

Article

# Advanced Bimodal Skew-Symmetric Distributions: Methodology and Application to Cancer Cell Protein Data

Gadir Alomair <sup>1,\*</sup> , Hugo S. Salinas <sup>2,\*</sup> , Hassan S. Bakouch <sup>3,4</sup> , Idika E. Okorie <sup>5</sup> and Olayan Albalawi <sup>6</sup> <sup>1</sup> Department of Quantitative Methods, School Business, King Faisal University, Al-Ahsa 31982, Saudi Arabia<sup>2</sup> Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Copiapó 1531772, Chile<sup>3</sup> Department of Mathematics, College of Science, Qassim University, Buraydah 51452, Saudi Arabia; h.bakouch@qu.edu.sa<sup>4</sup> Department of Mathematics, Faculty of Science, Tanta University, Tanta 31111, Egypt<sup>5</sup> Department of Mathematics, Khalifa University, Abu Dhabi 127788, United Arab Emirates; idika.okorie@ku.ac.ae<sup>6</sup> Department of Statistics, Faculty of Science, University of Tabuk, Tabuk 47512, Saudi Arabia; oalbalwi@ut.edu.sa

\* Correspondence: galomair@kfu.edu.sa (G.A.); hugo.salinas@uda.cl (H.S.S.)

**Abstract:** This paper explores bimodal skew-symmetric distributions, a versatile family of distributions characterized by parameters that control asymmetry and kurtosis. These distributions encapsulate both symmetrical and well-known asymmetrical behaviors. A simulation study evaluates the model's estimation accuracy, detailing the score function and the robustness of the observed information matrix, which is proven to be non-singular under specific conditions. We apply the bimodal skew-normal model to protein data from cancer cells, comparing its performance against four established distributions supported on the entire real line. Results indicate superior performance by the proposed model, underscoring its potential for enhancing analytical precision in biological research.

**Keywords:** skew-normal distribution; statistical model; bimodality; observed information matrix; protein in cancer cells; simulation

**MSC:** 60E05; 62E15; 62F10

**Citation:** Alomair, G.; Salinas, H.S.; Bakouch, H.S.; Okorie, I.E.; Albalawi, O. Advanced Bimodal Skew-Symmetric Distributions: Methodology and Application to Cancer Cell Protein Data. *Symmetry* **2024**, *16*, 985. <https://doi.org/10.3390/sym16080985>

Academic Editor: Calogero Vetro

Received: 8 June 2024

Revised: 23 July 2024

Accepted: 30 July 2024

Published: 2 August 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The family of skew-symmetric distributions has been increasingly recognized for its flexibility and efficacy in modeling real-world data by transforming symmetric probability density functions (PDFs) with specific generators. This family is defined by the following PDF:

$$2f(z)G(w(z)), \quad z \in \mathbb{R}, \quad (1)$$

where  $f$  is a symmetric PDF centered at 0 and  $G$  is the cumulative distribution function (CDF) of a continuous random variable that is symmetric around 0. The function  $w$  is required to be odd and continuous, meaning  $w(-z) + w(z) = 0$ .

This framework was initially developed by Azzalini [1,2], who introduced the skew-normal distribution by setting  $f = G' = \phi$ , the PDF of the standard normal distribution, and  $w(z) = \lambda z$ . This construction allows the skew-symmetric distribution to encapsulate both symmetric and skewed data through the parameter  $\lambda$ . Over time, various researchers (e.g., Gupta et al. [3], Ma and Genton [4], and Arellano-Valle et al. [5]) have expanded this model to include different forms of  $w$ , such as  $w(z) = \lambda_1 z + \lambda_2 z^3$  and  $w(z) = \lambda_1 z / \sqrt{1 + \lambda_2^2 z^2}$ , broadening its applicability within the skew-symmetric framework.

The general form (1) encompasses a wide range of submodels, from the symmetric density  $f$  (when  $w(z) = 0$ ) to the highly skewed half- $f$  densities (as  $w(z) \rightarrow \pm\infty$ ). These models can capture varying degrees of skewness in data, making them valuable in many

statistical applications. For example, Pewsey [6] examined a subfamily where  $w(z) = \lambda z$  for  $\lambda \in \mathbb{R}$ . However, the model  $2\eta^{-1}f((x - \xi)/\eta)G(\lambda(x - \xi)/\eta)$ ,  $x \in \mathbb{R}$ , with location parameter  $\xi$  and scale parameter  $\eta$ , encounters significant challenges in maximum likelihood estimation (MLE) when  $G' = f = \phi$ . Specifically, when  $\lambda = 0$ , the expected information matrix becomes singular, complicating the estimation process. Pewsey also noted that the observed information matrix fails to have an inverse for the CDF  $G$  when  $G''(0) = g'(0) = 0$ .

To address these challenges and further enhance the applicability of skew-symmetric distributions, we propose a novel  $w$  function that not only avoids the singularity issues at  $\lambda = 0$  but also enables the modeling of bimodal data, which is crucial in many practical fields, including medicine. Specifically, we consider  $w(z) = \lambda z^2$  for  $z > 0$  and  $w(z) = -\lambda z^2$  for  $z < 0$ . This function, equivalent to  $w(z) = \lambda \text{sign}(z)z^2 \equiv \lambda z|z|$ , where  $\text{sign}(\cdot)$  denotes the sign function, meets the necessary properties of  $w$  and introduces bimodality into the model. Some models in the literature that do not present singularity issues in the information matrix are as follows: Bakouch et al. [7] introduce a family of skewed distributions and explore the bimodal skew-normal distribution; Salinas et al. [8] present a two-piece normal distribution for modeling biaxial fatigue data; and Khorsheed et al. [9] propose a flexible form of three-parameter skew-normal distributions, enhancing flexibility for practical and industrial applications.

This proposed  $w$  function ensures that the model satisfies the regular conditions required for deriving the asymptotic distribution of the parameter vector. Importantly, the observed information matrix remains non-singular when  $\lambda = 0$ , overcoming a significant limitation of previous models. The motivation for introducing this function is to address the non-singularity issue at  $\lambda = 0$  and to incorporate bimodality, which enhances the model's capability to accurately represent complex data distributions, particularly in fields such as medical research.

The primary objectives of the proposed bimodal skew-symmetric distributions are to provide a flexible framework for accurately modeling data with bimodal and skewed characteristics, which are common in various practical applications. The goals include extending the existing skew-normal distributions to accommodate bimodal features, enhancing the ability to model asymmetric data, and developing practical tools for parameter estimation and goodness-of-fit tests. Additionally, we aim to demonstrate the effectiveness of the proposed model through empirical analyses, such as on protein data from cancer cells, to illustrate its practical value and encourage its adoption in relevant fields.

These distributions are highly relevant in various fields where data exhibit bimodal and skewed characteristics. For example, in biology, they can model the distribution of protein expression levels in cancer cells, where distinct subpopulations of cells exhibit different expression patterns. In finance, these distributions can describe the returns of assets that have two predominant regimes, such as bullish and bearish market conditions, while also accounting for skewness due to market asymmetries. In environmental science, they are useful for modeling pollutant concentrations that show bimodal behavior due to varying sources and conditions. By providing a flexible framework that captures these complex data structures, the proposed distribution offers significant advantages for accurate modeling and inference in real-world applications.

This paper is structured as follows. Section 2 defines the bimodal skew-symmetric distribution and examines its key probabilistic properties as well as certain inferential issues. Section 3 introduces the bimodal skew-normal (BSN) distribution as a special case of the bimodal skew-symmetric family and discusses its properties and estimation. In Section 4, we demonstrate the adaptability of this class of distributions by analyzing data on proteins in cancer cells. Finally, Section 5 provides concluding remarks.

## 2. Bimodal Skew-Symmetric Family

In this paper, we investigate a family of distributions, called bimodal skew-symmetric distributions, that is generated by Equation (1) using  $w(z) = \lambda z|z|$ , where  $\lambda \in \mathbb{R}$ . We start

by presenting a lemma that characterizes this class of distributions and then proceed to derive several important properties. These properties are relevant for understanding the behavior of this family of distributions and for developing inferential procedures for fitting the model to data.

**Lemma 1.** Let  $f$  be a symmetric PDF about 0 and let  $G$  be the CDF of a continuous random variable that is symmetric around 0. We define the function

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

which is a PDF for any value of  $\lambda \in \mathbb{R}$ . A random variable  $Z$  with a bimodal skew-symmetric distribution and a PDF given by (2) is denoted by  $Z \sim BSf(\lambda; G)$ .

