Journal of Multivariate Analysis 100 (2009) 91–101

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Shape mixtures of multivariate skew-normal distributions

Reinaldo B. Arellano-Valle [a], Marc G. Genton [b,c,*], Rosangela H. Loschi [d]

[a] *Facultad de Matematicas, Pontificia Universidad Catolica de Chile, Casilla 306, Santiago 22, Chile*

[b] *Department of Econometrics, University of Geneva, Bd du Pont-d'Arve 40, CH-1211 Geneva 4, Switzerland*

[c] *Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA*

[d] *Departamento de Estatística, Universidade Federal de Minas Gerais, 31270-901 - Belo Horizonte - MG, Brazil*

ARTICLE INFO

Article history:
Received 17 January 2007
Available online 7 April 2008

AMS 1991 subject classifications:
62H05
62E15

Keywords:
Bayes
Conjugacy
Shape parameter
Skewness
Skew-normal distribution
Regression model
Robustness

ABSTRACT

Classes of shape mixtures of independent and dependent multivariate skew-normal distributions are considered and some of their main properties are studied. If interpreted from a Bayesian point of view, the results obtained in this paper bring tractability to the problem of inference for the shape parameter, that is, the posterior distribution can be written in analytic form. Robust inference for location and scale parameters is also obtained under particular conditions.

1. Introduction

The empirical distribution of data sets often exhibits skewness and tails that are lighter or heavier than the normal distribution. For this reason, the construction of flexible parametric non-normal multivariate distributions has received renewed attention in recent years. An interesting approach is to multiply a symmetric (for example normal or Student's t) probability density function (pdf) by a function that introduces skewness in the resulting pdf. This idea was first formalized by Azzalini [6] (see also [7]), who defined a univariate skew-normal random variable $Y \sim SN(\mu, \sigma^2, \alpha)$ with pdf:

$$f(y \mid \mu, \sigma, \alpha) = (2/\sigma)\,\phi(z)\,\Phi(\alpha z), \quad y \in \mathbb{R}, \tag{1}$$

where $z = (y - \mu)/\sigma$, and ϕ and Φ denote the standard normal pdf and cumulative distribution function (cdf), respectively. Here, $\mu \in \mathbb{R}$, $\sigma > 0$ and $\alpha \in \mathbb{R}$ denote the location, scale and shape parameters, respectively. The symmetric normal pdf $N(\mu, \sigma^2)$ is retrieved by setting $\alpha = 0$ in (1), whereas skewness is obtained whenever $\alpha \neq 0$.

An extension of the univariate SN distribution (1) has been introduced by Arellano-Valle et al. [5]. Specifically, they defined a univariate skew-generalized normal random variable $Y \sim SGN(\mu, \sigma^2, \alpha_1, \alpha_2)$ with pdf:

$$f(y \mid \mu, \sigma, \alpha_1, \alpha_2) = (2/\sigma)\phi(z)\,\Phi\left(\alpha_1 z/\sqrt{1 + \alpha_2 z^2}\right), \quad y \in \mathbb{R}, \tag{2}$$

* Corresponding author at: Department of Econometrics, University of Geneva, Bd du Pont-d'Arve 40, CH-1211 Geneva 4, Switzerland.
E-mail address: Marc.Genton@metri.unige.ch (M.G. Genton).

0047-259X/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2008.03.009

where $\alpha_1 \in \mathbb{R}$ and $\alpha_2 > 0$. It follows from (2) that $SGN(\mu, \sigma^2, 0, \alpha_2) = N(\mu, \sigma^2)$ for any $\alpha_2 > 0$, $SGN(\mu, \sigma^2, \alpha_1, \infty) = N(\mu, \sigma^2)$ for any $\alpha_1 \in \mathbb{R}$, and $SGN(\mu, \sigma^2, \alpha_1, 0) = SN(\mu, \sigma^2, \alpha_1)$. Another important special case is the skew-curved normal distribution obtained by letting $\alpha_2 = \alpha_1^2 = \alpha$ and denoted by $Y \sim SCN(\mu, \sigma^2, \alpha)$. Further properties and applications of the *SGN* model were considered in [5]. In particular, it was established that the *SGN* distribution can be represented as a shape mixture of the *SN* distribution (1) by taking a normal mixing distribution for the shape parameter, that is:

$$[Y \mid S = s] \sim SN(\mu, \sigma^2, s) \quad \text{and} \quad S \sim N(\alpha_1, \alpha_2). \tag{3}$$

Moreover, from (3) it can be observed that the conditional distribution of the shape random variable S (controlling the skewness) given $Y = y$ depends on the data y as well as the location–scale parameter (μ, σ^2) through $z = (y - \mu)/\sigma$ only. This conditional distribution has pdf:

$$\pi(s \mid z, \alpha_1, \alpha_2) = \frac{\phi(t)\,\Phi(zs)}{\sqrt{\alpha_2}\,\Phi\left(\alpha_1 z/\sqrt{1+\alpha_2 z^2}\right)}, \quad s \in \mathbb{R}, \tag{4}$$

where $t = (s - \alpha_1)/\sqrt{\alpha_2}$. Notice that, for $\alpha_1 = 0$, the pdf in (4) reduces to the *SN* pdf in (1). Consequently, this conditional distribution provides another extended *SN* distribution. Moreover, if in (3) the location–scale parameter (μ, σ^2) is known, for instance $(\mu, \sigma^2) = (0, 1)$, and if the hyperparameters α_1 and α_2 are fixed, from a Bayesian point of view the distribution in (4) is the posterior pdf for the shape variable S. In this case, the distribution in (2) is the predictive distribution. It is important to notice that such a predictive distribution reduces to the normal distribution if in (3) we set $\alpha_1 = 0$, that is, if we elicit a proper $N(0, \alpha_2)$ prior distribution for the shape parameter S.

An extension of (1) to the multivariate case was introduced by Azzalini and Dalla Valle [10]. Denote by $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^{\mathrm{T}}$ the location vector, $\Sigma = (\sigma_{ij})$ an $n \times n$ positive definite scale matrix in which $\sigma_{ii} = \sigma_i^2$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^{\mathrm{T}}$ a shape vector controlling skewness. Consider the $n \times n$ diagonal matrix $D(\boldsymbol{\sigma})$ whose diagonal entries are the components of the vector $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)^{\mathrm{T}}$. Azzalini and Dalla Valle [10] defined an n-dimensional multivariate skew-normal random vector $\mathbf{Y} \sim SN_n(\boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha})$ with pdf:

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha}) = 2\phi_n(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma)\,\Phi(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{z}), \quad \mathbf{y} \in \mathbb{R}^n, \tag{5}$$

where $\mathbf{z} = D(\boldsymbol{\sigma})^{-1}(\mathbf{y} - \boldsymbol{\mu})$ and $\phi_n(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma)$ denotes the pdf of the multivariate normal $N_n(\boldsymbol{\mu}, \Sigma)$ distribution. The multivariate normal distribution is obtained by setting $\boldsymbol{\alpha} = \mathbf{0}$. Statistical applications of the multivariate *SN* distribution have been investigated by Azzalini and Capitanio [9], who also suggested extensions to elliptically contoured pdf's. Since then, many authors have tried to generalize these ideas to skewing arbitrary multivariate symmetric pdf's with very general forms of multiplicative functions; see, for example, [17,21,16,3,1,11]. An overview of these proposals can be found in the book edited by Genton [15], in [8], and from a unified point of view in [2].

This paper considers shape mixtures of two types of multivariate distributions: independent multivariate *SN* distributions obtained as the product of univariate *SN* marginal pdf's from (1); and dependent multivariate *SN* distributions defined by (5). These types of multivariate distributions were first presented by Arellano-Valle et al. [4], but only with a multivariate normal distribution as mixing measure for the shape vector of parameters $\mathbf{S}$, say. The goal of this paper is to extend these results by assuming mixing multivariate *SN* distributions for the shape vector of parameters. These extensions are developed in Section 2. Properties of the conditional distribution of the shape parameter are developed in Section 3. Some important results for Bayesian inference are then obtained because predictive distributions and posterior distributions for the shape parameter $\mathbf{S}$ can be computed easily as a consequence. In particular, it can be shown that the posteriors for the shape parameter are members of the unified skew-normal (SUN) family of distributions introduced by Arellano-Valle and Azzalini [1] who provided many of their main properties. It is also noteworthy that the predictive distributions are in the fundamental skew-normal class introduced by Arellano-Valle and Genton [3]. Even more importantly, we show that, although the resulting family of distributions for $\mathbf{S}$ is not conjugate in the usual sense, it brings tractability to the problem of inference on $\mathbf{S}$ (that is, the posteriors are obtained analytically) and simple interpretation of the results. For some special cases, we also prove that the posterior inference for the location and scale parameters is robust in the sense defined by O'Hagan [18], namely that the posterior is the same for a normal or an *SN* likelihood. These results are discussed in Section 4. We conclude the paper in Section 5.

2. Shape mixtures of multivariate *SN* distributions

We derive various multivariate extensions of the *SN* type pdf's in (2) and (4). In all these cases, the idea of shape mixture described in (3) is adapted for both independent and dependent multivariate *SN* distributions, with *SN* distributions as mixing measures.

There are many possibilities for defining multivariate extensions of (2) and (4) following the idea of shape mixture explained by (3). Indeed, consider an observable random vector $\mathbf{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ whose distribution is specified within the *SN* family. Since the components $Y_1, \ldots, Y_n$ can be independent or dependent, conditionally on the shape parameter, and because we can assume different or common shape parameters for the Y_i's, the following extensions of (3) can be considered:

(A) Assume a mixing (unobserved) shape random vector $\mathbf{S} = (S_1, \ldots, S_n)^{\mathrm{T}}$. Then, shape mixture distributions are obtained in any of the four following cases:

(A1) $[Y_i \mid \mathbf{S}=\mathbf{s}] \overset{\text{ind.}}{\sim} SN(\mu_i, \sigma_i^2, s_i)$ and $S_i \overset{\text{ind.}}{\sim} SN(\eta_i, \omega_i^2, \alpha_i)$, $i = 1, \ldots, n$;

(A2) $[Y_i \mid \mathbf{S}=\mathbf{s}] \overset{\text{ind.}}{\sim} SN(\mu_i, \sigma_i^2, s_i)$, $i = 1, \ldots, n$, and $\mathbf{S} \sim SN_n(\boldsymbol{\eta}, \Omega, \boldsymbol{\alpha})$;

(A3) $[\mathbf{Y} \mid \mathbf{S}=\mathbf{s}] \sim SN_n(\boldsymbol{\mu}, \Sigma, \mathbf{s})$ and $S_i \overset{\text{ind.}}{\sim} SN(\eta_i, \omega_i^2, \alpha_i)$, $i = 1, \ldots, n$;

(A4) $[\mathbf{Y} \mid \mathbf{S}=\mathbf{s}] \sim SN_n(\boldsymbol{\mu}, \Sigma, \mathbf{s})$ and $\mathbf{S} \sim SN_n(\boldsymbol{\eta}, \Omega, \boldsymbol{\alpha})$.

