

Article

Multivariate Skew-Power-Normal Distributions: Properties and Associated Inference

Guillermo Martínez-Flórez ^{1,†}, Artur J. Lemonte ^{2,*,†} and Hugo S. Salinas ^{3,†}

¹ Departamento de Matemáticas y Estadística, Universidad de Córdoba, Montería 230002, Colombia; gmartinezflorez11@gmail.com

² Departamento de Estatística, Universidade Federal do Rio Grande do Norte, Natal 59078970, RN, Brazil

³ Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Copiapó 1530000, Chile; hugo.salinas@uda.cl

* Correspondence: arturlemonte@gmail.com

† These authors contributed equally to this work.

Received: 8 November 2019; Accepted: 7 December 2019; Published: 12 December 2019



Abstract: The univariate power-normal distribution is quite useful for modeling many types of real data. On the other hand, multivariate extensions of this univariate distribution are not common in the statistic literature, mainly skewed multivariate extensions that can be bimodal, for example. In this paper, based on the univariate power-normal distribution, we extend the univariate power-normal distribution to the multivariate setup. Structural properties of the new multivariate distributions are established. We consider the maximum likelihood method to estimate the unknown parameters, and the observed and expected Fisher information matrices are also derived. Monte Carlo simulation results indicate that the maximum likelihood approach is quite effective to estimate the model parameters. An empirical application of the proposed multivariate distribution to real data is provided for illustrative purposes.

Keywords: distribution theory; maximum likelihood estimation; multivariate models; parametric inference; skewed distributions

1. Introduction

Asymmetric univariate distributions that can be used for explaining real data which are not adequately fitted by the usual normal distribution were studied in Azzalini [1], Fernández and Steel [2], Mudholkar and Hutson [3], Durrans [4], Pewsey et al. [5], and Martínez-Flórez et al. [6], among others. In particular, Azzalini [1] has considered a general structure for asymmetric distributions in the univariate setting, which is given by

$$\varphi_A(z; \lambda) = 2f(z)G(\lambda z), \quad z \in \mathbb{R}, \quad (1)$$

where $\lambda \in \mathbb{R}$ controls the amount of asymmetry in the distribution, f is a symmetric (around zero) probability density function (PDF), and G is an absolutely continuous cumulative distribution function (CDF). The special case $f = \phi$ and $G = \Phi$ corresponds to the well-known skew-normal (SN) distribution, where ϕ and Φ are the PDF and CDF of the standard normal distribution, respectively. We have that $\varphi_A(z; \lambda = 0) = f(z)$. The skew-normal distribution have been extensively studied in the statistic literature. To mention a few, but not limited to, the reader is referred to Henze [7], Chiogna [8], Pewsey [9], Gómez et al. [10], and Gómez et al. [11], among many others. Another important reference

regarding univariate asymmetric distributions is the work of Durrans [4], who introduced the fractional order statistics distribution with PDF given by

$$\varphi_D(z; \alpha) = \alpha f(z) \{F(z)\}^{\alpha-1}, \quad z \in \mathbb{R},$$

where $\alpha > 0$ is a shape parameter that controls the amount of asymmetry in the distribution, F is an absolutely continuous CDF, and $f = dF$ is the corresponding PDF. The special case $F = \Phi$ corresponds to the well-known power-normal (PN) distribution (Gupta and Gupta [12]). Note also that $\varphi_D(z; \alpha = 1) = f(z)$. It is worth emphasizing that the fractional order statistics distribution studied by Durrans [4] is very flexible and, in addition, the corresponding expected Fisher information is not singular at $\alpha = 0$. On the other hand, the expected Fisher information matrix for the SN distribution introduced by Azzalini [1] is singular at $\lambda = 0$. Therefore, the fractional order statistics distribution has this interesting advantage in relation to the SN distribution.

The univariate models previously mentioned are only adequate for fitting unimodal data. A univariate bimodal distribution was introduced in Bolfarine et al. [13], whose PDF is given by

$$\varphi_B(z; \alpha, \beta) = 2\alpha c_\alpha f(z) \{F(|z|)\}^{\alpha-1} G(\beta z), \quad z \in \mathbb{R},$$

where $\alpha > 0$, $\beta \in \mathbb{R}$, F is an absolutely continuous CDF with PDF $f = dF$ symmetric around zero, G is an absolutely continuous CDF which is symmetric around zero, and $c_\alpha = 2^{\alpha-1}/(2^\alpha - 1)$ is the normalizing constant. In particular, if $F = G = \Phi$, then we have the univariate asymmetric bimodal power-normal (ABPN) distribution. We use the notation $ABPN(\beta, \alpha)$ to refer to this univariate asymmetric bimodal distribution. It follows that the ABPN distribution is bimodal for $\alpha > 1$, and unimodal otherwise. Additionally, $\beta = 0$ leads to a symmetric (around zero) bimodal distribution, and we use the notation $BPN(\alpha)$ to refer to this univariate symmetric bimodal distribution. Note also that $\varphi_B(z; \alpha = 1, \beta) = 2f(z)G(\beta z)$ and hence the ABPN distribution reduces to the PDF in (1) when $\alpha = 1$. Also, $\varphi_B(z; \alpha = 1, \beta = 1) = f(z)$.

In this paper, we generalize the univariate ABPN distribution to the multivariate setting by using the approach in Arnold et al. [14]. The new multivariate distributions we propose are quite flexible and therefore can be very useful in analyzing many types of multivariate real data which occurs frequently in practice. Additionally, maximum likelihood (ML) estimation is implemented, and (observed and expected) Fisher information matrices are derived. Finally, extensions to multivariate elliptical scenarios are also discussed. It is worth emphasizing that an important claim to introduce new multivariate distributions relies on the fact that the practitioners will have new multivariate models to use in multivariate settings. Additionally, the formulae related with the new multivariate models are manageable and with the use of modern computer resources and its numerical capabilities, the proposed multivariate models may prove to be an useful addition to the arsenal of applied statisticians. We hope that the new multivariate distributions introduced in this paper may serve as alternative multivariate models to some well-known multivariate models available in the statistical literature as, for example, the multivariate SN distribution [15], and the conditional multivariate SN distribution [14], among others. We also hope that the new multivariate models may work better (at least in terms of model fitting) than some multivariate distributions available in the literature in certain practical situations, although it cannot always be guaranteed.

The paper is organized as follows. Section 2 presents a short revision of existing asymmetric multivariate models, which are used to fit unimodal data. In Section 3, we consider the symmetric multivariate PN extension with possible bimodality. ML estimation is also implemented in this section. Section 4 is devoted to the study of the asymmetric multivariate PN model with extension implemented to the location-scale situation. ML estimation is also considered in this section. Real data application is presented in Section 5. The real data illustration shows that the new multivariate distribution can be better in terms of model fitting than unimodal alternative multivariate models for

analyzing multivariate data. Multivariate extension of the univariate PN model in the elliptical context is considered in Section 6. In Section 7 some concluding remarks are provided.

2. Multivariate Skew Models

In the last two decades, statistical literature has given great emphasis on multivariate extensions of the SN model. Important extensions are the multivariate SN distributions (Azzalini and Dalla Valle, [15]), the conditionally specified multivariate SN distribution (Arnold et al. [14]), and the multivariate alpha-power model (Martínez-Flórez et al. [6]), among others. The multivariate SN distribution has been studied by several authors, including Azzalini and Capitanio [16], Gupta and Huang [17], Gupta et al. [18], and Genton [19]. Extensions of this model have been the subject of study in Arrellano-Valle and Azzalini [20,21], Arrellano-Valle and Genton [22], Gupta and Chen [23], and Gupta and Chang [24]. The d -dimensional SN PDF can be expressed as

$$2\phi_d(\mathbf{x} - \boldsymbol{\mu}; \boldsymbol{\Omega})\Phi(\boldsymbol{\lambda}'\boldsymbol{\omega}^{-1}(\mathbf{x} - \boldsymbol{\mu})), \quad \mathbf{x} \in \mathbb{R}^d,$$

where $\phi_d(\mathbf{x} - \boldsymbol{\mu}; \boldsymbol{\Omega})$ is the joint PDF of a multivariate normal distribution, $\boldsymbol{\Omega}$ is a $d \times d$ positive definite variance-covariance matrix, $\boldsymbol{\mu} \in \mathbb{R}^d$ is the location parameter, $\boldsymbol{\lambda} \in \mathbb{R}^d$ is a parameter vector which controls skewness, and $\boldsymbol{\omega} = \text{diag}\{\omega_1, \dots, \omega_d\}$ is a diagonal matrix composed by the standard deviations from the variance-covariance matrix $\boldsymbol{\Omega}$. We use the notation $\text{SN}_d(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\lambda})$ to refer to this distribution; see Azzalini and Capitanio [25] for more details about multivariate SN models. Another useful multivariate model based on conditional SN distributions was studied in Arnold et al. [14], whose joint PDF takes the form

$$2 \left[\prod_{j=1}^d \frac{1}{\omega_j} \phi \left(\frac{x_j - \mu_j}{\omega_j} \right) \right] \Phi \left(\boldsymbol{\lambda} \prod_{j=1}^d \left(\frac{x_j - \mu_j}{\omega_j} \right) \right), \quad \mathbf{x} \in \mathbb{R}^d,$$

where $\mu_j \in \mathbb{R}$ and $\omega_j > 0$ (for $j = 1, \dots, d$) are location and scale parameters, respectively, and λ is a shape parameter. We use the notation $\text{SNC}_d(\boldsymbol{\mu}, \boldsymbol{\omega}, \lambda)$ to refer to this distribution. Similarly to the univariate setting, the expected Fisher information matrix for the multivariate SN distribution is singular at $\boldsymbol{\lambda} = \mathbf{0}_d$, where $\mathbf{0}_k$ denotes a k -vector of zeros. On the other hand, expected Fisher information matrix for the conditional multivariate SN distribution is not singular at $\lambda = 0$.

