quad x \in \mathbb{R}. \tag{10}$$

Let us assume that  $\mathbf{x} = (x_1, x_2, \dots, x_n)^\top$  is a random sample of size  $n$  from the distribution  $\text{BSN}(\mu, \sigma, \alpha, \lambda)$ . From (10), the log-likelihood function is given as

$$\ell(\boldsymbol{\theta}) \propto -3n \log(\sigma) - n \log(1 + \alpha) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 + \sum_{i=1}^n \log(\sigma^2 + \alpha(x_i - \mu)^2) + \sum_{i=1}^n \log \Phi\left(\lambda \frac{x_i - \mu}{\sigma}\right),$$

where  $\boldsymbol{\theta} = (\mu, \sigma, \alpha, \lambda)^\top$ , which is a continuous function of each parameter. Then, differentiating the log-likelihood function, we obtain that the elements of the score vector,  $\mathbf{S}(\boldsymbol{\theta}) = (S_\mu, S_\sigma, S_\alpha, S_\lambda)^\top$ , where  $S_a = \partial \ell / \partial a$ , are given by

$$\begin{aligned} S_\mu &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) - 2\alpha \sum_{i=1}^n \frac{x_i - \mu}{\sigma^2 + \alpha(x_i - \mu)^2} - \frac{1}{\sigma^2} \sum_{i=1}^n \zeta\left(\lambda \frac{x_i - \mu}{\sigma}\right), \\ S_\sigma &= -\frac{3n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu)^2 + 2\sigma \sum_{i=1}^n \frac{1}{\sigma^2 + \alpha(x_i - \mu)^2} - \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \zeta\left(\lambda \frac{x_i - \mu}{\sigma}\right), \\ S_\alpha &= -\frac{n}{1 + \alpha} + \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma + \alpha(x_i - \mu)^2}, \\ S_\lambda &= \frac{1}{\sigma} \sum_{i=1}^n (x_i - \mu) \zeta\left(\lambda \frac{x_i - \mu}{\sigma}\right), \end{aligned}$$

where  $\zeta(a) = \phi(a) / \Phi(a)$ . Therefore, the ML estimator of  $\hat{\boldsymbol{\theta}}$  is the solution of the system  $\mathbf{S}(\boldsymbol{\theta}) = \mathbf{0}$ , which must be solved numerically.

## 4. Simulation Study

In this section we report results of a small scale Monte Carlo simulation study conducted to evaluate the performance of parameter estimation by ML. Results are based on 1000 samples generated from the BSN model for several sample sizes. Use was made of the stochastic representation presented in Remark 6(1) implemented in R software [34]. For each sample generated, the likelihood function was maximized using the `optim` function of the R software. Given that the Fisher information matrix is singular when in the vicinity of symmetry ( $\lambda = 0$ ), the use of algorithms such as Fisher-Scoring is not advisable. We therefore used the Nelder and Mead [35] algorithm, a direct search method that works satisfactorily for non-differentiable functions.

### 4.1. Nelder-Mead Method

For each sample size and parameter configuration, the ML estimators were computed leading to an empirical mean and empirical standard deviation (SD). Results are depicted in Tables 4 and 5. The main conclusion is that estimation was satisfactorily stable for moderate and large sample sizes.

**Table 4.** Empirical means and standard deviations for ML estimators using Nelder-Mead method (with  $\mu = 0$  and  $\sigma = 1$ ).

