Bivariate Power-Skew-Elliptical Distribution

Guillermo Martínez-Flórez [1,2,†], Roger Tovar-Falón [1,†] and Héctor W. Gómez [3,*,†]

[1] Departamento de Matemáticas y Estadística, Facultad de Ciencias Básicas, Universidad de Córdoba, Montería 230027, Colombia; guillermomartinez@correo.unicordoba.edu.co (G.M.-F.); rjtovar@correo.unicordoba.edu.co (R.T.-F.)

[2] Programa de Pós-Graduação em Modelagem e Métodos Quantitativos, Universidade Federal de Ceará, Fortaleza 60020-181, Brazil

[3] Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta 1240000, Chile

* Correspondence: hector.gomez@uantof.cl

† These authors contributed equally to this work.

check for
updates

Abstract: In this article, we introduce a power-skew-elliptical (PSE) distribution in the bivariate setting. The new bivariate model arises in the context of conditionally specified distributions. The proposed bivariate model is an absolutely continuous distribution whose marginals are univariate PSE distributions. The special case of the bivariate power-skew-normal (BPSN) distribution is studied in details. General properties of the BPSN distribution are derived and the estimation of the unknown parameters by maximum pseudo-likelihood is discussed. Further, a sandwich type matrix, which is a consistent estimator for the asymptotic covariance matrix of the maximum likelihood (ML) estimator is determined. Two applications for real data of the proposed bivariate distribution is provided for illustrative purposes.

1. Introduction

Family of distributions that unify the main characteristics of other families are not very common in distributions theory. In this sense, an asymmetric family widely used in many situations and different areas of knowledge is the skew-normal (SN) distribution of [1], that is characterized because it has a wide range of asymmetry. Another well-known family in this same area is the fractional order statistical distribution, also known as exponentiated distribution or power-normal distribution introduced by [2], which is characterized by having a wide range of kurtosis. The unification of these two families was studied by [3] and was called power-skew-normal distribution. The resulting family contains, such as special cases the normal, skew-normal and power-normal distributions, and the ranges of asymmetry and kurtosis are greater than any of these three families of distributions. This type of unification of distribution families is important in the literature of distribution theory because it introduces great flexibility to the resulting models.

This model was extended to the case of elliptical distributions and was considered the case of the Birnbaum-Saunders (BS) family by [4], resulting in a large unification of useful distributions for example, to model the lifetime in survival analysis or applications in reliability theory. The model is called Birnbaum-Saunders power skew elliptical (BSPSE), and it contains as special cases as a large number of extensions of the lifetime Birnbaum-Saunders model, for both elliptical distributions and skew-elliptical families. In the univariate case, this type of results suggests that multivariate

distributions constructed of unifications of distributions, it will induce multivariate models with great flexibility.

In this paper, we study a new family of bivariate distributions whose conditional densities follow a power-skew-elliptical distribution, which extends the power-skew normal model to the case of bivariate distributions, becoming a new alternative to the existing asymmetric bivariate models in literature such as the bivariate skew-normal distribution by [5] and the conditionally specified bivariate skewed distribution of [6].

A brief description of the some elliptical distributions is presented below.

1.1. Elliptical Distributions

A continuous one-dimensional random variable is said to have an elliptical distribution, if its distribution function is symmetric with support in the real number set. Specifically, a random variable X has a symmetric distribution if its probability density function (PDF) is given by

$$f(x) = \frac{c}{\eta} g\left(z^2\right), \tag{1}$$

for some non-negative function $g(u)$, with $u > 0$ and corresponds to the kernel of the PDF such that $\int_0^\infty u^{-\frac{1}{2}} g(u) du = 1/c$, where $z = (x - \xi)/\eta$ and c is a normalizing constant. The function $g(\cdot)$ is known as the density-generating function. An elliptically distributed random variable X with location and scale parameters ξ and η, respectively, and density-generating function, say g is denoted by $X \sim \mathrm{EC}(\xi, \eta; g)$. If $\xi = 0$ and $\eta = 1$, then X has spherical distribution, which is denoted as $X \sim \mathrm{EC}(0, 1; g)$.

Properties of this family have been studied in [7–11], among others. Particular cases of the $X \sim \mathrm{EC}(0, 1; g)$ distribution are the Pearson type VII distribution, the type Kotz distribution, the Student-t distribution with v degrees of freedom, the Cauchy distribution and the normal distribution, among others. The density-generating function of the generalized normal, Cauchy, Student-t, type I logistic, type II logistic and power exponential are, respectively, given by $g(u) = (2\pi)^{-1/2} \exp(-u/2)$, $g(u) = \{\pi(1+u)\}^{-1}$, $g(u) = v^{v/2} B(1/2, v/2)^{-1} (v+u)^{-(v+1)/2}$, where $v > 0$ and $B(\cdot, \cdot)$ is the beta function, $g(u) = c \exp(-u)(1 + \exp(-u))^{-2}$, where $c \approx 1.484300029$ is the normalizing constant obtained from $\int_0^\infty u^{-1/2} g(u) du = 1$, $g(u) = \exp(-\sqrt{u})(1 + \exp(-\sqrt{u}))^{-2}$ and $g(u) = c(k) \exp(-\frac{1}{2} u^{1/(1+k)})$, $-1 < k \leq 1$, where $c(k) = \Gamma(1 + (k+1)/2) 2^{1+(1+k)/2}$.

1.2. Skew-Elliptical Distribution

An extension of the elliptical model to the asymmetric case is the standard elliptical asymmetric (skew-elliptical) model defined as

$$h_Y(y; \lambda) = 2f(y)\mathcal{F}(\lambda y); \quad y, \lambda \in \mathbb{R}, \tag{2}$$

where $f(\cdot)$ is given in Equation (1), $\mathcal{F}(\cdot)$ is its respective cumulative distribution function (CDF), and λ is an asymmetry parameter. We use the notation $Y \sim \mathrm{SE}(0, 1; g, \lambda)$. The CDF for this model is given by

$$\mathcal{H}_Y(y) = 2 \int_{-\infty}^y f(t)\mathcal{F}(\lambda t) dt. \tag{3}$$

Skew-elliptical distributions are discussed in [12–16], among others.

To follow we present some distributions belonging to this family.

1.2.1. Skew-Normal Distribution

A particular case of model in Equation (3) is the SN distribution introduced by [1], which is obtained by letting $f(\cdot) = \phi(\cdot)$ and $\mathcal{F}(\cdot) = \Phi(\cdot)$, that is, the PDF and CDF of the standard normal distribution. The PDF and CDF of the SN model are given by

$$h(x) = f_{SN}(x) = 2\phi(x)\Phi(\lambda x); \quad x \in \mathbb{R}, \tag{4}$$

and

$$\mathcal{H}(x) = \mathcal{F}_{SN}(x) = \Phi(x) - 2T(x, \lambda); \quad x \in \mathbb{R}, \tag{5}$$

respectively, where $T(\cdot, \cdot)$ is the Owen's function, see [17] for more details.

1.2.2. Skew-Student-*t* Distribution

The skew-Student-*t* (SST) distribution (or skew-Pearson type IV) has the PDF given by

$$h(x) = f_{SST}(x) = 2\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{2\pi}\Gamma\left(\frac{v}{2}\right)}\left(1 + \frac{x^2}{v}\right)^{-\frac{v+1}{2}}\left[\frac{1}{2} + \lambda x \Gamma\left(\frac{v+1}{2}\right)\frac{F_{2-1}\left(\frac{1}{2}, \frac{v+1}{2}, -\frac{3}{2}, -\frac{\lambda^2 x^2}{v}\right)}{\sqrt{v\pi}\Gamma\left(\frac{v}{2}\right)}\right], \tag{6}$$

where $v > 0$ represents the degrees of freedom and $F_{2-1}(\cdot, \cdot, \cdot, \cdot)$ is the hypergeometric function, see [18].

1.2.3. Skew-Cauchy Distribution

A continuous random variable X following skew-Cauchy (SC) distribution has PDF given by

$$h(x) = f_{SC}(x) = \frac{2}{\pi}\left(1 + x^2\right)^{-1}\left[\frac{1}{2} + \frac{1}{\pi}\arctan(\lambda x)\right]. \tag{7}$$

1.2.4. Skew-Logistic Distribution

The PDF of the skew-logistic (SLOG) distribution is given by

$$h(x) = f_{SLOG}(x) = \frac{2\exp(-x)}{(1 + \exp(-x))^2}\frac{1}{1 + \exp(-\lambda x)}. \tag{8}$$

1.2.5. Skew-Laplace Distribution

The skew-Laplace (SL) distribution has PDF given by

$$h(x) = f_{SL}(x) = \begin{cases} \frac{1}{2}\exp\big((1+\lambda)x\big), & \text{if } x < 0, \\ \exp(-x)\big(1 - \frac{1}{2}\exp(-\lambda x)\big), & \text{if } x \geq 0. \end{cases} \tag{9}$$

1.3. Power-Skew-Elliptical Distribution

An alternative to the SN distribution of [1], was studied by [2] by introducing the fractional order statistical model, also known as alpha-power (AP) model, which has PDF given by

$$f_{AP}(z; \alpha) = \alpha h(z)\{\mathcal{H}(z)\}^{\alpha-1}, \quad z \in \mathbb{R}, \tag{10}$$

where $\mathcal{H}(\cdot)$ is an absolutely continuous CDF with PDF $h(\cdot)$, and $\alpha > 0$ is a parameter that controls the distributional shape. The case $\mathcal{H}(\cdot) = \Phi(\cdot)$ is called the power-normal (PN) distribution and has PDF given by

$$f_{PN}(z; \alpha) = \alpha\phi(z)\{\Phi(z)\}^{\alpha-1}, \quad z \in \mathbb{R}, \tag{11}$$

The PN model is denoted by $Z \sim \mathrm{PN}(\alpha)$ and is considered an alternative distribution for modeling data with asymmetry and kurtosis above (or below) the expected for the normal distribution. If PDF $h(\cdot)$ in model (10) has the form as in Equation (2), then the model is called PSE and its PDF is given by

$$f_{\mathrm{PSE}}(z; \lambda, \alpha) = \alpha h(z; \lambda)\{\mathcal{H}(z; \lambda)\}^{\alpha-1}, \quad z \in \mathbb{R}. \tag{12}$$

We will use the notation $Z \sim \mathrm{PSE}(0, 1; g, \lambda, \alpha)$. The case of this family of distributions for $h(\cdot; \lambda) = f_{\mathrm{SN}}(\cdot, \lambda)$ and $\mathcal{H}(\cdot; \lambda) = \mathcal{F}_{\mathrm{SN}}(\cdot, \lambda)$ was studied by [3] and it is called power-skew-normal distribution which is denoted by $\mathrm{PSN}(\lambda, \alpha)$. Some contributions to this family have been made by [19–22], among other.

Some additional works on distributions include those of [23,24], which the possibility of applying the analytical expressions for the calculation of the correct detection probability of the signal time window at synchronization has been proved.

The rest of the paper is organized as follows: Section 2 we introduce the new bivariate power-skew-elliptical family of distributions, several properties are derived and we consider the ML method for estimating the model parameters. In Section 3, we study the particular case of the bivariate BPSN model. ML estimation of the model is discussed and a reparameterization of the BPSN model is presented and we derive the information matrix. In Section 4, two applications is presented illustrating the good performance of the approaches developed in the paper.

2. Bivariate Power-Skew-Elliptical Distribution

In this section, we extend the PSE model to the bivariate case, this new model is a great extension because it is the bivariate unification of two families of distributions, on the one hand, the skew-elliptical family and on the other hand, the alpha-power family. The unification will generate a distribution with great flexibility in both asymmetry and kurtosis.