**Proof.** Let  $\mu(\lambda) = \int_{-\infty}^{\infty} 2f(z)G(\lambda z|z|)dz$ . We aim to prove that  $\mu(\lambda) = 1$  for all  $\lambda \in \mathbb{R}$ . In fact,

$$\begin{aligned} \mu(\lambda) &= \int_{-\infty}^0 2f(z)G(-\lambda z^2)dz + \int_0^{\infty} 2f(z)G(\lambda z^2)dz \\ &= \int_0^{\infty} 2f(z)G(-\lambda z^2)dz + \int_0^{\infty} 2f(z)G(\lambda z^2)dz \\ &= \int_0^{\infty} 2f(z)[G(-\lambda z^2) + G(\lambda z^2)]dz \\ &= \int_0^{\infty} 2f(z)dz = 1 \end{aligned}$$

□

### 2.1. Cumulative Distribution Function

The cumulative distribution function corresponding to the density in (2) is given by

$$H_Z(z; \lambda) = \int_{-\infty}^z 2f(t)G(\lambda t|t|)dt, \quad z \in \mathbb{R}. \quad (3)$$

**Proposition 1.** Suppose  $H_Z(z; \lambda)$  as given in (3), then the following properties are obtained:

- (i)  $H_Z(z; 0) = F(z)$ , where  $F$  is the CDF of  $f$ .
- (ii)  $H_Z(-z; \lambda) = 1 - H_Z(z; -\lambda)$ .
- (iii)  $\lim_{\lambda \rightarrow +\infty} H_Z(z; \lambda) = (2F(z) - 1)I(z \geq 0)$ .
- (iv)  $\lim_{\lambda \rightarrow -\infty} H_Z(z; \lambda) = 2F(z)I(z < 0)$ ,

where  $I(\cdot)$  is the indicator function.

**Proof.**

- (i)  $H_Z(z; 0) = \int_{-\infty}^z 2f(t)G(0)dt = \int_{-\infty}^z f(t)dt = F(z)$ , where  $F$  is the CDF of  $f$ .
- (ii)

$$\begin{aligned} H_Z(-z; \lambda) &= \int_{-\infty}^{-z} 2f(t)G(\lambda t|t|)dt = - \int_{\infty}^z 2f(-u)G(-\lambda u|-u|)du \\ &= \int_z^{\infty} 2f(u)G(-\lambda u|u|)du = 1 - \int_{-\infty}^z 2f(u)G(-\lambda u|u|)du \\ &= 1 - H_Z(z; -\lambda). \end{aligned}$$

(iii) Suppose  $z \geq 0$ , then

$$\begin{aligned}\lim_{\lambda \rightarrow +\infty} H_Z(z; \lambda) &= \lim_{\lambda \rightarrow +\infty} \int_{-\infty}^z 2f(t)G(\lambda t|t)dt \\ &= \lim_{\lambda \rightarrow +\infty} \left[ \int_{-\infty}^0 2f(t)G(-\lambda t^2)dt + \int_0^z 2f(t)G(\lambda t^2)dt \right] \\ &= \int_{-\infty}^0 2f(t)G(-\infty)dt + \int_0^z 2f(t)G(+\infty)dt \\ &= 2 \int_0^z f(t)dt = 2 \left[ \int_{-\infty}^z f(t)dt - 1/2 \right] = 2F(z) - 1.\end{aligned}$$

(iv) Suppose  $z < 0$ , then

$$\begin{aligned}\lim_{\lambda \rightarrow -\infty} H_Z(z; \lambda) &= \lim_{\lambda \rightarrow -\infty} \int_{-\infty}^z 2f(t)G(\lambda t|t)dt \\ &= \lim_{\lambda \rightarrow -\infty} \int_{-\infty}^z 2f(t)G(-\lambda t^2)dt = \lim_{\lambda \rightarrow -\infty} \int_{-\infty}^z 2f(t)(1 - G(\lambda t^2))dt \\ &= \int_{-\infty}^z 2f(t)dt - \int_{-\infty}^z 2f(t)G(-\infty)dt = 2 \int_{-\infty}^z f(t)dt = 2F(z).\end{aligned}$$

□

## 2.2. Properties

### 2.2.1. Basic Properties

The following properties are directly derived from Lemma 1.

**Proposition 2.** Using the previous notations, the following properties hold:

- (i)  $h_Z(z; 0) = f(z)$ .
- (ii)  $-Z \sim BSf(-\lambda; G)$ .
- (iii)  $Y = Z^2 \sim y^{-1/2}f(y^{1/2})I(y \geq 0)$ .
- (iv)  $Y = |Z| \sim 2f(y)I(y \geq 0)$ .
- (v)  $\lim_{\lambda \rightarrow +\infty} h_Z(z; \lambda) = 2f(z)I(z \geq 0)$ .
- (vi)  $\lim_{\lambda \rightarrow -\infty} h_Z(z; \lambda) = 2f(z)I(z < 0)$ .

**Proof.** Property (i) of Proposition 2 shows that the  $f$  distribution belongs to the family of  $BSf$  distributions. Properties (ii)–(iv) indicate the distributions of the variables  $-Z$ ,  $Z^2$ , and  $|Z|$ , respectively. Properties (v) and (vi) show the distributions that follow by considering the limiting values of  $\lambda$ . □

### 2.2.2. Bimodality Property

The bimodality property of the random variable  $Z$  when it follows a  $BSf$  distribution with  $\lambda \neq 0$  is presented in Proposition 3. To prove this, we differentiate Equation (2) with respect to  $z$  and equate it to zero, which yields two different solutions,  $z_1$  and  $z_2$ . The first solution  $z_1$  corresponds to a negative modal point, and the second solution  $z_2$  corresponds to a positive modal point. Thus, the random variable  $Z$  is a bimodal with two distinct modes at  $z_1 \in \mathbb{R}^-$  and  $z_2 \in \mathbb{R}^+$ . This property is useful in modeling real-life situations that exhibit two distinct peaks in their data distribution.

**Proposition 3.** Suppose  $Z \sim BSf(\lambda; G)$ , then the random variable  $Z$  is a bimodal for  $\lambda \neq 0$ .

**Proof.** Differentiating Equation (2) with respect to  $z$  and equating to zero implies

$$\begin{aligned} z_1 &= \frac{f'(z_1)\{1 - G(\lambda z_1^2)\}}{2\lambda f(z_1)g(\lambda z_1^2)}, \quad \text{if } z_1 < 0, \\ z_2 &= -\frac{f'(z_2)G(\lambda z_2^2)}{2\lambda f(z_2)g(\lambda z_2^2)}, \quad \text{if } z_2 \geq 0, \end{aligned}$$

where  $G' = g$  and  $\lambda \neq 0$ . Therefore,  $z_1 \in \mathbb{R}^-$  and  $z_2 \in \mathbb{R}^+$  are different modal points. Therefore, the random variable  $Z$  is a bimodal.  $\square$

### 2.3. Stochastic Representation of the Random Variable

**Proposition 4.** Suppose  $Z \sim BSf(\lambda; G)$  with  $\lambda \in \mathbb{R}$ . Then  $Z$  can be represented as  $Z = SY$ , where  $S$  and  $Y$  are dependent random variables with  $f_Y(y) = 2f(y)I(y \geq 0)$  and  $P(S = 1|Y = y) = 1 - P(S = -1|Y = y) = G(\lambda y^2)$ .

**Proof.** Let  $S$  and  $Y$  be defined as in the statement of the proposition. Using the joint distribution of  $(Z, S)$  and the Jacobian method, the marginal distribution of  $Z$  is obtained as follows:

If  $Z \geq 0$ , then  $Z = Y$  and  $S = 1$ . Therefore, we have

$$h_Z(z; \lambda) = f_Y(z)P(S = 1|Y = z) = f_Y(z)G(\lambda z^2) = 2f(z)G(\lambda z^2). \quad (4)$$

On the other hand, if  $Z < 0$ , then  $Z = -Y$  and  $S = -1$ ,

$$\begin{aligned} h_Z(z; \lambda) &= f_Y(-z)P(S = -1|Y = -z) = f_Y(-z)(1 - G(\lambda z^2)) \\ &= 2f(-z)G(-\lambda z^2) = 2f(z)G(-\lambda z^2). \end{aligned} \quad (5)$$

Therefore, from (4) and (5), we obtain  $h_Z(z; \lambda) = 2f(z)G(\lambda \text{sgn}(z)z^2) = 2f(z)G(\lambda z|z|)$ .  $\square$

This proof shows that a random variable  $Z$  that follows a BSf distribution with location parameter  $\lambda$  can be represented as a combination of two dependent random variables  $S$  and  $Y$ . The variable  $Y$  has a density function that is twice the absolute value of the density function of  $Z$  for positive values of  $Z$  and is zero for negative values of  $Z$ . The variable  $S$  takes the value of 1 with the probability given by the value of the cumulative distribution function of  $G$  evaluated at  $\lambda Y^2$ , and the value of  $-1$  with the complement of this probability.