<i>n</i>	Parameters		Estimations	
	$\alpha$	$\lambda$	$\hat{\alpha}(SD)$	$\hat{\lambda}(SD)$
25	0	−3	0.1366(0.0840)	−2.9256(0.0961)
		0	0.0540(0.0329)	−0.0028(0.0290)
	3	3	0.1318(0.0854)	3.1444(0.1066)
		−3	3.1032(0.1155)	−2.9445(0.1313)
		0	3.1444(0.1055)	0.0141(0.1221)
		3	3.0789(0.0915)	3.1387(0.1327)
50	0	−3	0.1235(0.0755)	−2.9242(0.0999)
		0	0.0472(0.0356)	0.0059(0.0304)
	3	3	0.1182(0.0756)	3.1525(0.1125)
		−3	3.1063(0.1172)	−2.9412(0.1346)
		0	3.1504(0.1116)	−0.0009(0.1039)
		3	3.0822(0.1071)	3.1359(0.1475)
75	0	−3	0.1152(0.0713)	−2.9171(0.1015)
		0	0.0449(0.0336)	0.0099(0.0312)
	3	3	0.1101(0.0727)	3.1529(0.1136)
		−3	3.1104(0.1131)	−2.9445(0.1420)
		0	3.1429(0.1221)	0.0056(0.0910)
		3	3.0903(0.1110)	3.1184(0.1533)

**Table 5.** Empirical means and standard deviations for ML estimators using Nelder-Mead method (with  $\mu = 0$  and  $\sigma = 1$ ).

<i>n</i>	Parameters		Estimations	
	$\alpha$	$\lambda$	$\hat{\alpha}(SD)$	$\hat{\lambda}(SD)$
100	0	−3	0.1029(0.0681)	−2.9175(0.1031)
		0	0.0423(0.0344)	0.0116(0.0304)
	3	3	0.1051(0.0697)	3.1535(0.1190)
		−3	3.1027(0.1151)	−2.9356(0.1471)
		0	3.1511(0.1141)	0.0000(0.0840)
		3	3.0890(0.1138)	3.1243(0.1534)
125	0	−3	0.0693(0.0644)	−2.9221(0.1051)
		0	0.0394(0.0333)	0.0132(0.0307)
	3	3	0.1009(0.0683)	3.1454(0.1175)
		−3	3.1041(0.1110)	−2.9393(0.1414)
		0	3.1511(0.1179)	−0.0024(0.0726)
		3	3.0913(0.1127)	3.1210(0.1528)
150	0	−3	0.0955(0.0631)	−2.9259(0.1023)
		0	0.0399(0.0334)	0.0161(0.0305)
	3	3	0.0973(0.0623)	3.1578(0.1177)
		−3	3.1123(0.1120)	−2.9415(0.1405)
		0	3.1418(0.1208)	−0.0026(0.0703)
		3	3.0921(0.1196)	3.1230(0.1546)
200	0	−3	0.0807(0.0527)	−2.9239(0.1012)
		0	0.0348(0.0305)	0.0197(0.0300)
	3	3	0.0912(0.0614)	3.1470(0.1147)
		−3	3.1076(0.1123)	−2.9329(0.1419)
		0	3.1374(0.1216)	−0.0035(0.0601)
		3	3.0962(0.1233)	3.1087(0.1625)

The simulation study also indicated certain difficulties in the estimation approach for situations near singularity, see Table 6 below. Table 6 below reports the results.

**Table 6.** Percentage of samples that presented convergence problems.

<i>n</i>	Percentage
25	21.27%
50	19.64%
75	18.79%

With  $n \geq 100$ , convergence problems were less frequent ( $\approx 5\%$ ), and for  $n \geq 200$ , very few cases presented problems.

#### 4.2. Fisher-Scoring Method

Table 7 shows simulation results using the Fisher-Scoring algorithm. It was obtained that for  $\alpha = 0$  and  $\lambda > 5$  (and  $\lambda < -5$ ), for  $n = 300$ , non-convergence was down to  $\approx 25\%$ . For  $n = 600$  some improvement was noted in the convergence rate although the standard error remained large as  $\lambda$  increased. Parameter recovery was satisfactory even for moderate sample sizes.

**Table 7.** Empirical means and standard deviations for ML estimators using Fisher-scoring method (with  $\mu = 0$  and  $\sigma = 1$ ).