For the construction of bivariate power-skew-elliptical (BPSE) family of distributions, we will use the approach discussed in [25] which is based on conditional distributions. According to [25] a two-dimensional random vector (X_1, X_2) is conditionally specified, if for any random variable X_2, the random variable $X_1 \mid X_2 = x_2$ is a member of a parametric family. Suppose that the joint PSE distribution function $\mathcal{H}_{\mathrm{BPSE}}(x_1, x_2)$, of the random vector (X_1, X_2) is such that, the conditional distribution of X_1 given $X_2 = x_2$ and the conditional distribution of X_2 given $X_1 = x_1$ are members of the PSE family of distributions with respect to a Lebesgue measure. We denote this by writing

$$X_1 \mid X_2 = x_2 \sim \mathrm{PSE}_1(\theta_1; g, \lambda_1, \underline{\omega}(x_2)) \tag{13}$$

and

$$X_2 \mid X_1 = x_1 \sim \mathrm{PSE}_2(\theta_2; g, \lambda_2, \underline{\tau}(x_1)), \tag{14}$$

where $\underline{\omega}, \underline{\tau}$ are positive dependence functions which are to be determined.

In such case, we have conditionals in a given exponential family and we can identify the corresponding joint density. We can argue as follows. If $h_{X_1}(x_1)$ and $h_{X_2}(x_2)$ are marginal densities for a joint PSE density $h_{\mathrm{BPSE}}(x_1, x_2)$ with conditional densities given by Equations (13) and (14), then it follows that

$$
\begin{aligned}
h_{\mathrm{BPSE}}(x_1, x_2) &= \tau(x_1) h_{X_1}(x_1) h_2(x_2; \lambda_2)\{\mathcal{H}_2(x_2; \lambda_2)\}^{\tau(x_1)-1}, \\
&= \omega(x_2) h_{X_2}(x_2) h_1(x_1; \lambda_1)\{\mathcal{H}_1(x_1; \lambda_1)\}^{\omega(x_2)-1}.
\end{aligned} \tag{15}
$$

Following [26] the solutions for dependence functions are given by

$$\omega(x_2) = \alpha_1 - \alpha_{12} \ln[\mathcal{H}_2(x_2; \lambda_2)] \tag{16}$$

and

$$\tau(x_1) = \alpha_2 - \alpha_{12} \ln[\mathcal{H}_1(x_1; \lambda_1)], \tag{17}$$

with α_1, α_2, positive real constants and $\alpha_{12} = \alpha_{21} \geq 0$.

Then, by using theorems in [27] and ([28], Chap. 2), we have that

$$
\begin{aligned}
h_{\mathrm{BPSE}}(x_1, x_2) = {} & c(\underline{\lambda}, \underline{\alpha}) h_1(x_1; \lambda_1) h_2(x_2; \lambda_2) \{\mathcal{H}_1(x_1; \lambda_1)\}^{\alpha_1 - 1} \{\mathcal{H}_2(x_2; \lambda_2)\}^{\alpha_2 - 1} \\
& \times \exp\{\alpha_{12} \ln[\mathcal{H}_1(x_1; \lambda_1)] \ln[\mathcal{H}_2(x_2; \lambda_2)]\},
\end{aligned} \tag{18}
$$

where $\lambda_1, \lambda_2 \in \mathbb{R}$, the constants $\alpha_1, \alpha_2 > 0$ and $\alpha_{12} \geq 0$ in order to guarantee $\int_{\mathbb{R}} \int_{\mathbb{R}} h(x_1, x_2) dx_1 dx_2 < \infty$, and $c(\underline{\lambda}, \underline{\alpha})$ is a normalizing constant with $\underline{\lambda} = (\lambda_1, \lambda_2)^{\top}$ and $\underline{\alpha} = (\alpha_1, \alpha_2, \alpha_{12})^{\top}$.

The independence case, where the joint distribution is the product of two PSE densities, is followed by taking $\alpha_{12} = 0$, with $c(\underline{\lambda}, \underline{\alpha}) = \alpha_1 \alpha_2$.

Different conditional bivariate skew-elliptical distributions can be obtained from the generating function $g(\cdot)$, such as presented above, skew-normal, skew-Cauchy, skew-Student-*t*, skew-logistic and skew-Laplace, among others. This new family, which we will denote by BPSE$_g$ generates a large number of bivariate distributions according to generating function $g(\cdot)$, in addition to those already mentioned, one could also talk about the distributions: special case, type Kotz, Bessel, and the many representations of the Pearson family of distributions different to the Cauchy and the Student-*t*. Thus, we define a broad flexible family of asymmetric bivariate distributions.

According to Equations (16) and (17), it follows that conditional distributions are given by

$$h_{X_1 | X_2}(x_1 \mid x_2) = \omega(x_2) h_1(x_1; \lambda_1) \{\mathcal{H}_1(x_1; \lambda_1)\}^{\alpha_1 - 1} \exp\{\alpha_{12} \ln[\mathcal{H}_1(x_1; \lambda_1) \ln[\mathcal{H}_2(x_2; \lambda_2)]]\}, \tag{19}$$

and

$$h_{X_2 | X_1}(x_2 \mid x_1) = \tau(x_1) h_2(x_2; \lambda_2) \{\mathcal{H}_2(x_2; \lambda_2)\}^{\alpha_2 - 1} \exp\{\alpha_{12} \ln[\mathcal{H}_1(x_1; \lambda_1) \ln[\mathcal{H}_2(x_2; \lambda_2)]]\}, \tag{20}$$

and hence, it follows that (19) and (20) belong to the exponentiated families of densities (13) and (14), where $h_i(\cdot)$ and $\mathcal{H}_i(\cdot)$, for $i = 1, 2$, are known density and distribution functions, respectively, while the marginal densities are given by

$$h_{X_1}(x_1) = c(\underline{\lambda}, \underline{\alpha}) \frac{h_1(x_1; \lambda_1) \{\mathcal{H}_1(x_1; \lambda_1)\}^{\alpha_1 - 1}}{\alpha_2 - \alpha_{12} \ln[\mathcal{H}_1(x_1; \lambda_1)]} \tag{21}$$

and

$$h_{X_2}(x_2) = c(\underline{\lambda}, \underline{\alpha}) \frac{h_2(x_2; \lambda_2) \{\mathcal{H}_2(x_2; \lambda_2)\}^{\alpha_2 - 1}}{\alpha_1 - \alpha_{12} \ln[\mathcal{H}_2(x_2; \lambda_2)]} \tag{22}$$

It follows that the CDFs of the conditioned PDFs given in Equations (19) and (20) are given by

$$\mathcal{H}_{X_1 | X_2}(x_1 \mid x_2) = \{\mathcal{H}_1(x_1; \lambda_1)\}^{\alpha_1 - \alpha_{12} \ln[\mathcal{H}_2(x_2; \lambda_2)]}, \tag{23}$$

and

$$\mathcal{H}_{X_2 | X_1}(x_2 \mid x_1) = \{\mathcal{H}_2(x_2; \lambda_2)\}^{\alpha_2 - \alpha_{12} \ln[\mathcal{H}_1(x_1; \lambda_1)]}. \tag{24}$$

The *r*th moment can be calculated by using

$$\mathbb{E}\left[X_1^r \mid X_2 = x_2\right] = (\alpha_1 - \alpha_{12} \ln[\mathcal{H}_2(x_2; \lambda_2)]) \int_0^1 [\mathcal{H}_1^{-1}(v_1; \lambda_1)]^r v_1^{\alpha_1 - 1 - \alpha_{12} \ln[\mathcal{H}_2(x_2; \lambda_2)]} dv_1,$$

$$\mathbb{E}\left[X_2^r \mid X_1 = x_1\right] = (\alpha_2 - \alpha_{12} \ln[\mathcal{H}_1(x_1; \lambda_1)]) \int_0^1 [\mathcal{H}_2^{-1}(v_2; \lambda_2)]^r v_2^{\alpha_2 - 1 - \alpha_{12} \ln[\mathcal{H}_1(x_1; \lambda_1)]} dv_2.$$

where $\mathcal{H}_i^{-1}(\cdot;\cdot)$ is the inverse function of the CDF $\mathcal{H}_i(\cdot;\cdot)$, $i = 1, 2$. To study the correlation between X_1 and X_2, we can compute the measure

$$\rho_{X_1 X_2} = \frac{\mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]}{\sigma_{X_1}\sigma_{X_2}} = \frac{\sigma_{X_1 X_2}}{\sigma_{X_1}\sigma_{X_2}},$$

where

$$\sigma_{X_1 X_2} = \int_0^1 \int_0^1 \mathcal{H}_1^{-1}(v_1;\lambda_1)\mathcal{H}_2^{-1}(v_2;\lambda_2) v_1^{\alpha_1 - 1} v_2^{\alpha_2 - 1}$$

$$\times \left[v_1^{-\alpha_{12}\ln(v_2)} - \frac{1}{(\alpha_2 - \alpha_{12}\ln(v_1))(\alpha_1 - \alpha_{12}\ln(v_2))} \right] dv_1 dv_2,$$

$$\sigma_{X_1}^2 = \int_0^1 \frac{[\mathcal{H}_1^{-1}(v_1;\lambda_1)]^2 v_1^{\alpha_1 - 1}}{\alpha_2 - \alpha_{12}\ln(v_1)} dv_1 - \left\{ \int_0^1 \frac{\mathcal{H}_1^{-1}(v_1;\lambda_1) v_1^{\alpha_1 - 1}}{\alpha_2 - \alpha_{12}\ln(v_1)} dv_1 \right\}^2,$$

$$\sigma_{X_2}^2 = \int_0^1 \frac{[\mathcal{H}_2^{-1}(v_2;\lambda_2)]^2 v_2^{\alpha_2 - 1}}{\alpha_1 - \alpha_{12}\ln(v_2)} dv_2 - \left\{ \int_0^1 \frac{\mathcal{H}_2^{-1}(v_2;\lambda_2) v_2^{\alpha_2 - 1}}{\alpha_1 - \alpha_{12}\ln(v_2)} dv_2 \right\}^2.$$

Moreover, we can consider

$$\widehat{\rho}_{X_1 X_2} = \frac{\widehat{\sigma}_{X_1 X_2}}{\widehat{\sigma}_{X_1}\widehat{\sigma}_{X_2}}.$$

as an estimator of ρ_{X_1, X_2}.

Statistical Inference for the Bpse Model

The normalization constant $c(\underline{\lambda}, \underline{\alpha})$ in the joint PDF $h(x_1, x_2)$ makes difficult the parameters estimation by maximizing the likelihood function. As alternative, we will follow the proposal of [29] for the parameters estimation of multivariate distributions, that is, we will maximize the pseudo-likelihood function. The pseudo-likelihood function is defined as the product of conditional functions. In this case, as in the ML method, the logarithm of the product of conditional distributions is maximized and this eliminates the logarithm of the normalization constant $c(\underline{\lambda}, \underline{\alpha})$ within the estimation process, which, as in this case will contain multiple integrals in its structure. Another important characteristic of this estimation process is that the maximum pseudo-likelihood estimators vector of the model parameters is consistent and converges asymptotically to a multivariate normal distribution.