This representation is useful because it provides a way to generate random samples from the BSf distribution using the joint distribution of  $S$  and  $Y$ . Additionally, it allows for the computation of various statistics and moments of the distribution using the properties of  $S$  and  $Y$ .

### 2.4. Calculation of Moments for the BSf Distribution

The random variable  $Z$  can be represented as a combination of two dependent random variables  $S$  and  $Y$ , as shown in Proposition 4. In this section, we derive a formula for computing the  $r$ -th moment of a random variable  $X$  that follows the  $BSf(\theta; G)$  distribution, where  $\theta = (\xi, \eta, \lambda)'$  and  $X = \xi + \eta Z$ , with  $Z \sim BSf(\lambda; G)$ .

**Proposition 5.** The  $r$ -th moment of  $X$  is given by

$$\mathbb{E}(X^r) = \sum_{k=0}^r \binom{r}{k} \xi^{r-k} \eta^k \mathbb{E}(Z^k), \quad (6)$$

where  $\mathbb{E}(Z^k)$  is given by

$$\mathbb{E}(Z^k) = \begin{cases} \mathbb{E}(Y^k), & \text{if } k \text{ is even,} \\ 2\mathbb{E}(Y^k G(\lambda Y^2)) - \mathbb{E}(Y^k), & \text{if } k \text{ is odd,} \end{cases} \quad (7)$$

and  $Y \sim 2f(y)I(y \geq 0)$  is the random variable in the stochastic representation of  $Z$  as given in Proposition 4.

**Proof.** By utilizing the stochastic representation provided in Proposition 4 and applying the properties of conditional expectation, we can derive the required expression.

$$\begin{aligned} \mathbb{E}(Z^k) &= \mathbb{E}(\mathbb{E}(Z^k|Y)) \\ &= \mathbb{E}(Y^k G(\lambda Y^2) + (-1)^k Y^k (1 - G(\lambda Y^2))) \\ &= \mathbb{E}((1 - (-1)^k) Y^k G(\lambda Y^2) + (-1)^k Y^k). \end{aligned}$$

The above leads to the conclusion that if  $k$  is even, then  $\mathbb{E}(Z^k) = \mathbb{E}(Y^k)$ . On the other hand, if  $k$  is odd, then  $\mathbb{E}(Z^k) = 2\mathbb{E}(Y^k G(\lambda Y^2)) - \mathbb{E}(Y^k)$ . To obtain  $\mathbb{E}(X^k)$ , it is possible to apply the binomial theorem along with the basic properties of the expectation.  $\square$

The mean and variance of a random variable  $X$  with BSf distribution can be easily calculated using the following corollary:

**Corollary 1.** Suppose  $Z \sim BSf(\lambda; G)$  and  $X = \zeta + \eta Z \sim BSf(\theta; G)$ . Then, the mean and variance of  $X$  are given by

$$\mathbb{E}(X) = \zeta + \eta(2b_1 - a_1) \quad \text{and} \quad \text{Var}(X) = \eta^2 a_2 - (2b_1 - a_1)^2, \quad (8)$$

where  $a_r = \int_0^\infty 2y^r f(y) dy$  and  $b_r = \int_0^\infty 2y^r f(y) G(\lambda y^2) dy$  for  $r = 1, 2$ .

This result provides a straightforward way to compute the expected value and variance of a BSf-distributed random variable  $X$ , where  $\zeta$ ,  $\eta$ ,  $\lambda$ , and  $G$  are parameters of the distribution. The integrals  $a_r$  and  $b_r$  can be numerically evaluated, making the calculation of  $\mathbb{E}(X)$  and  $\text{Var}(X)$  feasible in practice.

### 2.5. Observed Information Matrix for the Location–Scale BSf Distribution

Proposition 6 states that if  $\mathbf{x}$  is a random sample from a  $BSf(\theta; G)$  distribution with a continuous and differentiable symmetric univariate probability density function  $f$  and cumulative distribution function  $G$ , where  $f = \phi$ , then the solution to the score equations is  $\lambda = 0$ ,  $\zeta = \bar{x}$ , and  $\eta^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / n$ , and the observed information matrix is non-singular when  $\lambda = 0$  and  $f''$  and  $G''$  are continuous functions.

**Proposition 6.** Let  $\mathbf{x} = (x_1, x_2, \dots, x_n)'$  be a realization of the random sample  $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ , where  $X_1, X_2, \dots, X_n$  are independent and identically distributed random variables following a  $BSf(\theta; G)$  distribution. Assume that  $f$  and  $G$  are continuous and a differentiable symmetric univariate probability density function and cumulative distribution function, respectively, with  $f = \phi$ .

- (i) The solution to the score equations is  $\lambda = 0$ ,  $\zeta = \bar{x}$ , and  $\eta^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / n$ .
- (ii) When  $\lambda = 0$ , the observed information matrix is non-singular.

**Proof.** (i) Let  $l(\theta, \mathbf{x})$  be the log-likelihood function. Assuming  $f'$  exists, and denoting  $G' = g$ , the first-order partial derivatives of the log-likelihood are as follows:

$$\begin{aligned} \frac{\partial l(\theta, \mathbf{x})}{\partial \xi} &= -\frac{1}{\eta} \left\{ \sum_{i=1}^n \frac{f'(z_i)}{f(z_i)} + 2\lambda \sum_{i=1}^n \text{sgn}(z_i) z_i \frac{g(\lambda z_i |z_i|)}{G(\lambda z_i |z_i|)} \right\}, \\ \frac{\partial l(\theta, \mathbf{x})}{\partial \eta} &= -\frac{1}{\eta} \left\{ n + \sum_{i=1}^n z_i \frac{f'(z_i)}{f(z_i)} + 2\lambda \sum_{i=1}^n \text{sgn}(z_i) z_i^2 \frac{g(\lambda z_i |z_i|)}{G(\lambda z_i |z_i|)} \right\}, \\ \frac{\partial l(\theta, \mathbf{x})}{\partial \lambda} &= \sum_{i=1}^n \text{sgn}(z_i) z_i^2 \frac{g(\lambda z_i |z_i|)}{G(\lambda z_i |z_i|)}, \end{aligned}$$

where  $z_i = (x_i - \xi) / \eta$  and  $\text{sgn}(z_i) = \text{sign}(z_i)$ , which is defined to be  $|z_i| / z_i$ . Note that the log-likelihood function  $l(\theta, \mathbf{x})$  depends on the parameter  $\lambda$ . Therefore, the partial derivative  $\frac{\partial l}{\partial \lambda}$  measures the sensitivity of the log-likelihood with respect to changes in  $\lambda$ .

The score equations for the family  $SBf(\theta; G)$  are given by

$$\begin{aligned} \bar{v} + 2\lambda \overline{\text{sgn}(z)zw} &= 0, \\ 1 + \bar{v} + 2\lambda \overline{\text{sgn}(z)z^2w} &= 0, \\ \overline{\text{sgn}(z)z^2w} &= 0, \end{aligned}$$

where  $v_i = f'(z_i) / f(z_i)$  and  $w_i = g(\lambda z_i |z_i|) / G(\lambda z_i |z_i|)$ .

Solving these equations yields  $\bar{v} = -1$ ,  $\eta = \xi \bar{v} - \bar{v} \bar{x}$ , and  $\lambda = -\bar{v} / \overline{\text{sgn}(z)zw}$  for any solution. If  $\lambda = 0$ , then the score equations require  $\bar{v} = 0$ . In this case, we have  $\bar{w} = 2g(0)$  and  $\eta = -\bar{v} \bar{x}$ . Thus,  $\lambda = 0$ ,  $\xi = \bar{x}$ , and  $\eta^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / n$  are a solution to the score equations of the family  $SB\phi(\theta; G)$ , regardless of the choice of  $G$ .