3. The Symmetric Multivariate PN Distribution

The multivariate $\text{SN}_d(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\lambda})$ and $\text{SNC}_d(\boldsymbol{\mu}, \boldsymbol{\omega}, \lambda)$ models presented in Section 2 are alternative models to the multivariate normal distribution for fitting multivariate asymmetric data. However, these multivariate distributions are only adequate for fitting unimodal data. Therefore, it is interesting to develop new multivariate models (and simple as well) that are able to adequately fit data with possible bimodality.

3.1. The New Model

Initially, we define the symmetric multivariate PN (MBPN) distribution, whose joint PDF is given by

$$f(\mathbf{x}; \boldsymbol{\alpha}) = \phi_d(\mathbf{x}) \prod_{j=1}^d \alpha_j c_{\alpha_j} \{\Phi(|x_j|)\}^{\alpha_j - 1}, \quad \mathbf{x} \in \mathbb{R}^d,$$

where $\phi_d(\mathbf{x}) := \phi_d(\mathbf{x}; \mathbf{I}_d)$ is the joint PDF of a d -dimensional multivariate normal distribution with standardized marginals, \mathbf{I}_d is the identity matrix of order d , $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)'$ is the d -vector of shape parameters, and $c_{\alpha_j} = 2^{\alpha_j - 1} / (2^{\alpha_j} - 1)$ is a normalizing constant. We use the notation $\text{MBPN}_d(\boldsymbol{\alpha})$ to refer to this new multivariate distribution. We have the following proposition.

Proposition 1. Let $\mathbf{X} = (X_1, \dots, X_d)' \sim \text{MBPN}_d(\boldsymbol{\alpha})$. We have that

1. $X_j \sim \text{BPN}(\alpha_j)$ for $j = 1, \dots, d$.
2. The joint PDF of \mathbf{X} is symmetric.
3. The product moment of \mathbf{X} is

$$\mathbb{E}(X_1^{r_1} X_2^{r_2} \dots X_d^{r_d}) = \begin{cases} 0, & \text{if any } r_j \text{ is odd,} \\ 2^{\sum_{j=1}^d r_j} \prod_{j=1}^d \mathbb{E}(X_{j+}^{r_j}), & \text{if all } r_j \text{ are even,} \end{cases}$$

where $\mathbb{E}(X_{j+}^{r_j})$ is the moment of the positive part of $X_j \sim \text{BPN}(\alpha_j)$.

4. The joint PDF of \mathbf{X} is multimodal if $\alpha_j > 1$ for $j = 1, \dots, d$.

Let $d = 2$ and hence we have the bivariate BPN model. Thus, differentiating $f(x; \alpha)$ with respect to x_j for $j = 1, 2$ and equating to zero, we obtain

$$\alpha_j - 1 = \frac{|x_j| \Phi(|x_j|)}{\phi(x_j)}, \quad j = 1, 2. \quad (2)$$

If $0 < \alpha_j < 1$ in (2), then there is no solution. Moreover, for $\alpha_j = 1$ the bivariate normal distribution has solution $x_j = 0$, and for $\alpha_j > 1$ we obtain two solutions: $x_j = -(\alpha_j - 1)\phi(x_j)/\Phi(-x_j)$ for $x_j < 0$, and $x_j = (\alpha_j - 1)\phi(x_j)/\Phi(x_j)$ for $x_j \geq 0$. For these solutions we have the determinant

$$D = \begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{vmatrix},$$

where $f := f(x; \alpha)$. Evaluating D at the critical points it follows that

$$\begin{cases} D > 0 \quad \text{and} \quad \frac{\partial^2 f}{\partial x_1^2} > 0, & \text{for } \alpha_1 \leq 1 \text{ and } \alpha_2 \leq 1, \\ D > 0 \quad \text{and} \quad \frac{\partial^2 f}{\partial x_1^2} < 0, & \text{for } \alpha_1 > 1 \text{ or } \alpha_2 > 1, \end{cases}$$

and so we conclude that the joint bivariate PDF is bimodal if $\alpha_1 > 1$ or $\alpha_2 > 1$.

3.2. Inference

We consider the ML procedure to estimate the parameter vector $\alpha \in \mathbb{R}^d$ of the $\text{MBPN}_d(\alpha)$ distribution. Let $x_1 = (x_{11}, \dots, x_{1d})'$, \dots , $x_n = (x_{n1}, \dots, x_{nd})'$ be an observed sample of size n from $\mathbf{X} = (X_1, \dots, X_d)' \sim \text{MBPN}_d(\alpha)$. The log-likelihood function takes the form

$$\begin{aligned} \ell(\alpha) &= n \sum_{j=1}^d [\log(\alpha_j) + \log(c_{\alpha_j})] + \sum_{i=1}^n \sum_{j=1}^d \log(\phi(x_{ij})) \\ &\quad + \sum_{i=1}^n \sum_{j=1}^d (\alpha_j - 1) \log(\Phi(|x_{ij}|)). \end{aligned}$$

The ML estimate $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_d)'$ of $\alpha = (\alpha_1, \dots, \alpha_d)'$ is obtained by maximizing the log-likelihood function $\ell(\alpha)$ with respect to α . The maximization can be performed, for example, in the R software (R Core Team [26]) by using the `optim(...)` function. The partial derivatives of the log-likelihood function with respect to the model parameters become

$$\frac{\partial \ell(\alpha)}{\partial \alpha_j} = \frac{n}{\alpha_j} - \frac{n \log 2}{2^{\alpha_j} - 1} + \sum_{i=1}^n \log(\Phi(|x_{ij}|)), \quad j = 1, \dots, d.$$

The ML estimate $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \dots, \hat{\alpha}_d)'$ can also be obtained by solving simultaneously the nonlinear system of equations $\partial \ell(\boldsymbol{\alpha}) / \partial \alpha_j = 0$ for $j = 1, \dots, d$, which has no closed-form and, hence, the ML estimates need to be obtained through a numerical maximization of the log-likelihood function using nonlinear optimization algorithms.

Since the new MBPN distribution corresponds to a regular ML problem, we have that the standard asymptotics apply; that is, the ML estimators of the model parameters are asymptotically normal, asymptotically unbiased and have asymptotic variance-covariance matrix given by the inverse of the expected Fisher information matrix. Let $\boldsymbol{\Sigma}(\boldsymbol{\alpha})$ be the expected Fisher information matrix. So, when n is large and under some mild regularity conditions (Cox and Hinkley [27]), we have that $\sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \stackrel{a}{\sim} N_d(\mathbf{0}_d, \boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1})$, where $\stackrel{a}{\sim}$ means approximately distributed. It can be shown that

$$\boldsymbol{\Sigma}(\boldsymbol{\alpha}) = \text{diag} \left\{ \alpha_1^{-2} - 2^{\alpha_1} (2^{\alpha_1} - 1)^{-2} (\log 2)^2, \dots, \alpha_d^{-2} - 2^{\alpha_d} (2^{\alpha_d} - 1)^{-2} (\log 2)^2 \right\}.$$

Therefore, we immediately observe that the parameters α_j 's are globally orthogonal (Cox and Reid [28]). The above asymptotic normal distribution can be used to construct approximate confidence intervals (CI) for the model parameters. Let $0 < \vartheta < 1/2$ be the significance level. The asymptotic CI of α_j given by $\hat{\alpha}_j \pm \Phi^{-1}(1 - \vartheta/2) \text{se}(\hat{\alpha}_j)$ for $j = 1, \dots, d$, with asymptotic coverage of $100(1 - \vartheta)\%$. Here, $\text{se}(\cdot)$ is the square root of the diagonal element of $\boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1}$ corresponding to each parameter (i.e., the asymptotic standard error), and $\Phi^{-1}(\cdot)$ denotes the standard normal quantile function.