(B) Assume a mixing (unobserved) random variable S. Then, shape mixture distributions are obtained in any of the two following cases:

(B1) $[Y_i \mid S=s] \overset{\text{ind.}}{\sim} SN(\mu_i, \sigma_i^2, s)$, $i = 1, \ldots, n$, and $S \sim SN(\eta, \omega^2, \alpha)$;

(B2) $[\mathbf{Y} \mid S=s] \sim SN_n(\boldsymbol{\mu}, \Sigma, s\mathbf{1}_n)$ and $S \sim SN(\eta, \omega^2, \alpha)$, where $\mathbf{1}_n = (1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^n$.

In (A) and (B), the quantities $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^{\mathrm{T}}$ and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^{\mathrm{T}}$ denote location vectors, $\Sigma = (\sigma_{ij})$ and $\Omega = (\omega_{ij})$ are $n \times n$ positive definite scale matrices, and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^{\mathrm{T}}$ is a shape vector. If $\boldsymbol{\alpha} = \mathbf{0}$ and $\Omega = \mathrm{diag}\{\omega_1^2, \ldots, \omega_n^2\}$, then (A1) is equivalent to (A2) and (A3) is equivalent to (A4). In other words, for $\boldsymbol{\alpha} = \mathbf{0}$ the mixing random vector $\mathbf{S} \sim N_n(\boldsymbol{\eta}, \Omega)$ and, if Ω is a diagonal matrix, then the situations described previously in (A) reduce to (A2) and (A4). These two situations contain the simplest case in which the mixing components $S_1, \ldots, S_n$ are independent and identically distributed (i.i.d.) according to an $N(\eta, \omega^2)$ distribution. Notice also that (B1) and (B2) can be obtained from (A2) and (A4), respectively. In fact, one way to obtain such particular situations is to assume the following parametric structure for $\mathbf{S} \sim SN_n(\boldsymbol{\eta}, \Omega, \boldsymbol{\alpha})$:

$$\boldsymbol{\eta} = \eta \mathbf{1}_n, \quad \boldsymbol{\alpha} = \alpha \mathbf{1}_n, \quad \Omega = \omega^2\{(1-\rho)I_n + \rho \mathbf{1}_n \mathbf{1}_n^{\mathrm{T}}\}, \quad \rho \in [0, 1),$$

and then to let $\rho \to 1$. This structure is particularly interesting since it reduces the number of parameters to be estimated and also assumes exchangeability for the shape parameters $S_1, \ldots, S_n$, for all $\rho \in [0, 1)$.

From a Bayesian point of view, $Y_1, \ldots, Y_n$ can be interpreted as observations sampled from an SN type distribution, which are indexed either by different shape parameters represented by $S_1, \ldots, S_n$ (see situation (A)), or by a common shape parameter represented by S (see situation (B)). The structure considered in situation (A1) is the same as the structure initially assumed to identify multiple change points in the shape parameter via product partition models [12]. In addition, the results obtained for situation (B1) (see Proposition 3) are very important and useful to implement such a model. It will be shown in the following subsections that, in all of the scenarios in (A) and (B), both the unconditional distribution of $\mathbf{Y}$ and the conditional distribution of $\mathbf{S}$ (or S) given $\mathbf{Y} = \mathbf{y}$, have SN type distributions. In a Bayesian setup, these distributions correspond to predictive and posterior distributions when the location and scale parameters are known quantities.

The following notation will be considered throughout the paper. As mentioned before, for any n-dimensional vector $\mathbf{s} = (s_1, \ldots, s_n)^{\mathrm{T}}$, $D(\mathbf{s})$ denotes the $n \times n$ diagonal matrix whose diagonal entries are $s_1, \ldots, s_n$. Notice that, for any two n-dimensional vectors $\mathbf{s}$ and $\mathbf{z}$, $D(\mathbf{s})\mathbf{z} = D(\mathbf{z})\mathbf{s} = (s_1 z_1, \ldots, s_n z_n)^{\mathrm{T}}$. As in Section 1, denote by $\phi_n(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma)$ the pdf associated to the multivariate $N_n(\boldsymbol{\mu}, \Sigma)$ distribution, and by $\Phi_n(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma)$ the corresponding cdf. If $\boldsymbol{\mu} = \mathbf{0}$ (respectively, $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I_n$) these functions will be denoted by $\phi_n(\mathbf{y} \mid \Sigma)$ and $\Phi_n(\mathbf{y} \mid \Sigma)$ (respectively $\phi_n(\mathbf{y})$ and $\Phi_n(\mathbf{y})$). If we consider (1) and (5), all the scenarios described in (A) and (B) can be represented in terms of the corresponding pdf's of $[\mathbf{Y} \mid \mathbf{S} = \mathbf{s}]$ and $\mathbf{S}$. They are, respectively, given by:

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{s}) = 2^n |D(\boldsymbol{\sigma})|^{-1} \phi_n(\mathbf{z}) \Phi_n(D(\mathbf{s})\mathbf{z}),$$

and

$$f(\mathbf{s} \mid \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\alpha}) = 2^n |D(\boldsymbol{\omega})|^{-1} \phi_n(\mathbf{t}) \Phi_n(D(\boldsymbol{\alpha})\mathbf{t}),$$

where $\mathbf{z} = D(\boldsymbol{\sigma})^{-1}(\mathbf{y} - \boldsymbol{\mu})$ and $\mathbf{t} = D(\boldsymbol{\omega})^{-1}(\mathbf{s} - \boldsymbol{\eta})$, for the situations where the marginal components are independent. If the components are dependent, these pdf's become, respectively,

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, \mathbf{s}) = 2\phi_n(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma) \Phi(\mathbf{s}^{\mathrm{T}}\mathbf{z}),$$

and

$$f(\mathbf{s} \mid \boldsymbol{\eta}, \Omega) = 2\phi_n(\mathbf{s} \mid \boldsymbol{\eta}, \Omega) \Phi(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}).$$

The quantities $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)^{\mathrm{T}}$ and $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^{\mathrm{T}}$ represent scale vectors throughout the paper, for which σ_i^2 and ω_i^2 are interpreted as the (i, i) entry of Σ and Ω, respectively, in the dependent cases. The proof of the main results introduced in this paper are based essentially on the following well-known result. If $\mathbf{U} \sim N_k(\mathbf{c}, \mathbf{C})$ is a non-singular, multivariate normal random vector, then for any fixed m-dimensional vector $\mathbf{a}$ and $m \times k$ matrix A, we have that:

$$E[\Phi_m(A\mathbf{U} + \mathbf{a} \mid \mathbf{b}, B)] = \Phi_m(A\mathbf{c} + \mathbf{a} \mid \mathbf{b}, B + ACA^{\mathrm{T}}). \tag{6}$$

See, for example, [3] for a proof.

2.1. Mixtures on different shape parameters

Consider the situation where there are n observations $Y_1, \ldots, Y_n$ sampled from the SN distributions $SN(\mu_1, \sigma_1^2, s_1), \ldots, SN(\mu_n, \sigma_n^2, s_n)$, respectively. Assume that, given $\mathbf{S} = (S_1, \ldots, S_n)^{\mathrm{T}}$, the random quantities $Y_1, \ldots, Y_n$ are independent. Then, we have the following result.

Proposition 1. *Let* $[Y_i \mid \mathbf{S} = \mathbf{s}] \overset{\text{ind.}}{\sim} SN(\mu_i, \sigma_i^2, s_i)$, $i = 1, \ldots, n$, *where* $\mathbf{S} = (S_1, \ldots, S_n)^{\mathrm{T}}$.

(i) *If* $S_i \overset{\text{ind.}}{\sim} SN(\eta_i, \omega_i^2, \alpha_i)$, $i = 1, \ldots, n$, *then the pdf of* $\mathbf{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ *is*

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\alpha}) = 4^n |D(\boldsymbol{\sigma})|^{-1} \phi_n(\mathbf{z}) \prod_{i=1}^{n} \Phi_2\left(\begin{pmatrix}\eta_i z_i \\ 0\end{pmatrix} \middle| \begin{pmatrix}1 + \omega_i^2 z_i^2 & \alpha_i \omega_i z_i \\ \alpha_i \omega_i z_i & 1 + \alpha_i^2\end{pmatrix}\right),$$

and the conditional pdf of $[\mathbf{S}|\mathbf{Y} = \mathbf{y}]$ *is*

$$f(\mathbf{s} \mid \mathbf{z}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\alpha}) = \frac{|D(\boldsymbol{\omega})|^{-1} \phi_n(\mathbf{t})\, \Phi_n(D(\boldsymbol{\alpha})\mathbf{t})\, \Phi_n(D(\mathbf{z})\mathbf{s})}{\prod_{i=1}^{n} \Phi_2\left(\begin{pmatrix}\eta_i z_i \\ 0\end{pmatrix} \middle| \begin{pmatrix}1 + \omega_i^2 z_i^2 & \alpha_i \omega_i z_i \\ \alpha_i \omega_i z_i & 1 + \alpha_i^2\end{pmatrix}\right)},$$

where $\mathbf{z} = D(\boldsymbol{\sigma})^{-1}(\mathbf{y} - \boldsymbol{\mu})$ *and* $\mathbf{t} = D(\boldsymbol{\omega})^{-1}(\mathbf{s} - \boldsymbol{\eta})$, *i.e.,* $z_i = (y_i - \mu_i)/\sigma_i$ *and* $t_i = (s_i - \eta_i)/\omega_i$, $i = 1, \ldots, n$.

