Journal of Multivariate Analysis 166 (2018) 98–110

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Scale and shape mixtures of multivariate skew-normal distributions

Reinaldo B. Arellano-Valle [a], Clécio S. Ferreira [b], Marc G. Genton [c,*]

[a] Department of Statistics, Pontificia Universidad Católica de Chile, Santiago 22, Chile
[b] Department of Statistics, Federal University of Juiz de Fora, Juiz de Fora 36036-900, Brazil
[c] Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

ARTICLE INFO

Article history:
Received 24 June 2017
Available online 26 February 2018

AMS subject classifications:
62H05
62H12

Keywords:
EM-algorithm
Scale mixtures of normal distributions
Scale mixtures of skew-normal distributions
Shape mixtures of skew-normal distributions
Skew-normal distribution
Skew scale mixtures of normal distributions

ABSTRACT

We introduce a broad and flexible class of multivariate distributions obtained by both scale and shape mixtures of multivariate skew-normal distributions. We present the probabilistic properties of this family of distributions in detail and lay down the theoretical foundations for subsequent inference with this model. In particular, we study linear transformations, marginal distributions, selection representations, stochastic representations and hierarchical representations. We also describe an EM-type algorithm for maximum likelihood estimation of the parameters of the model and demonstrate its implementation on a wind dataset. Our family of multivariate distributions unifies and extends many existing models of the literature that can be seen as submodels of our proposal.

1. Introduction

In recent years, the use of the multivariate skew-normal (SN) distribution [13,14] in both theoretical and applied studies has been increasingly popular. In several of these studies, the location-scale version of the multivariate SN distribution has been implemented using different parametrizations but equivalent to the original one [3]; see also the books by Genton [22] and Azzalini and Capitanio [16]. In this paper, we consider the multivariate SN version with the parametrization proposed by Arellano-Valle and Genton [8]. According to these authors, a p-dimensional random vector $\mathbf{Y}$ follows a multivariate SN distribution with location parameter $\boldsymbol{\mu} \in \mathbb{R}^p$, scale parameter $\Sigma > 0$ (a $p \times p$ positive definite matrix) and skewness/shape parameter $\boldsymbol{\lambda} \in \mathbb{R}^p$, denoted by $\mathbf{Y} \sim \mathcal{SN}_p(\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda})$, if its probability density function (pdf) is given, for all $\mathbf{y} \in \mathbb{R}^p$, by

$$f_{\mathrm{SN}}(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda}) = 2\phi_p(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma)\Phi_1\left\{\boldsymbol{\lambda}^\top \Sigma^{-1/2}(\mathbf{y} - \boldsymbol{\mu})\right\}, \quad (1)$$

where $\phi_p(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma) = |\Sigma|^{-1/2}\phi_p(\mathbf{z})$, with $\phi_p(\mathbf{z}) = (2\pi)^{-p/2}\exp(-\mathbf{z}^\top\mathbf{z}/2)$ and $\mathbf{z} = \Sigma^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$, is the pdf of the p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ, denoted by $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$, Φ_1 denotes the cumulative distribution function (cdf) of the standard normal distribution $\mathcal{N}_1(0, 1)$, and $\Sigma^{-1/2}$ is the symmetric square root matrix of Σ^{-1}. When $\boldsymbol{\lambda} = \mathbf{0}$, the SN distribution reduces to the multivariate normal distribution, viz. $\mathbf{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$. Also, in terms of the "standardized" random vector $\mathbf{Z} \sim \mathcal{SN}_p(\boldsymbol{\lambda}) \equiv \mathcal{SN}_p(\mathbf{0}, \mathbf{I}_p, \boldsymbol{\lambda})$, where $\mathbf{I}_p$ is the identity matrix of dimension $p \times p$, the multivariate SN distribution can be represented stochastically as follows:

$$\mathbf{Y} = \boldsymbol{\mu} + \Sigma^{1/2}\mathbf{Z}, \quad \text{with} \quad \mathbf{Z} \stackrel{d}{=} \boldsymbol{\delta}|Z_0| + (\mathbf{I}_p - \boldsymbol{\delta}\boldsymbol{\delta}^\top)^{1/2}\mathbf{Z}_1, \quad (2)$$