Hence, given a random sample of vectors $(x_{11}, x_{21}), (x_{12}, x_{22}), \cdots, (x_{1n}, x_{2n})$, with bivariate joint distribution PSE, the pseudo-likelihood function based on the conditional densities of the BPSE distribution, is given by

$$L_p(\underline{\beta}) = h_{X_1|X_2}(x_1 \mid x_2) h_{X_2|X_1}(x_2 \mid x_1), \tag{25}$$

where $\underline{\beta} = (\underline{\theta_1}, \underline{\theta_2}, \lambda_1, \lambda_2, \alpha_1, \alpha_2, \alpha_{12})$, with $\underline{\theta_1}$ and $\underline{\theta_2}$ being the parameters of the $h_{X_1|X_2}(x_1 \mid x_2)$ and $h_{X_2|X_1}(x_2 \mid x_1)$ distributions, respectively. Thus, the maximum pseudo-likelihood estimator of $\underline{\beta}$ is defined as the value $\underline{\beta}_0$ of $\underline{\beta}$ which maximizes the pseudo-likelihood function.

The log-pseudo-likelihood function is defined as the logarithm of the pseudo-likelihood function and for BPSE model it is expressed by

$$
\begin{aligned}
\ell_P(\underline{\beta}) = {} & \sum_{i=1}^{n} \sum_{j=1}^{2} \ln\left(\alpha_j - \sum_{\substack{k=1 \\ k \neq j}}^{2} \alpha_{jk} \ln[\mathcal{H}_k(x_{ki}; \underline{\theta_k}, \lambda_k)]\right) + \sum_{i=1}^{n} \sum_{j=1}^{2} \ln\left[h_j(x_{ji}; \underline{\theta_j}, \lambda_j)\right] \\
& + \sum_{i=1}^{n} \sum_{j=1}^{2} (\alpha_j - 1) \ln\left[\mathcal{H}_j(x_{ji}; \underline{\theta_j}, \lambda_j)\right] \\
& + \sum_{i=1}^{n} \sum_{j=1}^{2} \sum_{\substack{k=1 \\ k \neq j}}^{2} \alpha_{jk} \ln[\mathcal{H}_j(x_{ji}; \underline{\theta_j}, \lambda_j)] \ln[\mathcal{H}_k(x_{ki}; \underline{\theta_k}, \lambda_k)].
\end{aligned}
\tag{26}
$$

The pseudo-score function is defined as the partial derivatives of the log-pseudo-likelihood function with respect to each of the parameters model, this is denoted by

$$
U_p(\underline{\beta}) = \left(U_p(\underline{\theta_1}), U_p(\underline{\theta_2}), U_p(\lambda_1), U_p(\lambda_2), U_p(\alpha_1), U_p(\alpha_2), U_p(\alpha_{12})\right)^{\top}.
$$

In accordance with [30], the pseudo likelihood estimator $\widetilde{\underline{\beta}}$ of $\underline{\beta}$ is consistent, asymptotically normally distributed with covariance matrix given by

$$
\mathbf{\Sigma}(\widetilde{\underline{\beta}}) = \frac{1}{n} \mathbf{\Gamma}^{-1}(\underline{\beta}) \mathbf{\Psi}(\underline{\beta}) \mathbf{\Gamma}^{-1}(\underline{\beta}).
$$

Cheng, C. and Riu, J. [31] consider a sandwich type estimator, which is consistent for the asymptotic covariance matrix to estimate $\mathbf{\Sigma}(\widetilde{\underline{\beta}})$ which is given by

$$
\widehat{\mathbf{\Sigma}}(\widetilde{\underline{\beta}}) = \frac{1}{n} \widehat{\mathbf{\Gamma}}_n^{-1}(\widetilde{\underline{\beta}}) \widehat{\mathbf{\Psi}}_n(\widetilde{\underline{\beta}}) \widehat{\mathbf{\Gamma}}_n^{-1}(\widetilde{\underline{\beta}})^{\top},
$$

where

$$
\widehat{\mathbf{\Gamma}}_n(\underline{\beta}) = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \underline{\beta}^{\top}} U_i(\underline{\beta}) \bigg|_{\widetilde{\underline{\beta}}} \quad \text{and} \quad \widehat{\mathbf{\Psi}}_n(\underline{\beta}) = \frac{1}{n} \sum_{i=1}^{n} U_i(\underline{\beta}) U_i(\underline{\beta})^{\top} \bigg|_{\widetilde{\underline{\beta}}},
$$

with $U_i(\underline{\beta}) = \frac{\partial}{\partial \underline{\beta}} \ell_P(\underline{\beta})$, is the score vector for the pseudo-likelihood function.

The structure of the $\mathbf{\Gamma}$ and $\mathbf{\Psi}$ matrices are going to depend on the PDF in the BPSE model, specifically of the $g(\cdot)$ generating function. The non-singularity of $\mathbf{\Sigma}$ must be treated in each particular case, as well as its possible solution in case of singularity of this matrix.

3. Bivariate Power-Skew-Normal Model

In this section, we study the bivariate PSE distribution when density functions $h_1(x_1; \lambda_1)$ and $h_2(x_2; \lambda_2)$ correspond to the SN density of [1], that is,

$$
h_i(x_i; \lambda_i) = f_{\mathrm{SN}}(x_i; \lambda_i) = 2\phi(x_i)\Phi(\lambda_i x_i) \quad \text{and} \quad \mathcal{H}_i(x_i; \lambda_i) = \mathcal{F}_{\mathrm{SN}}(x_i; \lambda_i) = \Phi(x_i) - 2T(x_i; \lambda_i).
$$

In this case, the BPSN probability density function is given by

$$
\begin{aligned}
h_{\mathrm{BPSN}}(x_1, x_2) = {} & c(\underline{\lambda}, \underline{\alpha}) f_{\mathrm{SN}}(x_1; \lambda_1) f_{\mathrm{SN}}(x_2; \lambda_2) \{\mathcal{F}_{\mathrm{SN}}(x_1; \lambda_1)\}^{\alpha_1 - 1} \{\mathcal{F}_{\mathrm{SN}}(x_2; \lambda_2)\}^{\alpha_2 - 1} \\
& \times \exp\{\alpha_{12} \ln[\mathcal{F}_{\mathrm{SN}}(x_1; \lambda_1)] \ln[\mathcal{F}_{\mathrm{SN}}(x_2; \lambda_2)]\},
\end{aligned}
\tag{27}
$$

with $\lambda_1, \lambda_2 \in \mathbb{R}$, and $\alpha_1, \alpha_2, \alpha_{12} > 0$. The normalization constant can be written by using the transformation $v_j = \mathcal{F}_{\mathrm{SN}}(x_j; \lambda_j)$ for $j = 1, 2$ as

$$c(\underline{\boldsymbol{\lambda}}, \underline{\boldsymbol{\alpha}}) = c(\underline{\boldsymbol{\alpha}}) = \left(\int_0^1 \int_0^1 v_1^{\alpha_1 - 1} v_2^{\alpha_2 - 1} \exp(\alpha_{12} \ln(v_1) \ln(v_2)) \right)^{-1}.$$

This standard BPSN model will be denoted by $\mathrm{BPSN}(\lambda_1, \lambda_2, \alpha_1, \alpha_2, \alpha_{12})$. For $\lambda_1 = \lambda_2 = 0$ the bivariate conditional exponentiated normal model, studied by [26] is obtained, while for $\lambda_1 = \lambda_2 = \alpha_{12} = 0$, the bivariate joint distribution of independent power-skew-normal random variables is obtained; and for $\lambda_1 = \lambda_2 = \alpha_{12} = 0$ and $\alpha_1 = \alpha_2 = 1$, the bivariate joint distribution of independent normal random variables is obtained. Note that, if $\alpha_1 = \alpha_2 = 1$, then it would have a type of bivariate conditional SN distribution.

The location-scale extension of the BPSN model can be written as

$$\begin{aligned}
h_{\mathrm{BPSN}}(x_1, x_2) &= \frac{c(\underline{\boldsymbol{\lambda}}, \underline{\boldsymbol{\alpha}})}{\eta_1 \eta_2} f_{\mathrm{SN}}(z_1; \lambda_1) f_{\mathrm{SN}}(z_2; \lambda_2) \{\mathcal{F}_{\mathrm{SN}}(z_1; \lambda_1)\}^{\alpha_1 - 1} \{\mathcal{F}_{\mathrm{SN}}(z_2; \lambda_2)\}^{\alpha_2 - 1} \\
&\quad \times \exp\{\alpha_{12} \ln[\mathcal{F}_{\mathrm{SN}}(z_1; \lambda_1)] \ln[\mathcal{F}_{\mathrm{SN}}(z_2; \lambda_2)]\},
\end{aligned} \tag{28}$$

where $z_j = (x_j - \xi_j)/\eta_j$, with $-\infty < \xi_j < \infty$ and $\eta_j > 0$, for $j = 1, 2$. We will denote it by $\mathrm{BPSN}\big((\xi_1, \eta_1), (\xi_2, \eta_2), \lambda_1, \lambda_2, \alpha_1, \alpha_2, \alpha_{12}\big)$. Figure 1 presents the contour graphs of the BPSN model for some selected values of the parameters.
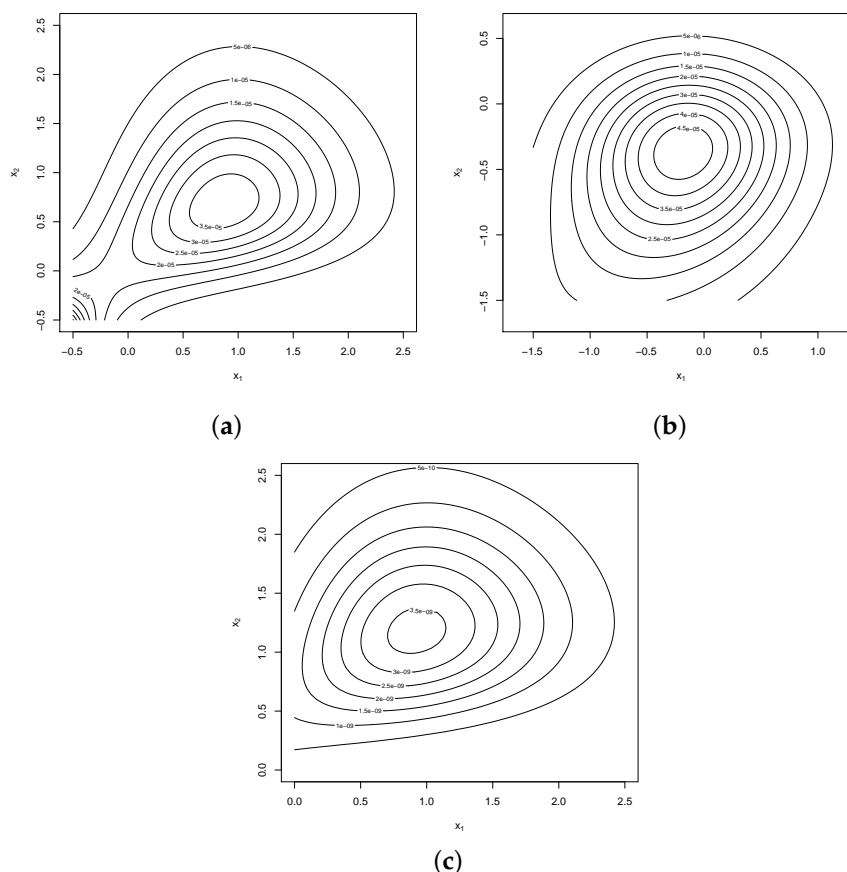

(a)

(b)

(c)

Figure 1. Graphs of contorns for the BPSN model (**a**) BPSN $((0,1), (0,1), 1.25, 2.75, 2.25, 1.75, 0.75)$, (**b**) BPSE $((0,1), (0,1), -1.25, -2.75, 2.25, 1.75, 0.75)$ and (**c**) BPSE $((0,1), (0,1), 1.25, 2.75, 2.25, 3.75, 1.25)$.