(ii) *If* $\mathbf{S} \sim SN_n(\boldsymbol{\eta}, \Omega, \boldsymbol{\alpha})$, *then the pdf of* $\mathbf{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ *is*

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\eta}, \Omega, \boldsymbol{\alpha}) = 2^{n+1} |D(\boldsymbol{\sigma})|^{-1} \phi_n(\mathbf{z}) \times \Phi_{n+1}\left(\begin{pmatrix}D(\boldsymbol{\eta})\mathbf{z} \\ 0\end{pmatrix} \middle| \begin{pmatrix}I_n + D(\mathbf{z})\Omega D(\mathbf{z}) & D(\mathbf{z})\Omega D(\boldsymbol{\omega})^{-1}\boldsymbol{\alpha} \\ \boldsymbol{\alpha}^{\mathrm{T}} D(\boldsymbol{\omega})^{-1} \Omega D(\mathbf{z}) & 1 + \boldsymbol{\alpha}^{\mathrm{T}} D(\boldsymbol{\omega})^{-1} \Omega D(\boldsymbol{\omega})^{-1} \boldsymbol{\alpha}\end{pmatrix}\right),$$

and the conditional pdf of $[\mathbf{S} \mid \mathbf{Y} = \mathbf{y}]$ *is*

$$f(\mathbf{s} \mid \mathbf{z}, \boldsymbol{\eta}, \Omega, \boldsymbol{\alpha}) = \frac{\phi_n(\mathbf{s} \mid \boldsymbol{\eta}, \Omega)\, \Phi(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t})\, \Phi_n(D(\mathbf{z})\mathbf{s})}{\Phi_{n+1}\left(\begin{pmatrix}D(\boldsymbol{\eta})\mathbf{z} \\ 0\end{pmatrix} \middle| \begin{pmatrix}I_n + D(\mathbf{z})\Omega D(\mathbf{z}) & D(\mathbf{z})\Omega D(\boldsymbol{\omega})^{-1}\boldsymbol{\alpha} \\ \boldsymbol{\alpha}^{\mathrm{T}} D(\boldsymbol{\omega})^{-1} \Omega D(\mathbf{z}) & 1 + \boldsymbol{\alpha}^{\mathrm{T}} D(\boldsymbol{\omega})^{-1} \Omega D(\boldsymbol{\omega})^{-1} \boldsymbol{\alpha}\end{pmatrix}\right)},$$

where $\mathbf{z} = D(\boldsymbol{\sigma})^{-1}(\mathbf{y} - \boldsymbol{\mu})$ *and* $\mathbf{t} = D(\boldsymbol{\omega})^{-1}(\mathbf{s} - \boldsymbol{\eta})$.

Proof. For case (i), since $\mathbf{s} = D(\boldsymbol{\omega})\mathbf{t} + \boldsymbol{\eta}$, it follows from the hypotheses and using the Jacobian method that

$$\begin{aligned} f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\alpha}) &= 4^n |D(\boldsymbol{\sigma})|^{-1} \phi_n(\mathbf{z}) \prod_{i=1}^{n} \int_{-\infty}^{\infty} \phi(t_i)\, \Phi(t_i z_i \omega_i + \eta_i z_i)\, \Phi(\alpha_i t_i) \mathrm{d}t_i \\ &= 4^n |D(\boldsymbol{\sigma})|^{-1} \phi_n(\mathbf{z}) \prod_{i=1}^{n} \mathrm{E}\left\{\Phi_2\left(\begin{pmatrix}z_i \omega_i \\ \alpha_i\end{pmatrix} t_i + \begin{pmatrix}\eta_i z_i \\ 0\end{pmatrix}\right)\right\} \\ &= 4^n |D(\boldsymbol{\sigma})|^{-1} \phi_n(\mathbf{z}) \prod_{i=1}^{n} \mathrm{E}\left\{\Phi_2\left(\mathbf{A} t_i + \mathbf{a}\right)\right\}. \end{aligned}$$

Consequently, the result follows from (6) by assuming $\mathbf{c} = 0$, $\mathbf{C} = 1$, $\mathbf{b} = \mathbf{0}$ and $B = I_2$. The conditional distribution of $\mathbf{S}$ given $\mathbf{z}$ is obtained by a straightforward application of Bayes' theorem. The proof of case (ii) follows similarly. □

Now consider the situation in which the vector of observable quantities $\mathbf{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ is sampled from a dependent SN distribution. Then, we have the following result.

Proposition 2. *Assume that* $[\mathbf{Y} \mid \mathbf{S} = \mathbf{s}] \sim SN_n(\boldsymbol{\mu}, \Sigma, \mathbf{s})$ *whose pdf is given by* $f(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, \mathbf{s}) = 2\phi_n(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma)\, \Phi(\mathbf{s}^{\mathrm{T}}\mathbf{z})$ *with* $\mathbf{z} = D(\boldsymbol{\sigma})^{-1}(\mathbf{y} - \boldsymbol{\mu})$, $\mathbf{y} \in \mathbb{R}^n$.

(i) *If* $S_i \overset{\text{ind.}}{\sim} SN(\eta_i, \omega_i^2, \alpha_i)$, $i = 1, \ldots, n$, *then the pdf of* $\mathbf{Y}$ *is*

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\alpha}) = 2^{n+1} \phi_n(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma)\, \Phi_{n+1}\left(\begin{pmatrix}\boldsymbol{\eta}^{\mathrm{T}}\mathbf{z} \\ \mathbf{0}\end{pmatrix} \middle| \begin{pmatrix}1 + \mathbf{z}^{\mathrm{T}} D^2(\boldsymbol{\omega})\mathbf{z} & \mathbf{z}^{\mathrm{T}} D(\boldsymbol{\omega}) D(\boldsymbol{\alpha}) \\ D(\boldsymbol{\alpha}) D(\boldsymbol{\omega})\mathbf{z} & I_n + D^2(\boldsymbol{\alpha})\end{pmatrix}\right),$$

and the conditional pdf of $[\mathbf{S} \mid \mathbf{Y} = \mathbf{y}]$ *is*

$$f(\mathbf{s} \mid \mathbf{z}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\alpha}) = \frac{|D(\boldsymbol{\omega})|^{-1} \phi_n(\mathbf{t})\, \Phi_n(D(\boldsymbol{\alpha})\mathbf{t})\, \Phi(\mathbf{z}^{\mathrm{T}}\mathbf{s})}{\Phi_{n+1}\left(\begin{pmatrix}\boldsymbol{\eta}^{\mathrm{T}}\mathbf{z} \\ \mathbf{0}\end{pmatrix} \middle| \begin{pmatrix}1 + \mathbf{z}^{\mathrm{T}} D^2(\boldsymbol{\omega})\mathbf{z} & \mathbf{z}^{\mathrm{T}} D(\boldsymbol{\omega}) D(\boldsymbol{\alpha}) \\ D(\boldsymbol{\alpha}) D(\boldsymbol{\omega})\mathbf{z} & I_n + D^2(\boldsymbol{\alpha})\end{pmatrix}\right)},$$

where $\mathbf{t} = D(\boldsymbol{\omega})^{-1}(\mathbf{s} - \boldsymbol{\eta})$.

(ii) *If* $\mathbf{S} \sim SN_n(\boldsymbol{\eta}, \Omega, \boldsymbol{\alpha})$, *then the pdf of* $\mathbf{Y}$ *is*

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, \boldsymbol{\eta}, \Omega, \boldsymbol{\alpha}) = 4\phi_n(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma)\,\Phi_2\left(\begin{pmatrix}\boldsymbol{\eta}^{\mathrm{T}}\mathbf{z}\\ 0\end{pmatrix}\middle|\begin{pmatrix}1+\mathbf{z}^{\mathrm{T}}\Omega\mathbf{z} & \mathbf{z}^{\mathrm{T}}\Omega D(\boldsymbol{\omega})^{-1}\boldsymbol{\alpha}\\ \boldsymbol{\alpha}^{\mathrm{T}}D(\boldsymbol{\omega})^{-1}\Omega\mathbf{z} & 1+\boldsymbol{\alpha}^{\mathrm{T}}D(\boldsymbol{\omega})^{-1}\Omega D(\boldsymbol{\omega})^{-1}\boldsymbol{\alpha}\end{pmatrix}\right),$$

and the conditional pdf of $[\mathbf{S} \mid \mathbf{Y} = \mathbf{y}]$ *is*

$$f(\mathbf{s} \mid \mathbf{z}, \boldsymbol{\eta}, \Omega, \boldsymbol{\alpha}) = \frac{\phi_n(\mathbf{s} \mid \boldsymbol{\eta}, \Omega)\,\Phi(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t})\,\Phi(\mathbf{z}^{\mathrm{T}}\mathbf{s})}{\Phi_2\left(\begin{pmatrix}\boldsymbol{\eta}^{\mathrm{T}}\mathbf{z}\\ 0\end{pmatrix}\middle|\begin{pmatrix}1+\mathbf{z}^{\mathrm{T}}\Omega\mathbf{z} & \mathbf{z}^{\mathrm{T}}\Omega D(\boldsymbol{\omega})^{-1}\boldsymbol{\alpha}\\ \boldsymbol{\alpha}^{\mathrm{T}}D(\boldsymbol{\omega})^{-1}\Omega\mathbf{z} & 1+\boldsymbol{\alpha}^{\mathrm{T}}D(\boldsymbol{\omega})^{-1}\Omega D(\boldsymbol{\omega})^{-1}\boldsymbol{\alpha}\end{pmatrix}\right)},$$

where $\mathbf{t} = D(\boldsymbol{\omega})^{-1}(\mathbf{s} - \boldsymbol{\eta})$.

The proof is similar to the one of Proposition 1 and thus it is omitted.