* Corresponding author.
E-mail address: marc.genton@kaust.edu.sa (M.G. Genton).

https://doi.org/10.1016/j.jmva.2018.02.007

where "$\overset{d}{=}$" means "equal in distribution", $\boldsymbol{\delta} = \boldsymbol{\lambda}/(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}$, $|Z_0|$ denotes the absolute value of Z_0, and $Z_0 \sim \mathcal{N}_1(0, 1)$ is independent of $\mathbf{Z}_1 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$. This representation is very useful to derive most of the main properties of the multivariate SN distribution. It is equivalent to the following hierarchical representation as location mixture of the multivariate normal distribution, which is of great utility in the formulation of the SN statistical model

$$\mathbf{Y} \mid U = u \sim \mathcal{N}_p(\boldsymbol{\mu} + \bar{\boldsymbol{\lambda}}u, \boldsymbol{\Psi}), \tag{3}$$

where $U = \gamma\,|Z_0| \sim \mathcal{HN}_1(0, \gamma^2)$, a half-normal random variable, $\bar{\boldsymbol{\lambda}} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}$, $\boldsymbol{\Psi} = \boldsymbol{\Sigma} - \gamma^2\bar{\boldsymbol{\lambda}}\bar{\boldsymbol{\lambda}}^\top$ and $\gamma = 1/(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}$. Also, from (2) or (3) it follows straightforwardly that the expectation and variance of $\mathbf{Y}$ are, respectively,

$$\mathrm{E}(\mathbf{Y}) = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\,\gamma\bar{\boldsymbol{\lambda}} \quad \text{and} \quad \mathrm{var}(\mathbf{Y}) = \boldsymbol{\Sigma} - \frac{2}{\pi}\gamma^2\bar{\boldsymbol{\lambda}}\bar{\boldsymbol{\lambda}}^\top. \tag{4}$$

Currently the multivariate SN distribution is the most popular member of the so called "skew-symmetric" (SS) class introduced by Azzalini and Capitanio [13] in their Lemma 1. In the location-scale case, it is defined in terms of its pdf, given, for all $\mathbf{y} \in \mathbb{R}^p$, by

$$f_{\mathrm{SS}}(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = 2|\boldsymbol{\Sigma}|^{-1/2} f_0\{\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\} G_0[w\{\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\}], \tag{5}$$

where f_0 is the pdf of a p-variate centrally symmetric distribution, i.e., $f_0(-\mathbf{z}) = f_0(\mathbf{z})$ for all $\mathbf{z} \in \mathbb{R}^p$, G_0 is a univariate cdf such that $G_0(-z) = 1 - G_0(z)$ for all $z \in \mathbb{R}$ (symmetric about zero), and w is an odd real-valued function, so $w(-\mathbf{z}) = -w(\mathbf{z})$. The SS class is equivalent to the one studied by Wang et al. [34] for which $G_0\{w(\mathbf{z})\}$ is replaced by a nonnegative function $\Pi(\mathbf{z})$ such that $\Pi(\mathbf{z}) + \Pi(-\mathbf{z}) = 1$ for all $\mathbf{z} \in \mathbb{R}^p$. The SS class includes the generalized skew-elliptical (GSE) distributions studied by Genton and Loperfido [23] in which f_0 is a p-variate spherical density generator and $\Pi(\mathbf{z})$ is used. As with the SN case, the linear function $w(\mathbf{z}) = \boldsymbol{\lambda}^\top\mathbf{z}$ with $\boldsymbol{\lambda} \in \mathbb{R}^p$ in the SS pdf given in (5) is interpreted as a skewness/shape parameter vector. Although it has been the most common specification of the function $w(\mathbf{z})$, one can find many other options such as $w(\mathbf{z}) = \boldsymbol{\lambda}^\top\mathbf{z}/(1+\tau\mathbf{z}^\top\mathbf{z})^{1/2}$ and $w(\mathbf{z}) = \boldsymbol{\lambda}^\top\mathbf{z} + \tau(\boldsymbol{\lambda}^\top\mathbf{z})^3$, where $\tau > 0$, among others. For a more comprehensive and detailed discussion on SS distributions, see the book of Azzalini and Capitanio [16].