We observe that the condition  $\lambda = 0$  and  $\eta = -\bar{v} \bar{x}$  are a solution to the score equations only if we can select a density  $f$  such that  $\sum_{i=1}^n f'(z_i) / f(z_i) = 0$ . Therefore, we conclude that the estimators of the family  $BSf$  for  $\lambda = 0$  and  $f = \phi$  coincide with the class  $2\eta^{-1} \phi((x - \xi) / \eta) G(\lambda(x - \xi) / \eta)$  studied by Pewsey [6].

(ii) Assuming that  $G'' = g'$  and  $f''$  exist, we can obtain the second-order partial derivatives of the log-likelihood by defining  $u_i = f''(z_i) / f(z_i)$  and  $t_i = g'(\lambda z_i |z_i|) / G(\lambda z_i |z_i|)$ . With these definitions, the partial derivatives can be computed as follows:

$$\begin{aligned} \frac{\partial^2 l(\theta, \mathbf{x})}{\partial \xi^2} &= -\frac{n}{\eta^2} \left\{ \bar{v}^2 - \bar{u} - 2\lambda \overline{\text{sgn}(z)w} + 4\lambda^2 (\bar{z}^2 \bar{w}^2 - \bar{z}^2 \bar{t}) \right\}, \\ \frac{\partial^2 l(\theta, \mathbf{x})}{\partial \xi \partial \eta} &= -\frac{n}{\eta^2} \left\{ -\bar{v} + \bar{z} \bar{v}^2 - \bar{z} \bar{u} - 4\lambda \overline{\text{sgn}(z)zw} + 4\lambda^2 (\bar{z}^3 \bar{w}^2 - \bar{z}^3 \bar{t}) \right\}, \\ \frac{\partial^2 l(\theta, \mathbf{x})}{\partial \xi \partial \lambda} &= -\frac{n}{\eta} \left\{ 2\overline{\text{sgn}(z)zw} - 2\lambda (\bar{z}^3 \bar{w}^2 - \bar{z}^3 \bar{t}) \right\}, \\ \frac{\partial^2 l(\theta, \mathbf{x})}{\partial \eta^2} &= -\frac{n}{\eta^2} \left\{ \bar{z}^2 \bar{v}^2 - \bar{z}^2 \bar{u} - 2\bar{z} \bar{v} - 6\lambda \overline{\text{sgn}(z)z^2w} - 4\lambda^4 (\bar{z}^4 \bar{w}^2 - \bar{z}^4 \bar{t}) - 1 \right\}, \\ \frac{\partial^2 l(\theta, \mathbf{x})}{\partial \eta \partial \lambda} &= -\frac{n}{\eta} \left\{ 2\overline{\text{sgn}(z)z^2w} - 2\lambda (\bar{z}^4 \bar{w}^2 - \bar{z}^4 \bar{t}) \right\}, \\ \frac{\partial^2 l(\theta, \mathbf{x})}{\partial \lambda^2} &= -n (\bar{z}^4 \bar{w}^2 - \bar{z}^4 \bar{t}). \end{aligned}$$

From the score equations, we can see that  $\overline{\text{sgn}(z)z^2w} = 0$  and  $\bar{v} = -1$ . Moreover, if there exists a solution to these equations such that  $\lambda = 0$ , then we have  $\bar{v} = 0$ ,  $\bar{w} = 2g(0)$ ,  $\bar{t} = 2g'(0)$ , and  $\eta = -\bar{v} \bar{x}$  for any solution.

$$\frac{\partial^2 l(\theta, \mathbf{x})}{\partial \xi^2} = -\frac{n}{\eta^2} \left\{ \bar{v}^2 - \bar{u} \right\}, \quad \frac{\partial^2 l(\theta, \mathbf{x})}{\partial \xi \partial \eta} = -\frac{n}{\eta^2} \left\{ \bar{z} \bar{v}^2 - \bar{z} \bar{u} \right\},$$



$$\frac{\partial^2 l(\theta, \mathbf{x})}{\partial \xi \partial \lambda} = -\frac{4ng(0)}{\eta} \overline{\text{sgn}(z)z}, \quad \frac{\partial^2 l(\theta, \mathbf{x})}{\partial \eta^2} = -\frac{n}{\eta^2} \left\{ \overline{z^2 v^2} - \overline{z^2 u} + 1 \right\},$$

$$\frac{\partial^2 l(\theta, \mathbf{x})}{\partial \eta \partial \lambda} = 0, \quad \frac{\partial^2 l(\theta, \mathbf{x})}{\partial \lambda^2} = -2n(2g^2(0) - g'(0)) \overline{z^4}.$$

Note that many symmetric densities around zero are differentiable at this point, including popular ones such as the normal, logistic, and Student's t densities. This means that for these distributions, we have  $g'(0) = 0$ . However, there are exceptions to this rule, such as the double exponential density, which is not differentiable at zero. When we set  $\lambda = 0$  and  $f = \phi$ , we can calculate that  $\xi = \bar{x}$  and  $\eta^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / n$ . We can then define standardized scores as  $z_i = (x_i - \bar{x}) / \eta$ , which have a mean of zero and a variance of one:  $\bar{z} = 0$  and  $\overline{z^2} = 1$ .

Using these standardized scores, we can express the first derivative of  $\phi$  as  $v_i = \phi'(z_i) / \phi(z_i) = -z_i$ . This gives us  $\bar{v} = 0$ ,  $\overline{v^2} = 1$ ,  $\overline{zv^2} = \overline{z^3}$ , and  $\overline{z^2 v^2} = \overline{z^4}$ .

Conversely, we can find the second derivative of  $\phi$  by using the formula  $u_i = \phi''(z_i) / \phi(z_i) = z_i^2 - 1$ . We can calculate that  $\overline{u^2} - 1 = 0$  and  $\overline{zu} = \overline{z(z^2 - 1)} = \overline{z^3} - \bar{z} = \overline{z^3}$ . Additionally, we have  $\overline{z^2 u} = \overline{z^4} - 1$ .

The second-order partial derivatives for this solution are given by

$$\frac{\partial^2 l(\theta, \mathbf{x})}{\partial \xi^2} = -\frac{n}{\eta^2}, \quad \frac{\partial^2 l(\theta, \mathbf{x})}{\partial \xi \partial \eta} = 0, \quad \frac{\partial^2 l(\theta, \mathbf{x})}{\partial \xi \partial \lambda} = -\frac{4ng(0)\delta}{\eta^2},$$

$$\frac{\partial^2 l(\theta, \mathbf{x})}{\partial \eta^2} = -\frac{2n}{\eta^2}, \quad \frac{\partial^2 l(\theta, \mathbf{x})}{\partial \eta \partial \lambda} = 0, \quad \frac{\partial^2 l(\theta, \mathbf{x})}{\partial \lambda^2} = -2n(2g^2(0) - g'(0))\kappa,$$

where  $\delta = \sum_{i=1}^n |x_i - \bar{x}| / n$  is the mean absolute deviation and  $\kappa = n \sum_{i=1}^n (x_i - \bar{x})^4 / (\sum_{i=1}^n (x_i - \bar{x})^2)^2$  is the kurtosis. This leads to the observed information matrix:

$$n \begin{pmatrix} \frac{1}{\eta^2} & 0 & \frac{4g(0)\delta}{\eta^2} \\ 0 & \frac{2}{\eta^2} & 0 \\ \frac{4g(0)\delta}{\eta^2} & 0 & 2(2g^2(0) - g'(0))\kappa \end{pmatrix}$$

which is always non-singular, except when  $g$  is not differentiable at the origin. This result ensures the regularity conditions necessary to obtain the asymptotic distribution of the MLE for  $\theta$ . It should be noted that this condition was not met with the distribution  $2\eta^{-1}\phi((x - \xi)/\eta)G(\lambda(x - \xi)/\eta)$  studied by Azzalini and others.  $\square$

**Remark 1.** The functions discussed in Proposition 6 are crucial for addressing the singularity issue in statistical models. They are designed to ensure the non-singularity of the observed information matrix, which is essential for accurate parameter estimation and model performance.

The function introduced in this research paper, denoted as  $w(z)$ , is defined as  $w(z) = \lambda z^2$  for  $z > 0$  and  $w(z) = -\lambda z^2$  for  $z < 0$ . This function is equivalent to  $w(z) = \lambda \text{sign}(z)z^2 \equiv \lambda z|z|$ , where  $\text{sign}(\cdot)$  denotes the sign function.

The significance of this function lies in its ability to introduce bimodality into the model while avoiding singularity issues at  $\lambda = 0$ . By incorporating bimodality, the model can more accurately represent complex data distributions, particularly in fields like medical research.