3.3. A Short Simulation Study

We conduct Monte Carlo simulation experiments in order to explore the performance of the ML method in estimating the MBPN_d($\boldsymbol{\alpha}$) model parameters in finite-samples in the bivariate case (i.e., $d = 2$). Parameter values to generate the data are $\alpha_1 = 0.5, 1.5$ and 2.5 , and $\alpha_2 = 0.25, 2.5$ and 4.75 . Sample sizes considered were $n = 50, 250, 500$ and 1500 . In this study, 10,000 random samples were generated for each sample size. To generate random variates from $(X_1, X_2)' \sim \text{MBPN}_2(\boldsymbol{\alpha})$ distribution, we generate two independent uniform random variables, say $U_1 \sim U(0, 1)$ and $U_2 \sim U(0, 1)$, such that

$$(U_{1i}, U_{2i}) = \begin{cases} U_1 < 1/2 \text{ and } U_2 < 1/2, \\ U_1 < 1/2 \text{ and } U_2 \geq 1/2, \\ U_1 \geq 1/2 \text{ and } U_2 < 1/2, \\ U_1 \geq 1/2 \text{ and } U_2 \geq 1/2, \end{cases}$$

for $i = 1, \dots, n$, leading to the following relation

$$(X_{1i}, X_{2i}) = \begin{cases} \left(-\Phi^{-1} (1 - 2(1 - 2^{-\alpha_1})U_1)^{1/\alpha_1}, -\Phi^{-1} (1 - 2(1 - 2^{-\alpha_2})U_2)^{1/\alpha_2} \right), \\ \left(-\Phi^{-1} (1 - 2(1 - 2^{-\alpha_1})U_1)^{1/\alpha_1}, \Phi^{-1} (2^{-\alpha_2} (2(2^{\alpha_2} - 1)(U_2 - 2^{-1})) + 1)^{1/\alpha_2} \right), \\ \left(\Phi^{-1} (2^{-\alpha_1} (2(2^{\alpha_1} - 1)(U_1 - 2^{-1})) + 1)^{1/\alpha_1}, -\Phi^{-1} (1 - 2(1 - 2^{-\alpha_2})U_2)^{1/\alpha_2} \right), \\ \left(\Phi^{-1} (2^{-\alpha_1} (2(2^{\alpha_1} - 1)(U_1 - 2^{-1})) + 1)^{1/\alpha_1}, \Phi^{-1} (2^{-\alpha_2} (2(2^{\alpha_2} - 1)(U_2 - 2^{-1})) + 1)^{1/\alpha_2} \right). \end{cases}$$

The ML estimates are evaluated by considering the following quantities: the empirical mean, and squared root of the mean squared error (\sqrt{MSE}), which are computed from 10,000 Monte Carlo replications. All simulations were performed using the R language with the optimization of the log-likelihood function obtained by using the `optim(...)` function. From Tables 1 and 2, it is evident that the performance of the ML estimators of the MBPN model parameters is good; that is, they are very close to the true parameter values in all cases, and the \sqrt{MSE} decreases as the sample size increases, as expected, since the ML estimators are consistent. In short, the numerical results provide a clear indication that the ML method can be used quite effectively to estimate the MBPN model parameters.

Table 1. Empirical mean.

α_1	α_2	$\hat{\alpha}_1$				$\hat{\alpha}_2$			
		$n = 50$	$n = 250$	$n = 500$	$n = 1500$	$n = 50$	$n = 250$	$n = 500$	$n = 1500$
0.5	0.25	0.5914	0.5416	0.5201	0.5158	0.5286	0.3117	0.2644	0.2462
	2.5	0.5776	0.5501	0.5328	0.5180	2.5586	2.5168	2.4898	2.4932
	4.75	0.5932	0.5571	0.5386	0.5226	4.8880	4.7870	4.7658	4.7476
1.5	0.25	1.5205	1.5172	1.5123	1.5048	0.5386	0.3188	0.2675	0.2465
	2.5	1.5551	1.5156	1.5083	1.5058	2.5399	2.5151	2.5067	2.5046
	4.75	1.5492	1.5200	1.5131	1.5119	4.8714	4.7635	4.7601	4.7526
2.5	0.25	2.5295	2.5212	2.5108	2.5093	0.5439	0.2921	0.2740	0.2497
	2.5	2.5234	2.5207	2.5174	2.5078	2.5258	2.5071	2.5063	2.5054
	4.75	2.5300	2.4895	2.5027	2.5008	4.7690	4.7672	4.7608	4.7467

Table 2. Squared root of the mean squared error (\sqrt{MSE}).

α_1	α_2	$\hat{\alpha}_1$				$\hat{\alpha}_2$			
		$n = 50$	$n = 250$	$n = 500$	$n = 1500$	$n = 50$	$n = 250$	$n = 500$	$n = 1500$
0.5	0.25	0.5875	0.2920	0.1951	0.1212	0.5204	0.2621	0.1967	0.1274
	2.5	0.5788	0.2910	0.2010	0.1196	0.8049	0.3493	0.2366	0.1381
	4.75	0.5689	0.2846	0.1874	0.1091	0.8985	0.4181	0.2827	0.1633
1.5	0.25	0.7280	0.3054	0.2272	0.1282	0.5066	0.2632	0.1933	0.1254
	2.5	0.6811	0.3112	0.2146	0.1244	0.7990	0.3535	0.2453	0.1363
	4.75	0.7376	0.3048	0.2087	0.1220	0.9571	0.4146	0.2867	0.1655
2.5	0.25	0.7814	0.3280	0.2308	0.1291	0.4912	0.2560	0.1994	0.1269
	2.5	0.7382	0.3387	0.2322	0.1339	0.7812	0.3373	0.2445	0.1437
	4.75	0.7911	0.3263	0.2348	0.1340	0.9670	0.4077	0.2900	0.1685

3.4. Location-Scale Extension

The joint PDF of the MBPN model in the location-scale context is simply given by

$$f(x; \xi, \eta, \alpha) = \phi_d(x - \xi; \Lambda) \prod_{j=1}^d \alpha_j c_{\alpha_j} \left\{ \Phi \left(\left| \frac{x_j - \xi_j}{\eta_j} \right| \right) \right\}^{\alpha_j - 1}, \quad x \in \mathbb{R}^d,$$

where $\Lambda = \text{diag}\{\eta_1, \dots, \eta_d\}$ with $\eta_j > 0$ (for $j = 1, \dots, d$) being scale parameters, and $\xi = (\xi_1, \dots, \xi_d)' \in \mathbb{R}^d$ is a d -vector of location parameters. We shall use the notation $\text{MBPN}_d(\xi, \eta, \alpha)$, where $\eta = (\eta_1, \dots, \eta_d)'$.

4. The Asymmetric Multivariate PN Model

4.1. The New Model

Although the MBPN model defined in Section 3 can present multimodality, it corresponds to a symmetric multivariate distribution. An immediate extension for fitting asymmetric (possibly multimodal) multivariate data is given by the joint PDF

$$f(x; \alpha, \beta) = 2\phi_d(x) \prod_{j=1}^d \alpha_j c_{\alpha_j} \{ \Phi(|x_j|) \}^{\alpha_j - 1} \Phi(\beta'x), \quad x \in \mathbb{R}^d,$$

where $\beta = (\beta_1, \dots, \beta_d)' \in \mathbb{R}^d$ is a parameter vector which controls skewness. The above joint PDF corresponds to the multivariate extension of the univariate ABPN model, and we shall use the notation $\text{MABPN}_d(\alpha, \beta)$ to refer to this multivariate distribution. Let $d = 2$ and hence we have the bivariate ABPN model with parameter vector $(\alpha_1, \alpha_2, \beta_1, \beta_2)$. Some contour plots of the joint

bivariate PDF are presented in Figure 1. Note that the joint PDF can take different forms and will therefore be useful in analyzing bivariate data. Additionally, note that it can be unimodal or bimodal. According to an anonymous referee, the MABPN distribution has a straightforward utilization within the errors-in-variables models, especially for an application in calibration; see, for example, [29].

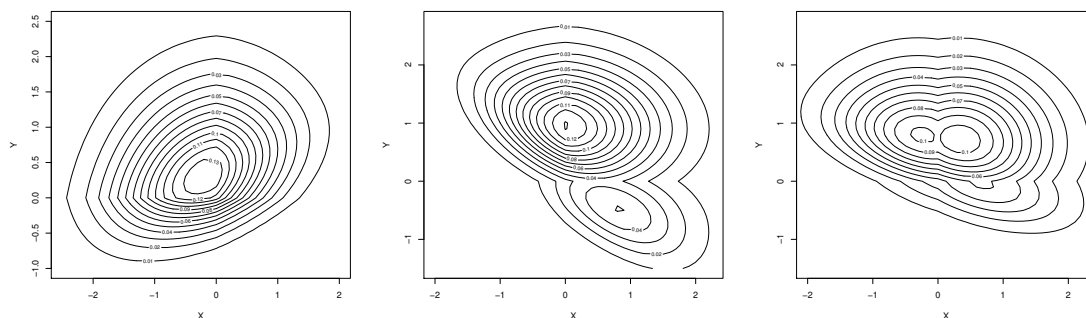


Figure 1. Contour plots for some parameter values $(\alpha_1, \alpha_2, \beta_1, \beta_2)$: $(0.75, 0.25, -0.85, 2.0)$ (left), $(0.75, 3.75, 1.85, 2.0)$ (middle), and $(1.5, 2.25, 1.0, 2.5)$ (right).

We have the following proposition.

Proposition 2. Let $\mathbf{X} = (X_1, \dots, X_d)' \sim \text{MABPN}_d(\boldsymbol{\alpha}, \boldsymbol{\beta})$. We have that

1. If $\boldsymbol{\beta} = \mathbf{0}_d$, then the MABPN model reduces to the MBPN model.
2. $-\mathbf{X} \sim \text{MABPN}_d(\boldsymbol{\alpha}, -\boldsymbol{\beta})$.
3. If $\boldsymbol{\alpha} = \mathbf{1}_d$, then the MABPN model reduces to the $\text{SN}_d(\mathbf{0}_d, \mathbf{I}_d, \boldsymbol{\beta})$ model, where $\mathbf{1}_d = (1, \dots, 1)'$ is a d -vector of ones.
4. The product moment of \mathbf{X} is given by

$$\mathbb{E}(X_1^{r_1} X_2^{r_2} \dots X_d^{r_d}) = \begin{cases} 0, & \text{if any } r_j \text{ is odd,} \\ \prod_{j=1}^d \mathbb{E}(Y_j^{r_j}), & \text{if all } r_j \text{ are even,} \end{cases}$$

where $Y_j \sim \text{BPN}(\alpha_j)$ for $j = 1, \dots, d$.