Among the different extensions of the multivariate SN distribution, we emphasize here the rich families obtained as scale and/or shape mixtures of the SN distribution, which have been initially studied in [7,9,17,20,21]. Most of these distributions are part of the SS class, and have hierarchical representations that are useful in statistical analysis based on these models. In this work, we introduce a multivariate class of scale and shape mixtures of the SN (hereinafter SSMSN) distributions which contains all of the aforementioned subfamilies as special cases, thus extending some related results presented firstly within a univariate context by Arellano-Valle et al. [7]. The relevance of such distributions is due to the fact that they can be represented hierarchically in terms of the multivariate normal distribution. This facilitates substantially the implementation of both classical and Bayesian statistical analysis. After introducing the SSMSN class, we present the main probabilistic properties of this family and we establish the theoretical foundations for subsequent inference with this model.

2. Scale and shape mixtures of multivariate SN distributions

2.1. Background

The symmetric heavy tails family obtained as scale mixtures of normal (SMN) distributions [1] has attracted much attention during the last two decades mainly because they allow for robust modeling of symmetric empirical distributions; see, e.g., [2,4,5,18,19,27]. This motivated Branco and Dey [17] to introduce the scale mixtures of skew-normal (SMSN) distributions, thus providing a more flexible family of robust statistical models. With similar motivations, [7,9] introduced the shape mixtures of skew-normal (SHMSN) distributions, [20,21] studied the so-called skew-scale mixtures of normal (SSMN) and [33] considered the SMSN subclass of skew-normal generalized hyperbolic (SNGH) distributions. More recently, other forms of mixtures have emerged such as the variance–mean (or location-scale) mixtures of skew-normal distributions [12] and the location-scale mixtures of skew-elliptical distributions [31], both including the SMSN family as a particular case, but not the SHMSN and SSMN families.

In this section, we introduce the SSMSN family of multivariate skewed distributions which is obtained as scale and shape/skewness mixtures of the multivariate SN distribution. The SSMSN family contains as special cases the SMSN class, hence also the asymmetric SNGH and symmetric SMN classes, the SHMSN, SSMN, and also the scale mixtures of skew-Normal-Cauchy distributions considered recently by [25]. For a better understanding of the new SSMSN family, we start recalling the definitions of the SMSN and SHMSN classes of distributions.

Let $\mathbf{Y}$ be a p-dimensional continuous random vector. The distribution of $\mathbf{Y}$ is in the multivariate SMSN class with location parameter $\boldsymbol{\mu}$, dispersion matrix $\boldsymbol{\Sigma}$ and skewness/shape parameter $\boldsymbol{\lambda}$, if there is a random variable $V \sim H(\cdot \mid \boldsymbol{\nu})$, a univariate cdf indexed by the parameter vector $\boldsymbol{\nu}$, such that $\mathbf{Y} \mid V = v \sim \mathcal{SN}_p[\boldsymbol{\mu}, \kappa(v)\boldsymbol{\Sigma}, \boldsymbol{\lambda}]$. That is, for $\boldsymbol{\Sigma} > 0$ the pdf of $\mathbf{Y}$ is given, for all $\mathbf{y} \in \mathbb{R}^p$, by