Notice that all pdf's obtained in Propositions 1 and 2 define different multivariate families of fundamental *SN* distributions [3]. Some of them admit, as special cases, those situations where some of the original parameters, i.e., $\boldsymbol{\alpha}$, $\boldsymbol{\eta}$ and $\boldsymbol{\omega}$ or Ω, are equal to zero. However, if we assume that $\boldsymbol{\omega} = \mathbf{0}$ or $\Omega = O$, the zero matrix, then the corresponding distributions for the shape parameter ($\mathbf{S}$ or S) are degenerate. If we consider, for instance, the situation presented in part (ii) of Proposition 2 assuming $\Omega = O$, we have that $P(\mathbf{S} = \boldsymbol{\eta}) = 1$. In this case, the pdf of $\mathbf{Y}$ is obtained from a discrete mixture of distributions and becomes $SN_n(\boldsymbol{\mu}, \Sigma, \boldsymbol{\eta})$, that is,

$$\begin{aligned} f(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, \boldsymbol{\eta}, \Omega, \boldsymbol{\alpha}) &= f(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, \mathbf{S} = \boldsymbol{\eta})P(\mathbf{S} = \boldsymbol{\eta})\\ &= 2\phi_n(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma)\,\Phi(\boldsymbol{\eta}^{\mathrm{T}}\mathbf{z}). \end{aligned}$$

In this case, the conditional distribution of $\mathbf{S}$ given $\mathbf{Y} = \mathbf{y}$ is also degenerate. From a Bayesian point of view, it means that eliciting a degenerate prior distribution for $\mathbf{S}$ makes it impossible to update using the data; see [18] for details on Cromwell's law.

The special cases where $\boldsymbol{\alpha} = \mathbf{0}$ were first considered by Arellano-Valle et al. [4]. They assumed an $N_n(\boldsymbol{\eta}, \Omega)$ prior distribution for $\mathbf{S}$. The results obtained by Arellano-Valle et al. [4] are summarized in the following corollary.

Corollary 1. *Assume that* $\boldsymbol{\alpha} = \mathbf{0}$, *that is,* $\mathbf{S} \sim N_n(\boldsymbol{\eta}, \Omega)$.

(i) *If* $[Y_i \mid \mathbf{S} = \mathbf{s}] \overset{\text{ind.}}{\sim} SN(\mu_i, \sigma_i^2, s_i)$, $i = 1, \ldots, n$, *then*

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\eta}, \Omega) = 2^n|D(\boldsymbol{\sigma})|^{-1}\phi_n(\mathbf{z})\,\Phi_n(D(\boldsymbol{\eta})\mathbf{z} \mid I_n + D(\mathbf{z})\Omega D(\mathbf{z})).$$

This pdf reduces to $N_n(\boldsymbol{\mu}, D^2(\boldsymbol{\sigma}))$ *if* $\boldsymbol{\eta} = \mathbf{0}$ *and* Ω *is a diagonal matrix. If* $\Omega = O$ *it becomes the following independent multivariate SN pdf:* $f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\eta}) = 2^n|D(\boldsymbol{\sigma})|^{-1}\phi_n(\mathbf{z})\,\Phi_n(D(\boldsymbol{\eta})\mathbf{z})$. *Moreover, the conditional distribution of* $\mathbf{S}$ *given* $\mathbf{Y} = \mathbf{y}$ *is*

$$f(\mathbf{s} \mid \mathbf{z}, \boldsymbol{\eta}, \Omega) = \frac{\phi_n(\mathbf{s} \mid \boldsymbol{\eta}, \Omega)\,\Phi_n(D(\mathbf{z})\mathbf{s})}{\Phi_n\left(D(\boldsymbol{\eta})\mathbf{z} \mid I_n + D(\mathbf{z})\Omega D(\mathbf{z})\right)}.$$

(ii) *If* $[\mathbf{Y} \mid \mathbf{S} = \mathbf{s}] \sim SN_n(\boldsymbol{\mu}, \Sigma, \mathbf{s})$, *then*

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, \boldsymbol{\eta}, \Omega) = 2\phi_n(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma)\,\Phi\left(\frac{\boldsymbol{\eta}^{\mathrm{T}}\mathbf{z}}{\sqrt{1+\mathbf{z}^{\mathrm{T}}\Omega\mathbf{z}}}\right).$$

This pdf reduces to $N_n(\boldsymbol{\mu}, \Sigma)$ *if* $\boldsymbol{\eta} = \mathbf{0}$. *If* $\Omega = O$ *it becomes the following dependent multivariate SN pdf:* $f(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, \boldsymbol{\eta}) = 2\phi_n(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma)\,\Phi(\boldsymbol{\eta}^{\mathrm{T}}\mathbf{z})$. *Moreover, the conditional distribution of* $\mathbf{S}$ *given* $\mathbf{Y} = \mathbf{y}$ *is*

$$f(\mathbf{s} \mid \mathbf{z}, \boldsymbol{\eta}, \Omega) = \frac{\phi_n(\mathbf{s} \mid \boldsymbol{\eta}, \Omega)\,\Phi(\mathbf{z}^{\mathrm{T}}\mathbf{s})}{\Phi_2\left(\frac{\boldsymbol{\eta}^{\mathrm{T}}\mathbf{z}}{\sqrt{1+\mathbf{z}^{\mathrm{T}}\Omega\mathbf{z}}}\right)}.$$

Another special and useful situation, because it significantly reduces the number of parameters involved in $\boldsymbol{\eta}$ and Ω, is obtained from part (ii) of Propositions 1 and 2 and Corollary 1. Consider the exchangeable structure

$$\boldsymbol{\eta} = \eta\mathbf{1}_n, \quad \Omega = \omega^2\{(1-\rho)I_n + \rho\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\}, \quad \rho \in [0, 1). \tag{7}$$

If, in addition, it is assumed that $\boldsymbol{\alpha} = \alpha\mathbf{1}_n$, it follows immediately that all *SN* type distributions obtained in part (ii) of Proposition 2 will be exchangeable.

2.2. Mixtures on a common shape parameter

All models introduced in Section 2.1 assume that the observable quantities $Y_1, \ldots, Y_n$ have *SN* distributions with different shape parameters. However, in many situations, the shape parameters are common. Shape mixture models for such situations are introduced next.

Proposition 3. *Let* $[Y_i \mid S = s] \overset{\text{ind.}}{\sim} SN(\mu_i, \sigma_i^2, s), i = 1, \ldots, n$, *and suppose that* $S \sim SN(\eta, \omega^2, \alpha)$. *Then, the pdf of* $\mathbf{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ *is*

$$f(\mathbf{y} \mid \eta, \omega^2, \alpha) = 2^{n+1}|D(\boldsymbol{\sigma})|^{-1}\phi_n(\mathbf{z})\,\Phi_{n+1}\left(\begin{pmatrix}\eta\mathbf{z}\\0\end{pmatrix}\middle|\begin{pmatrix}I_n+\omega^2\mathbf{z}\mathbf{z}^{\mathrm{T}} & \alpha\omega\mathbf{z}\\ \alpha\omega\mathbf{z}^{\mathrm{T}} & 1+\alpha^2\end{pmatrix}\right),$$

where $\mathbf{z} = D(\boldsymbol{\sigma})^{-1}(\mathbf{y}-\boldsymbol{\mu})$. *For* $\alpha = 0$ *this pdf reduces to*

$$f(\mathbf{y} \mid \eta, \omega^2) = 2^n|D(\boldsymbol{\sigma})|^{-1}\phi_n(\mathbf{z})\,\Phi_n(\eta\mathbf{z} \mid I_n + \omega^2\mathbf{z}\mathbf{z}^{\mathrm{T}}).$$

Moreover, the conditional pdf of $[S \mid \mathbf{Y} = \mathbf{y}]$ *is*

$$f(s \mid \mathbf{z}, \eta, \omega, \alpha) = \frac{\omega^{-1}\phi(t)\,\Phi(\alpha t)\,\Phi_n(\mathbf{z}s)}{\Phi_{n+1}\left(\begin{pmatrix}\eta\mathbf{z}\\0\end{pmatrix}\middle|\begin{pmatrix}I_n+\omega^2\mathbf{z}\mathbf{z}^{\mathrm{T}} & \alpha\omega\mathbf{z}\\ \alpha\omega\mathbf{z}^{\mathrm{T}} & 1+\alpha^2\end{pmatrix}\right)}.$$

For $\alpha = 0$ *this pdf reduces to*

$$f(s \mid \mathbf{z}, \eta, \omega) = \frac{\omega^{-1}\phi(t)\,\Phi_n(\mathbf{z}s)}{\Phi_n(\eta\mathbf{z} \mid I_n + \omega^2\mathbf{z}\mathbf{z}^{\mathrm{T}})},$$

where $t = (s-\eta)/\omega$.

Proposition 4. *Let* $[\mathbf{Y} \mid S = s] \sim SN_n(\boldsymbol{\mu}, \Sigma, s\mathbf{1}_n)$ *whose conditional pdf is given by* $2\phi_n(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma)\Phi(sz)$ *with* $z = \sum_{i=1}^n z_i$, $z_i = (y_i - \mu_i)\sigma_i^{-1}$. *Assume that* $S \sim SN(\eta, \omega^2, \alpha)$. *Then, the pdf of* $\mathbf{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ *is*

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, \eta, \omega^2, \alpha) = 4\phi_n(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma)\,\Phi_2\left(\begin{pmatrix}\eta z\\0\end{pmatrix}\middle|\begin{pmatrix}1+\omega^2z^2 & \alpha\omega z\\ \alpha\omega z & 1+\alpha^2\end{pmatrix}\right). \tag{8}$$

For $\alpha = 0$ *this pdf reduces to* $f(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, \eta, \omega) = 2\phi_n(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma)\,\Phi\left(\frac{\eta z}{\sqrt{1+\omega^2z^2}}\right)$. *Moreover, the conditional pdf of* $[S|\mathbf{Y} = \mathbf{y}]$ *is*

$$f(s \mid z, \eta, \omega, \alpha) = \frac{\omega^{-1}\phi(t)\,\Phi(\alpha t)\,\Phi(zs)}{\Phi_2\left(\begin{pmatrix}\eta z\\0\end{pmatrix}\middle|\begin{pmatrix}1+\omega^2z^2 & \alpha\omega z\\ \alpha\omega z & 1+\alpha^2\end{pmatrix}\right)}, \tag{9}$$

where $t = (s-\eta)/\omega$. *For* $\alpha = 0$ *this pdf reduces to*

$$f(s \mid z, \eta, \omega) = \frac{\omega^{-1}\phi(t)\,\Phi(zs)}{\Phi\left(\frac{\eta z}{\sqrt{1+\omega^2z^2}}\right)}.$$

The proofs are similar to the one of Proposition 1 and thus are omitted. It is worth noticing that the results in Propositions 3 and 4 can be obtained as particular cases of part (ii) of Propositions 1 and 2, respectively, if the exchangeable structure in (7) with $\rho = 1$ is considered and also if it is assumed that $\alpha\omega^{-1} = \sum_{i=1}^n \alpha_i\omega_i^{-1}$. For illustration, we plot in Fig. 1 the pdf (8) of $\mathbf{Y} = (Y_1, Y_2)^{\mathrm{T}}$ and the conditional pdf (9) of $[S \mid \mathbf{Y} = (1, 2)^{\mathrm{T}}]$ from Proposition 4 for $n = 2$, $\boldsymbol{\mu} = \mathbf{0}$, $\Sigma = I_2$, $\omega = 1$, $\alpha = 3$, $z = y_1 + y_2$, and $\eta = 1, 3$. As expected, the pdf (8) is more skewed for $\eta = 3$ than for $\eta = 1$, whereas it is the reverse for the pdf (9).