From Proposition 2, note that even moments of $\mathbf{X} = (X_1, \dots, X_d)' \sim \text{MABPN}_d(\boldsymbol{\alpha}, \boldsymbol{\beta})$ do not depend on the parameter vector $\boldsymbol{\beta}$. It implies that the correlation coefficient between the random variables X_j and $X_{j'}$, where $j \neq j'$ for $j, j' = 1, \dots, d$, depends basically on the parameter vector $\boldsymbol{\beta}$. Note that the covariance between X_j and $X_{j'}$, say $\tau(j, j')$, reduces to

$$\tau(j, j') = \mathbb{E}(X_j X_{j'}) - \mathbb{E}(X_j)\mathbb{E}(X_{j'}) = -\mathbb{E}(X_j)\mathbb{E}(X_{j'}),$$

where $\mathbb{E}(X_j)$ and $\mathbb{E}(X_{j'})$ are computed under the marginal PDF of X_j and $X_{j'}$, respectively. Therefore, the parameter vector $\boldsymbol{\beta}$ also governs the correlation. To illustrate it, we compute the correlation in the bivariate case (i.e., $d = 2$), where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)'$ and $\boldsymbol{\beta} = (\beta_1, \beta_2)'$. The parameter values we consider are $\alpha_1 = 0.25, 1.75, 3.5, 7.0$ and 15 , $\alpha_2 = 0.75, 2.5$ and 5.0 , $\beta_1 = -1.5, 0.0$ and 1.5 , and $\beta_2 = -2.5, 0.0$ and 2.5 . Table 3 lists the results for the correlation. Note that if β_1 or β_2 goes to zero, independently of the values for α_1 and α_2 , the correlation tends to zero, illustrating the fact that the parameter vector $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ dominates the dependence between the random variables. It is interesting to note that values of β_1 and β_2 with the same sign lead to a negative correlation and in situations where their signs are opposite, the correlation is positive.

Table 3. Correlation for the bivariate ABPN distribution.

		$\beta_1 = -1.5$			$\beta_1 = 0$			$\beta_1 = 1.5$		
α_1	α_2	$\beta_2 = -2.5$	$\beta_2 = 0$	$\beta_2 = 2.5$	$\beta_2 = -2.5$	$\beta_2 = 0$	$\beta_2 = 2.5$	$\beta_2 = -2.5$	$\beta_2 = 0$	$\beta_2 = 2.5$
0.25		-0.348	0.000	0.350	-0.001	2×10^{-4}	0.000	0.347	0.000	-0.348
1.75		-0.349	0.001	0.349	0.000	2×10^{-4}	-0.001	0.345	0.001	-0.348
3.5	0.75	-0.331	0.000	0.333	-0.002	4×10^{-4}	0.002	0.331	0.000	-0.331
7.0		-0.284	0.001	0.285	0.000	0×10^0	0.003	0.280	-0.000	-0.279
15		-0.189	0.002	0.186	-0.000	0×10^0	-0.002	0.183	0.002	-0.186
0.25		-0.384	5×10^{-4}	0.385	-0.000	2×10^{-4}	-0.001	0.382	0.001	-0.382
1.75		-0.383	-1×10^{-4}	0.383	0.000	0×10^0	-0.002	0.385	-0.001	-0.382
3.5	2.5	-0.369	-1×10^{-4}	0.367	0.000	3×10^{-4}	0.001	0.367	-0.001	-0.368
7.0		-0.319	-1×10^{-3}	0.321	-0.001	4×10^{-5}	0.002	0.316	-0.001	-0.316
15		-0.223	9×10^{-4}	0.220	-0.001	6×10^{-4}	-0.002	0.221	0.002	-0.219
0.25		-0.418	0.001	0.419	-0.000	4×10^{-4}	-0.000	0.414	-0.002	-0.417
1.75		-0.417	-0.001	0.419	-0.000	3×10^{-4}	-0.001	0.418	0.001	-0.418
3.5	5.0	-0.401	0.000	0.403	-0.001	1×10^{-4}	-0.001	0.402	0.003	-0.405
7.0		-0.352	0.001	0.354	-0.002	4×10^{-4}	-0.001	0.352	-0.001	-0.353
15		-0.255	0.002	0.258	-0.000	6×10^{-4}	-0.004	0.257	0.001	-0.253

4.2. Location-Scale Extension

The location-scale version of the MABPN model has joint PDF in the form

$$f(x; \zeta, \eta, \alpha, \beta) = 2\phi_d(x - \zeta, \Lambda) \prod_{j=1}^d \alpha_j c_{\alpha_j} \left\{ \Phi \left(\left| \frac{x_j - \zeta_j}{\eta_j} \right| \right) \right\}^{\alpha_j - 1} \Phi \left(\beta' \Lambda^{-1} (x - \zeta) \right), \quad x \in \mathbb{R}^d,$$

where the $d \times d$ matrix Λ and the d -vector ζ were previously defined. We shall use the notation $MABPN_d(\zeta, \eta, \alpha, \beta)$ to refer to this location-scale MABPN distribution.

4.3. Inference

Let $\theta = (\zeta', \eta', \alpha', \beta)'$ be the parameter vector of interest. The log-likelihood function for θ , given the observed sample $x_1 = (x_{11}, \dots, x_{1d})', \dots, x_n = (x_{n1}, \dots, x_{nd})'$ of size n from $X = (X_1, \dots, X_d)' \sim MABPN_d(\theta)$, is given by

$$\begin{aligned} \ell(\theta) = & n \sum_{j=1}^d [\log(\alpha_j) + \log(c_{\alpha_j}) - \log(\eta_j)] + \sum_{i=1}^n \sum_{j=1}^d \log(\phi(z_{ij})) \\ & + \sum_{i=1}^n \sum_{j=1}^d (\alpha_j - 1) \log(\Phi(|z_{ij}|)) + \sum_{i=1}^n \log \left(\Phi \left(\sum_{j=1}^d \beta_j z_{ij} \right) \right), \end{aligned} \tag{3}$$

where $z_{ij} = (x_{ij} - \zeta_j) / \eta_j$ for $i = 1, \dots, n$ and $j = 1, \dots, d$. The ML estimate $\hat{\theta} = (\hat{\zeta}', \hat{\eta}', \hat{\alpha}', \hat{\beta})'$ of $\theta = (\zeta', \eta', \alpha', \beta)'$, where $\hat{\zeta} = (\hat{\zeta}_1, \dots, \hat{\zeta}_d)'$, $\hat{\eta} = (\hat{\eta}_1, \dots, \hat{\eta}_d)'$, $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_d)'$ and $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_d)'$, is obtained by maximizing the log-likelihood function $\ell(\theta)$ with respect to θ by using, for example, the R function `optim(...)`. The first-order partial derivatives of (3) become ($j = 1, \dots, d$)

$$\begin{aligned} \frac{\partial \ell(\theta)}{\partial \zeta_j} &= n \{ \bar{z}_j - (\alpha_j - 1) \overline{\text{sgn}(z_j) w_j} - \beta_j \overline{z_j w_{1j}} \} / \eta_j, \\ \frac{\partial \ell(\theta)}{\partial \eta_j} &= n \{ \bar{z}_j^2 - 1 - (\alpha_j - 1) \overline{|z_j| w_j} - \beta_j \overline{z_j w_{1j}} \} / \eta_j, \\ \frac{\partial \ell(\theta)}{\partial \alpha_j} &= n \{ \bar{u}_j + 1/\alpha_j - \log 2 / (2^{\alpha_j} - 1) \}, \quad \frac{\partial \ell(\theta)}{\partial \beta_j} = n \bar{z}_j \overline{w_{1j}}, \end{aligned}$$

where $u_{ij} = \log(\Phi(|z_{ij}|))$, $w_{ij} = \phi(|z_{ij}|)/\Phi(|z_{ij}|)$, $w_{1ij} = \phi(z_{1ij})/\Phi(z_{1ij})$ and $z_{1ij} = \sum_{j=1}^d \beta_j z_{ij}$ for $i = 1, \dots, n$ and $j = 1, \dots, d$. We are using the notation $\bar{u}_j = n^{-1} \sum_{i=1}^n u_{ij}$, $\bar{w}_{1j} = n^{-1} \sum_{i=1}^n w_{1ij}$, $\overline{|z_j|w_j} = n^{-1} \sum_{i=1}^n |z_{ij}|w_{ij}$, and so on. The ML estimates can also be obtained by solving simultaneously the nonlinear system of equations $\partial \ell(\boldsymbol{\theta})/\partial \xi_j = 0$, $\partial \ell(\boldsymbol{\theta})/\partial \eta_j = 0$, $\partial \ell(\boldsymbol{\theta})/\partial \alpha_j = 0$ and $\partial \ell(\boldsymbol{\theta})/\partial \beta_j = 0$ for $j = 1, \dots, d$, which has no closed-form and, hence, the ML estimates need to be obtained through a numerical maximization of the log-likelihood function using nonlinear optimization algorithms.