$$f_{\mathrm{SMSN}}(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \int_{\mathcal{S}(H)} f_{\mathrm{SN}}\{\mathbf{y} \mid \boldsymbol{\mu}, \kappa(v)\boldsymbol{\Sigma}, \boldsymbol{\lambda}\}\, dH(v \mid \boldsymbol{\nu}), \tag{6}$$

for some positive scale (weight) function $\kappa(v)$, where for any distribution F, $\mathcal{S}(F)$ denotes the support of F, and $f_{\rm SN}(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ with $\mathbf{y} \in \mathbb{R}^p$ is the multivariate skew-normal pdf in (1). We denote this family by $\mathbf{Y} \sim \mathcal{SMSN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, H, \kappa)$. In (6), it is also usual to assume that $\mathcal{S}(H) \subseteq (0, \infty)$ and $\kappa(v) = 1/v$.

Similarly, the multivariate SHMSN class, in which $\mathbf{Y} \mid S = s \sim \mathcal{SN}_p[\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta(s)\boldsymbol{\lambda}]$, for some mixing random variable $S \sim G(\cdot \mid \boldsymbol{\tau})$ and shape function $\eta(s)$ such that the distribution of $\eta(S)$ is not symmetric about zero, leads to the pdf defined, for all $\mathbf{y} \in \mathbb{R}^p$, by

$$f_{\rm SHMSN}(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\tau}) = \int_{\mathcal{S}(G)} f_{\rm SN}\{\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta(s)\boldsymbol{\lambda}\} dG(s \mid \boldsymbol{\tau}).$$

To refer to this family, we use the notation $\mathbf{Y} \sim \mathcal{SHMSN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, G, \eta)$. For $\boldsymbol{\lambda} = \mathbf{0}$, we have that the SMSN pdf (6) reduces to the SMN pdf given, for all $\mathbf{y} \in \mathbb{R}^p$, by

$$f_{\rm SMN}(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) = \int_{\mathcal{S}(H)} \phi_p\{\mathbf{y} \mid \boldsymbol{\mu}, \kappa(v)\boldsymbol{\Sigma}\} dH(v \mid \boldsymbol{\nu}), \tag{7}$$

where in (7) we simply get the p-variate normal pdf $\phi_p(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\mathbf{y} \in \mathbb{R}^p$.

2.2. Definition and preliminary properties

We explore the idea of mixing simultaneously the scale and shape/skewness parameters of the SN distribution with different, but not necessarily independent, mixing distributions. We thus develop the SSMSN family of distributions defined as follows.

Definition 1. A p-dimensional random vector $\mathbf{Y}$ follows a scale and shape mixture of multivariate skew-normal distributions with a location parameter $\boldsymbol{\mu} \in \mathbb{R}^p$, a $p \times p$ scale matrix $\boldsymbol{\Sigma} > 0$ and shape parameter $\boldsymbol{\lambda} \in \mathbb{R}^p$, if there are two random variables V and S, which have joint distribution $Q(v, s \mid \boldsymbol{\nu}, \boldsymbol{\tau})$ indexed by a vector of parameters $(\boldsymbol{\nu}, \boldsymbol{\tau})$, such that

$$\mathbf{Y} \mid V = v, S = s \sim \mathcal{SN}_p[\boldsymbol{\mu}, \kappa(v)\boldsymbol{\Sigma}, \eta(v, s)\boldsymbol{\lambda}], \tag{8}$$

for some scale (weight) function $\kappa(v)$ and real-valued shape function $\eta(v, s)$ such that the distribution of $\eta(V, S)$ is not symmetric about zero.