3. Properties of the conditional distribution of the shape parameter

Conditional distributions for the shape parameter, given $\mathbf{Y} = \mathbf{y}$, were provided in the previous section. It can be noticed that such distributions are members of the unified skewed-normal (SUN) family introduced by Arellano-Valle and Azzalini [1]. In that paper, Arellano-Valle and Azzalini [1] unify several coexisting proposals for the multivariate skew-normal clarifying their connections. They also provide many of the main properties of this unified class of distributions. In particular, from some properties of the SUN family, we can compute the mean and the covariance matrix of the shape parameter, given $\mathbf{Y} = \mathbf{y}$.

Let $\mathbf{X} \sim SUN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\gamma}, \Psi, \Lambda, \Delta)$ be a d-dimensional SUN random vector. Then, its density can be obtained by the following conditioning mechanism: $\mathbf{X} \overset{d}{=} (\mathbf{V} \mid \mathbf{U} > \mathbf{0})$, where

$$\begin{pmatrix}\mathbf{U}\\ \mathbf{V}\end{pmatrix} \sim N_{m+d}\left(\begin{pmatrix}\boldsymbol{\gamma}\\ \boldsymbol{\xi}\end{pmatrix}, \begin{pmatrix}\Lambda & \Delta^{\mathrm{T}}\\ \Delta & \Psi\end{pmatrix}\right).$$

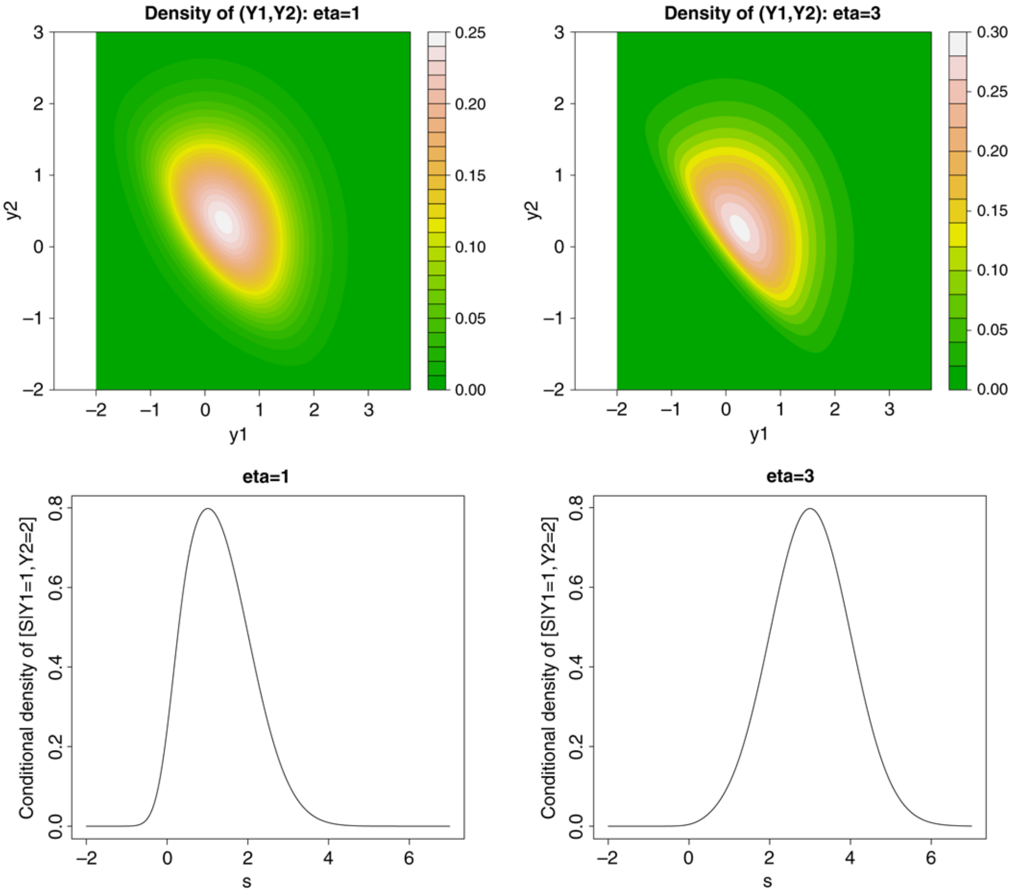

Fig. 1. Densities (8) (top row) and (9) (bottom row) from Proposition 4 for $n = 2$, $\boldsymbol{\mu} = \mathbf{0}$, $\Sigma = I_2$, $\omega = 1$, $\alpha = 3$, $z = y_1 + y_2$, and: $\eta = 1$ (left column); $\eta = 3$ (right column).

Let $\mathbf{V}_0 = \mathbf{V} - E(\mathbf{V} \mid \mathbf{U}) = \mathbf{V} - \Delta\Lambda^{-1}(\mathbf{U} - \boldsymbol{\gamma})$. Since it can be proved that $\mathbf{V}_0$ is independent of $\mathbf{U}$, it follows that $\mathbf{X} \stackrel{d}{=} \mathbf{V}_0 + \Delta\Lambda^{-1}(\mathbf{U}_0 - \boldsymbol{\gamma})$, where the random vector $\mathbf{U}_0$ is such that $\mathbf{U}_0 \stackrel{d}{=} (\mathbf{U} \mid \mathbf{U} > \mathbf{0})$ and it is independent of $\mathbf{V}_0$. Thus, the mean and the variance of $\mathbf{X}$ are given, respectively, by:

$$E(\mathbf{X}) = \Delta\Lambda^{-1}(\boldsymbol{\gamma}_0 - \boldsymbol{\gamma}) \quad \text{and} \quad V(\mathbf{X}) = \Psi - \Delta\Lambda^{-1}(\Lambda - \Lambda_0)\Lambda^{-1}\Delta^{\mathrm{T}},$$

where $\boldsymbol{\gamma}_0 = E(\mathbf{U}_0)$ and $\Lambda_0 = V(\mathbf{U}_0)$. Expressions for $\boldsymbol{\gamma}_0$ and Λ_0 can be obtained from [20] and also from [14].

For instance, from [14] it follows that:

$$\boldsymbol{\gamma}_0 = \boldsymbol{\gamma} + \Lambda \frac{\Phi_m^*(\boldsymbol{\gamma} \mid \Lambda)}{\Phi_m(\boldsymbol{\gamma} \mid \Lambda)},$$

and, consequently, the mean of $\mathbf{X}$ is

$$E(\mathbf{X}) = \Delta \frac{\Phi_m^*(\boldsymbol{\gamma} \mid \Lambda)}{\Phi_m(\boldsymbol{\gamma} \mid \Lambda)},$$

where $\Phi_m^*(\mathbf{u} - \boldsymbol{\gamma} \mid \Lambda) = \frac{\partial}{\partial \mathbf{u}}\Phi_m(\mathbf{u} - \boldsymbol{\gamma} \mid \Lambda)$ is the m-dimensional vector whose ith component is given by:

$$\frac{\partial}{\partial \mathbf{u}_i}\Phi_m(\mathbf{u} - \boldsymbol{\gamma} \mid \Lambda) = \phi\left(\gamma_i\lambda_{ii}^{-1/2}\right)\Phi_{m-1}(\boldsymbol{\gamma}_{-i} + \Lambda_{i-i}\lambda_{ii}^{-1}\gamma_i \mid \Lambda_{-i-i}).$$

Here, the notations for a random vector $\mathbf{U} \sim N_m(\boldsymbol{\gamma}, \Lambda)$ are: $\mathbf{U}_{-i} = (U_1, \ldots, U_{i-1}, U_{i+1}, \ldots, U_m)^{\mathrm{T}}$, $\boldsymbol{\gamma}_{-i} = E(\mathbf{U}_{-i})$, $\Lambda_{-i} = V(\mathbf{U}_{-i})$ and $\Lambda_{i-i} = \mathrm{Cov}(\mathbf{U}_i, \mathbf{U}_{-i})$.

In order to provide an example, consider the situation (ii) presented in Proposition 2 in which we assume $[\mathbf{Y} \mid \mathbf{S} = \mathbf{s}] \sim SN_n(\boldsymbol{\mu}, \Sigma, \mathbf{s})$ and a dependent prior for the shape parameter, that is, $\mathbf{S} \sim SN_n(\boldsymbol{\eta}, \Omega, \boldsymbol{\alpha})$. In that case, we have that $\boldsymbol{\xi} = \boldsymbol{\eta}$, $\mathbf{z} = D(\boldsymbol{\sigma})^{-1}(\mathbf{y} - \boldsymbol{\mu})$, $\boldsymbol{\gamma} = (\mathbf{z}^{\mathrm{T}}\boldsymbol{\eta}, 0)^{\mathrm{T}}$, $\Delta = (\Psi\mathbf{z},\ \Psi D(\boldsymbol{\omega})^{-1}\boldsymbol{\alpha})$ and

$$\Lambda = \begin{pmatrix} 1 + \mathbf{z}^{\mathrm{T}}\Omega\mathbf{z} & \mathbf{z}^{\mathrm{T}}\Omega D(\boldsymbol{\omega})^{-1}\boldsymbol{\alpha} \\ \boldsymbol{\alpha}^{\mathrm{T}}D(\boldsymbol{\omega})^{-1}\Omega\mathbf{z} & 1 + \boldsymbol{\alpha}^{\mathrm{T}}D(\boldsymbol{\omega})^{-1}\Omega D(\boldsymbol{\omega})^{-1}\boldsymbol{\alpha} \end{pmatrix}.$$