4.4. Observed Information Matrix

Let $\mathbf{H}(\boldsymbol{\theta}) = -\partial^2 \ell(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$ be the $4d \times 4d$ observed information matrix, whose elements are denoted by $k_{\theta_j \theta_{j'}}$ for $j, j' = 1, \dots, d$. After some algebraic manipulations, we have that

$$\begin{aligned}
 k_{\xi_j \xi_j} &= n\{1 + (\alpha_j - 1)(\overline{|z_j|w_j} + \overline{w_j^2}) + \beta_j^2(\overline{z_{1j}w_{1j}} + \overline{w_{1j}^2})\}/\eta_j^2, \\
 k_{\xi_j \xi_{j'}} &= n \beta_j \beta_{j'} (\overline{z_{1j}w_{1j}} + \overline{w_{1j}^2})/\eta_j \eta_{j'}, \\
 k_{\xi_j \eta_j} &= n\{2\bar{z}_j + (\alpha_j - 1)(\overline{\text{sgn}(z_j)z_j^2 w_j} + \overline{z_j w_j^2} - \overline{\text{sgn}(z_j)w_j}) - \beta_j \overline{z_j w_{1j}} \\
 &\quad + \beta_j^2(\overline{z_j z_{1j} w_{1j}} + \overline{z_j w_{1j}^2})\}/\eta_j^2, \\
 k_{\xi_j \eta_{j'}} &= n \beta_j \beta_{j'} (\overline{z_{j'} z_{1j} w_{1j}} + \overline{z_{j'} w_{1j}^2})/\eta_j \eta_{j'}, \quad k_{\xi_j \alpha_j} = n \overline{\text{sgn}(z_j)w_j}/\eta_j, \\
 k_{\xi_j \beta_j} &= n\{\overline{w_{1j}} - \beta_j(\overline{z_j z_{1j} w_{1j}} + \overline{z_j w_{1j}^2})\}/\eta_j, \quad k_{\xi_j \beta_{j'}} = -n\{\beta_{j'}(\overline{z_{j'} z_{1j} w_{1j}} + \overline{z_{j'} w_{1j}^2})\}/\eta_j, \\
 k_{\eta_j \eta_j} &= n\{3\bar{z}_j^2 - 1 - (\alpha_j - 1)(2\overline{|z_j|w_j} - \overline{|z_j|^3 w_j} - \overline{z_j^2 w_j^2}) - 2\beta_j \overline{z_j w_{1j}} \\
 &\quad + \beta_j^2(\overline{z_j^2 z_{1j} w_{1j}} + \overline{z_j^2 w_{1j}^2})\}/\eta_j^2, \\
 k_{\eta_j \eta_{j'}} &= n \beta_j \beta_{j'} (\overline{z_j z_{j'} z_{1j} w_{1j}} + \overline{z_j z_{j'} w_{1j}^2})/\eta_j \eta_{j'}, \quad k_{\eta_j \alpha_j} = n \overline{|z_j|w_j}/\eta_j, \\
 k_{\eta_j \beta_j} &= n\{\overline{z_j w_{1j}} - \beta_j(\overline{z_j^2 z_{1j} w_{1j}} + \overline{z_j^2 w_{1j}^2})\}/\eta_j, \quad k_{\eta_j \beta_{j'}} = -n\beta_{j'}(\overline{z_j z_{j'} z_{1j} w_{1j}} + \overline{z_j z_{j'} w_{1j}^2})/\eta_j, \\
 k_{\alpha_j \alpha_j} &= n\{\alpha_j^{-2}(2^{\alpha_j} - 1)^2 - 2^{\alpha_j}(\log 2)^2\}/(2^{\alpha_j} - 1)^2, \quad k_{\beta_j \beta_j} = n\{\overline{z_j^2 z_{1j} w_{1j}} + \overline{z_j^2 w_{1j}^2}\}, \\
 k_{\beta_j \beta_{j'}} &= n\{\overline{z_j z_{j'} z_{1j} w_{1j}} + \overline{z_j z_{j'} w_{1j}^2}\}, \quad k_{\xi_j \alpha_{j'}} = k_{\eta_j \alpha_{j'}} = k_{\alpha_j \alpha_{j'}} = k_{\alpha_j \beta_j} = k_{\alpha_j \beta_{j'}} = 0.
 \end{aligned}$$

4.5. Expected Information Matrix

Let $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathbb{E}(\mathbf{H}(\boldsymbol{\theta}))$ be the $4d \times 4d$ expected information matrix, and $\sigma_{\theta_j \theta_{j'}} = \mathbb{E}(k_{\theta_j \theta_{j'}})$ be the corresponding elements for $j, j' = 1, \dots, d$. These elements can be computed by using numerical procedures. When n is large and under some mild regularity conditions, we have that $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \overset{d}{\sim} N_{4d}(\mathbf{0}_d, \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1})$. From this asymptotic normal distribution, approximate CIs for the model parameters are computed in the usual manner. In particular, for $\alpha_j = 1$ and $\lambda_j = 0$ ($j = 1, \dots, d$), the expected Fisher information matrix can be expressed as

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Lambda}_1^{-1} & \mathbf{0}'_d & \delta_0 \boldsymbol{\Lambda}^{-1} & (\sqrt{2/\pi}) \boldsymbol{\Lambda}^{-1} \\ \mathbf{0}'_d & 2\boldsymbol{\Lambda}_1^{-1} & \delta_1 \boldsymbol{\Lambda}^{-1} & \mathbf{0}'_d \\ \delta_0 \boldsymbol{\Lambda}^{-1} & \delta_1 \boldsymbol{\Lambda}^{-1} & [1 - 2(\log 2)^2] \mathbf{I}_d & \mathbf{0}'_d \\ \sqrt{2/\pi} \boldsymbol{\Lambda}^{-1} & \mathbf{0}'_d & \mathbf{0}'_d & (2/\pi) \mathbf{I}_d \end{bmatrix},$$

where $\Lambda_1 = \text{diag}\{\eta_1^2, \dots, \eta_d^2\}$, $\delta_r = \mathbb{E}(\text{sgn}(Z_j)Z_j^r\phi(|Z_j|)/\Phi(|Z_j|))$ with $Z_j \sim N(0,1)$ and $r = 0,1$, and $\text{sgn}(\cdot)$ denotes the signum function. After some algebraic manipulations, it can be shown that

$$|\Sigma(\theta)| = (-1)^d \frac{(4\delta_0^2)^d}{\prod_{j=1}^d \eta_j^4 \pi^j} \neq 0.$$

Therefore, the expected Fisher information matrix of the MABPN distribution is nonsingular at the vicinity of symmetry. This is not the case, however, with the multivariate SN distribution (Azzalini and Dalla-Valle [15]), whose expected Fisher information matrix is singular at the vicinity of symmetry, i.e., $\lambda = \mathbf{0}_d$.

5. Numerical Illustration

In this section, we present an application of the proposed bivariate ABPN distribution to real data for illustrative purposes. For the sake of comparison, we also consider the bivariate SN distribution (Azzalini and Dalla Valle [15]), and the conditional bivariate SN distribution (Arnold et al. [14]). All bivariate distributions were fitted using the location-scale extension. We shall consider the real data (see, for example, Fisher [30]) corresponding to measurements of the flowers of fifty plants each of the three species found growing together in the same colony, namely: Iris-setosa, Iris-versicolor, and Iris-virginica. Two flower measurements are considered: sepal length, and sepal width. 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 $n = 150$; that is, 150 sepal lengths, and 150 sepal widths. The observed value of the Shapiro-Wilk test for multivariate normality (see Villasenor and González [31]) is 0.9845 (p -value = 0.04). We also compute the multivariate skewness based on Mardia [32]. The observed value of the multivariate skewness is 0.37 (p -value = 0.055), which clearly suggests the presence of skewness and hence of nonnormality. 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 now consider the bivariate SN distribution, the conditional bivariate SN distribution, and the bivariate ABPN distribution to fit these real data. In order to compare the model fitting of these competing bivariate distributions, we make use of the Akaike information criterion (AIC). For fitting the bivariate SN model, we use the R function `msn.m1e`, leading to the following ML estimates (standard errors in parentheses): $\hat{\mu}_1 = 4.915(0.196)$, $\hat{\mu}_2 = 3.051(0.196)$, $\hat{\lambda}_1 = 2.554(0.972)$ and $\hat{\lambda}_2 = 0.220(0.279)$. The estimated variance-covariance matrix is

$$\hat{\Omega} = \begin{bmatrix} 1.543 & -0.037 \\ -0.037 & 0.189 \end{bmatrix}.$$

For the the bivariate SN model, we obtain $\text{AIC} = 550.56$. The ML estimates of the conditional bivariate SN model parameters (standard errors in parentheses) are: $\hat{\mu}_1 = 5.867(0.065)$, $\hat{\mu}_2 = 3.055(0.036)$, $\hat{\sigma}_1 = 0.794(0.043)$, $\hat{\sigma}_2 = 0.438(0.026)$ and $\hat{\lambda} = -0.224(0.110)$. For the the bivariate conditional SN model, we obtain $\text{AIC} = 555.10$. Also, the ML estimates of the proposed bivariate ABPN model parameters (standard errors in parentheses) are: $\hat{\zeta}_1 = 5.917(0.051)$, $\hat{\zeta}_2 = 2.372(0.056)$, $\hat{\eta}_1 = 0.729(0.050)$, $\hat{\eta}_2 = 0.644(0.041)$, $\hat{\alpha}_1 = 2.378(0.605)$, $\hat{\alpha}_2 = 3.718(0.845)$, $\hat{\beta}_1 = -0.481(0.300)$ and $\hat{\beta}_2 = 3.414(0.860)$. For the the proposed bivariate ABPN model, we have that $\text{AIC} = 546.80$, which indicates that the proposed bivariate ABPN model outperforms the bivariate SN distribution and the conditional bivariate SN distribution to model the bivariate data. Finally, Figure 2 displays the scatter plot of real data and contour levels of the fitted bivariate PDFs. Visual inspection reveals a satisfactory fit of the bivariate ABPN PDF to the real bivariate data.