According to (8), the pdf of $\mathbf{Y}$ is given, for all $\mathbf{y} \in \mathbb{R}^p$, by

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\tau}) = \int_{\mathcal{S}(Q)} f_{\rm SN}\{\mathbf{y} \mid \boldsymbol{\mu}, \kappa(v)\boldsymbol{\Sigma}, \eta(v, s)\boldsymbol{\lambda}\} dQ(v, s \mid \boldsymbol{\nu}, \boldsymbol{\tau}). \tag{9}$$

For a SSMSN random vector $\mathbf{Y}$ with SN conditional distribution as in (8) or (9), we denote it $\mathbf{Y} \sim \mathcal{SSMSN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, Q, \kappa, \eta)$ or $\mathbf{Y} \sim \mathcal{SSMSN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\tau}, \kappa, \eta)$. We use $\mathbf{Y} \sim \mathcal{SSMSN}_p(\boldsymbol{\lambda}, Q, \kappa, \eta)$ or $\mathbf{Y} \sim \mathcal{SSMSN}_p(\boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\tau}, \kappa, \eta)$ when $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_p$.

The SSMSN family includes finite mixtures either in the scale SN parameter or the shape SN parameter or in both, depending on whether the mixing random variables S or V or both are discrete. Although it is not seen explicitly from (9), the SSMSN pdf in (9) belongs to the GSE subclass of the SS class defined in (5) as established by the following proposition.

Proposition 1. *The SSMSN pdf in (9) is equivalent to the GSE class given, for all $\mathbf{y} \in \mathbb{R}^p$, by*

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\tau}) = 2 f_{SMN}(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) \Pi\{\boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})\}, \tag{10}$$

where $\Pi(\mathbf{z}) = \int_{\mathcal{S}(Q_0)} \Phi_1\{\kappa(v)^{-1/2}\eta(v, s)A(\mathbf{z})\} dQ_0(v, s \mid \mathbf{y}, \boldsymbol{\nu}, \boldsymbol{\tau})$, with $\mathbf{z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$, $A(\mathbf{z}) = \boldsymbol{\lambda}^\top \mathbf{z}$ and $Q_0(v, s \mid \mathbf{y}, \boldsymbol{\nu}, \boldsymbol{\tau})$ being the conditional distribution of $(V, S) \mid \mathbf{Y}_0 = \mathbf{y}$ in which $\mathbf{Y}_0$ denotes the symmetric SMN random vector obtained when $\boldsymbol{\lambda} = \mathbf{0}$, i.e., with pdf $f_{SMN}(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu})$.

Proof. First, note for $\boldsymbol{\lambda} = \mathbf{0}$ that the SSMSN pdf in (9) reduces to the symmetric SMN pdf $f_{SMN}(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu})$ given in (7). Denote by $\mathbf{Y}_0$ the corresponding SMN random vector. In this case, $f_0(\mathbf{y} \mid v, s, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \phi_p\{\mathbf{y} \mid \boldsymbol{\mu}, \kappa(v)\boldsymbol{\Sigma}\}$ becomes the conditional pdf of $\mathbf{Y}_0 \mid V = v, S = s$, which obviously does not depend on s. Let $Q_0(v, s \mid \mathbf{y}, \boldsymbol{\nu}, \boldsymbol{\tau})$ be the conditional distribution of $(V, S) \mid \mathbf{Y}_0 = \mathbf{y}$. This conditional distribution depends on $\mathbf{y}$ only through $\mathbf{z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$ since $\{(V, S) \mid \mathbf{Y}_0 = \mathbf{y}\} \stackrel{d}{=} \{(V, S) \mid \mathbf{Z}_0 = \mathbf{z}\}$, where $\mathbf{Z}_0 = \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}_0 - \boldsymbol{\mu})$ has a spherical SMN distribution; more specifically, $\mathbf{Z}_0 = \kappa(V)^{1/2}\mathbf{U}_0$, where $\mathbf{U}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$. Using these facts in (9) and writing $Q_0(v, s \mid \mathbf{y}, \boldsymbol{\nu}, \boldsymbol{\tau})$ as $Q_0(v, s \mid \mathbf{z}, \boldsymbol{\nu}, \boldsymbol{\tau})$ and $A(\mathbf{z}) = \boldsymbol{\lambda}^\top \mathbf{z} = \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$, we then have