Consequently, the conditional expectation of $\mathbf{S}$ given $\mathbf{Y} = \mathbf{y}$ is

$$E(\mathbf{S} \mid \mathbf{z}, \boldsymbol{\eta}, \Omega, \boldsymbol{\alpha}) = \left(\Psi\mathbf{z},\ \ \Psi D(\boldsymbol{\omega})^{-1}\boldsymbol{\alpha}\right) \frac{\begin{pmatrix} \phi\left(\mathbf{z}^{\mathrm{T}}\boldsymbol{\eta}(1+\mathbf{z}^{\mathrm{T}}\Omega\mathbf{z})^{-1/2}\right)\Phi\left(\dfrac{-\mathbf{z}^{\mathrm{T}}\Omega D(\boldsymbol{\omega})^{-1}\boldsymbol{\alpha}\mathbf{z}^{\mathrm{T}}\boldsymbol{\eta}}{(1+\mathbf{z}^{\mathrm{T}}\Omega\mathbf{z})(1+\boldsymbol{\alpha}^{\mathrm{T}}D(\boldsymbol{\omega})^{-1}\Omega D(\boldsymbol{\omega})^{-1}\boldsymbol{\alpha})^{1/2}}\right) \\ (2\pi)^{-1/2}\Phi\left(-\mathbf{z}^{\mathrm{T}}\boldsymbol{\eta}(1+\mathbf{z}^{\mathrm{T}}\Omega\mathbf{z})^{-1/2}\right) \end{pmatrix}}{\Phi_2\left(\begin{pmatrix}\boldsymbol{\eta}^{\mathrm{T}}\mathbf{z} \\ 0\end{pmatrix}\middle| \begin{pmatrix} 1+\mathbf{z}^{\mathrm{T}}\Omega\mathbf{z} & \mathbf{z}^{\mathrm{T}}\Omega D(\boldsymbol{\omega})^{-1}\boldsymbol{\alpha} \\ \boldsymbol{\alpha}^{\mathrm{T}}D(\boldsymbol{\omega})^{-1}\Omega\mathbf{z} & 1+\boldsymbol{\alpha}^{\mathrm{T}}D(\boldsymbol{\omega})^{-1}\Omega D(\boldsymbol{\omega})^{-1}\boldsymbol{\alpha}\end{pmatrix}\right)}.$$

4. Bayesian inference on shape mixtures of *SN* distributions

From a Bayesian point of view, some important results for inference on the shape, location and scale parameters are obtained as a byproduct from the propositions given in Section 2. As mentioned before, if the location and scale parameters are known, the results in Section 2 provide explicit expressions for the predictive and posterior distributions of the observable vector $\mathbf{Y}$ and the shape parameter, respectively. Otherwise, if the location and scale parameters are unknown, the conditional distribution of $\mathbf{S}$ given $\mathbf{Y} = \mathbf{y}$ provides the full conditional distribution for $\mathbf{S}$ only.

Nevertheless, it is more important to notice that some simplicity is introduced in the inference on the shape parameter by these results. Robust inference is also obtained for the location and scale parameters, in some special cases. These subjects are discussed next.

4.1. On conjugacy in shape mixtures of SN distributions

The key problem in implementing the Bayesian paradigm for inference is the calculation of the required integrals. Many computational methods such as stochastic simulation techniques and many others have been proposed in the literature in order to solve this issue. Another route for tackling this problem is to find a class $\mathcal{P}$ of probability distributions rich enough to represent well a wide range of prior opinions and that permit a tractable implementation and simple interpretation of the results. Here, tractability means to be able to easily evaluate analytically the integrals required in the posterior. In general, this is achieved if $\mathcal{P}$ is closed under sampling and also closed under products thus generating posterior and prior distributions in the same family. That property is named natural conjugacy. Formally, denote by $\mathcal{F} = \{f(\mathbf{x}|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ the family of sampling distributions indexed by $\boldsymbol{\theta}$ and by $\mathcal{P} = \{\pi(\boldsymbol{\theta}|\mathbf{a}) : \mathbf{a} \in \mathcal{A}\}$ the family of distributions defined on Θ for which $\mathcal{A}$ is the set of hyperparameters. The families $\mathcal{P}$ and $\mathcal{F}$ are natural conjugates if:

- $\mathcal{P}$ is closed under sampling of $\mathcal{F}$, that is $f(\mathbf{x}|\boldsymbol{\theta}) \propto \pi(\boldsymbol{\theta}|\mathbf{a})$, for each $\mathbf{x}$ and for any $\pi(\boldsymbol{\theta}|\mathbf{a}) \in \mathcal{P}$;
- $\mathcal{P}$ is closed under products, that is, $\forall \mathbf{a}_0, \mathbf{a}_1 \in \mathcal{A}$, there exists $\mathbf{a}_2 \in \mathcal{A}$ such that $\pi(\boldsymbol{\theta}|\mathbf{a}_2) \propto \pi(\boldsymbol{\theta}|\mathbf{a}_0)\pi(\boldsymbol{\theta}|\mathbf{a}_1)$.

According to Bernardo and Smith [13], a natural conjugate family can be identified only if the likelihood admits sufficient statistics of fixed dimension. However, sufficiency can be too strong an assumption for the likelihood families considered in this paper. More details on conjugacy and natural conjugacy can be found in [13,19], for instance.

We start by focusing our attention on the estimation of $\mathbf{S} = (S_1, \ldots, S_n)^{\mathrm{T}}$ in the situation discussed in part (i) of Proposition 1. Assume that $[Y_i \mid \mathbf{S} = \mathbf{s}] \overset{\text{ind.}}{\sim} SN(\mu_i, \sigma_i^2, s_i)$, where μ_i and σ_i^2 are known parameters, for all $i = 1, \ldots, n$. Consequently, the likelihood is given by:

$$f(\mathbf{y} \mid \mathbf{S} = \mathbf{s}) = 2^n|D(\boldsymbol{\sigma})|^{-1}\phi_n(\mathbf{z})\Phi_n(D(\mathbf{z})\mathbf{s}) \propto \Phi_n(D(\mathbf{z})\mathbf{s}). \tag{10}$$

Looking for natural conjugacy, if $\mathcal{P}$ is considered as the class of probability distributions which are proportional to the likelihood, then tractability is reached. Indeed, the posterior distribution is proportional to the product of cdf's $\Phi_n(D(\mathbf{z})\mathbf{s})\Phi_n(D(\boldsymbol{\alpha})\mathbf{s})$, where $\boldsymbol{\alpha}$ denotes the prior shape parameter. In this case, to the contrary of what is observed in the usual natural conjugate analysis, such a family is too restrictive (that is, it is not rich enough in forms) and cannot represent well the prior opinion in many circumstances.

A tractable implementation of the posterior (that is, it can be put in analytic form) can still be achieved if a more general class of distributions is considered. For instance, assume, as in part (i) of Proposition 1, that $S_i \overset{\text{ind.}}{\sim} SN(\eta_i, \omega_i^2, \alpha_i)$, $i = 1, \ldots, n$. Thus, the prior pdf of $\mathbf{S}$ is

$$f(\mathbf{s} \mid \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\alpha}) = \frac{|D(\boldsymbol{\omega})|^{-1}\phi_n(\mathbf{t})\Phi_n(D(\boldsymbol{\alpha})\mathbf{t})}{\prod_{i=1}^{n}\Phi(0)}. \tag{11}$$

From Proposition 1, it follows that the posterior density of $\mathbf{S}$ is:

$$f(\mathbf{s} \mid \mathbf{z}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\alpha}) = \frac{|D(\boldsymbol{\omega})|^{-1}\phi_n(\mathbf{t})\Phi_n(D(\boldsymbol{\alpha})\mathbf{t})\Phi_n(D(\mathbf{z})\mathbf{s})}{\prod_{i=1}^{n}\Phi_2\left(\begin{pmatrix}\eta_i z_i \\ 0\end{pmatrix}\middle| \begin{pmatrix}1+\omega_i^2 z_i^2 & \alpha_i\omega_i z_i \\ \alpha_i\omega_i z_i & 1+\alpha_i^2\end{pmatrix}\right)}, \tag{12}$$

which depends on $(\mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\sigma})$ through $\mathbf{z} = D(\boldsymbol{\sigma})^{-1}(\mathbf{y} - \boldsymbol{\mu})$ only. Notice from (11) and (12) that both the prior and the posterior belong to the following family of probability distributions:

$$\mathcal{P} = \left\{ f(\mathbf{s} \mid \boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\alpha}) = K|D(\boldsymbol{\omega})|^{-1}\phi_n(\mathbf{t})\,\Phi_n(D(\boldsymbol{\alpha})\mathbf{t})\,\Phi_n(D(\boldsymbol{\psi})\mathbf{s}) : (\boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\alpha}) \in \mathcal{A} \right\}, \tag{13}$$

where $\mathbf{t} = D(\boldsymbol{\omega})^{-1}(\mathbf{s} - \boldsymbol{\eta})$, $K^{-1} = \int |D(\boldsymbol{\omega})|^{-1}\phi_n(\mathbf{t})\,\Phi_n(D(\boldsymbol{\alpha})\mathbf{t})\,\Phi_n(D(\boldsymbol{\psi})\mathbf{s})\mathrm{d}\mathbf{s}$ and $\mathcal{A}$ denotes the set of labels of the distributions. Notice that the prior in (11) is obtained eliciting $\psi_i = 0$ or $\psi_i \to \infty$, for all $i = 1, \ldots, n$.

The family of probability $\mathcal{P}$ in (13) is not closed under sampling of the family of distributions associated to the likelihood in (10) but if the prior is chosen in $\mathcal{P}$, the posterior will also be in this family. Although the family $\mathcal{P}$ is not a natural conjugate family with respect to the likelihood in (10), this family yields interesting results related to the computation of the posterior of the shape parameter. It brings tractability to the problem of inference on $\mathbf{S}$ and also leads to a simple interpretation of the results since the posterior is obtained from the prior by updating the part of the distribution which introduces the skewness only.