<i>n</i>	Parameters		Estimations	
	$\alpha$	$\lambda$	$\hat{\alpha}(SD)$	$\hat{\lambda}(SD)$
300	0	−3	0.1094(0.2647)	−3.2822(0.9544)
		0	0.0114(0.1809)	0.0062(39.9985)
		3	0.1584(0.2484)	3.3967(0.8984)
	3	−3	6.6575(3.3495)	−3.6730(2.2947)
		0	3.2757(0.9631)	−0.0041(0.0592)
		3	5.1002(6.3787)	3.4440(11.2287)
400	0	−3	0.0907(0.2379)	−3.2340(0.8117)
		0	0.0027(0.1858)	0.0009(31.9078)
		3	0.1269(0.2220)	3.3254(0.8207)
	3	−3	3.4261(2.9044)	−3.0799(1.4723)
		0	3.1452(0.7781)	−0.0011(0.0507)
		3	3.9208(3.3462)	3.0523(1.2339)
500	0	−3	0.0871(0.2150)	−3.2148(0.7567)
		0	0.0002(0.1713)	0.0000(42.6352)
		3	0.0959(0.1935)	3.2200(0.6685)
	3	−3	3.7597(3.5400)	−3.1587(1.6946)
		0	3.0440(0.6729)	−0.0101(0.0458)
		3	4.0856(3.7022)	3.4352(1.6144)
600	0	−3	0.0898(0.1911)	−3.1157(0.6401)
		0	0.0001(0.1331)	0.0000(27.9890)
		3	0.0613(0.1786)	3.1242(0.6368)
	3	−3	4.1054(3.5126)	−3.2793(1.3546)
		0	2.9885(0.5965)	0.0035(0.0417)
		3	3.3072(3.1817)	3.0986(1.4427)

### 5. Application

We next present an application of the  $BN(\mu, \sigma, \alpha)$  model to a real data situation in which comparisons are made with the mixture of normals model and a flexible asymmetric model. We fitted these models to the variable ultrasound weight, i.e., fetus weight before birth, in 500 observations of the variable  $z = \text{b.weight}$ , which is the ultrasound weight (fetal weight in grams). These data are available at <http://www.mat.udc.cl/hsalinas/cursos/2011/R/weight.rar>. Given the difference in weight of males and females in the gestation stage, it is clear that these data present bimodality.

The descriptive statistics are given in Table 8, where  $\sqrt{b_1}$  and  $b_2$  denote the asymmetry and kurtosis coefficients respectively.

The fit of this data set with the BSN model is compared with the fit given by the mixture of two normals (MN) model, and the flexible skew-normal (FSN) model introduced by Ma and Genton [16], for which the respective densities are given by:

1.  $f(z; \mu_1, \mu_2, \sigma_1, \sigma_2, p) = \frac{p}{\sigma_1} \phi\left(\frac{z-\mu_1}{\sigma_1}\right) + \frac{1-p}{\sigma_2} \phi\left(\frac{z-\mu_2}{\sigma_2}\right)$
2.  $f(z; \mu, \sigma, \alpha, \beta) = \frac{2}{\sigma} \phi\left(\frac{z-\mu}{\sigma}\right) \Phi\left(\alpha\left(\frac{z-\mu}{\sigma}\right) + \beta\left(\frac{z-\mu}{\sigma}\right)^3\right)$

where  $\phi(\cdot)$  and  $\Phi(\cdot)$  denote the density and distribution functions of the standard normal distribution,  $\mu, \mu_1, \mu_2, \alpha, \beta \in \mathbb{R}, \sigma, \sigma_1, \sigma_2 \in \mathbb{R}^+$  and  $0 \leq p \leq 1$ . ML estimation, Akaike’s information criterion (AIC) (see Akaike [36]) and Bayesian information criterion (BIC) (see Schwarz [37]) are reported in Table 9.