$$\begin{aligned} f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\tau}) &= 2 f_{SMN}(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) \int_{\mathcal{S}(Q_0)} \Phi_1\{\kappa(v)^{-1/2}\eta(v, s)A(\mathbf{z})\} dQ_0(v, s \mid \mathbf{z}, \boldsymbol{\nu}, \boldsymbol{\tau}) \\ &= 2 f_{SMN}(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) \int_{\mathcal{S}(Q_0)} \Pr\{\kappa(v)^{1/2} U_0 \le \eta(v, s)A(\mathbf{z})\} dQ_0(v, s \mid \mathbf{z}, \boldsymbol{\nu}, \boldsymbol{\tau}) \\ &= 2 f_{SMN}(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) \mathrm{E}[\Pr\{\kappa(V)^{1/2} U_0 \le \eta(V, S)A(\mathbf{z}) \mid \mathbf{Z}_0 = \mathbf{z}\}], \end{aligned}$$

where $U_0 \sim \mathcal{N}_1(0, 1)$ is independent of (V, S) and $\mathbf{Z}_0 = \kappa(V)^{1/2}\mathbf{U}_0$, and so of $\mathbf{U}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$. Let $Z_0 = \kappa(V)^{1/2}U_0$ and $W = \eta(V, S)$, and denote by $\Pi(\mathbf{z}) = \mathrm{E}[\Pr\{Z_0 \le WA(\mathbf{z}) \mid \mathbf{Z}_0 = \mathbf{z}\}]$, where the last expectation is with respect to the conditional distribution of $W \mid \mathbf{Z}_0 = \mathbf{z}$ (or $(V, S) \mid \mathbf{Z}_0 = \mathbf{z}$). Then, the last expression above for the SSMSN pdf becomes

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\tau}) = 2f_{SMN}(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, \boldsymbol{\nu})\Pi(\mathbf{z}),$$

in which

$$0 \le \Pi(\mathbf{z}) = \mathrm{E}[F_{Z_0|W,\mathbf{Z}_0=\mathbf{z}}\{WA(\mathbf{z})\}] = F_{\{Z_0-WA(\mathbf{z})\}|\mathbf{Z}_0=\mathbf{z}}(0) \neq 1/2$$

since, by assumption, the distribution of W is not symmetric (at zero). Also, since $A(-\mathbf{z}) = -A(\mathbf{z})$ and $(Z_0, \mathbf{Z}_0) = \kappa(V)^{1/2}(U_0, \mathbf{U}_0)$ is a spherical random vector, hence also centrally symmetric, we have

$$\begin{aligned}
\Pi(-\mathbf{z}) &= \mathrm{E}[\Pr\{Z_0 \le WA(-\mathbf{z}) \mid \mathbf{Z}_0 = -\mathbf{z}\}] = \mathrm{E}[\Pr\{Z_0 \le -WA(\mathbf{z}) \mid \mathbf{Z}_0 = -\mathbf{z}\}] \\
&= \mathrm{E}[\Pr\{-Z_0 \ge WA(\mathbf{z}) \mid -\mathbf{Z}_0 = \mathbf{z}\}] = \mathrm{E}[\Pr\{Z_0 \ge WA(\mathbf{z}) \mid \mathbf{Z}_0 = \mathbf{z}\}] \\
&= 1 - \Pi(\mathbf{z}).
\end{aligned}$$

That is, $\Pi(\mathbf{z}) \ge 0$ and $\Pi(\mathbf{z}) + \Pi(-\mathbf{z}) = 1$ for all $\mathbf{z} \in \mathbb{R}^p$. Thus the proposition is proved. □