The proposed  $w$  function ensures that the model satisfies the regular conditions required for deriving the asymptotic distribution of the parameter vector. It effectively overcomes the singularity issue at  $\lambda = 0$ , a significant limitation in previous models.

These functions provide a robust solution to the singularity problem by maintaining the non-singularity of the observed information matrix except when  $g$  is not differentiable at the origin.

By resolving the singularity issue, these functions enhance the model's reliability and accuracy in estimating parameters, making it a valuable tool for analyzing complex datasets, such as the protein data from cancer cells studied in this research paper.



Overall, these functions not only address the singularity problem but also contribute to the model's capability to handle bimodal data effectively, showcasing their significance in statistical modeling and data analysis.

### 3. Bimodal Skew-Normal Model

In this section, we provide a detailed description of the BSN distribution and investigate some of its key properties. To evaluate the performance of the resulting estimate, we conduct a simulation study that examines the basic inference obtained through the maximum likelihood approach.

#### 3.1. Shape Case

In the shape case, we can derive the probability density function using Lemma 1 with  $f = G' = \phi$ , which yields

$$f_Z(z; \lambda) = 2\phi(z)\Phi(\lambda z|z|), \quad z, \lambda \in \mathbb{R}, \quad (9)$$

where  $\phi$  and  $\Phi$  are the PDF and CDF of the standard normal distribution, respectively. The PDF (9) can be represented as a composition of two functions, that is,

$$f_Z(z; \lambda) = \begin{cases} 2\phi(z)(1 - \Phi(\lambda z^2)), & z < 0, \\ 2\phi(z)\Phi(\lambda z^2), & z \geq 0, \end{cases} \quad (10)$$

where  $\lambda \in \mathbb{R}$ .

**Proposition 7.** If  $Z \sim \text{BSN}(\lambda)$ , then the CDF is given by

$$F_Z(z; \lambda) = \begin{cases} 2\Phi(z) - c_\lambda^{-1}\Phi_\lambda(z), & z < 0, \\ c_\lambda^{-1}(\Phi_\lambda(0) + \int_0^z \phi_\lambda(t)dt), & z \geq 0, \end{cases} \quad (11)$$

where

$$\Phi_\lambda(z) = \int_{-\infty}^z \phi_\lambda(t)dt, \quad \phi_\lambda(t) = 2c_\lambda\phi(t)\Phi(\lambda t^2) \quad \text{and} \quad c_\lambda^{-1} = \int_{-\infty}^{\infty} \phi_\lambda(t)dt.$$

**Proof.** For  $z < 0$ ,

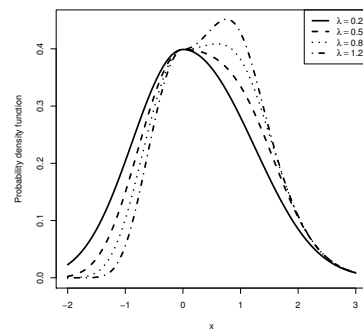
$$\begin{aligned} F_Z(z; \lambda) &= \int_{-\infty}^z 2\phi(t)dt - \int_{-\infty}^z 2\phi(t)\Phi(\lambda t^2)dt \\ &= 2\Phi(z) - c_\lambda^{-1} \int_{-\infty}^z \phi_\lambda(t)dt \\ &= 2\Phi(z) - c_\lambda^{-1}\Phi_\lambda(z). \end{aligned}$$

and for  $z \geq 0$ ,

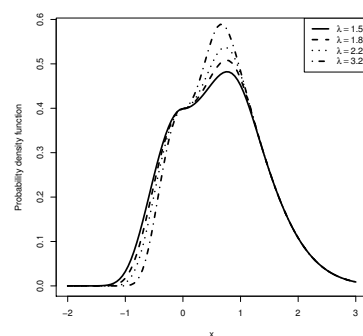
$$\begin{aligned} F_Z(z; \lambda) &= \int_{-\infty}^0 2\phi(t)\Phi(\lambda t^2)dt - \int_0^z 2\phi(t)\Phi(\lambda t^2)dt \\ &= c_\lambda^{-1}\Phi_\lambda(0) - c_\lambda^{-1} \int_0^z \phi_\lambda(t)dt \\ &= c_\lambda^{-1}(\Phi_\lambda(0) + \int_0^z \phi_\lambda(t)dt). \end{aligned}$$

□

Figures 1 and 2 show different densities that can be derived from Equation (9).



**Figure 1.** Plot of density function of BSN for different values of  $\lambda$ .



**Figure 2.** Plot of density function of BSN for different values of  $\lambda$ .

The BSN distribution, denoted by  $Z \sim \text{BSN}(\lambda)$ , is a distribution that is gaining attention as a valid competitor to the skew-normal model (SN) [1]. This is because both models control the degree of skewness using the same scalar parameter,  $\lambda \in \mathbb{R}$ . If  $\lambda = 0$ , then the model is reduced to the symmetric normal model. However, as demonstrated in this study, a significant advantage of the BSN model over the SN model is that in the presence of the location parameter, the BSN information matrix is non-singular at  $\lambda = 0$ . Consequently, under the null hypothesis of normality given by  $H_0 : \lambda = 0$ , the conventional regularity conditions leading to the ordinary asymptotic normal distribution of the MLE hold.

### 3.2. Some Basic Properties

Proposition 2 yields several important properties of the BSN distribution. First, if  $Z \sim \text{BSN}(\lambda)$ , then  $-Z \sim \text{BSN}(-\lambda)$ . This means that the distribution is symmetric with respect to the origin. Second,  $Z^2 \sim \chi_1^2$ , which indicates that the distribution of the squared BSN variable follows a chi-squared distribution with one degree of freedom. Finally,  $|Z| \sim \text{HN}(0,1)$ , where  $\text{HN}$  denotes the half-normal distribution. This property implies that the absolute value of a BSN variable follows an HN distribution with mean zero and variance one. These properties are useful for analyzing and interpreting data modeled with the BSN distribution.

#### 3.2.1. Quantile Function

The quantile function of the BSN distribution is derived by inverting Equation (11), as follows:

$$Q_p = F_Z^{-1}(p; \lambda), \quad 0 < p < 1. \quad (12)$$

This inverse does not have a closed expression and has to be calculated using some suitable numerical method. Note that  $Q_{0.5}$ ,  $Q_{0.25}$ , and  $Q_{0.75}$  stand for median, first quartile, and third quartile of the BSN distribution, correspondingly.

### 3.2.2. Moments, Skewness, and Kurtosis

The even moments of  $Z \sim \text{BSN}(\lambda)$  are equal to the corresponding even moments of a standardized normal random variable. The odd moments can be computed from the result in Proposition 5, as follows:

$$\mathbb{E}(Z^{2k+1}) = 2\mathbb{E}(Y^{2k+1}\Phi(\lambda Y^2)) - \frac{2^{1/2+k}}{\sqrt{\pi}}\Gamma(k+1), \quad k \in \{0, 1, 2, \dots\}, \quad (13)$$

where  $Y \sim \text{HN}(0, 1)$ . In particular, the first four moments are

$$\begin{aligned} \mathbb{E}(Z) &= \frac{2}{\pi}\rho(\lambda), \quad \mathbb{E}(Z^2) = 1, \\ \mathbb{E}(Z^3) &= \frac{2}{\pi\lambda^2}\left[\lambda + \frac{4\lambda^2 - 1}{2}\rho(\lambda)\right], \quad \mathbb{E}(Z^4) = 3, \end{aligned}$$

where

$$\rho(\lambda) = \frac{\text{sgn}(\lambda)}{\phi\left(\frac{1}{2|\lambda|}\right)}\left(1 - \Phi\left(\frac{1}{2|\lambda|}\right)\right).$$

By using the moments above and the standard definitions, we can calculate the skewness and kurtosis of the BSN distribution directly.