Table 1 shows the correlation coefficients of the BPSN model for values λ_1 ranging from -2.5 to 2.5 and from 0.5 to 0.5; λ_2 ranging from -1.5 to 2.0 and from 0.5 to 0.5; $\alpha_1 = 3.25$, $\alpha_2 = 2.75$ and

$\alpha_{12} = 0.5$. It can be observed that BPSN model is very flexible in terms of correlation, since $\rho \in (-0.9933, 1.0)$, which contains the correlation coefficients of the bivariate conditional exponentiated: normal, logistic and Student-t models, for some values of $\alpha_1 \in (0.5, 10)$, $\alpha_2 \in (0.25, 10)$ and $\alpha_{12} \in (0.5, 2.5)$, see [26], for more details. According to [26], for bivariate conditional exponentiated normal model with $\alpha_1, \alpha_2 \in (0.4, 100)$ and $\alpha_{12} \in (0.2, 100)$, the range of possible values for the correlation coefficient is on interval $(-0.8634, 0.9247)$. Likewise, the range of possible values for the correlation coefficient for the BPSN model contains the respective range of possible values of the bivariate conditional exponentiated model studied by [28] which is $(0.20, 0.60)$ for $\alpha_{12} \in (0, 1000)$, and finally this range also contains the range of possible values for the correlation coefficient of the conditionally specified bivariate skewed model of [6], which is $(-0.63662, 0.63662)$.

Table 1. Correlation coefficient for the BPSN distribution.

λ_2/λ_1	-1.5	-1.0	-0.5	0	0.5	1.0	1.5	2.0
-2.5	0.8033	0.3730	-0.2686	-0.6699	-0.8444	-0.9252	-0.9680	-0.9933
-2.0	0.7518	0.3559	-0.2385	-0.6124	-0.7759	-0.8519	-0.8924	-0.9164
-1.5	0.6025	0.2984	-0.1661	-0.4625	-0.5938	-0.6556	-0.6888	-0.7087
-1.0	0.2105	0.1324	-0.0046	-0.1014	-0.1479	-0.1713	-0.1845	-0.1927
-0.5	-0.2853	-0.0921	0.1722	0.3245	0.3855	0.4117	0.4246	0.4317
0	-0.5511	-0.2190	0.2543	0.5385	0.6573	0.7105	0.7377	0.7534
0.5	-0.6643	-0.2756	0.2846	0.6244	0.7679	0.8327	0.8662	0.8856
1.0	-0.7183	-0.3035	0.2971	0.6631	0.8185	0.8888	0.9254	0.9467
1.5	-0.7478	-0.3193	0.3031	0.6834	0.8451	0.9186	0.9569	0.9791
2.0	-0.7657	-0.3291	0.3062	0.6951	0.8609	0.9362	0.9755	0.9985
2.5	-0.7776	-0.3358	0.3081	0.7026	0.8709	0.9476	0.9876	1.0000

3.1. Statistical Inference

We consider a random sample of vectors following a BPSN distribution. The corresponding log-pseudo-likelihood function for the parameter vector $\underline{\beta} = ((\xi_1, \eta_1), (\xi_2, \eta_2), \lambda_1, \lambda_2, \alpha_1, \alpha_2, \alpha_{12})^\top$, is given by

$$
\begin{aligned}
\ell_P(\underline{\beta}) = & \sum_{i=1}^{n} \ln(\alpha_1 - \alpha_{12} \ln[\mathcal{F}_{\mathrm{SN}}(z_{2i}; \lambda_2)]) + \sum_{i=1}^{n} \ln[f_{\mathrm{SN}}(z_{1i}; \lambda_1)] \\
& + \sum_{i=1}^{n} \ln(\alpha_2 - \alpha_{12} \ln[\mathcal{F}_{\mathrm{SN}}(z_{1i}; \lambda_1)]) + \sum_{i=1}^{n} \ln[f_{\mathrm{SN}}(z_{2i}; \lambda_2)] \\
& + \sum_{i=1}^{n} (\alpha_1 - 1) \ln[\mathcal{F}_{\mathrm{SN}}(z_{1i}; \lambda_1)] + \sum_{i=1}^{n} (\alpha_2 - 1) \ln[\mathcal{F}_{\mathrm{SN}}(z_{2i}; \lambda_2)] \\
& - 2 \sum_{i=1}^{n} \alpha_{12} \ln[\mathcal{F}_{\mathrm{SN}}(z_{1i}; \lambda_1)] \ln[\mathcal{F}_{\mathrm{SN}}(z_{2i}; \lambda_2)],
\end{aligned}
\tag{29}
$$

where $z_{ji} = (x_{ji} - \xi_j)/\eta_j$. Then, the pseudo-score function which is denoted by

$$
U_P(\underline{\beta}) = (U_p(\xi_1), U_p(\eta_1), U_p(\xi_2), U_p(\eta_2), U_p(\lambda_1), U_p(\lambda_2), U_p(\alpha_1), U_p(\alpha_2), U_p(\alpha_{12}))^\top.
$$

has elements given by

$$
\begin{aligned}
U_P(\xi_j) = & \frac{\alpha_{12}}{\eta_j} \sum_{i=1}^{n} \frac{W_{(1)ji}}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\mathrm{SN}}(z_{ji}; \lambda_j))]} + \frac{1}{\eta_j} \sum_{i=1}^{n} \left[z_{ji} - \sqrt{\frac{2}{\pi}} \lambda_j W_{(2)ji} \right] \\
& - \frac{1}{\eta_j} \sum_{i=1}^{n} [(\alpha_j - 1) - 2\alpha_{12} \ln(\mathcal{F}_{\mathrm{SN}}(z_{j'i}; \lambda_{j'}))] W_{(1)ji}, \quad j = 1, 2,
\end{aligned}
$$

$$U_P(\eta_j) = \frac{\alpha_{12}}{\eta_j} \sum_{i=1}^{n} \frac{z_{ji}W_{(1)ji}}{[\alpha_j - \alpha_{12}\ln(\mathcal{F}_{\mathrm{SN}}(z_{ji};\lambda_j))]} + \frac{1}{\eta_j} \sum_{i=1}^{n} \left[z_{ji}^2 - 1 - \sqrt{\frac{2}{\pi}}\lambda_j z_{ji}W_{(2)ji} \right]$$

$$- \frac{1}{\eta_j} \sum_{i=1}^{n} [(\alpha_j - 1) - 2\alpha_{12}\ln(\mathcal{F}_{\mathrm{SN}}(z_{j'i};\lambda_{j'}))]z_{ji}W_{(1)ji}, \quad j = 1, 2,$$

$$U_P(\lambda_j) = \sqrt{\frac{2}{\pi}} \frac{\alpha_{12}}{1 + \lambda_j^2} \sum_{i=1}^{n} \frac{W_{(3)ji}}{[\alpha_j - \alpha_{12}\ln(\mathcal{F}_{\mathrm{SN}}(z_{ji};\lambda_j))]} + \sqrt{\frac{2}{\pi}} \sum_{i=1}^{n} z_{ji}W_{(2)ji}$$

$$- \sqrt{\frac{2}{\pi}} \frac{1}{1 + \lambda_j^2} \sum_{i=1}^{n} [(\alpha_j - 1) - 2\alpha_{12}\ln(\mathcal{F}_{\mathrm{SN}}(z_{j'i};\lambda_{j'}))]W_{(3)ji}, \quad j = 1, 2,$$

$$U_P(\alpha_j) = \sum_{i=1}^{n} \frac{1}{[\alpha_j - \alpha_{12}\ln(\mathcal{F}_{\mathrm{SN}}(z_{ji};\lambda_j))]} + \sum_{i=1}^{n} \ln(\mathcal{F}_{\mathrm{SN}}(z_{ji};\lambda_j)), \quad j = 1, 2,$$

and

$$U_P(\alpha_{12}) = \sum_{i=1}^{n}\sum_{j=1}^{2} \frac{-\ln(\mathcal{F}_{\mathrm{SN}}(z_{ji};\lambda_j))}{[\alpha_j - \alpha_{12}\ln(\mathcal{F}_{\mathrm{SN}}(z_{ji};\lambda_j))]} - 2\sum_{i=1}^{n} \ln(\mathcal{F}_{\mathrm{SN}}(z_{1i};\lambda_1))\ln(\mathcal{F}_{\mathrm{SN}}(z_{2i};\lambda_2))$$

where $W_{(1)ji} = \frac{f_{\mathrm{SN}}(z_{ji};\lambda_j)}{\mathcal{F}_{\mathrm{SN}}(z_{ji};\lambda_j)}$, $W_{(2)ji} = \frac{\phi\left(\sqrt{1+\lambda_j^2}\,z_{ji}\right)}{f_{\mathrm{SN}}(z_{ji};\lambda_j)}$, $W_{(3)ji} = \frac{\phi\left(\sqrt{1+\lambda_j^2}\,z_{ji}\right)}{\mathcal{F}_{\mathrm{SN}}(z_{ji};\lambda_j)}$.

The solution to the system of non-linear equations $U_p(\xi_1) = 0$, $U_p(\eta_1) = 0$, $U_p(\xi_2) = 0$, $U_p(\eta_2) = 0$, $U_p(\lambda_1) = 0$, $U_p(\lambda_2) = 0$, $U_p(\alpha_1) = 0$, $U_p(\alpha_2) = 0$, $U_p(\alpha_{12}) = 0$, leads to pseudo-likelihood estimates of the parameter vector of the BPSN model, this system must be solved by using iterative numerical algorithms.

For estimating the covariance matrix, we have from [31] that, thte $\boldsymbol{\Psi}(\cdot)$ matrix will be estimated from the pseudo-score function given above, while the $\boldsymbol{\Gamma}(\cdot)$ component of the $\boldsymbol{\Sigma}(\cdot)$ matrix, we will write it in the form $\boldsymbol{\Gamma} = (\gamma_{\beta_j\beta_{j'}}) = -\frac{1}{n}\mathbf{K}$ where $\mathbf{K} = (\kappa_{\beta_j\beta_{j'}})$ is a matrix of second derivatives of the pseudo-likelihood function, with respect to the parameters of the model, the elements $\kappa_{\beta_j\beta_{j'}}$ of this matrix are presented in the Appendix A.

3.2. Reparameterization for the Bpsn Model

As well known in the literature of distributions theory, the SN model has a singular information matrix for $\lambda = 0$, however, it has been proposed to perform a reparameterization of the model parameters, [32], this problem is presented by several extensions of the SN model, including for example, the power-skew-normal model whose information matrix is singular for $\lambda = 0$ and $\alpha = 1$, here, [33] present a reparameterization of the models parameters for this case.

Another solution to the problem of the singularity of the SN model when $\lambda = 0$ was presented by [1], which consists of a representation of the form $Y = \mu + \sigma\frac{Z - \mathbb{E}[Z]}{\sqrt{\mathrm{Var}[Z]}}$, where $\mu \in \mathbb{R}$ and $\sigma > 0$ are parameters of the random variable Y, and $Z \sim \mathrm{SN}(\lambda)$. This representation is called centered parametrization, since $\mathbb{E}[Y] = \mu$ and $\mathrm{Var}[Y] = \sigma^2$. The new representation has parameters vector $\boldsymbol{\theta} = (\mu, \sigma, \gamma_1)^\top$, where $-0.9953 \leq \gamma_1 \leq 0.9953$ represents the asymmetry coefficient of the random variable Y. Under this representation, the information matrix of the new parameters vector turns out to be non-singular. Thus, the information matrix is written in the form $\mathbf{D}\mathbf{I}_\lambda\mathbf{D}^\top$ where $\mathbf{I}_\lambda$ is the information matrix of the model with parameters vector $(\xi, \eta, \lambda)^\top$ and $\mathbf{D}$ is a matrix of derivatives of

the parameters vector $(\xi, \eta, \lambda)^\top$ with respect to the parameters vector $\boldsymbol{\theta}$. Under this reparameterization, we have the relationship

$$\xi = \mu - c\sigma\gamma_1^{1/3}, \quad \eta = \sigma\sqrt{1 + c^2\gamma_1^{2/3}} \quad \text{and} \quad \lambda = \frac{c\gamma_1^{1/3}}{\sqrt{b^2 + c^2(b^2-1)\gamma_1^{2/3}}}, \tag{30}$$

with $b = \sqrt{\frac{2}{\pi}}$ and $c = \{2/(4-\pi)\}^{1/3}$.