From the proof of Proposition 1, for any given $\mathbf{z} = \Sigma^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$ we have in (10) that $\Pi(\mathbf{z}) = \mathrm{E}[F_{Z_{0\mathbf{z}}|W_\mathbf{z}}\{W_\mathbf{z}A(\mathbf{z})\}]$ in which the expectation is with respect to $W_\mathbf{z} = \eta(V_\mathbf{z}, S_\mathbf{z})$, with $Z_{0\mathbf{z}} \stackrel{d}{=} (Z_0 \mid \mathbf{Z}_0 = \mathbf{z})$, $(V_\mathbf{z}, S_\mathbf{z}) \stackrel{d}{=} \{(V, S) \mid \mathbf{Z}_0 = \mathbf{z}\}$ and $(Z_0, \mathbf{Z}_0) = \kappa(V)(U_0, \mathbf{U}_0)$, with $U_0 \sim \mathcal{N}_1(0, 1)$ and $\mathbf{U}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$ which are independent.

Other preliminary properties of the SSMSN distributions are given in the following proposition. They are proved using (4) with γ replaced by $\gamma(v, s) = \eta(v, s)/\{1 + \eta(v, s)^2\boldsymbol{\lambda}\boldsymbol{\lambda}^\top\}^{1/2}$ and the basic properties of the conditional expectation.

Proposition 2. *Let* $\mathbf{Y} \sim \mathcal{SSMSN}_p(\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda}, Q, \kappa, \eta)$, $\gamma(v, s) = \eta(v, s)/\{1 + \eta(v, s)^2\boldsymbol{\lambda}\boldsymbol{\lambda}^\top\}^{1/2}$ *and* $\bar{\boldsymbol{\lambda}} = \Sigma^{1/2}\boldsymbol{\lambda}$. *The following statements hold true.*

(i) *If* $\mathrm{E}\{\kappa(V)^{1/2}\} < \infty$, *then* $\mathrm{E}(\mathbf{Y}) = \boldsymbol{\mu} + \sqrt{2/\pi}\,\mathrm{E}\{\kappa(V)^{1/2}\gamma(V, S)\}\bar{\boldsymbol{\lambda}}$.

(ii) *If additionally* $\mathrm{E}\{\kappa(V)\} < \infty$, *then* $\mathrm{var}(\mathbf{Y}) = \mathrm{E}\{\kappa(V)\}\Sigma - (2/\pi)\mathrm{E}\{\kappa(V)^{1/2}\gamma(V, S)\}^2\bar{\boldsymbol{\lambda}}\bar{\boldsymbol{\lambda}}^\top$.

In particular, if the mixing variables V *and* S *are independent and* $\eta(v, s) = \iota(s)$, *so that* $\gamma(s) = \iota(s)/\{1 + \iota(s)^2\boldsymbol{\lambda}^\top\boldsymbol{\lambda}\}^{1/2}$, *then in* (i) *and* (ii) *we have* $\mathrm{E}\{\kappa(V)^{1/2}\gamma(V, S)\} = \mathrm{E}\{\kappa(V)^{1/2}\}\mathrm{E}\{\gamma(S)\}$.

One important aspect of Definition 1 is that the hypothesis of independence between the mixing random variables (V, S) is not required for the construction of the SSMSN class. The assumption of independence between V and S is useful mainly to simplify the calculation of (9) and also to study some of its main properties. Indeed, if V and S are independent then $Q(v, s \mid \boldsymbol{\nu}, \boldsymbol{\tau}) = H(v \mid \boldsymbol{\nu})G(s \mid \boldsymbol{\tau})$, for all (v, s), where $H(v \mid \boldsymbol{\nu})$ and $G(s \mid \boldsymbol{\tau})$ are the marginal distributions of V and S, respectively. To refer to this case we use the notation $\mathbf{Y} \sim \mathcal{SSMSN}_p(\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda}, HG, \kappa, \eta)$. Another important simplification in the construction of SSMSN distributions occurs when the same mixing variable in the scale and shape functions is considered, i.e., when it is assumed that $S = V$ and so $H = G$. In the next subsection, we describe several special SSMSN subfamilies.