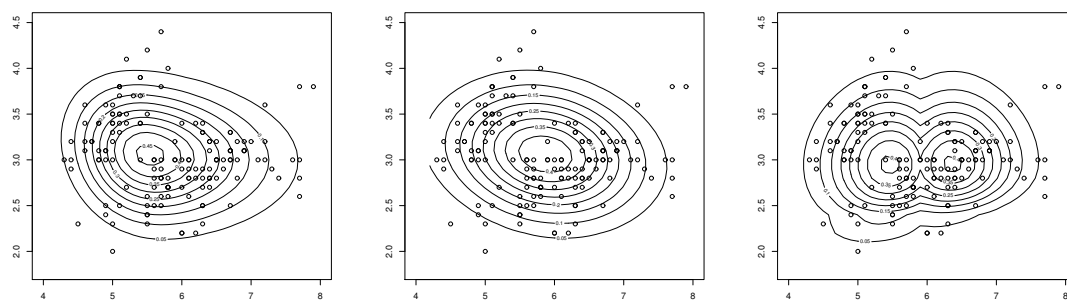


Figure 2. Scatter plot of real data and contour level of the fitted PDFs: (left) bivariate SN distribution, (middle) conditional SN distribution, and (right) bivariate ABPN distribution.

We would like to point out that there is some uncertainty as to what constitutes a “substantial” difference in AIC values in practical situations. The empirical evidence scale of [33] uses the AIC difference $Y_m = AIC_m - AIC_{\min}$ over all candidate models, where $m = 1, \dots, T$, and T denotes the total number of models considered. The models with values of $Y_m \in [0, 2]$ have substantial support to be considered as good as the best approximating model. Two additional measures are then introduced to provide the “strengths” of each model: the evidence ratio (ER_m) and the weight (w_m) of model m . These measures are defined as

$$ER_m = \frac{\exp(-Y_{\min}/2)}{\exp(-Y_m/2)} = \exp(Y_m/2), \quad w_m = \frac{\exp(-Y_m/2)}{\sum_{m=1}^T \exp(-Y_m/2)},$$

and $Y_{\min} = 0$. The values of ER_m can be interpreted as the greater likelihood of the best approximating model with respect to model m , whereas the values of w_m can be interpreted as the probability of a given model being the best approximating model. The values of these measures are given in Table 4. For example, we conclude from this table that the bivariate ABPN distribution is about 7 and 63 times more likely to be the best approximating model than the bivariate SN distribution and conditional SN distribution, respectively. Additionally, the chance of these models with respect to the bivariate ABPN distribution is also non-existent. The best bivariate ABPN distribution has a weight approximately 0.856; that is, there is (approximately) a 86% chance that it really is the best approximating model among the current models to describe these data. Notice that, by definition, the strength of evidence in favor of model i over the model j can be obtained simply by considering w_i/w_j .

Table 4. Some measures for the fitted models.

Link Model	Y_m	ER_m	w_m
SN	3.76	6.55	0.131
Conditional SN	8.30	63.43	0.013
ABPN	0.00	1.00	0.856

6. Elliptical Family Extension

In the previous sections, the d -dimensional normal distribution played an important role in deriving the multivariate ABPN distribution. In this section, we extend those results by considering the elliptical family of distributions. For the case of a random variable (one-dimensional case), elliptical distributions correspond to all symmetric distributions in \mathbb{R} . Specifically, a random variable X follows an symmetric distribution if its PDF is given by

$$f_X(x) = \frac{c}{\eta} g\left(\frac{(x - \xi)^2}{\eta^2}\right), \quad x \in \mathbb{R}, \tag{4}$$

where $g(u)$ (for $u \geq 0$) is a nonnegative real function and corresponds to the kernel of the PDF, and c is a normalizing constant such that $f_X(x)$ is a PDF. We use the notation $EC(\xi, \eta; g)$ according to (4). In general, $\xi \in \mathbb{R}$ is the location parameter and coincides with the mean if the first moment of the distribution exists, and $\eta > 0$ is the scale parameter.

First, we shall introduce a multivariate elliptical PN family of distributions. A random vector follows a multivariate elliptical PN distribution if its joint PDF is given by

$$\varphi_F(\mathbf{x}; \boldsymbol{\alpha}) = \prod_{j=1}^d \alpha_j c_{\alpha_j} f_j(x_j) \{F_j(|x_j|)\}^{\alpha_j-1}, \quad \mathbf{x} \in \mathbb{R}^d, \quad (5)$$

where f_j is defined in (4), and F_j corresponds to its CDF for $j = 1, \dots, d$. Clearly, since f_j belongs to the elliptical family, then it can be easily shown that $\varphi_F(-\mathbf{x}; \boldsymbol{\alpha}) = \varphi_F(\mathbf{x}; \boldsymbol{\alpha})$ so that φ_F is a symmetric (around zero) PDF. We use the notation $MES_F(\boldsymbol{\alpha})$ to refer to this distribution. From Lemma 1 of Gupta and Chang [24], we have the following generalization.

Proposition 3. Let $\mathbf{Y} = (Y_1, \dots, Y_d)'$ be a random vector with joint PDF φ_F , and $\mathbf{Z} = (Z_1, \dots, Z_d)'$ be a random vector with absolutely continuous CDF G . Then

$$\varphi_{GF}(\mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}) = 2 \prod_{j=1}^d \alpha_j c_{\alpha_j} f_j(x_j) \{F_j(|x_j|)\}^{\alpha_j-1} G(\boldsymbol{\beta}'\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d,$$

is a joint PDF of a random vector $\mathbf{X} = (X_1, \dots, X_d)'$ for any $\boldsymbol{\beta} \in \mathbb{R}^d$.

Proof. We will use the methodology given in Gupta and Chang [24]. Given that $\varphi_F(\mathbf{x}; \boldsymbol{\alpha}) \geq 0$ and $G(\boldsymbol{\beta}'\mathbf{x}) \geq 0$ for any $\mathbf{x} \in \mathbb{R}^d$, it then follows that $\varphi_{GF}(\mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}) \geq 0$. Let

$$h(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} 2 \prod_{j=1}^d \alpha_j c_{\alpha_j} f_j(x_j) \{F_j(|x_j|)\}^{\alpha_j-1} G(\boldsymbol{\beta}'\mathbf{x}) \prod_{j=1}^d dx_j.$$

Lebesgue's dominated convergence theorem implies that

$$\begin{aligned} \frac{\partial h(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \beta_k} &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} 2 \prod_{j=1}^d \alpha_j c_{\alpha_j} f_j(x_j) \{F_j(|x_j|)\}^{\alpha_j-1} \frac{\partial G(\boldsymbol{\beta}'\mathbf{x})}{\partial \beta_k} \prod_{j=1}^d dx_j \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} 2x_k \prod_{j=1}^d \alpha_j c_{\alpha_j} f_j(x_j) \{F_j(|x_j|)\}^{\alpha_j-1} G'(\boldsymbol{\beta}'\mathbf{x}) \prod_{j=1}^d dx_j. \end{aligned}$$

Moreover, given that $2x_k \prod_{j=1}^d \alpha_j c_{\alpha_j} f_j(x_j) \{F_j(|x_j|)\}^{\alpha_j-1} G'(\boldsymbol{\beta}'\mathbf{x})$ is an odd function, we have

$$\frac{\partial h(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \beta_k} = 0, \quad j = 1, \dots, d.$$

Therefore, it allows us to conclude that $h(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is a constant and, hence, given that $h(\boldsymbol{\alpha}, \mathbf{0}) = 1$, we have that $\varphi_{GF}(\mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta})$ is a joint PDF. \square

The new multivariate distribution defined in Proposition 3 will be denoted by $MESS_{GF}(\boldsymbol{\alpha}, \boldsymbol{\beta})$. We have the following proposition.

Proposition 4. Let $\mathbf{X} = (X_1, \dots, X_d)' \sim MESS_{GF}(\boldsymbol{\alpha}, \boldsymbol{\beta})$. We have that

1. If $\boldsymbol{\beta} = \mathbf{0}_d$, then $\mathbf{X} = (X_1, \dots, X_d)' \sim MES_F(\boldsymbol{\alpha})$.
2. $-\mathbf{X} \sim MESS_{GF}(\boldsymbol{\alpha}, -\boldsymbol{\beta})$.

3. For $d = 2$, regression functions are of linear type

$$\mathbb{E}(X_j|X_{j'} = x_{j'}) = \frac{M_j(x_j)}{N_{j'}(x_{j'})},$$

where

$$M_j(x_j) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_j f_j(x_j) \{F_j(|x_j|)\}^{\alpha_j-1} G(\beta_1 x_1 + \beta_2 x_2) dx_1 dx_2,$$

$$N_j(x_j) = \int_{-\infty}^{\infty} f_j(x_j) \{F_j(|x_j|)\}^{\alpha_j-1} G(\beta_1 x_1 + \beta_2 x_2) dx_1 dx_2.$$

The product moment of $\mathbf{X} = (X_1, \dots, X_d)' \sim \text{MESS}_{GF}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ are provided in the next proposition.

Proposition 5. If $\mathbf{X} = (X_1, \dots, X_d)' \sim \text{MESS}_{GF}(\boldsymbol{\alpha}, \boldsymbol{\beta})$, then

$$\mathbb{E}[X_1^{r_1} X_2^{r_2} \dots X_d^{r_d}] = \mathbb{E}[Y_1^{r_1} Y_2^{r_2} \dots Y_d^{r_d}],$$

where $\mathbf{Y} \sim \text{MES}_F(\boldsymbol{\alpha})$.