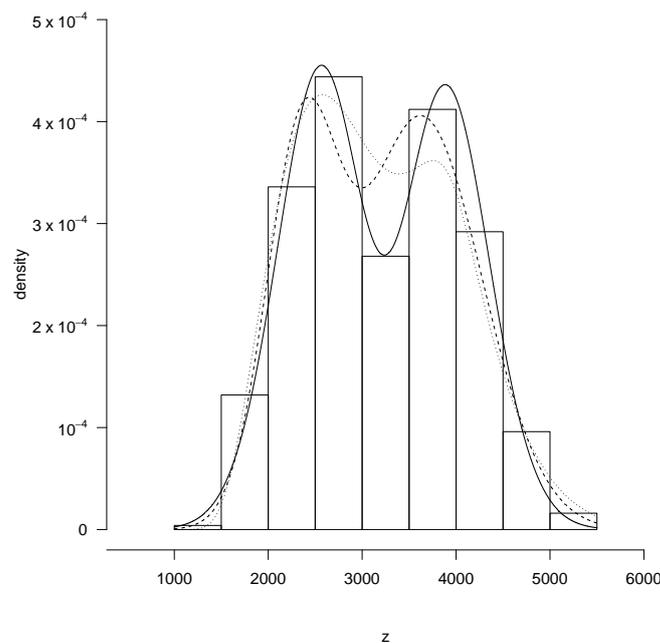
**Table 8.** Summary statistics for variable fetal weight in grams.

Data	$n$	$\bar{Z}$	$S$	$\sqrt{b_1}$	$b_2$
Fetal weight in grams	500	3210.356	834.092	0.071	2.068

**Table 9.** ML estimates and Akaike’s information criterion (AIC) values for the three models.

Model	ML Estimates	AIC	BIC
BSN	$\hat{\mu} = 3232.843, \hat{\sigma} = 555.657, \hat{\alpha} = 1.668, \hat{\lambda} = -0.023$	<b>8095.392</b>	<b>8112.250</b>
MN	$\hat{\mu}_1 = 2350.058, \hat{\mu}_2 = 3632.354, \hat{\sigma}_1 = 381.034, \hat{\sigma}_2 = 647.793, \hat{p} = 0.343$	8107.554	8128.627
FSN	$\hat{\mu} = 2740.386, \hat{\sigma} = 956.663, \hat{\alpha} = -0.299, \hat{\beta} = 0.861$	8111.558	8128.416

Notice that the BSN model presents the best fit (smallest AIC and BIC values). This is also corroborated by Figure 3, which shows the curves corresponding to the fitted models (parameters estimated using the ML approach), overlaid by the data set histogram.



**Figure 3.** Models fitted by the ML approach for the fetal weight in grams data set: BSN (solid line), MN (dashed line) and FSN (dotted line).

## 6. Concluding Remarks

In this paper we have studied further properties of the so called BSN model. Stochastic representations were studied for the model itself and for some sub-models. Derivation of the FIM and verification of its non-singularity is in preparation and will be the subject of a future paper. Satisfactory results are obtained for the BSN model in a real data application. Some additional features of the BSN model are:

- The BSN model presents quite good flexibility in the modes, as shown in Figures 1 and 2.
- The BSN models present no identifiability problems, as shown by Proposition 11.
- The moments, moment generating function, asymmetry and kurtosis coefficients have closed form expressions.
- The simulation study shows that, although there is a little irregularity in consistency for values  $\alpha = \lambda = 0$ , the ML estimators are consistent. This point of irregularity can be explained by the singularity that exists in this point of the information matrix; we will address this situation in a future work.
- We use a model which is a mixture of non-identifiable normal distributions for comparison with the BSN model.
- In the application, two statistical criteria for comparing models were used. The two criteria indicate that the BSN model provides the best fit for these data.