Also it has that, when $\lambda \to 0$, the information matrix converges to the diagonal matrix $\Sigma_c = \text{diag}(\sigma^2, \sigma^2/2, 6)$. This guarantees the existence and uniqueness of the MLE of ξ and η, for each fixed value of λ, see [1].

The BPSN model also inherits the singularity problem in the matrix $\boldsymbol{\Gamma}(\cdot)$ for values close to $\lambda_1 = \lambda_2 = 0$ and $\alpha_1 = \alpha_2 = 1$, specifically, for values close $\lambda_1 = 0$, the columns corresponding to the elements $k_{\xi_1\beta_j}$ and $k_{\lambda_1\beta_j}$, where β_j is related to the rest of the parameters, are linearly dependent. The same way, it happens for values close $\lambda_2 = 0$. This situation also occurs in the pseudo-score function, leading to problems to guarantee the existence and uniqueness of the pseudo-estimated values. This leads to the non-existence of the inverse of the $\boldsymbol{\Gamma}(\cdot)$ matrix and therefore, the covariance matrix $\boldsymbol{\Sigma}(\cdot)$ can not be calculated.

Following [1], we define $Y_j = \mu_j + \sigma_j \frac{Z_j - \mathbb{E}[Z_j]}{\sqrt{\text{Var}[(Z_j)]}}$ for $j = 1, 2$, where $Z_j \sim \text{SN}(\lambda_j)$, then we arrive to representation of the BPSN with parameters vector $\underline{\boldsymbol{\theta}} = (\mu_1, \mu_2, \sigma_1, \sigma_2, \gamma_{j1}, \gamma_{j2}, \alpha_1, \alpha_2, \alpha_{12})^\top$ where, for $j = 1, 2$, it has $-0.9953 \le \gamma_{j1} \le 0.9953$. As showed by [1], this representation can be written in the form $Y_j = \xi_j + \eta_j Z$, for $j = 1, 2$, that is, $Y_j \sim \text{SN}(\xi_j, \eta_j, \lambda_j)$ where

$$\xi_j = \mu_j - c\sigma_j\gamma_{j1}^{1/3}, \quad \eta_j = \sigma_j\sqrt{1 + c^2\gamma_{j1}^{2/3}} \quad \text{and} \quad \lambda_j = \frac{c\gamma_{j1}^{1/3}}{\sqrt{b^2 + c^2(b^2-1)\gamma_{j1}^{2/3}}}, \tag{31}$$

for $j = 1, 2$. We denote it by $\text{SN}_c(\mu_j, \sigma_j, \gamma_{j1})$. Here, the new bivariate centered power-skew-normal model (BPSN$_c$) is defined just as the PDF defined by model in Equation (28), where ξ_j, η_j and λ_j are defined as in Equation (31). Given the relationship in Equation (31), ML estimates for the vector $\underline{\boldsymbol{\theta}} = (\mu_1, \mu_2, \sigma_1, \sigma_2, \gamma_{j1}, \gamma_{j2}, \alpha_1, \alpha_2, \alpha_{12})^\top$ can be obtained from the estimates of the original model, that is, the estimates of maximum pseudo-likelihood for α_1, α_2 and α_{12} are the same as the model without reparametrizating, while for the parameters subvector $(\mu_1, \mu_2, \sigma_1, \sigma_2, \gamma_{j1}, \gamma_{j2})^\top$, it can be obtained by using the inverse relationships of (31), that is, from:

$$\mu_j = \xi_j + b\sigma_j\delta_j, \quad \sigma_j = \eta_j\frac{[1 + \eta_j(1-b^2)]^{1/2}}{(1+\lambda_j^2)^{1/2}}, \quad \gamma_{j1} = \frac{4-\pi}{2}(b\lambda_j)^3\left[1 + \lambda_j^2(1-b^2)\right]^{-3/2}, \tag{32}$$

where $\delta_j = \lambda_j/\sqrt{1+\lambda_j^2}$. From (32), if $\lambda_j \approx 0$, then $\gamma_{j1} \approx 0$, where γ_{j1} corresponds to the asymmetry coefficient of the random variable Y_j. In this way, it is important to analyze the magnitude of the sample asymmetry coefficient of each variable.

The covariance matrix of [31] is obtained in the same way as in BPSN model, however, it can be demonstrated that the covariance matrix in the BPSN$_c$ model lets herself be written as

$$\boldsymbol{\Sigma}_c(\underline{\tilde{\boldsymbol{\theta}}}) = \frac{1}{n}\left(\mathbf{D}\boldsymbol{\Gamma}(\underline{\boldsymbol{\beta}})\mathbf{D}^\top\right)^{-1}\mathbf{D}\boldsymbol{\Psi}(\underline{\boldsymbol{\beta}})\mathbf{D}^\top\left(\mathbf{D}\boldsymbol{\Gamma}(\underline{\boldsymbol{\beta}})\mathbf{D}^\top\right)^{-1}.$$

where $\mathbf{D} = \dfrac{\partial\underline{\boldsymbol{\beta}}}{\partial\underline{\boldsymbol{\theta}}}$.

It can be shown that $\mathbf{D}$ is a block-diagonal matrix, given by

$$\mathbf{D} = \text{diag}(\mathbf{D}_1, \mathbf{D}_2, \mathbf{I}_3),$$

where $\mathbf{I}_3$ is an identity matrix of dimension 3×3, and $\mathbf{D}_j$ for $j = 1, 2$ is the derivative matrix of the parameter vector $(\xi_j, \eta_j, \lambda_j)^\top$ with respect to the vector $(\mu_j, \sigma_j, \gamma_{1j})^\top$. The elements of this matrix can be found in their general form in ([34], p. 68). When $\gamma_{1j} \to 0$, then $\mathbf{D}_j \boldsymbol{\Gamma}(\underline{\boldsymbol{\beta}_j}) \mathbf{D}_j^\top$ tends to the diagonal matrix $\mathrm{diag}(\sigma_j^2, \sigma_j^2/2, 6)$ which is non-singular, see [1,34].

According to [1], rewriting $\underline{\boldsymbol{\beta}} = (\underline{\boldsymbol{\beta}_1}, \underline{\boldsymbol{\beta}_2}, \underline{\boldsymbol{\alpha}})^\top$, where $\underline{\boldsymbol{\beta}_j} = (\xi_j, \eta_j, \lambda_j)^\top$, it can be shown that the $\mathbf{D}_j \boldsymbol{\Gamma}(\underline{\boldsymbol{\beta}_j}) \mathbf{D}_j^\top$ sub matrices for $j = 1, 2$, of the $\mathbf{D}\boldsymbol{\Gamma}(\underline{\boldsymbol{\beta}}) \mathbf{D}^\top$ matrix are non-singular and therefore, the rows of the $\mathbf{D}\boldsymbol{\Gamma}(\underline{\boldsymbol{\beta}}) \mathbf{D}^\top$ matrix are linearly independent, that is, this matrix is invertible, which guarantees the existence of the $\boldsymbol{\Sigma}_c(\widetilde{\boldsymbol{\theta}})$ matrix. The estimator of the covariance matrix $\boldsymbol{\Sigma}_c(\widetilde{\boldsymbol{\theta}})$ is obtained by replacing $\underline{\boldsymbol{\beta}}$ and $\underline{\boldsymbol{\theta}}$ by their respective estimators.

4. Numerical Illustrations

4.1. Illustration 1

We consider an illustration where we study the fit of the BPSN model for a data set studied in [35]. We pooled together, the 50 Iris-setosa data points, the 50 Iris-versicolor data points and the 50 Iris-virginica data points, to get a total sample size of $n = 150$. Descriptive statistics for the data set are presented in Table 2. Quantities $\sqrt{b_1}$ and b_2 correspond to sample asymmetry and kurtosis coefficients.

Table 2. Summary statistics for the data set.

Variable	$\overline{x}_j$	s_j	$\sqrt{b_{j1}}$	b_{j2}
x_1	5.843	0.828	0.308	-0.605
x_2	3.057	0.435	0.312	0.138

Table 2 shows that variables x_1 and x_2 present high asymmetry, so a BPSN model could fit this data set. Under normality assumption, hypothesis tests for the asymmetry coefficient of x_1 and x_2, ($H_0 : \sqrt{\beta_{j1}} = 0$, $j = 1, 2$.) show test statistics 7.716 and 7.815, respectively, values well above the percentile of $\chi^2_{(1)}$ distribution, at level 5% whose value is 3.84, which indicates that the asymmetry in each variable is very important. Likewise, the univariate normality tests of [36] show p-values of 0.0225 and 0.0202, concluding that the distributions are asymmetric.

The bivariate normality tests of [37–40] yielded test statistics (with p-values in parentheses), of 2.8909 (0.0000), 9.3258 (0.0094) and 62.2958 (0.0000), respectively, hence, it is concluded that the bivariate observations vector does not follow a bivariate normal distribution. Thus, an asymmetric bivariate distribution, such as the BPSN model may be useful to fit this data set. Hence, the bivariate normal distribution is not a tenable model for the data under study, and an alternative model that is able to incorporate some degree of asymmetry would probably fit the data better. We fitted the conditional bivariate skew-normal model (see, [6]), the bivariate power-normal (BPN) model and the BPSN model. To compare fitted model, we make use of the Akaike Information Criterion (AIC), by [41] and the corrected AIC (CAIC) [42]. These measures are defined as follows

$$AIC = -2\ell(\widehat{\boldsymbol{\theta}}) + 2p \quad \text{and} \quad CAIC = -2\ell(\widehat{\boldsymbol{\theta}}) + \frac{2n(p+1)}{n-p-2}.$$

We used the `optim` fuction of statistical package [43] for fitting the bivariate model. To choose the initial values in the iterative estimation process, for α_1, α_2, α_{12}, in the BPSN model we use the transformation $Y_j = -\ln\left(\mathcal{F}_{\mathrm{SN}}(z_{1i}; \lambda_j)\right)$ for $j = 1, 2$ which yields the bivariate exponential conditionals model discussed in detail by [28].

$$f_{Y_1 Y_2}(y_1, y_2) = \mathbf{k}(\alpha_1, \alpha_2, \alpha_{12}) \exp\left(-\alpha_1 y_1 - \alpha_2 y_2 - \alpha_{12} y_1 y_2\right) \tag{33}$$

This transformation leads to obtaining estimates by the method of the moments of α_1, α_2 and α_{12}, which are consistent and asymptotically normal, see [28]. These estimators are given by:

$$\widetilde{\alpha}_1 = \frac{\widetilde{\gamma}}{\overline{y}_1(\widetilde{\gamma} + I(\widetilde{\gamma}-1))}, \quad \widetilde{\alpha}_2 = \frac{\widetilde{\gamma}}{\overline{y}_2(\widetilde{\gamma} + I(\widetilde{\gamma}-1))} \quad \text{and} \quad \widetilde{\alpha}_{12} = \frac{\widetilde{\gamma}(\widetilde{\gamma}-1)}{\overline{y}_1\overline{y}_2(\widetilde{\gamma} + I(\widetilde{\gamma}-1))},$$

where $\widetilde{\gamma} = \frac{I}{1+\rho_{Y_1Y_2}I}c$ with $\rho_{Y_1Y_2} = cor(y_1,y_2)$ and $I = cv(y_1)cv(y_2)$, where *cor* is the usual Pearson correlation between Y_1 and Y_2, and $cv(y) = \sqrt{S_y^2}/\overline{y}$.