Proof. Let $\mathbf{X} = (X_1, \dots, X_d)' \sim \text{MESS}_{GF}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and $\mathbf{t} = (t_1, \dots, t_d) \in \mathbb{R}^d$. Also, let $\psi_{\mathbf{X}}(\mathbf{t})$ be the characteristic function of \mathbf{X} . We have that

$$\psi_{\mathbf{X}}(\mathbf{t}) + \psi_{\mathbf{X}}(-\mathbf{t}) = 2\psi_{\mathbf{Y}}(\mathbf{t}) = 2 \prod_{j=1}^d \psi_{Y_j}(t_j),$$

where $Y_j \sim \alpha_j c_{\alpha_j} f_j(y_j) (F(|y_j|))^{\alpha_j-1}$ is independent of $\boldsymbol{\beta}$. Thus,

$$\frac{\partial^m \psi_{\mathbf{Y}}(\mathbf{t})}{\partial t_1^{r_1} \partial t_2^{r_2} \dots \partial t_d^{r_d}} < \infty,$$

for $m = r_1 + r_2 + \dots + r_d$ even, so that

$$\mathbb{E}[X_1^{r_1} X_2^{r_2} \dots X_d^{r_d}] = \mathbb{E}[Y_1^{r_1} Y_2^{r_2} \dots Y_d^{r_d}].$$

The previous result implies that even moments of \mathbf{X} do not depend on $\boldsymbol{\beta}$, so that

$$\mathbb{E}(X_1^{r_1} X_2^{r_2} \dots X_d^{r_d}) = \begin{cases} 0, & \text{if any } r_j \text{ is odd,} \\ 2^{1+\sum_{j=1}^d r_j} \prod_{j=1}^d \mathbb{E}(Y_{j+}^{r_j}), & \text{if all } r_j \text{ are even,} \end{cases}$$

where $\mathbb{E}(Y_{j+}^{r_j})$ is the moment of the positive part of the variable Y_j as defined before. \square

6.1. ML Estimation

Let $\mathbf{x}_1 = (x_{11}, \dots, x_{1d})'$, \dots , $\mathbf{x}_n = (x_{n1}, \dots, x_{nd})'$ be an observed sample of size n from $\mathbf{X} = (X_1, \dots, X_d)' \sim \text{MESS}_{GF}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \boldsymbol{\beta})$. The log-likelihood function takes the form

$$\ell(\boldsymbol{\theta}) = n \sum_{j=1}^d (\log(\alpha_j) + \log(c_{\alpha_j}) - \log(\eta_j)) + \sum_{i=1}^n \sum_{j=1}^d \log(f_j(z_{ij}))$$

$$+ \sum_{i=1}^n \sum_{j=1}^d (\alpha_j - 1) \log(F_j(|z_{ij}|)) + \sum_{i=1}^n \log \left(G \left(\sum_{j=1}^d \beta_j z_{ij} \right) \right),$$

where $z_{ij} = (x_{ij} - \xi_j)/\eta_j$ for $i = 1, \dots, n$ and $j = 1, \dots, d$, and $\theta = (\zeta', \eta', \alpha', \beta')'$. The first-order partial derivatives are given by

$$\frac{\partial \ell(\theta)}{\partial \xi_j} = -\frac{1}{\eta_j} \left\{ \sum_{i=1}^n \frac{f'_j(z_{ij})}{f_j(z_{ij})} + (\alpha_j - 1) \sum_{i=1}^n \operatorname{sgn}(z_{ij}) \frac{f_j(|z_{ij}|)}{F_j(|z_{ij}|)} + \beta_j \sum_{i=1}^n \frac{G'(z_{1ij})}{G(z_{1ij})} \right\},$$

$$\begin{aligned} \frac{\partial \ell(\theta)}{\partial \eta_j} &= -\frac{1}{\eta_j} \left\{ n + \sum_{i=1}^n z_{ij} \frac{f'_j(z_{ij})}{f_j(z_{ij})} + (\alpha_j - 1) \sum_{i=1}^n |z_{ij}| \frac{f_j(|z_{ij}|)}{F_j(|z_{ij}|)} \right. \\ &\quad \left. + \beta_j \sum_{i=1}^n z_{ij} \frac{G'(z_{1ij})}{G(z_{1ij})} \right\}, \end{aligned}$$

$$\frac{\partial \ell(\theta)}{\partial \alpha_j} = \frac{n}{\alpha_j} - \frac{n \log 2}{2^{\alpha_j} - 1} + \sum_{i=1}^n \log(F_j(|z_{ij}|)), \quad \frac{\partial \ell(\theta)}{\partial \beta_j} = \sum_{i=1}^n z_{ij} \frac{G'(z_{1ij})}{G(z_{1ij})},$$

where $G'(u) = dG(u)/du$.

6.2. Expected Information Matrix

Let $\sigma_{\xi_j \xi_j}$, $\sigma_{\xi_j \xi_{j'}}$, \dots , $\sigma_{\beta_j \beta_j}$ be the elements of the expected information matrix. Also, define $U_1 = f'_j(Z_j)/f_j(Z_j)$, $U_2 = f''_j(Z_j)/f_j(Z_j)$, $W_1 = f_j(|Z_j|)/F_j(|Z_j|)$, $W_2 = f'_j(|Z_j|)/F_j(|Z_j|)$, $K_1 = G'(Z_{1j})/G(Z_{1j})$ and $K_2 = G''(Z_{1j})/G(Z_{1j})$, where $Z_{1j} = \sum_{i=1}^d \beta_j Z_j$, $f'(u) = df(u)/du$ and $f''(u) = d^2f(u)/du^2$. We have

$$\begin{aligned} \sigma_{\xi_j \xi_j} &= \{ \mathbb{E}(U_1^2) - \mathbb{E}(U_2) + (\alpha_j - 1)[\mathbb{E}(W_1^2) - \mathbb{E}(W_2)] + \beta_j^2[E(K_1^2) - E(K_2)] \} / \eta_j^2, \\ &= \{ (a_{02} - a_0) + (\alpha_j - 1)(b_{02} - b_0) + \beta_j^2(c_{02} - c_0) \} / \eta_j^2, \end{aligned}$$

$$\sigma_{\xi_j \xi_{j'}} = \beta_j \beta_{j'} (c_{02} - c_0) / \eta_j \eta_{j'},$$

$$\begin{aligned} \sigma_{\xi_j \eta_j} &= \left\{ \mathbb{E}(Z_j U_1^2) - \mathbb{E}(Z_j U_2) - \mathbb{E}(U_1) - \beta_j \mathbb{E}(K_1) + \beta_j^2[\mathbb{E}(Z_j K_1) - \mathbb{E}(Z_j K_2)] \right. \\ &\quad \left. + (\alpha_j - 1)[\mathbb{E}(Z_j W_1) - \mathbb{E}(Z_j W_2) - \mathbb{E}(\operatorname{sgn}(Z_j) W_1)] \right\} / \eta_j^2, \\ &= \{ (a_{12} - a_1 - a_{01}) + (\alpha_j - 1)(b_{12} - b_1 - d_0) - \beta_j c_{01} + \beta_j^2(c_{12} - c_1) \} / \eta_j^2, \end{aligned}$$

$$\sigma_{\xi_j \eta_{j'}} = \beta_j \beta_{j'} (c_{12} - c_1) / \eta_j \eta_{j'},$$

$$\sigma_{\xi_j \alpha_j} = \{ \mathbb{E}(\operatorname{sgn}(Z_j) W_1) \} / \eta_j = d_0 / \eta_j,$$

$$\sigma_{\xi_j \beta_j} = \{ \mathbb{E}(K_1) - \beta_j [\mathbb{E}(Z_j K_1^2) - \mathbb{E}(Z_j K_2)] \} / \eta_j = \{ c_{01} - \beta_j (c_{12} - c_1) \} / \eta_j,$$

$$\sigma_{\xi_j \beta_{j'}} = -\beta_j \{ \mathbb{E}(Z_{j'} K_1^2) - \mathbb{E}(Z_{j'} K_2) \} / \eta_j,$$

$$\begin{aligned} \sigma_{\eta_j \eta_j} &= \left\{ -1 - 2\mathbb{E}(Z_j U_1) + \mathbb{E}(Z_j^2 U_1^2) - \mathbb{E}(Z_j^2 U_2) - 2\beta_j \mathbb{E}(Z_j K_1) + \beta_j^2 \mathbb{E}(Z_j^2 K_1^2) \right. \\ &\quad \left. - \beta_j^2 \mathbb{E}(Z_j^2 K_2) + (\alpha_j - 1)[\mathbb{E}(Z_j^2 W_1^2) - \mathbb{E}(Z_j^2 W_2) - 2\mathbb{E}(|Z_j| W_1)] \right\} / \eta_j^2, \\ &= \left\{ (a_{22} - 1 - 2a_{11} - a_2) + (\alpha_j - 1)(b_{22} - b_2 - 2d_1) - 2\beta_j c_{11} \right. \\ &\quad \left. + \beta_j^2 (c_{22} - c_2) \right\} / \eta_j^2, \end{aligned}$$