**Author Contributions:** Conceptualization, D.E.-O., H.B. and H.W.G.; Data curation, J.F.O.-P.; Formal analysis, D.E.-O., J.F.O.-P. and O.V.; Funding acquisition, H.W.G.; Investigation, D.E.-O., O.V. and H.W.G.; Methodology, H.B.; Resources, O.V.; Software, J.F.O.-P.; Writing—original draft, O.V. and H.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research of H. W. Gómez was supported by Grant SEMILLERO UA-2020 (Chile). The research of O. Venegas was supported by Vicerrectoría de Investigación y Postgrado de la Universidad Católica de Temuco, Proyecto interno FEQUIP 2019-INRN-03.

**Acknowledgments:** We acknowledge the referees' suggestions that helped us improve this work.

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

## References

1. Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
2. Henze, N. A probabilistic representation of the skew-normal distribution. *Scand. J. Stat.* **1986**, *13*, 271–275.
3. Azzalini, A. Further results on a class of distributions which includes the normal ones. *Statistica* **1986**, *46*, 199–208.
4. Arnold, B.C.; Beaver, R.J.; Groeneveld, R.A.; Meeker, W.Q. The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrika* **1993**, *58*, 471–478. [[CrossRef](#)]
5. Pewsey, A. Problems of inference for Azzalini's skew-normal distribution. *J. Appl. Stat.* **2000**, *27*, 859–870. [[CrossRef](#)]
6. Gupta, A.K.; Chen, T. Goodness-of-fit tests for the skew-normal distribution. *Commun. Stat. Simula C* **2001**, *30*, 907–930.
7. Gupta, R.C.; Brown, N. Reliability studies of the skew-normal distribution and its application to a strength-stress model. *Commun. Stat. Theory Methods* **2001**, *30*, 2427–2445. [[CrossRef](#)]
8. Arellano-Valle, R.B.; Gómez, H.W.; Quintana, F.A. A New Class of Skew-Normal Distributions. *Commun. Stat. Theory Methods* **2004**, *33*, 1465–1480. [[CrossRef](#)]
9. Azzalini, A. The skew-normal distribution and related multivariate families (with discussion). *Scand. J. Stat.* **2005**, *32*, 159–188. [[CrossRef](#)]
10. Gómez, H.W.; Salinas, H.S.; Bolfarine, H. Generalized skew-normal models: Properties and inference. *Statistics* **2006**, *40*, 495–505. [[CrossRef](#)]
11. Gómez, H.W.; Torres, F.J.; Bolfarine, H. Large-Sample Inference for the Epsilon-Skew-t Distribution. *Commun. Stat. Theory Methods* **2007**, *36*, 73–81. [[CrossRef](#)]