Similar results can be obtained for the other cases discussed in Section 2. For example, tractable implementation of the posteriors for the cases presented in part (ii) of Proposition 1, which considers correlations among the S_i's, and in parts (i) and (ii) of Proposition 2, where it is assumed that the Y_i's are dependent, are, respectively, reached if the following families are considered:

1(ii) $\mathcal{P} = \{f(\mathbf{s} \mid \boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\alpha}) = K\phi_n(\mathbf{s} \mid \boldsymbol{\eta}, \Omega)\,\Phi(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t})\,\Phi_n(D(\boldsymbol{\psi})\mathbf{s}) : (\boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\alpha}) \in \mathcal{A}\}$, in which $K^{-1} = \int \phi_n(\mathbf{s} \mid \boldsymbol{\eta}, \Omega)\,\Phi(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t})\,\Phi_n(D(\boldsymbol{\psi})\mathbf{s})\mathrm{d}\mathbf{s}$ and $\mathcal{A}$ denotes the set of labels of the distributions.

2(i) $\mathcal{P} = \{f(\mathbf{s} \mid \boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\alpha}) = K|D(\boldsymbol{\omega})|^{-1}\phi_n(\mathbf{t})\,\Phi_n(D(\boldsymbol{\alpha})\mathbf{t})\,\Phi(\mathbf{s}^{\mathrm{T}}\boldsymbol{\psi}) : (\boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\alpha}) \in \mathcal{A}\}$, in which $K^{-1} = \int |D(\boldsymbol{\omega})|^{-1}\phi_n(\mathbf{t})\,\Phi_n(D(\boldsymbol{\alpha})\mathbf{t})\,\Phi(\mathbf{s}^{\mathrm{T}}\boldsymbol{\psi})\mathrm{d}\mathbf{s}$ and $\mathcal{A}$ denotes the set of labels of the distributions.

2(ii) $\mathcal{P} = \{f(\mathbf{s} \mid \boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\alpha}) = K\phi_n(\mathbf{s} \mid \boldsymbol{\eta}, \Omega)\,\Phi(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t})\,\Phi(\mathbf{s}^{\mathrm{T}}\boldsymbol{\psi}) : (\boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\alpha}) \in \mathcal{A}\}$, in which $K^{-1} = \int \phi_n(\mathbf{s} \mid \boldsymbol{\eta}, \Omega)\,\Phi(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{s})\,\Phi(\mathbf{s}^{\mathrm{T}}\boldsymbol{\psi})\mathrm{d}\mathbf{s}$ and $\mathcal{A}$ denotes the set of labels of the distributions.

Notice that for these families, the posterior mean and variance of the shape parameter can be obtained by applying the results presented in Section 3. For further properties, see [1].

4.2. *Robust inference for location and scale parameters*

Assume that the location and scale parameters (say, for instance, $\boldsymbol{\mu}$ and Σ, respectively) are unknown. Suppose that it is reasonable to assume that such parameters are independent of $\mathbf{S}$ and have a joint prior distribution $\pi(\boldsymbol{\mu}, \Sigma)$. Consequently, the joint posterior distribution for $(\boldsymbol{\mu}, \Sigma)$ is given by:

$$\pi(\boldsymbol{\mu}, \Sigma \mid \mathbf{y}) \propto \pi(\boldsymbol{\mu}, \Sigma) \int f(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, \mathbf{s})\mathrm{d}\lambda(\mathbf{s}).$$

For instance, assume that $[\mathbf{Y} \mid \mathbf{S} = \mathbf{s}, \boldsymbol{\mu}, \Sigma] \sim SN_n(\boldsymbol{\mu}, \Sigma, \mathbf{s})$ and that $S_i \overset{\text{ind.}}{\sim} SN(\eta_i, \omega_i^2, \alpha_i)$, $i = 1, \ldots, n$. From part (i) of Proposition 2, it follows that

$$\pi(\boldsymbol{\mu}, \Sigma \mid \mathbf{y}) \propto \pi(\boldsymbol{\mu}, \Sigma)\phi_n(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma)\,\Phi_{n+1}\left(\begin{pmatrix} \boldsymbol{\eta}^{\mathrm{T}}\mathbf{z} \\ \mathbf{0} \end{pmatrix} \middle| \begin{pmatrix} 1 + \mathbf{z}^{\mathrm{T}}D^2(\boldsymbol{\omega})\mathbf{z} & \mathbf{z}^{\mathrm{T}}D(\boldsymbol{\omega})D(\boldsymbol{\alpha}) \\ D(\boldsymbol{\alpha})D(\boldsymbol{\omega})\mathbf{z} & I_n + D^2(\boldsymbol{\alpha}). \end{pmatrix} \right). \tag{14}$$

From (14) it follows that the posterior for $(\boldsymbol{\mu}, \Sigma)$ is a skewed distribution. Moreover, a very important result arises from (14) if we assume that $\boldsymbol{\alpha} = \mathbf{0}$ and $\boldsymbol{\eta} = \mathbf{0}$, which corresponds to eliciting a centered normal prior distribution for $\mathbf{S}$, $\mathbf{S} \sim N_n(\mathbf{0}, \Omega)$ say. In this case the joint posterior distribution for $(\boldsymbol{\mu}, \Sigma)$ becomes:

$$\pi(\boldsymbol{\mu}, \Sigma \mid \mathbf{y}) \propto \pi(\boldsymbol{\mu}, \Sigma)\phi_n(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma). \tag{15}$$

Notice that the posterior in (15) – which is built assuming a dependent *SN* likelihood – is the same distribution as when we assume a dependent normal likelihood, that is, assuming $[\mathbf{Y} \mid \boldsymbol{\mu}, \Sigma] \sim N_n(\boldsymbol{\mu}, \Sigma)$. This invariance property can also be established for the independent *SN* model $[Y_i \mid \mathbf{S} = \mathbf{s}, \boldsymbol{\mu}, \boldsymbol{\sigma}] \sim N(\mu_i, \sigma_i^2, s_i)$, $i = 1, \ldots, n$, where the scale parameter is $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)^{\mathrm{T}}$. In fact, robust inference [18] for the location and scale parameters is obtained for the special cases described in the following corollaries. Assume only non-degenerate priors for the shape parameter.

Corollary 2. *If $(\boldsymbol{\mu}, \boldsymbol{\sigma})$ is independent of the shape parameter, the posterior distribution for $(\boldsymbol{\mu}, \boldsymbol{\sigma})$ is*

$$\pi(\boldsymbol{\mu}, \boldsymbol{\sigma} \mid \mathbf{Y} = \mathbf{y}) \propto \pi(\boldsymbol{\mu}, \boldsymbol{\sigma})|D(\boldsymbol{\sigma})|^{-1}\phi_n(\mathbf{z}),$$

where $\mathbf{z} = D(\boldsymbol{\sigma})^{-1}(\mathbf{y} - \boldsymbol{\mu})$ for each of the following situations:

(i) $[Y_i \mid \mathbf{S} = \mathbf{s}, \boldsymbol{\mu}, \boldsymbol{\sigma}] \overset{\text{ind.}}{\sim} SN(\mu_i, \sigma_i^2, s_i)$ *and* $S_i \overset{\text{ind.}}{\sim} SN(0, \omega_i^2, 0)$, $i = 1, \ldots, n$;

(ii) $[Y_i \mid \mathbf{S} = \mathbf{s}, \boldsymbol{\mu}, \boldsymbol{\sigma}] \overset{\text{ind.}}{\sim} SN(\mu_i, \sigma_i^2, s_i)$, $i = 1, \ldots, n$, *and* $\mathbf{S} \sim SN_n(\mathbf{0}, \Omega, \mathbf{0})$ *for a diagonal matrix* Ω;

(iii) $[Y_i \mid \boldsymbol{\mu}, \boldsymbol{\sigma}] \overset{\text{ind.}}{\sim} N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$.

Notice that if $[Y_i \mid S = s, \boldsymbol{\mu}, \boldsymbol{\sigma}] \overset{\text{ind.}}{\sim} SN(\mu_i, \sigma_i^2, s)$, $i = 1, \ldots, n$ (Case (B1), Section 2), the robustness properties mentioned in Corollary 2 will follow if we elicit a degenerate prior for S only, that is, if we declare $P(S = 0) = 1$. This case, however, corresponds to the normal statistical model.

Corollary 3. *Assume $(\boldsymbol{\mu}, \Sigma)$ and the shape parameter are independent. Then the posterior distribution for $(\boldsymbol{\mu}, \Sigma)$ is*

$$\pi(\boldsymbol{\mu}, \Sigma \mid \mathbf{Y} = \mathbf{y}) \propto \pi(\boldsymbol{\mu}, \Sigma)\phi_n(\mathbf{y}|\boldsymbol{\mu}, \Sigma)$$

for each of the following situations:

(i) $[\mathbf{Y} \mid \mathbf{S} = \mathbf{s}, \boldsymbol{\mu}, \Sigma] \sim SN_n(\boldsymbol{\mu}, \Sigma, \mathbf{S})$ *and* $S_i \overset{\text{ind.}}{\sim} SN(0, \omega_i^2, 0)$, $i = 1, \ldots, n$;
(ii) $[\mathbf{Y} \mid \mathbf{S} = \mathbf{s}, \boldsymbol{\mu}, \Sigma] \sim SN_n(\boldsymbol{\mu}, \Sigma, \mathbf{S})$ *and* $\mathbf{S} \sim SN_n(\mathbf{0}, \Omega, \mathbf{0})$;
(iii) $[\mathbf{Y} \mid S = s, \boldsymbol{\mu}, \Sigma] \sim SN_n(\boldsymbol{\mu}, \Sigma, s\mathbf{1}_n)$ *and* $S \sim SN(0, \omega^2, 0)$;
(iv) $[\mathbf{Y} \mid \boldsymbol{\mu}, \Sigma] \sim N_n(\boldsymbol{\mu}, \Sigma)$.

The proof of Corollary 2 follows straightforwardly from Propositions 1 and 3 and the results presented in Corollary 3 are a consequence of Propositions 2 and 4.