### 3.2.3. Entropy

A measure of the uncertainty's variation is the entropy of a random variable  $Z$  with a certain PDF. Greater data uncertainty is indicated by a high entropy value. The Rényi entropy [10],  $R_\alpha(Z)$ , for  $Z$  is defined as

$$R_\alpha(Z) = \frac{1}{1-\alpha} \log\left\{\int_{\mathbb{R}} f_Z^\alpha(z) dz\right\}, \quad (14)$$

where  $\alpha > 0$  and  $\alpha \neq 1$ . Suppose  $Z$  has the BSN distribution; then, by substituting (9) in (14), we obtain

$$\begin{aligned} \int_{\mathbb{R}} f_Z^\alpha(z) dz &= \int_{-\infty}^0 2^\alpha \phi^\alpha(z) \Phi^\alpha(-\lambda z^2) dz + \int_0^\infty 2^\alpha \phi^\alpha(z) \Phi^\alpha(\lambda z^2) dz \\ &= \int_0^\infty 2^\alpha \phi^\alpha(-z) \Phi^\alpha(-\lambda z^2) dz + \int_0^\infty 2^\alpha \phi^\alpha(z) \Phi^\alpha(\lambda z^2) dz \\ &= 2^\alpha \int_0^\infty \phi^\alpha(z) (\Phi^\alpha(-\lambda z^2) + \Phi^\alpha(\lambda z^2)) dz \end{aligned}$$

So, one obtains the Rényi entropy as follows:

$$R_\alpha(Z) = \frac{\alpha \log(2)}{1-\alpha} + \frac{1}{1-\alpha} \log\left\{\int_0^\infty \phi^\alpha(z) (\Phi^\alpha(-\lambda z^2) + \Phi^\alpha(\lambda z^2)) dz\right\}.$$

Shannon entropy [11] defined by  $S_\alpha(Z) = \mathbb{E}\{-\log(f_Z(z))\}$  is the particular case of Equation (14) when  $\alpha \rightarrow 1^+$ . Both  $R_\alpha(Z)$  and  $S_\alpha(Z)$  do not have closed expressions and must be calculated using some suitable numerical method.

### 3.2.4. Order Statistics

Let  $Z \sim \text{BSN}(\lambda)$  and  $Z_1, Z_2, \dots, Z_n$  be a random sample of independent and identically distributed variables with CDF  $F_Z$  given in (11) and PDF  $f_Z$  given in (9). Define the random variable  $Z_{(1)} = \min\{Z_1, Z_2, \dots, Z_n\}$ . It is known that the CDF of the sample minimum is given by

$$F_{Z_{(1)}}(y) = 1 - (1 - F_Z(y))^n$$

and its PDF is

$$f_{Z_{(1)}}(y) = n(1 - F_Z(y))^{n-1}f_Z(y).$$

On the other hand, define the random variable  $Z_{(n)} = \max\{Z_1, Z_2, \dots, Z_n\}$ . It is known that the CDF of the sample maximum is given by

$$F_{Z_{(n)}}(y) = (F_Z(y))^n$$

and its PDF is

$$f_{Z_{(n)}}(y) = n(F_Z(y))^{n-1}f_Z(y).$$

In general, the PDF of the  $k$ -th-order statistic from a random sample of size  $n$  drawn from the distribution of  $Z$  is

$$f_{Z_{(k)}}(y) = \frac{n!}{(k-1)!(n-k)!} (F_Z(y))^{k-1} (1 - F_Z(y))^{n-k} f_Z(y).$$

### 3.2.5. Maximum Likelihood Estimates for $\lambda$

The log-likelihood function for  $\lambda$  can be defined for a random sample of size  $n$  from  $Z \sim \text{BSN}(\lambda)$ , as given below:

$$l(\lambda, \mathbf{z}) = \frac{n}{2} \log\left(\frac{2}{\pi}\right) - \frac{1}{2} \sum_{i=1}^n z_i^2 + \sum_{i=1}^n \log(\Phi(\lambda z_i |z_i|)), \quad (15)$$

where  $\mathbf{z}$  denotes the sample data. This log-likelihood function helps to estimate the parameter  $\lambda$  of the BSN distribution. The likelihood equation induced from this function is given by

$$\frac{dl(\lambda, \mathbf{z})}{d\lambda} = \sum_{i=1}^n z_i |z_i| \frac{\phi(\lambda z_i |z_i|)}{\Phi(\lambda z_i |z_i|)} = 0. \quad (16)$$

The solution to this likelihood equation provides the MLE of  $\lambda$  for the bimodal skew-normal distribution. The numerical values of  $\hat{\lambda}$  can be determined via any statistical software.

### 3.3. Simulation Study

In this section, we evaluate the effectiveness of the MLE method for estimating the parameter  $\lambda$  in the BSN distribution. We generate random samples of various sizes:  $n = 50, 100, 150,$  and  $200,$  keeping  $\lambda$  fixed.

#### 3.3.1. Generating Samples from $\text{BSN}(\lambda)$

To generate random samples, follow these steps:

- (i) Set the parameter  $\lambda$  and choose the sample size  $n$ .
- (ii) Generate a standard normal random variable  $U \sim N(0, 1)$ .
- (iii) Compute  $Y = |U|$ .
- (iv) Generate a Bernoulli random variable  $S$  with success probability  $p = \Phi(\lambda Y^2)$ .
- (v) Set  $Z = Y$  if  $S = 1$ , otherwise  $Z = -Y$ .

#### 3.3.2. Analyzing $\lambda$ with MLE

After obtaining the samples,  $\lambda$  is estimated using the MLE method. We assess the performance of this estimation by calculating the standard error (SE), bias, and mean square error (MSE) in R programming language [12]:

- (i) For a chosen  $\lambda$ , simulate a sample of size  $n$  as described.
- (ii) Estimate  $\lambda$  using MLE.
- (iii) Repeat the above steps 1000 times.
- (iv) Compute the mean, SE, bias, and MSE of these 1000 estimates:

The mean, SE, bias, and MSE of  $\hat{\lambda}$  are given by

$$\begin{aligned}\bar{\lambda} &= \frac{1}{1000} \sum_{i=1}^{1000} \hat{\lambda}_i, \\ \text{SE}_{\bar{\lambda}} &= \sqrt{\frac{1}{1000} \sum_{i=1}^{1000} (\hat{\lambda}_i - \bar{\lambda})^2}, \\ \text{bias}_{\bar{\lambda}} &= \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\lambda}_i - \lambda)\end{aligned}$$

and

$$\text{MSE}_{\bar{\lambda}} = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\lambda}_i - \lambda)^2,$$

respectively. Here,  $\hat{\lambda}$  represents the MLE of  $\lambda$  for the  $i$ th iteration under a specific sample size  $n$ ,  $\bar{\lambda}$  corresponds to the mean of the parameter estimates obtained, for example,  $\hat{\lambda}_i$ 's, and  $\lambda$  denotes the actual value of the parameter.

Table 1 displays the mean estimates, SEs, biases, and MSEs of  $\lambda$  for various sample sizes. The performance of the MLE was investigated for a wide range of initial values for  $\lambda$ , and in all cases, the MLE converged well. The results in Table 1 were obtained by setting the initial value of  $\lambda$  to 1.00, irrespective of its actual value. Thus, any choice of initial value for  $\lambda$  is expected to yield similar results, as shown in Table 1.

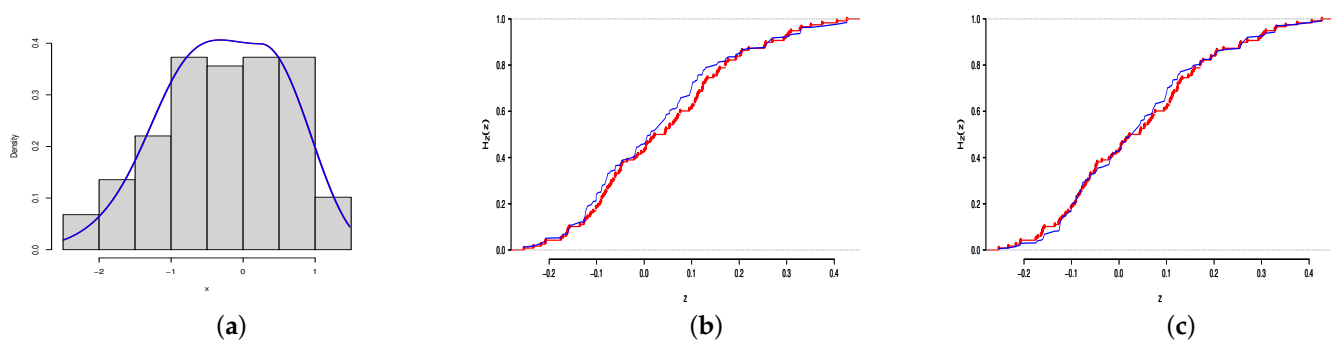
**Table 1.** Simulation results.