For the initial points of λ_1, λ_2, and the location-scale parameters, we fit the univariate SN distributions by using the *selm* function of [43], the moment estimators of these parameters given in [44], and the values of the means and standard deviations of the bivariate normal model. Finally with the obtained values for $((\widetilde{\xi}_1, \widetilde{\eta}_1), (\widetilde{\xi}_2, \widetilde{\eta}_2), \widetilde{\lambda}_1, \widetilde{\lambda}_2, \widetilde{\alpha}_1, \widetilde{\alpha}_2, \widetilde{\alpha}_{12})$, the iterative estimation process is started.

Table 3 presents ML estimates, AIC and CAIC values for BCSN, BPN and BPSN models, which is the one corresponding to the best (smallest AIC or CAIC) model fitting, which clearly indicates a better fit for BPSN model. Contour plots for the BPN, bivariate conditional skew-normal (BCSN) and BPSN distributions are presented in Figure 2.

Table 3. Estimated parameters (standard errors), of the BCSN, BPN and BPSN models.

Estimate	BCSN	BPN	BPSN
$\hat{\xi}_1$	5.867 (0.055)	3.746 (0.136)	4.119 (0.163)
$\hat{\xi}_2$	3.055 (0.035)	1.655 (0.082)	2.572 (0.159)
$\hat{\eta}_1$	0.794 (0.043)	1.384 (0.058)	1.417 (0.204)
$\hat{\eta}_2$	0.438 (0.026)	0.808 (0.055)	2.146 (0.390)
$\hat{\lambda}_1$	-0.224 (0.110)		11.147 (3.685)
$\hat{\lambda}_2$			-3.200 (0.504)
$\hat{\alpha}_1$		9.358 (0.613)	2.127 (0.192)
$\hat{\alpha}_2$		14.746 (0.374)	18.260 (3.158)
$\hat{\alpha}_{12}$		2.671 (0.715)	5.016 (1.473)
AIC	555.10	551.71	549.00
CAIC	557.68	554.73	552.59

Initially, the BPN model is compared to the BPSN model by the hypothesis tests

$$H_0 : (\lambda_1, \lambda_2) = (0,0) \quad \text{versus} \quad H_1 : (\lambda_1, \lambda_2) \neq (0,0).$$

by using the likelihood ratio statistic,

$$\Lambda = \frac{\mathcal{L}_{\text{BPN}}(\widehat{\boldsymbol{\theta}})}{\mathcal{L}_{\text{BPSN}}(\widehat{\boldsymbol{\theta}})}$$

where $\mathcal{L}_{\text{BPN}}(\cdot)$ and $\mathcal{L}_{\text{BPSN}}(\cdot)$ are the pseudo-likelihood function of the BPN ans BPSN model, respectively. We obtain

$$-2\log(\Lambda) = -2\big(\ell_{\text{BPN}}(\widehat{\boldsymbol{\theta}}) - \ell_{\text{BPSN}}(\widehat{\boldsymbol{\theta}})\big) = 6.704$$

which is greater than the value of the $\chi^2_{2,95\%} = 5.99$. Then the BPSN model is a good alternative for fitting the data set. This result suggests the importance of the parameters λ_1 and λ_2 in the good fit of the BPSN model, the effect of these two parameters can be seen in the graph (c) of the Figure 2. The contour graphs in the Figure show that the BPSN model manages to better capture the distribution of the data set under consideration, since more points are contained within the contours of the BPSN distribution compared to the BCSN and BPN models.

Another hypothesis of special interest is the significance of the parameter vector $(\alpha_1, \alpha_2, \alpha_{12})$, in this particular case there is interest in the hypothesis set

$$H_0 : (\alpha_1, \alpha_2, \alpha_{12}) = (1, 1, 0) \quad \text{versus} \quad H_1 : (\alpha_1, \alpha_2, \alpha_{12}) \neq (1, 1, 0)$$

So, under H_0, we have the independent bivariate SN distribution for (X_1, X_2). An appropriate test follows by using a statistic of type Wald which follows from the asymptotic normality of the maximum pseudo-likelihood estimator $\widehat{\underline{\beta}}$. This statistic can be defined as

$$W_n = (\mathbf{A}\widehat{\underline{\beta}} - m)^{\top} \left(\widehat{\Sigma}_3(\widehat{\underline{\beta}})\right)^{-1} (\mathbf{A}\widehat{\underline{\beta}} - m),$$

where $\widehat{\Sigma}_3(\widehat{\beta})$ is a submatrix of $\widehat{\Sigma}(\widetilde{\beta})$ corresponding to the vector $(\alpha_1, \alpha_2, \alpha_{12})$, $\mathbf{A} = (\mathbf{0}_{3\times 6} \ \mathbf{I}_3)$ and $m = (1, 1, 0)^{\top}$, which, under the null hypothesis follows a χ^2 distribution with 3 degrees of freedom.

Thus, we obtained that $W_n = 1229.17$ with $p_{value} = 0.0000$, that is, the null hypothesis is rejected, indicating that the exponentiated component is significant in the model, thus, both components of the unification, skew and power, are significant in the good fit of the BPSN model.
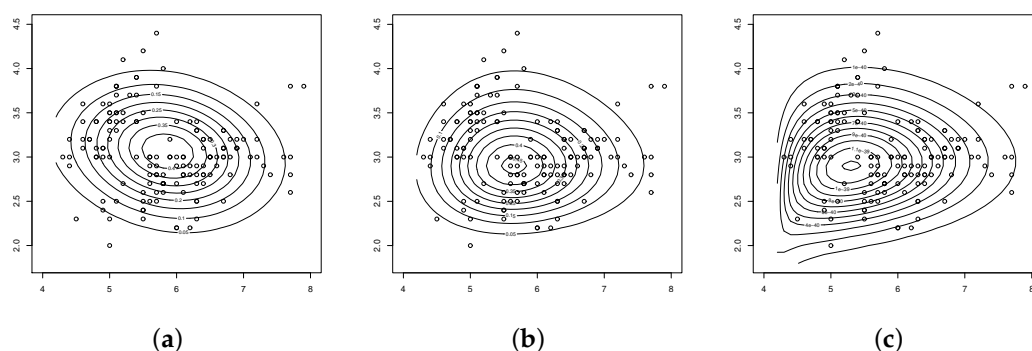

(a)   (b)   (c)

Figure 2. Contour plot of bivariate distributions for Iris data set: (**a**) BCSN model, (**b**) BPN model and (**c**) BPSN model.

The multivariate Kolmogorov-Smirnov test of goodness of fit proposed by [45], in special for the case of a bivariate distribution, which we denote by BKS (bivariate Kolmogorov-Smirnov), the statistic is given by

$$d_n = \sup_{(x_1, x_2) \in \mathbb{R}^2} \left| \mathcal{F}_n(x_1, x_2) - \mathcal{F}(x_1, x_2) \right|$$

where $\mathcal{F}_n$ is the empirical distribution function of the sample, and $\mathcal{F}$ is some specified distribution function. When $\mathcal{F}$ distribution is unknown, the Kolmogorov-Smirnov statistic is defined by

$$d_n(\mathcal{F}) = \max\{D^1, D^2\},$$

where

$$D^1 \sup_{(x_1, x_2) \in \mathbb{R}^2} \left| \mathcal{G}_n(y_1, y_2) - y_1 \times y_2 \right|$$

by using the transformation $y_1 = \mathcal{F}_{X_1}(x_1)$ and $y_2 = \mathcal{F}_{X_2|X_1}(x_2 \mid x_1)$, and

$$D^2 \sup_{(x_1, x_2) \in \mathbb{R}^2} \left| \mathcal{G}_n(y_2, y_1) - y_2 \times y_1 \right|$$

by using the transformation $y_2 = \mathcal{F}_{X_2}(x_2)$ and $y_1 = \mathcal{F}_{X_1|X_2}(x_1 \mid x_2)$, where $\mathcal{G}_n$ is the empirical distribution function of the sample. For the special case of the BPSN model,

$$d_n(\text{BPSN}) = \max\{0.05907079, 0.07485913\} = 0.07485913,$$

which is less than 0.1464, which is the critical value of Table 1 given by [45], at level of 5%. Therefore, it is concluded that, the BPSN model fits well with the iris data set.

It should be noted that, the BPSN model is compared to other models, in particular a model of the skew-normal family of [1] proposed by [6], and a model of the power-normal family of [2] studied by [26]. Our proposal better fitted the data set studied in [35]. This allows us to conclude that the BPSN model is a viable alternative to those existing in the literature when the data set presents degrees of asymmetry that are not captured by the multivariate normal model.

4.2. Illustration 2

In the second application we use data on measurements on air-pollution variables recorded at 12:00 noon in the Los Angeles area on different days, available in [46]. For this application, we use the variables $x = Wind$ and $y = NO_2$. For air pollutant concentrations, it is usually assumed that the data are uncorrelated and independent and thus do not require the diurnal or cyclic trend analysis [47]. The concentration of average air pollutants has been used in epidemiological surveillance as an indicator of the atmospheric contamination and its associated adverse effects in humans, causing diseases such as bronchitis.

The bivariate normality test of Royston returns a test statistic value of 13.55147 with p-value = 0.001141065, whereas the generalized Shapiro-Wilk test for multivariate normality returns a test statistic value of $MVW = 0.94391$, with p-value = 0.01166 rejecting the hypothesis of normality of the observations vector. Thus, a model like the BPSN is an alternative to fit the vector of observations.

In this case, the maximum pseudo-likelihood estimates for the parameter vector is given by $\hat{\xi}_1 = 5.8000(0.3802)$, $\hat{\xi}_2 = 2.8274(0.6279)$, $\hat{\eta}_1 = 4.9416(0.5998)$, $\hat{\eta}_2 = 6.1694(0.6590)$, $\hat{\lambda}_1 = -1.8609(0.1185)$, $\hat{\lambda}_2 = -0.5184(0.1263)$, $\hat{\alpha}_1 = 10.14438(5.0364124)$, $\hat{\alpha}_2 = 11.3732(6.5075)$ and $\hat{\alpha}_{12} = 15.4635(8.4410)$.

For the hypothesis

$$H_0 : (\alpha_1, \alpha_2, \alpha_{12}) = (1, 1, 0) \quad \text{versus} \quad H_1 : (\alpha_1, \alpha_2, \alpha_{12}) \neq (1, 1, 0)$$

nosotros obtuvimos que $W_n = 196.6627$ with $p_{value} = 0.0000$, es decir, se rechaza la hipótesis nula, indicando que la componente exponenciada es significativa en el modelo. For the The multivariate Kolmogorov-Smirnov test of goodness of fit,

$$d_n(\text{BPSN}) = \max\{0.1619804, 0.1066137\} = 0.1619804,$$

which is less than the values 0.2789 (for $n = 40$) and 0.2512 (for $n = 50$), then the BPSN model presents a good fit for the environmental pollution data in the city of Los Angeles. Contour plots for the bivariate PSN distributions is presented in Figure 3.