$$\begin{aligned}
\sigma_{\eta_j \eta_{j'}} &= \beta_j \beta_{j'} [\mathbb{E}(Z_j Z_{j'} K_1^2) - \mathbb{E}(Z_j Z_{j'} K_2)] / \eta_j \eta_{j'}, \\
\sigma_{\eta_j \alpha_j} &= \mathbb{E}(|Z_j| W_1) / \eta_j = d_1 / \eta_j, \\
\sigma_{\eta_j \beta_j} &= \{\mathbb{E}(Z_j K_1) - \beta_j [\mathbb{E}(Z_j^2 K_1^2) - \mathbb{E}(Z_j^2 K_2)]\} / \eta_j = \{c_{11} - \beta_j (c_{22} - c_2)\} / \eta_j, \\
\sigma_{\eta_j \beta_{j'}} &= -\beta_{j'} [\mathbb{E}(Z_j Z_{j'} K_1^2) - \mathbb{E}(Z_j Z_{j'} K_2)] / \eta_j, \\
\sigma_{\alpha_j \alpha_j} &= 1 / \alpha_j - (\log 2)^2 2^{\alpha_j} / (2^{\alpha_j} - 1)^2, \\
\sigma_{\beta_j \beta_j} &= \mathbb{E}(Z_j^2 K_1) - \mathbb{E}(Z_j^2 K_2) = c_{22} - c_2, \\
\sigma_{\beta_j \beta_{j'}} &= \mathbb{E}(Z_j Z_{j'} K_1) - \mathbb{E}(Z_j Z_{j'} K_2), \\
\sigma_{\xi_j \alpha_{j'}} &= \sigma_{\eta_j \alpha_{j'}} = \sigma_{\alpha_j \alpha_{j'}} = \sigma_{\alpha_j \beta_j} = \sigma_{\alpha_j \beta_{j'}} = 0,
\end{aligned}$$

where $a_{rs} = \mathbb{E}(Z_j^r (f_j'(Z_j) / f_j(Z_j))^s)$, $a_r = \mathbb{E}(Z_j^r f_j''(Z_j) / f_j(Z_j))$, $b_{rs} = \mathbb{E}(Z_j^r (f_j(|Z_j|) / F_j(|Z_j|))^s)$, $b_r = \mathbb{E}(Z_j^r f_j'(|Z_j|) / F_j(|Z_j|))$, $c_{rs} = \mathbb{E}(Z_j^r (G_j'(Z_{1j}) / G_j(Z_{1j}))^s)$, $c_r = \mathbb{E}(Z_j^r G_j''(Z_{1j}) / G_j(Z_{1j}))$ and $d_r = \mathbb{E}(\text{sgn}(Z_j) Z_j^r f_j(|Z_j|) / F_j(|Z_j|))$ for $r = 0, 1, 2$ and $s = 1, 2$.

7. Concluding Remarks

The univariate power-normal distribution has been quite useful in the modeling of many types of real data. On the other hand, extensions of the univariate power-normal distribution to the multivariate setup have been little explored in the statistic literature. We have proposed new multivariate power-normal distributions, which are quite simple and can be useful in the modeling of multivariate data. The new multivariate power-normal distributions are absolutely continuous distributions. The joint probability density functions of the new multivariate power-normal distributions do not involve any complicated function and, hence, they can be computed easily. By employing the frequentist approach, the estimation of the multivariate power-normal distribution parameters is conducted by the maximum likelihood method. We also provide closed-form expressions for the observed and expected Fisher information matrices. We illustrate the methodology developed in this paper by means of an application to real data. We verify through the real data application that the proposed bivariate power-normal distribution was superior to the well-known bivariate skew-normal distribution (Azzalini and Dalla Valle [15]), as well as conditional bivariate skew-normal distribution (Arnold et al. [14]). Finally, it is worth stressing that the formulas related with the multivariate power-normal distributions are manageable (such as log-likelihood function, score function, and observed and expected Fisher information matrices, etc), and with the use of modern computer resources and its numerical capabilities, the proposed multivariate distributions may prove to be an useful addition to the arsenal of applied statisticians. Finally, we have also introduced in this paper multivariate PN distributions by considering the elliptical family of distributions, and discussed some of its properties.

Author Contributions: Individual contributions to this article: conceptualization, G.M.-F., A.J.L., and H.S.S.; methodology, G.M.-F., A.J.L., and H.S.S.; validation, G.M.-F., A.J.L., and H.S.S.; software, G.M.-F., A.J.L., and H.S.S.; investigation, G.M.-F., A.J.L., and H.S.S.; resources, G.M.-F., A.J.L., and H.S.S.; writing—original draft preparation, G.M.-F., A.J.L., and H.S.S.; writing—review and editing, G.M.-F., A.J.L., and H.S.S.; funding acquisition, A.J.L.

Funding: This research was funded in part by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq/Brazil) under grant 301808/2016–3.

Acknowledgments: We thank the anonymous referees for helpful suggestions which improved the article.

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

References

1. Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **1985**, *12*, 171–178.

2. Fernandez, C.; Steel, M.F.J. On Bayesian modeling of fat tails and skewness. *J. Am. Stat. Assoc.* **1998**, *93*, 359–371.
3. Mudholkar, G.S.; Hutson, A.D. The epsilon-skew-normal distribution for analyzing near-normal data. *J. Stat. Plan. Inference* **2000**, *83*, 291–309. [[CrossRef](#)]
4. Durrans, S.R. Distributions of fractional order statistics in hydrology. *Water Resour. Res.* **1992**, *28*, 1649–1655. [[CrossRef](#)]
5. Pewsey, A.; Gómez, H.W.; Bolfarine, H. Likelihood-based inference for power distributions. *Test* **2012**, *21*, 775–789. [[CrossRef](#)]
6. 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](#)]
7. Henze, N. A probabilistic representation of the skew-normal distribution. *Scand. J. Stat.* **1986**, *13*, 271–275.
8. Chiogna, M. A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution. *Stat. Methods Appl.* **2005**, *14*, 331–341. [[CrossRef](#)]
9. Pewsey, A. Problems of inference for Azzalini’s skew-normal distribution. *J. Appl. Stat.* **2000**, *27*, 859–870. [[CrossRef](#)]
10. Gómez, H.W.; Salinas, H.S.; Bolfarine, H. Generalized skew-normal models: Properties and inference. *Statistics* **2006**, *40*, 495–505. [[CrossRef](#)]
11. Gómez, H.W.; Venegas, O.; Bolfarine, H. Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics* **2007**, *18*, 395–407.
12. Gupta R.D.; Gupta, R.C. Analyzing skewed data by power normal model. *Test* **2008**, *17*, 197–210. [[CrossRef](#)]
13. Bolfarine, H.; Martínez-Flórez, G.; Salinas, H.S. Bimodal symmetric-asymmetric power-normal families. *Commun. -Stat. -Theory Methods* **2018**, *47*, 259–276. [[CrossRef](#)]
14. Arnold, B.C.; Castillo, E.; Sarabia, J.M. Conditionally specified multivariate skewed distributions. *Sankhya A* **2002**, *64*, 206–226.
15. Azzalini, A.; Dalla Valle, A. The multivariate skew-normal distribution. *Biometrika* **1996**, *83*, 715–726. [[CrossRef](#)]
16. Azzalini, A.; Capitanio, A. Statistical applications of the multivariate skew normal distribution. *J. R. Stat. Soc.* **1999**, *61*, 579–602. [[CrossRef](#)]
17. Gupta, A.K.; Huang, W.J. Quadratic forms in skew normal variates. *J. Math. Anal. Appl.* **2002**, *273*, 558–564. [[CrossRef](#)]
18. Gupta, A.K.; Chang, F.C.; Huang, W.J. Some skew-symmetric models. *Random Oper. Stoch. Equ.* **2002**, *10*, 133–140. [[CrossRef](#)]
19. Genton, M.G. *Skew-Elliptical Distributions and Their Applications: A Journey beyond Normality*; Chapman and Hall/CRC: New York, NY, USA, 2004.
20. Arellano-Valle, R.B.; Azzalini, A. On the unification of families of skew-normal distributions. *Scand. J. Stat.* **2006**, *33*, 561–574. [[CrossRef](#)]
21. Arellano-Valle, R.B.; Azzalini, A. The centred parametrization for the multivariate skew-normal distribution. *J. Multivar. Anal.* **2008**, *99*, 1362–1382. [[CrossRef](#)]
22. Arellano-Valle, R.B.; Genton, M.G. On fundamental skew distributions. *J. Multivar. Anal.* **2005**, *96*, 93–116. [[CrossRef](#)]
23. Gupta, A.K.; Chen, J.T. A class of multivariate skew-normal models. *Ann. Inst. Stat. Math.* **2004**, *56*, 305–315. [[CrossRef](#)]
24. Gupta, A.K.; Chang, F.C. Multivariate skew-symmetric distributions. *Appl. Math. Lett.* **2003**, *16*, 643–646. [[CrossRef](#)]
25. Azzalini, A.; Capitanio, A. *The Skew-Normal and Related Families*; IMS Monographs Series; Cambridge University Press: UK, London, 2014.
26. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018.
27. Cox, D.R.; Hinkley, D.V. *Theoretical Statistics*; Chapman and Hall: London, UK, 1974.
28. Cox, D.R.; Reid, N. Parameter orthogonality and approximate conditional inference. *J. R. Stat. Soc. B* **1987**, *40*, 1–39. [[CrossRef](#)]
29. Pesta, M. Total least squares and bootstrapping with application in calibration. *Statistics* **2013**, *47*, 966–991. [[CrossRef](#)]

30. Fisher, R.A. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **1936**, *7*, 179–188. [[CrossRef](#)]
31. Villasenor, J.A.; González, E. A generalization of Shapiro-Wilk's test for multivariate normality. *Commun. -Stat. -Theory Methods* **2009**, *38*, 1870–1883. [[CrossRef](#)]
32. Mardia, K.V. Measures of multivariate skewness and kurtosis with applications. *Biometrika* **1970**, *57*, 519–530. [[CrossRef](#)]
33. Burnham, K.P.; Anderson, D.R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*; Springer: New York, NY, USA, 2002.



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).