Actual Value		Estimates			
$\lambda$	$n$	$\bar{\lambda}$	$\text{SE}_{\bar{\lambda}}$	$\text{bias}_{\bar{\lambda}}$	$\text{MSE}_{\bar{\lambda}}$
−0.45	50	−0.5181	0.2302	−0.0681	0.0576
	100	−0.4867	0.1460	−0.0367	0.0226
	150	−0.4736	0.1114	−0.0236	0.0130
	200	−0.4658	0.0951	−0.0158	0.0093
−3.00	50	−3.7459	2.3316	−0.7459	5.9874
	100	−3.2822	0.9619	−0.2822	1.0040
	150	−3.1787	0.7434	−0.1787	0.5841
	200	−3.1211	0.5792	−0.1211	0.3497
2.00	50	2.4008	1.2510	0.4008	1.7242
	100	2.1520	0.6375	0.1520	0.4291
	150	2.1191	0.4874	0.1191	0.2515
	200	2.0999	0.3808	0.0999	0.1548
1.30	50	1.4952	0.6503	0.1952	0.4606
	100	1.3923	0.3758	0.0923	0.1496
	150	1.3534	0.2686	0.0534	0.0749
	200	1.3426	0.2246	0.0426	0.0522
0.00	50	$6 \times 10^{-3}$	0.1333	$6 \times 10^{-3}$	0.0178
	100	$-1 \times 10^{-3}$	0.0832	$-1 \times 10^{-3}$	0.0069
	150	$-1 \times 10^{-3}$	0.0657	$-1 \times 10^{-3}$	0.0043
	200	$9 \times 10^{-4}$	0.0544	$9 \times 10^{-4}$	0.0030

Overall, Table 1 reveals that as  $n$  increases, the SEs, biases, and MSEs decrease, indicating that the MLE provides consistent estimates.

#### 4. Practical Data Analysis

In this section, we illustrate the modeling capabilities of the BSN distribution by fitting it to 118 observations of the Homo sapiens PIG7 data in Çankaya [13] using the MLE method. To ensure computational stability, we scaled the data by  $-0.171$  before fitting the distributions. The data are left-skewed with the Pearson's moment coefficient of skewness of  $-0.356576$  and appear to be bimodal, as seen in the empirical density plot in Figure 3a. Some of the descriptive statistics of the data include a minimum value of  $-2.4971$ , a maximum value of  $1.4795$ , a mean value of  $-0.2512$ , a variance value of  $0.8162$ , a median value of  $-0.1930$ , a first quartile value of  $-0.8348$ , and a third quartile value of  $0.4678$ . The resulting MLE of  $\lambda$  is  $-0.3676782$ , with a corresponding SE of  $0.1149803$ . To assess the goodness-of-fit of the BSN distribution to the empirical data, we employed the Kolmogorov–Smirnov (K-S) test, with a test statistic defined as  $D_n = \max_{1 \leq i \leq k} (i/k - z_i, z_i - (i-1)/k)$ , where  $z_i$  is the  $i$ th data value. For large sample size  $n$ , the  $p$ -value of the K-S test is given by  $P(\sqrt{n} D_n \leq x) \approx \frac{\sqrt{2\pi}}{\hat{F}(z)} \sum_{k=1}^{\infty} \exp\left\{-\frac{(2k-1)^2 \pi^2}{8[\hat{F}(z)]^2}\right\}$ , where  $\hat{F}(z)$  is the estimated CDF of the theoretical distribution, see Kolmogorov [14] and Smirnov [15]. The K-S test measures the disparity between the empirical and estimated cumulative distribution functions (CDFs), with a smaller difference indicating a better fit. In general, if the  $p$ -value of the K-S test is greater than  $0.05$ , we conclude that the model provides a good fit for the data. The fitted BSN distribution gives a K-S statistic of  $0.10706$  with a  $p$ -value of  $0.1337 (>0.05)$ . Therefore, based on this evidence and visual inspection through the plot of the CDFs in Figure 3b, we conclude that the one-parameter BSN distribution provides a good fit for the data.



**Figure 3.** The plot of the empirical (rectangular bars) and estimated (blue line) PDF (a), estimated CDF of the uncentered BSN distribution (blue line) with the empirical CDF (red line) (b), and estimated CDF of the centered BSN distribution (blue line) with the empirical CDF (red line) (c).

We compare the fit of the BSN distribution with that of four other distributions, namely, the normal distribution, double Lindley distribution [16], Laplace distribution, and Student's  $t$ -distribution. The PDFs of these distributions are as follows:

- (i) Normal distribution with PDF given by

$$f_X(x; \xi, \eta) = \frac{1}{\eta} \phi\left(\frac{x - \xi}{\eta}\right), \quad x, \xi \in \mathbb{R}, \eta > 0,$$

where  $\phi(x)$  is the standard normal PDF.

- (ii) Double Lindley distribution with PDF given by

$$f_X(x; \lambda) = \frac{\lambda^2(1 + |x|)}{2(\lambda + 1)} \exp\{-\lambda|x|\}, \quad x \in \mathbb{R}, \lambda > 0.$$

- (iii) Laplace distribution with PDF given by

$$f_X(x; \xi, \lambda) = \frac{1}{2\lambda} \exp\left\{-\frac{|x - \xi|}{\lambda}\right\}, \quad x, \xi \in \mathbb{R}, \lambda > 0.$$

(iv) Student's  $t$ -distribution with PDF given by

$$f_X(x; \lambda) = \frac{\left(1 + \frac{x^2}{\lambda}\right)^{-\frac{\lambda+1}{2}}}{\sqrt{\lambda} B\left(\frac{1}{2}, \frac{\lambda}{2}\right)}, \quad x \in \mathbb{R}, \lambda > 0,$$

where  $B(a, b)$  is the beta function.

We used the K-S test to compare the goodness-of-fit of these distributions with that of the BSN distribution. The K-S test statistic measures the maximum distance between the empirical CDF of the data and the CDF of the fitted distribution, with a smaller test statistic indicating a better fit. The  $p$ -values of the K-S tests for each distribution were computed, and if the  $p$ -value was larger than 0.05, we concluded that the distribution provided a good fit for the data.

To ensure a fair comparison between the fits of the normal distribution, Laplace distribution, and BSN distribution, it is important to center the BSN distribution about the mean ( $\xi$ ), as both the normal and Laplace distributions are centered around the mean. To accomplish this, we introduce an additional parameter  $\xi$  to the BSN distribution, resulting in a centered BSN distribution with PDF  $2\phi(x - \xi)\Phi(\lambda(x - \xi)|x - \xi|)$  for  $x, \xi, \lambda \in \mathbb{R}$ . To determine the best-fitting model for the data, we use the information criteria listed below along with the K-S test:

- (i) Akaike information criterion (AIC), given by  $AIC = -2l(\hat{\theta}) + 2k$ .
- (ii) Bayesian information criterion (BIC), given by  $BIC = -2l(\hat{\theta}) + k \log(n)$ .
- (iii) corrected AIC (AICc), given by  $AICc = AIC + \frac{2k(k+1)}{n-k-1}$ .

Here,  $l(\hat{\theta})$  denotes the estimated log-likelihood value,  $n$  represents the number of data points,  $\theta$  is the unknown parameter, and  $k$  indicates the number of parameters in the model. A smaller value of the information criterion indicates a better fit. Tables 2 and 3 present the results of the fitted distributions. Based on these tables, we can observe that the centered BSN distribution outperformed all other considered distributions, with the smallest K-S statistic, largest  $p$ -value of K-S, and smallest values of AIC, BIC, and AICc. This is also evident from Figure 3c, where we can see that the CDF of the estimated centered BSN distribution closely mimics the empirical CDF.

**Table 2.** Fit results for different distributions.

Distribution	Parameter Estimate [SE]			K-S Test	
	$\hat{\lambda}$	$\hat{\xi}$	$\hat{\eta}$	$D$	$p$ -Value
Normal	-	-0.2512 [0.0828]	0.8996 [0.0586]	0.0998	0.1903
BSN	-0.7720 [0.3029]	0.2444 [0.1130]	-	0.0837	0.3801
Double Lindley	1.7862 [0.1304]	-	-	0.1480	0.0114
Laplace	-	-0.1293 [0.0331]	0.7521 [0.0692]	0.1170	0.0791
Student's $t$	235,670.6 [104.4868]	-	-	0.1187	0.0721



**Table 3.** Model fit discrimination.

Distribution	$-l(\hat{\lambda})$	Information Criteria		
		AIC	BIC	AICc
Normal	154.9475	313.8951	319.4365	313.9994
BSN	151.2718	306.5435	312.0849	306.6479
Double Lindley	164.4999	330.9998	333.7704	331.0342
Laplace	166.1680	336.3361	341.8774	336.4404
Student's $t$	159.9040	321.8079	324.5786	321.8424

In Table 4, descriptive statistics obtained from the empirical distribution and the estimated centered BSN distribution are compared. From the results, we can conclude that the fitted centered BSN distribution accurately captured the important features of the empirical distribution, as the first three moments and the standard deviation (std) of the estimated centered BSN distribution are similar to those of the empirical distribution. It is noteworthy that the direction of skewness is the same for both distributions. However, a slight difference in skewness values is observed, which may be due to rounding errors in the numerical integration of the  $k$ -th-order moments. The R code used to compute the descriptive statistics is provided in Appendix A.