If the location and scale parameters are such that $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta})$ and $\Sigma = \Sigma(\boldsymbol{\theta})$ (respectively $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\theta})$), where $\boldsymbol{\theta}$ is a p-dimensional vector of unknown parameters, $p < n$, and is independent of the shape parameter $\mathbf{S}$, it follows from the previous results that robust inference for $\boldsymbol{\theta}$ is also obtained. Indeed, if $\boldsymbol{\theta}$ has prior distribution $\pi(\boldsymbol{\theta})$, the posterior distribution for $\boldsymbol{\theta}$ is

$$\pi(\boldsymbol{\theta} \mid \mathbf{y}) \propto \pi(\boldsymbol{\theta}) \int f(\mathbf{y} \mid \boldsymbol{\mu}(\boldsymbol{\theta}), \Sigma(\boldsymbol{\theta}), \mathbf{s}) \mathrm{d}\lambda(\mathbf{s}).$$

In this case, the number of parameters to be estimated is reduced. The important applications of such results are the analysis of independent and dependent SN regression models. In this case $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)$. For instance, first, assume that the response vector $\mathbf{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ given $\boldsymbol{\mu}$, Σ and $\mathbf{S} = \mathbf{s}$, has the $SN_n(\boldsymbol{\mu} = X\boldsymbol{\beta}, \Sigma = \sigma^2 I_n, \mathbf{s})$ distribution. Consider a dependent $SN_n(\boldsymbol{\eta}_0, \Omega_0, \boldsymbol{\alpha}_0)$ as the prior distribution for the shape vector of parameters $\mathbf{S} = (S_1, \ldots, S_n)^{\mathrm{T}}$. Suppose that it is reasonable to assume $\mathbf{S}$ independent of the regression model parameters $(\boldsymbol{\beta}, \sigma^2)$. From part (ii) of Proposition 2 it follows that the marginal likelihood function $f(\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2)$ and the posterior full conditional distribution for $\mathbf{S}$, which is denoted by $f(\mathbf{s} \mid \mathbf{y}, \boldsymbol{\beta}, \sigma^2) = f(\mathbf{s} \mid \mathbf{z})$, are given, respectively, by:

$$f(\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2) = 4\phi_n(\mathbf{y} \mid X\boldsymbol{\beta}, \sigma^2)\,\Phi_2\left(\begin{pmatrix}\boldsymbol{\eta}_0^{\mathrm{T}}\mathbf{z} \\ 0\end{pmatrix}\middle| \begin{pmatrix} 1 + \mathbf{z}^{\mathrm{T}}\Omega_0\mathbf{z} & \mathbf{z}^{\mathrm{T}}\Omega_0 D(\boldsymbol{\omega}_0)^{-1}\boldsymbol{\alpha}_0 \\ \boldsymbol{\alpha}_0^{\mathrm{T}} D(\boldsymbol{\omega}_0)^{-1}\Omega_0\mathbf{z} & 1 + \boldsymbol{\alpha}_0^{\mathrm{T}} D(\boldsymbol{\omega}_0)^{-1}\Omega_0 D(\boldsymbol{\omega}_0)^{-1}\boldsymbol{\alpha}_0 \end{pmatrix}\right), \tag{16}$$

and

$$f(\mathbf{s} \mid \mathbf{z}) = \frac{\phi_n(\mathbf{s} \mid \boldsymbol{\eta}_0, \Omega_0)\,\Phi(\boldsymbol{\alpha}_0^{\mathrm{T}}(\mathbf{s} - \boldsymbol{\eta}_0))\,\Phi(\mathbf{z}^{\mathrm{T}}\mathbf{s})}{\Phi_2\left(\begin{pmatrix}\boldsymbol{\eta}_0^{\mathrm{T}}\mathbf{z} \\ 0\end{pmatrix}\middle| \begin{pmatrix} 1 + \mathbf{z}^{\mathrm{T}}\Omega_0\mathbf{z} & \mathbf{z}^{\mathrm{T}}\Omega_0 D(\boldsymbol{\omega}_0)^{-1}\boldsymbol{\alpha}_0 \\ \boldsymbol{\alpha}_0^{\mathrm{T}} D(\boldsymbol{\omega}_0)^{-1}\Omega_0\mathbf{z} & 1 + \boldsymbol{\alpha}_0^{\mathrm{T}} D(\boldsymbol{\omega}_0)^{-1}\Omega_0 D(\boldsymbol{\omega}_0)^{-1}\boldsymbol{\alpha}_0 \end{pmatrix}\right)},$$

where $\mathbf{z} = (\mathbf{y} - X\boldsymbol{\beta})\sigma^{-1}$. An important simplification occurs if $\boldsymbol{\alpha}_0 = \boldsymbol{\eta}_0 = \mathbf{0}$, that is, if the dependent multivariate $N_n(\mathbf{0}, \Omega_0)$ prior is considered for the shape vector of parameters $\mathbf{S}$. In particular, when eliciting such a prior, the marginal likelihood in (16) reduces to the likelihood function for the usual normal regression model $N_n(X\boldsymbol{\beta}, \sigma^2 I_n)$, that is, $f(\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2) = \phi_n(\mathbf{y} \mid X\boldsymbol{\beta}, \sigma^2 I_n)$ in both cases. Consequently, independently of the prior specification for $(\boldsymbol{\beta}, \sigma^2)$, the joint posterior distribution for such parameters is the same as if we were to assume normality for the regression error terms (see Corollary 3).

As a consequence of part (i) of Proposition 1 and Corollary 2, we have, for instance, that robust inference for $\boldsymbol{\beta}$ and σ^2 is also obtained for the independent SN regression model $[Y_i \mid \boldsymbol{\beta}, \sigma^2, S_i = s_i] \overset{\text{ind.}}{\sim} SN(\mu_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}, \sigma^2, s_i)$, $i = 1, \ldots, n$, where $\mathbf{x}_1^{\mathrm{T}}, \ldots, \mathbf{x}_n^{\mathrm{T}}$ denote the rows of the design matrix X. In this case it is necessary to assume that (a) the shape parameters $S_1, \ldots, S_n$ are independent; (b) they are independent of $(\boldsymbol{\beta}, \sigma^2)$; and (c) they are normally distributed with zero mean, i.e., $S_i \overset{\text{ind.}}{\sim} N(0, \omega_{0i}^2)$, $i = 1, \ldots, n$. In fact, robust inference for $\boldsymbol{\beta}$ and σ^2, in the sense discussed in this paper, follows for all cases discussed in Corollaries 2 and 3.

5. Conclusion

In this paper, classes of shape mixtures of both dependent and independent multivariate SN distributions were considered and some aspects on Bayesian inference in these classes were discussed. For particular situations, the results obtained provide a tractable implementation for the posterior of the shape parameters. In addition, robust inference for the location and scale parameters was also obtained under special conditions.

Despite the importance, from a Bayesian point of view, of the obtained results, only a partial answer for the problem of inference for location, scale and shape parameters has been provided. One of the greatest challenges related to this problem is to construct posterior distributions for such parameters under more general conditions and, if at all possible, to derive the conjugate family in these cases.

Acknowledgments

The authors thank two referees whose comments and suggestions have contributed to the improvement of the paper. The research of R.B. Arellano-Valle was supported in part by FONDECYT (Chile), grants 1040865, 7060133 and 7070137. The work of M.G. Genton was partially supported by NSF grant DMS-0504896 and by a grant from the Swiss National Science Foundation. R.H. Loschi acknowledges CNPq (*Conselho Nacional de Desenvolvimento Científico e Tecnológico*) of the Ministry for Science and Technology of Brazil, grants 3004505/2006-4, 472066/2004-8 and 472877/2006-2, for a partial allowance to her research.

References

[1] R.B. Arellano-Valle, A. Azzalini, On the unification of families of skew-normal distributions, Scand. J. Statist. 33 (2006) 561–574.
[2] R.B. Arellano-Valle, M.D. Branco, M.G. Genton, A unified view on skewed distributions arising from selections, Canad. J. Statist. 34 (2006) 581–601.
[3] R.B. Arellano-Valle, M.G. Genton, On fundamental skew distributions, J. Multivariate Anal. 96 (2005) 93–116.
[4] R.B. Arellano-Valle, L.M. Castro, M.G. Genton, H.W. Gómez, Bayesian inference for shape mixtures of skewed distributions, with application to regression analysis, 2007, preprint.
[5] R.B. Arellano-Valle, H.W. Gómez, F.A. Quintana, A new class of skew-normal distributions, Comm. Statist. Theory Methods 33 (2004) 1465–1480.
[6] A. Azzalini, A class of distributions which includes the normal ones, Scand. J. Statist. 12 (1985) 171–178.
[7] A. Azzalini, Further results on a class of distributions which includes the normal ones, Statistica 44 (1986) 199–208.
[8] A. Azzalini, The skew-normal distribution and related multivariate families. With discussion by Marc G. Genton and a rejoinder by the author, Scand. J. Statist. 32 (2005) 159–200.
[9] A. Azzalini, A. Capitanio, Statistical applications of the multivariate skew normal distribution, J. R. Statist. Soc. Ser. B 61 (1999) 579–602.
[10] A. Azzalini, A. Dalla Valle, The multivariate skew-normal distribution, Biometrika 83 (1996) 715–726.
[11] A. Azzalini, M.G. Genton, Robust likelihood methods based on the skew-*t* and related distributions, Internat. Statist. Rev. 76 (2008) 106–129.
[12] D. Barry, J.A. Hartigan, Product partition models for change point problems, Ann. Statist. 20 (1992) 260–279.
[13] J.M. Bernardo, A.F.M. Smith, Bayesian Theory, John Wiley & Sons, New York, 1994.
[14] J.A. Dominguez-Molina, G. González-Farías, R. Ramos-Quiroga, Skew-normality in stochastic frontier analysis, in: M.G. Genton (Ed.), Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality, Chapman & Hall/CRC, Boca Raton, FL, 2004, pp. 223–241.
[15] M.G. Genton, Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality, Edited Volume, Chapman & Hall/CRC, Boca Raton, FL, 2004, 416 pp.
[16] M.G. Genton, N. Loperfido, Generalized skew-elliptical distributions and their quadratic forms, Ann. Inst. Statist. Math. 57 (2005) 389–401.
[17] Y. Ma, M.G. Genton, A flexible class of skew-symmetric distributions, Scand. J. Statist. 31 (2004) 459–468.
[18] A. O'Hagan, Bayesian Inference, in: Kendall's Advanced Theory of Statistics 2A, John Wiley & Sons, New York, USA, 1994.
[19] C.D. Paulino, M.A.A. Turkeman, B. Murteira, Estatística Bayesiana, Fundaçao Calouste Gulbenkian, Lisboa, Portugal, 2003.
[20] G.M. Tallis, The moment generating function of the truncated multi-normal distribution, J. R. Statist. Soc. Ser. B 23 (1961) 223–229.
[21] J. Wang, J. Boyer, M.G. Genton, A skew-symmetric representation of multivariate distributions, Statist. Sinica 14 (2004) 1259–1270.