Figure 3. Contour plots for the bivariate PSN distributions.
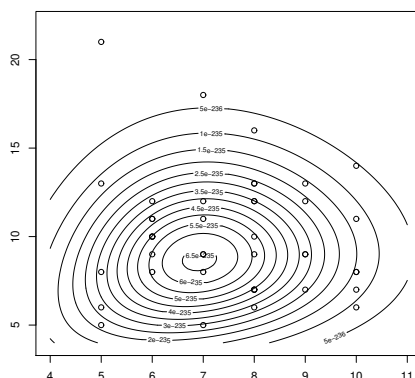
5. Concluding Remarks

In this article, on the basis of conditionally specified distributions, we have introduced a new bivariate PSE distribution which is very general, quite flexible and widely applicable. The new bivariate model is an absolutely continuous bivariate distribution whose marginals are univariate PSE distributions. We have derived several properties of the bivariate PSE distribution and special attention is centered in the particular case of de bivariate PSN distribution. The estimation of the unknown parameters of the new bivariate model is approached by using the proposal of [29] by maximization of the pseudo-likelihood function and the observed information matrix is determined. LR tests for some hypotheses of interest are also considered. As remarked, the new bivariate PSN model proposed in this article can be skewed and correlated, and therefore is much more flexible than other bivariate skew models available in the literature for analysing bivariate data. This is supported in the application to real data which is verified that the new bivariate PSN model provides consistently a better fit than the bivariate model proposed by [6].

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

Appendix A

The elements of the **K** matrix can be written as:

$$
\begin{aligned}
\kappa_{\zeta_j \zeta_j} = & \frac{\alpha_{12}}{\eta_j^2} \sum_{i=1}^{n} \left[\frac{z_{ji} W_{(1)ji} - \sqrt{\frac{2}{\pi}} \lambda_j W_{(3)ji} + W_{(1)ji}^2}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\mathrm{SN}}(z_{ji}; \lambda_j))]} - \frac{\alpha_{12} W_{(1)ji}^2}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\mathrm{SN}}(z_{ji}; \lambda_j))]^2} \right] \\
& - \frac{1}{\eta_j^2} \sum_{i=1}^{n} \left[1 + \sqrt{\frac{2}{\pi}} \lambda_j^3 z_{ji} W_{(2)ji} + \frac{2\lambda_j^2}{\pi} W_{(2)ji}^2 \right] \\
& - \frac{1}{\eta_j^2} \sum_{i=1}^{n} [(\alpha_j - 1) - 2\alpha_{12} \ln(\mathcal{F}_{\mathrm{SN}}(z_{j'i}; \lambda_{j'}))] \left[z_{ji} W_{(1)ji} - \sqrt{\frac{2}{\pi}} \lambda_j W_{(3)ji} + W_{(1)ji}^2 \right],
\end{aligned}
$$

$$
\begin{aligned}
\kappa_{\zeta_j \eta_j} = & \frac{\alpha_{12}}{\eta_j^2} \sum_{i=1}^{n} \left[\frac{(z_{ji}^2 - 1) W_{(1)ji} - \sqrt{\frac{2}{\pi}} \lambda_j z_{ji} W_{(3)ji} + z_{ji} W_{(1)ji}^2}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\mathrm{SN}}(z_{ji}; \lambda_j))]} - \frac{\alpha_{12} z_{ji} W_{(1)ji}^2}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\mathrm{SN}}(z_{ji}; \lambda_j))]^2} \right] \\
& + \frac{1}{\eta_j^2} \sum_{i=1}^{n} \left[-2z_{ji} + \sqrt{\frac{2}{\pi}} \lambda_j (1 - \lambda_j^2 z_{ji}^2) W_{(2)ji} - \frac{2\lambda_j^2 z_{ji}}{\pi} W_{(2)ji}^2 \right] \\
& - \frac{1}{\eta_j^2} \sum_{i=1}^{n} [(\alpha_j - 1) - 2\alpha_{12} \ln(\mathcal{F}_{\mathrm{SN}}(z_{j'i}; \lambda_{j'}))] \left[(z_{ji}^2 - 1) W_{(1)ji} - \sqrt{\frac{2}{\pi}} \lambda_j z_{ji} W_{(3)ji} + z_{ji} W_{(1)ji}^2 \right],
\end{aligned}
$$

$$
\kappa_{\xi_j \lambda_j} = \sqrt{\frac{2}{\pi}} \frac{1}{1+\lambda_j^2} \frac{\alpha_{12}}{\eta_j} \sum_{i=1}^{n} \left[\frac{(1+\lambda_j^2) z_{ji} W_{(3)ji} + W_{(1)ji} W_{(3)ji}}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\text{SN}}(z_{ji};\lambda_j))]} - \frac{\alpha_{12} W_{(1)ji} W_{(3)ji}}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\text{SN}}(z_{ji};\lambda_j))]^2} \right]
$$

$$
+ \frac{1}{\eta_j} \sqrt{\frac{2}{\pi}} \sum_{i=1}^{n} \left[(\lambda_{ji}^2 z_{ji}^2 - 1) W_{(2)ji} + \sqrt{\frac{2}{\pi}} \lambda_j z_{ji} W_{(2)ji}^2 \right]
$$

$$
- \sqrt{\frac{2}{\pi}} \frac{1}{1+\lambda_j^2} \frac{1}{\eta_j} \sum_{i=1}^{n} [(\alpha_j - 1) - 2\alpha_{12} \ln(\mathcal{F}_{\text{SN}}(z_{j'i};\lambda_{j'}))][(1+\lambda_j^2) z_{ji} W_{(3)ji} + W_{(1)ji} W_{(3)ji}],
$$

$$
\kappa_{\xi_j \alpha_j} = \frac{-1}{\eta_j} \sum_{i=1}^{n} \left[\frac{\alpha_{12}}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\text{SN}}(z_{ji};\lambda_j))]^2} + 1 \right] W_{(1)ji},
$$

$$
\kappa_{\xi_j \alpha_{12}} = \frac{1}{\eta_j} \sum_{i=1}^{n} \left[\frac{\alpha_j}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\text{SN}}(z_{ji};\lambda_j))]^2} + 2\ln(\mathcal{F}_{\text{SN}}(z_{j'i};\lambda_j)) \right] W_{(1)ji},
$$

$$
\kappa_{\eta_j \eta_j} = \frac{\alpha_{12}}{\eta_j^2} \sum_{i=1}^{n} \left[\frac{(z_{ji}^2 - 2) z_{ji} W_{(1)ji} - \sqrt{\frac{2}{\pi}} \lambda_j z_{ji}^2 W_{(3)ji} + z_{ji}^2 W_{(1)ji}^2}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\text{SN}}(z_{ji};\lambda_j))]} - \frac{\alpha_{12} z_{ji}^2 W_{(1)ji}^2}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\text{SN}}(z_{ji};\lambda_j))]^2} \right]
$$

$$
\frac{1}{\eta_j^2} \sum_{i=1}^{n} \left[1 - 3z_{ji}^2 + \sqrt{\frac{2}{\pi}} \lambda_j z_{ji} (2 - \lambda_j^2 z_{ji}^2) W_{(2)ji} - \frac{2}{\pi} \lambda_j^2 z_{ji}^2 W_{(2)ji}^2 \right]
$$

$$
- \frac{1}{\eta_j^2} \sum_{i=1}^{n} [(\alpha_j - 1) - 2\alpha_{12} \ln(\mathcal{F}_{\text{SN}}(z_{j'i};\lambda_{j'}))] \left[(z_{ji}^2 - 2) z_{ji} W_{(1)ji} - \sqrt{\frac{2}{\pi}} \lambda_j z_{ji}^2 W_{(3)ji} + z_{ji}^2 W_{(1)ji}^2 \right],
$$

$$
\kappa_{\eta_j \lambda_j} = \sqrt{\frac{2}{\pi}} \frac{1}{1+\lambda_j^2} \frac{\alpha_{12}}{\eta_j} \sum_{i=1}^{n} \left[\frac{z_{ji}^2 (1+\lambda_j^2) W_{(3)ji} + z_{ji} W_{(1)ji} W_{(3)ji}}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\text{SN}}(z_{ji};\lambda_j))]} - \frac{\alpha_{12} z_{ji} W_{(1)ji} W_{(3)ji}}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\text{SN}}(z_{ji};\lambda_j))]^2} \right]
$$

$$
- \sqrt{\frac{2}{\pi}} \frac{1}{\eta_j} \sum_{i=1}^{n} z_{ji} W_{(2)ji} \left[1 - \lambda_j^2 z_{ji}^2 - \sqrt{\frac{2}{\pi}} \lambda_j z_{ji} W_{(2)ji} \right]
$$

$$
- \sqrt{\frac{2}{\pi}} \frac{1}{1+\lambda_j^2} \frac{1}{\eta_j} \sum_{i=1}^{n} [(\alpha_j - 1) - 2\alpha_{12} \ln(\mathcal{F}_{\text{SN}}(z_{j'i};\lambda_{j'}))][z_{ji}^2 (1+\lambda_j^2) + z_{ji} W_{(1)ji}] W_{(3)ji},
$$

$$
\kappa_{\eta_j \alpha_j} = -\frac{1}{\eta_j} \sum_{i=1}^{n} \left[\frac{\alpha_{12} z_{ji}}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\text{SN}}(z_{ji};\lambda_j))]^2} + z_{ji} \right] W_{(1)ji},
$$

$$
\kappa_{\eta_j \alpha_{12}} = \frac{1}{\eta_j} \sum_{i=1}^{n} \left[\frac{\alpha_j z_{ji}}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\text{SN}}(z_{ji};\lambda_j))]^2} + 2z_{ji} \ln(\mathcal{F}_{\text{SN}}(z_{j'i};\lambda_j)) \right] W_{(1)ji},
$$

$$
\kappa_{\lambda_j \lambda_j} = \sqrt{\frac{2}{\pi}} \frac{\alpha_{12}}{(1+\lambda_j^2)^2} \sum_{i=1}^{n} \left[\frac{\left[-\lambda_j (1+\lambda_j^2) z_{ji}^2 - 2\lambda_j + \sqrt{\frac{2}{\pi}} W_{(3)ji} \right] W_{(3)ji}}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\text{SN}}(z_{ji};\lambda_j))]} - \sqrt{\frac{2}{\pi}} \frac{\alpha_{12} W_{(3)ji}^2}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\text{SN}}(z_{ji};\lambda_j))]^2} \right]
$$

$$
+ \sqrt{\frac{2}{\pi}} \sum_{i=1}^{n} \left[-\lambda_j z_{ji}^3 W_{(2)ji} - \sqrt{\frac{2}{\pi}} z_{ji}^2 W_{(2)ji}^2 \right]
$$

$$
- \sqrt{\frac{2}{\pi}} \frac{1}{(1+\lambda_j^2)^2} \sum_{i=1}^{n} [(\alpha_j - 1) - 2\alpha_{12} \ln(\mathcal{F}_{\text{SN}}(z_{j'i};\lambda_{j'}))] \left[-\lambda_j (1+\lambda_j^2) z_{ji}^2 - 2\lambda_j + \sqrt{\frac{2}{\pi}} W_{(3)ji} \right] W_{(3)ji},
$$

$$
\kappa_{\lambda_j \alpha_j} = -\sqrt{\frac{2}{\pi}} \frac{1}{1+\lambda_j^2} \sum_{i=1}^{n} \left[\frac{\alpha_{12}}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\text{SN}}(z_{ji};\lambda_j))]^2} + 1 \right] W_{(3)ji},
$$

$$\kappa_{\lambda_j \alpha_{12}} = \sqrt{\frac{2}{\pi}} \frac{1}{1 + \lambda_j^2} \sum_{i=1}^{n} \left[\frac{\alpha_j}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\mathrm{SN}}(z_{ji}; \lambda_j))]^2} + 2\ln(\mathcal{F}_{\mathrm{SN}}(z_{j'i}; \lambda_j)) \right] W_{(3)ji},$$

$$\kappa_{\alpha_j \alpha_j} = -\sum_{i=1}^{n} \frac{1}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\mathrm{SN}}(z_{ji}; \lambda_j))]^2},$$