**Table 4.** Some descriptive statistics for the empirical and centered BSN distributions.

	Moments			Std	Skewness
	$\mu'_1$	$\mu'_2$	$\mu'_3$		
Empirical	−0.2512142	0.8723585	−0.8886568	0.903419	−0.356576
BSN	−0.2645494	0.8109423	−0.7589403	0.860788	−0.238893

## 5. Concluding Remarks

In this study, we introduced a new family of continuous distributions known as bimodal skew-symmetric distributions. The BSN distribution, which is essential to this family, is distinguished by the single parameter that causes its asymmetry. The statistical properties of this distribution have been thoroughly discussed, emphasizing its flexibility and applicability.

Utilizing the MLE method, we estimated this sole asymmetry parameter, demonstrating the practicality and effectiveness of the BSN model when applied to real-world data. The analysis highlights the BSN distribution's capability to adeptly model data features, such as skewness and bimodality, which are often encountered in practical datasets but are challenging to address with more traditional models.

To enhance the utility of the BSN distribution and facilitate its comparison with more conventional distributions like the normal and Laplace distributions, both of which are two-parameter models centered about the mean, we plan to extend the BSN distribution by centering it about the mean in future applications. This adjustment will allow the BSN distribution to be directly comparable to these models, providing a fair basis for performance evaluation.

The results from this study are promising, showing that the two-parameter BSN distribution not only meets but exceeds the performance of the four considered competing distributions for the dataset in question. This superior performance underscores the potential of the BSN distribution as a robust and versatile tool in statistical modeling, particularly suitable for complex real-world data that exhibit asymmetry and bimodality.

This study contributes to an application of bimodal skew-symmetric distributions to the analysis of cancer cell protein data, addressing the inherent bimodality and asymmetry of such data. The proposed model enhances the flexibility and accuracy of statistical representations, leading to improved parameter estimation and robust analysis even in the presence of noise. By incorporating regularization techniques to prevent singularity

issues and leveraging the model's adaptability to capture complex biological variability, this research provides an effective tool for identifying subpopulations and characterizing protein profiles in cancer cells. These contributions not only advance the field of statistical modeling in bioinformatics but also have practical implications for biomarker discovery and proteomics analysis, paving the way for more precise and meaningful insights into cancer biology.

The implications of these findings are significant, suggesting that the BSN distribution can serve as an alternative to traditional models, offering enhanced flexibility and better fit for specific types of data. Future studies will focus on further developing this model, improving its statistical inference procedures, and extending its application to a broader range of datasets.

**Author Contributions:** Conceptualization, H.S.S., H.S.B. and I.E.O.; methodology, H.S.S., H.S.B. and I.E.O.; software, H.S.B., I.E.O. and G.A.; validation, H.S.S., H.S.B., I.E.O., and G.A.; writing—original draft preparation, H.S.S. and I.E.O.; writing—review and editing, H.S.S., H.S.B., I.E.O., G.A. and O.A.; visualization, H.S.S., H.S.B. and I.E.O.; funding acquisition, G.A. and O.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by King Faisal University, Saudi Arabia [GRANT KFU241133].

**Data Availability Statement:** Data are accessible from the authors upon request.

**Acknowledgments:** This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [GRANT KFU241133].

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

## Appendix A

### R Codes

#### $k$ –order ordinary moments for the centered BSN distribution

```
moments<-function(z,k)
{
lambda<--0.7720
xi<-0.2444
f<-function(z)z^k*2*dnorm(z-xi)*pnorm(lambda*(z-xi)*abs(z-xi))
out<-integrate(f,lower=-Inf,upper=Inf)$value
out
}
first<-moments(z,1)
second<-moments(z,2)
third<-moments(z,3)
variance<-second-first^2
std<-sqrt(var)
skew<-(third-3*first*vvariance-first^3)/std^3
cbind(first,second,third,variance,std,skew)
```

#### Simulation

```
sim<-function(lambda,nn,N)
{
m<-length(nn)
esthat<-0
sdhat<-0
bias<-0
mse<-0
for(n in 1:m)
```

```

{
loop<-function(nn,N)
{
t<-0
result<-matrix(0,N,1)
while(t<N)
{
t<-t+1
sim<-function(lambda,n)
{
t<-0
z<-0
while(t<n)
{
t<-t+1
u<-rnorm(1)
y<-abs(u)
p<-pnorm(lambda*y^2)
s<-rbinom(1,1,p)
if(s==1)
{
z[t]<-y
}else{
z[t]<--y
}
}
z
}
x<-sim(lambda,nn[n])
ff=function (q)
{
tt=1.0e20
lambda=q[1]
tt=-sum(log(2*dnorm(x)*pnorm(lambda*x*abs(x))))#MLE
if (is.na(tt)) tt=1.0e20
if (abs(tt)>1.0e20) tt=1.0e20
return(tt)
}
est=nlm(ff,p=c(1))
result[,1][t]=est$estimate[1]#lambda
}
result
}
est<-loop(nn,N)[,1]
esthat[n]=mean(est)
sdhat[n]=sd(est)
bias[n]=mean(est-lambda)
mse[n]=mean((est-lambda)**2)
}
data.frame(nn,esthat,sdhat,bias,mse)
}

```

## References

1. Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
2. Azzalini, A. Further results on a class of distributions which includes the normal ones. *Statistica* **1986**, *46*, 199–208.
3. Gupta, A.K.; Chang, F.C.; Huang, W.J. Some skew-symmetric models. *Random Oper. Stoch. Equ.* **2002**, *10*, 113–140. [[CrossRef](#)]
4. Ma, Y.; Genton, M.G. Flexible class of the skew-symmetric distributions. *Scand. J. Stat.* **2004**, *31*, 459–468. [[CrossRef](#)]
5. 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](#)]
6. Pewsey, A. Some observations on a simple means of generating skew distributions. In *Advances in Distribution Theory, Order Statistics, and Inference Part of the Series Statistics for Industry and Technology*; Springer: New York, NY, USA, 2006; pp. 75–84.
7. Bakouch, H.S.; Salinas, H.S.; Mamode Khan, N.; Chesneau, C. A new family of skewed distributions with application to some daily closing prices. *Comput. Math. Methods* **2021**, *3*, e1154. [[CrossRef](#)]
8. Salinas, H.; Bakouch, H.; Qarmalah, N.; Martínez-Flórez, G. A flexible class of two-piece normal distribution with a regression illustration to biaxial fatigue data. *Mathematics* **2023**, *11*, 1271. [[CrossRef](#)]
9. Khorsheed, E.; Salinas, H.S.; Bakouch, H.S. A new family of skew-normal lifetime distributions for industrial applications. In *Proceedings of the 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, Sakheer, Bahrain, 26–27 October 2020; pp. 1–4.
10. Rényi, A. On measures of information and entropy. In *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, Los Angeles, CA, USA, 20–30 June 1960*; Neymann, J., Ed.; University of California Press: Berkeley, CA, USA, 1961; pp. 547–561.
11. Shannon, C.E. A mathematical theory of communication. *Bell. Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
12. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2024. Available online: <https://www.R-project.org/> (accessed on 10 January 2024).
13. Çankaya, M.N. Asymmetric bimodal exponential power distribution on the real line. *Entropy* **2018**, *20*, 23. [[CrossRef](#)] [[PubMed](#)]
14. An, K. Sulla determinazione empirica di una legge didistribuzione. *Giorn. Dell'inst. Ital. Degli Att.* **1933**, *4*, 89–91.
15. Smirnov, N. Table for estimating the goodness of fit of empirical distributions. *Ann. Math. Stat.* **1948**, *19*, 279–281. [[CrossRef](#)]
16. Nitha, K.; Krishnarani, S. A new family of heavy tailed symmetric distribution for modeling financial data. *J. Stat. Appl. Probab.* **2017**, *6*, 577–586.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.