12. Venegas, O.; Sanhueza, A.I.; Gómez, H.W. An extension of the skew-generalized normal distribution and its derivation. *Proyecciones J. Math.* **2011**, *30*, 401–413. [[CrossRef](#)]
13. McLachlan, G.; Peel, D. *Finite Mixture Models*; John Wiley & Sons, Inc.: New York, NY, USA, 2000.
14. Marin, J.M.; Mengersen, K.; Robert, C. Bayesian modeling and inference on mixtures of distributions. In *Handbook of Statistics*; Elsevier: Amsterdam, The Netherlands, **2005**; Volume 25, pp. 459–503.
15. Azzalini, A.; Capitanio, A. Distributions generate by perturbation of symmetry with emphasis on a multivariate skew-t distribution. *J. R. Stat. Soc. Ser. B* **2003**, *65*, 367–389. [[CrossRef](#)]
16. Ma, Y.; Genton, M.G. Flexible class of skew-symmetric distributions. *Scand. J. Stat.* **2004**, *31*, 459–468. [[CrossRef](#)]
17. Arellano-Valle, R.B.; Gómez, H.W.; Quintana, F.A. Statistical inference for a general class of asymmetric distributions. *J. Stat. Plan. Inference* **2005**, *128*, 427–443. [[CrossRef](#)]
18. Kim, H.J. On a class of two-piece skew-normal distributions. *Statistics* **2005**, *39*, 537–553. [[CrossRef](#)]
19. Lin, T.I.; Lee, J.C.; Hsieh, W.J. Robust mixture models using the skew-t distribution. *Stat. Comput.* **2007**, *17*, 81–92. [[CrossRef](#)]
20. Lin, T.I.; Lee, J.C.; Yen, S.Y. Finite mixture modeling using the skew-normal distribution. *Stat. Sin.* **2007**, *17*, 909–927.
21. Elal-Olivero, D.; Gómez, H.W.; Quintana, F.A. Bayesian Modeling using a class of Bimodal skew-Elliptical distributions. *J. Stat. Plan. Inference* **2009**, *139*, 1484–1492. [[CrossRef](#)]
22. Arnold, B.C.; Gómez, H.W.; Salinas, H.S. On multiple constraint skewed models. *Statistics* **2009**, *43*, 279–293. [[CrossRef](#)]
23. Arellano-Valle, R.B.; Cortés, M.A.; Gómez, H.W. An extension of the epsilon-skew-normal distribution. *Commun. Stat. Theory Methods* **2010**, *39*, 912–922. [[CrossRef](#)]
24. Elal-Olivero, D. Alpha-Skew-Normal Distribution. *Proyecciones J. Math.* **2010**, *29*, 224–240. [[CrossRef](#)]
25. Gómez, H.W.; Elal-Olivero, D.; Salinas, H.S.; Bolfarine, H. Bimodal extension based on the skew-normal distribution with application to pollen data. *Environmetrics* **2011**, *22*, 50–62. [[CrossRef](#)]
26. Arnold, B.C.; Gómez, H.W.; Salinas, H.S. A doubly skewed normal distribution. *Statistics* **2015**, *49*, 842–858. [[CrossRef](#)]
27. Da Silva Braga, A.; Cordeiro, G.M.; Ortega, E.M.M. A new skew-bimodal distribution with applications. *Commun. Stat. Theory Methods* **2017**, *47*, 2950–2968. [[CrossRef](#)]
28. Venegas, O.; Salinas, H.S.; Gallardo, D.I.; Bolfarine, B.; Gómez, H.W. Bimodality based on the generalized skew-normal distribution. *J. Stat. Comput. Simul.* **2018**, *88*, 156–181. [[CrossRef](#)]
29. Shah, S.; Chakraborty, S.; Hazarika, P.J. A New One Parameter Bimodal Skew Logistic Distribution and its Applications. *arXiv* **2019**, arXiv:1906.04125.
30. Gómez-Déniz, E.; Pérez-Rodríguez, J.V.; Reyes, J.; Gómez, H.W. A Bimodal Discrete Shifted Poisson Distribution. A Case Study of Tourists' Length of Stay. *Symmetry* **2020**, *12*, 442. [[CrossRef](#)]
31. Esmaeili, H.; Lak, F.; Alizadeh, M.; Monfared, M.E.D. The Alpha-Beta Skew Logistic Distribution: Properties and Applications. *Stat. Optim. Inf. Comput.* **2020**, *8*, 304–317. [[CrossRef](#)]
32. Imani, M.; Ghoreishi, S.F. Bayesian optimization objective-based experimental design. In Proceedings of the 2020 American Control Conference (ACC 2020), Denver, CO, USA, 1–3 July 2020.
33. Maleki, M.; Wraith, D.; Arellano-Valle, R.B. Robust finite mixture modeling of multivariate unrestricted skew-normal generalized hyperbolic distributions. *Stat. Comput.* **2019**, *29*, 415–428. [[CrossRef](#)]
34. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2017.
35. Nelder, J.A.; Mead, R. A simplex algorithm for function minimization. *Comput. J.* **1965**, *7*, 308–313. [[CrossRef](#)]
36. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control.* **1974**, *19*, 716–723. [[CrossRef](#)]
37. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [[CrossRef](#)]