$$\kappa_{\alpha_j \alpha_{12}} = \sum_{i=1}^{n} \frac{\ln(\mathcal{F}_{\mathrm{SN}}(z_{ji}; \lambda_j))}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\mathrm{SN}}(z_{ji}; \lambda_j))]^2},$$

$$\kappa_{\alpha_{12} \alpha_{12}} = -\sum_{i=1}^{n} \sum_{j=1}^{2} \frac{\ln^2(\mathcal{F}_{\mathrm{SN}}(z_{ji}; \lambda_j))}{[\alpha_j - \alpha_{12} \ln(\mathcal{F}_{\mathrm{SN}}(z_{ji}; \lambda_j))]^2},$$

$$\kappa_{\alpha_j \alpha_{j'}} = 0, \quad \kappa_{\alpha_j \lambda_{j'}} = 0, \quad \kappa_{\alpha_j \eta_{j'}} = 0, \quad \kappa_{\alpha_j \xi_{j'}} = 0,$$

$$\kappa_{\lambda_j \lambda_{j'}} = -\frac{4}{\pi} \frac{\alpha_{12}}{(1 + \lambda_j^2)(1 + \lambda_{j'}^2)} \sum_{i=1}^{n} W_{(3)ji} W_{(3)j'i},$$

$$\kappa_{\lambda_j \eta_{j'}} = -2\sqrt{\frac{2}{\pi}} \frac{1}{\eta_{j'}} \frac{\alpha_{12}}{1 + \lambda_j^2} \sum_{i=1}^{n} z_{j'i} W_{(3)ji} W_{(1)j'i},$$

$$\kappa_{\lambda_j \xi_{j'}} = -2\sqrt{\frac{2}{\pi}} \frac{1}{\eta_{j'}} \frac{\alpha_{12}}{1 + \lambda_j^2} \sum_{i=1}^{n} W_{(3)ji} W_{(1)j'i},$$

$$\kappa_{\eta_j \eta_{j'}} = -\frac{2\alpha_{12}}{\eta_j \eta_{j'}} \sum_{i=1}^{n} z_{ji} z_{j'i} W_{(1)ji} W_{(1)j'i},$$

$$\kappa_{\eta_j \xi_{j'}} = -\frac{2\alpha_{12}}{\eta_j \eta_{j'}} \sum_{i=1}^{n} z_{ji} W_{(1)ji} W_{(1)j'i},$$

$$\kappa_{\xi_j \xi_{j'}} = -\frac{2\alpha_{12}}{\eta_j \eta_{j'}} \sum_{i=1}^{n} W_{(1)ji} W_{(1)j'i}.$$

References

1.　Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
2.　Durrans, S.R. Distributions of fractional order statistics in hydrology. *Water Resour. Res.* **1992**, *28*, 1649–1655. [CrossRef]
3.　Martínez-Flórez, G.; Bolfarine, H.; Gómez, H.W. Skew-normal alpha-power model. *Statistics* **2014**, *48*, 1414–1428. [CrossRef]
4.　Martínez-Flórez, G.; Bolfarine, H.; Gómez, Y.M.; Gómez, H.W. An Unification of Families of Birnbaum-Saunders Distributions with Applications. *Rev. Stat. Stat. J.* **2020**. Available online: https://www.ine.pt/revstat/pdf/ANUNIFICATIONOFFAMILIESOFBIRNBAUM-SAUNDERS.pdf (accessed on 8 August 2020).
5.　Azzalini, A.; Dalla-Valle, A. The multivariate skew-normal distribution. *Biometrika* **1996**, *83*, 715–726. [CrossRef]
6.　Arnold, B.C.; Castillo, E.; Sarabia, J.M. Conditionally specified multivariate skewed distributions. *Sankhya Indian J. Stat. Ser. A* **2002**, *64*, 206–226.
7.　Arellano-Valle, R.; Bolfarine, H. On some characterizations of the *T*-Distribution. *Stat. Probab. Lett.* **1995**, *25*, 79–85. [CrossRef]
8.　Cambanis, S.; Huang, S.; Simons, G. On the theory of elliptically contoured distributions. *J. Multivar. Anal.* **1981**, *11*, 368–385. [CrossRef]
9.　Fang, K.T.; Kotz, S.; Ng, K.W. *Symmetric Multivariate and Related Distributions*, 3rd ed.; Chapman & Hall: London, UK, 1990.
10.　Gupta, A.K.; Varga, T. *Elliptically Contoured Models in Statistics*; Kluwer Academic Publishers: Boston, MA, USA, 1993.

11. Kelker, D. Distribution theory of spherical distributions and location scale parameters generalization. *Sankhya Indian J. Stat. Ser. A* **1970**, *32*, 419–430.

12. Azzalinio, A.; Capitanio, A. Statistical applications of the multivariate skew normal distribution. *J. R. Stat. Soc. Ser. B* **1999**, *61*, 579–602. [CrossRef]

13. Branco, M.D.; Dey, D.K. A general class of multivariate skew-elliptical distributions. *J. Multivar. Anal.* **2001**, *79*, 99–113. [CrossRef]

14. Genton, M.G.; Loperfido, N.M. Generalized skew-elliptical distributions and their quadratic forms. *Ann. Inst. Stat. Math.* **2005**, *57*, 389–401. [CrossRef]

15. Shushi, T. Generalized skew-elliptical distributions are closed under affine transformations. *Stat. Probab. Lett.* **2018**, *134*, 1–4. [CrossRef]

16. Adcock, C.; Azzalini, A. A Selective Overview of Skew-Elliptical and Related Distributions and of Their Applications. *Symmetry* **2020**, *12*, 118. [CrossRef]

17. Owen, D.B. Tables for computing bivariate normal probabilities. *Ann. Math. Stat.* **1956**, *27*, 1075–1090. [CrossRef]

18. Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*, 2nd ed.; Wiley-Blackwell; John Wiley & Sons: Hoboken, NJ, USA, 1995; Volume 2.

19. Martínez-Flórez, G.; Pacheco, M.; Giraldo, R. Inference in log-alpha-power and log-skew-normal multivariate models. *Commun. Stat. Theory Methods* **2016**, *45*, 4397–4415. [CrossRef]

20. Martínez-Flórez, G.; Farias, R.B.A.; Moreno-Arenas, G. Multivariate log-Birnbaum-Saunders regression models. *Commun. Stat. Theory Methods* **2017**, *46*, 10166–10178. [CrossRef]

21. Martínez-Flórez, G.; Lemonte, A.J.; Salinas, H.S. Multivariate Skew-Power-Normal Distributions: Properties and Associated Inference. *Symmetry* **2019**, *11*, 1509. [CrossRef]

22. Lemonte, A.J.; Martínez-Flórez, G.; Moreno-Arenas, G. Multivariate Birnbaum-Saunders distribution: Properties and associated inference. *J. Stat. Comput. Simul.* **2016**, *85*, 374–392. [CrossRef]

23. Pljonkin, A.P. Features of the Photon Pulse Detection Algorithm in the Quantum Key Distribution System. In Proceedings of the 2017 International Conference on Cryptography, Security and Privacy, Goa, India, 15–17 December 2017; pp. 81–84.

24. Pljonkin, A.P. Vulnerability of the Synchronization Process in the Quantum Key Distribution System. *Int. J. Cloud Appl. Comput.* **2019**, *9*, 50–58. [CrossRef]

25. Arnold, B.C.; Castillo, E.; Sarabia, J.M. Conditionally specified distributions. In *Lecture Notes in Statistics*; Berger, J., Fienberg, J., Gani, J., Krickeberg, I., Singer, B., Eds.; Springer: New York, NY, USA, 1992; Volume 73.

26. Martínez-Flórez, G.; Arnold, B.C.; Bolfarine, H.; Gómez, H.W. The multivariate alpha-power model. *J. Stat. Plan. Inference* **2013**, *143*, 1244–1255. [CrossRef]

27. Arnold, B.C.; Strauss, D. Bivariate distributions with conditionals in prescribed exponential families. *J. Roy. Stat. Soc. Ser. B* **1991**, *53*, 365–375. [CrossRef]

28. Arnold, B.C.; Castillo, E.; Sarabia, J.M. *Conditionally Specification of Statistical Models*; Springer Series in Statistics; Springer: New York, NY, USA, 1999.

29. Besag, J. Statistical analysis of non-lattice data. *J. Roy. Stat. Soc. Ser. D* **1975**, *24*, 179–195. [CrossRef]

30. Arnold, B.C.; Strauss, D. Pseudolikelihood Estimation: Some Examples. *Sankhya Indian J. Stat. Ser. B* **1991**, *53*, 233–2435.

31. Cheng, C.; Riu, J. On Estimating Linear Relationships When Both Variables Are Subject to Heteroscedastic Measurement Errors. *Technometrics* **2006**, *48*, 511–519. [CrossRef]

32. Rotnitziky, A.; Cox, D.R.; Bottai, M.; Robins, J. Likelihood-based inference with singular information matrix. *Bernoulli* **2000**, *6*, 243–284. [CrossRef]

33. Salinas, H.S.; Gómez, H.W.; Martínez-Flórez, G.; Bolfarine, H. Skew-normal alpha-power model [Statistics 48(2014) 1414–1428]. *Statistics* **2018**, *52*, 950–953. [CrossRef]

34. Azzalini, A.; Capitanio, A. *The Skew-Normal and Related Families*, 1st ed.; Cambridge University Press: Cambridge, UK, 2014.

35. Fisher, R.A. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **1936**, *7*, 179–188. [CrossRef]

36. Anderson, T.W.; Darling, D.A. A test of goodness of fit. *J. Am. Stat. Assoc.* **1954**, *49*, 765–769. [CrossRef]

37. Doornik, J.A.; Hansen, H. An Omnibus Test for Univariate and Multivariate Normality. *Oxf. Bull. Econ. Stat.* **2008**, *70*, 927–939. [CrossRef]

38. Henze, N.; Zirkler, B. A Class of Invariant Consistent Tests for Multivariate Normality. *Commun. Stat. Theory Methods* **1990**, *19*, 3595–3617. [CrossRef]

39. Royston, J.P. Some Techniques for Assessing Multivarate Normality Based on the Shapiro-Wilk *W*. *J. Roy. Stat. Soc. Ser. C* **1983**, *32*, 121–133. [CrossRef]

40. Royston, J.P. Remark AS R94: A Remark on Algorithm AS 181: The W-test for Normality. *J. Roy. Stat. Soc. Ser. C* **1995**, *44*, 547–551. [CrossRef]

41. Akaike, H. A new look at statistical model identification. *IEEE Trans. Autom. Contr.* **1974**, *19*, 716–722. [CrossRef]

42. Cavanaugh, J.E. Unifying the derivations for the Akaike and corrected Akaike information criteria. *Stat. Probab. Lett.* **1997**, *33*, 201–208. [CrossRef]

43. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019; Available online: http://www.R-project.org (accessed on 10 January 2020).

44. Pewsey, A. Problems of inference for Azzalini's skew-normal distribution. *J. Appl. Stat.* **2000**, *27*, 859–870. [CrossRef]

45. Justel, A.; Peña, D.; Zamar, Z. A multivariate Kolmogorov-Smirnov test of goodness of fit. *Stat. Probab. Lett.* **1997**, *35*, 251–259. [CrossRef]

46. Johnson, R.A.; Wichern, D.W. *Applied Multivariate Statistical Analysis*, 6th ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2007.

47. Gokhale, S.; Khare, M. Statistical behavior of carbon monoxide from vehicular exhausts in urban environments. *Environ. Model. Softw.* **2007**, *22*, 526–535. [CrossRef]