2.3. Special subclasses of SSMSN distributions

The family of SSMSN distributions introduced in Definition 1 is quite large, and contains several subfamilies of asymmetric distributions frequently considered in the literature due to their desirable properties. In fact, some well-known SSMSN subfamilies are obtained as follows:

1. If $\eta(v, s) = \kappa(v) = 1$, then the SSMSN pdf (9) becomes the SN pdf (1).
2. If $\eta(v, s) = 1$, then the SMSN distributions defined by (6) follow. As examples, for $\kappa(v) = 1/v$, we have the multivariate skew-t with $V \sim \mathcal{G}(\nu/2, \nu/2)$, and the multivariate skew-slash, with $V \sim \mathcal{B}(\nu, 1)$.
3. If $\kappa(v) = 1$ and $\eta(v, s) = \iota(s)$, then the SHMSN distributions follow, i.e., for all $\mathbf{y} \in \mathbb{R}^p$,

   $$f(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\tau}) = 2\phi_p(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma)\int_{\mathcal{S}(G)} \Phi_1\{\iota(s)\boldsymbol{\lambda}^\top\Sigma^{-1/2}(\mathbf{y} - \boldsymbol{\mu})\}dG(s \mid \boldsymbol{\tau}).$$

   As examples, for $\iota(S) = S \sim \mathcal{N}(1, \tau)$, we have a multivariate extension of the skew-generalized-normal (SGN) studied by [10], with pdf given by $2\phi_p(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma)\Phi_1\{w(\mathbf{z})\}$, $\mathbf{y} \in \mathbb{R}^p$, with $w(\mathbf{z}) = \boldsymbol{\lambda}^\top\mathbf{z}/\{1+\tau(\boldsymbol{\lambda}^\top\mathbf{z})^2\}^{1/2}$ and $\mathbf{z} = \Sigma^{-1/2}(\mathbf{y}-\boldsymbol{\mu})$. For $\eta(S) = |S| \sim \mathcal{HN}(0, 1)$, we obtain the multivariate skew-normal-Cauchy (SNC) distribution studied recently by [26].
4. If $\eta(v, s) = \kappa(v)^{1/2}$, then the SSMN distributions follow, i.e., for all $\mathbf{y} \in \mathbb{R}^p$,

   $$f(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\tau}) = 2\int_{\mathcal{S}(H)} \phi_p\{\mathbf{y} \mid \boldsymbol{\mu}, \kappa(v)\Sigma\}dH(v \mid \boldsymbol{\nu})\Phi_1\{\boldsymbol{\lambda}^\top\Sigma^{-1/2}(\mathbf{y} - \boldsymbol{\mu})\}.$$

   In this case, for $\kappa(v) = 1/v$, two examples are the multivariate skew-t-normal (STN) and the multivariate skew-slash-normal (SLSN), for which $V \sim \mathcal{G}(\nu/2, \nu/2)$ and $V \sim \mathcal{B}(\nu, 1)$, respectively.

However, the SSMSN family allows us to go beyond the above distributions by varying the scale and shape functions $\kappa(v)$ and $\eta(v,s)$ for a given distribution of (V,S), or alternatively by fixing the scale and shape functions and varying the distribution of (V,S). In the particular case where the mixing variables V and S are independent, it is also convenient to distinguish the following two types of SSMSN subfamilies:

I. **Shape mixtures of SMSN (hereinafter SHSMSN)**: If $\eta(v,s)=\iota(s)$, then, for all $\mathbf{y}\in\mathbb{R}^p$,

$$f(\mathbf{y}\mid\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\boldsymbol{\nu},\boldsymbol{\tau})=2\int_{\mathcal{S}(H)}\phi_p\{\mathbf{y}\mid\boldsymbol{\mu},\kappa(v)\boldsymbol{\Sigma}\}\int_{\mathcal{S}(G)}\Phi_1\{\kappa(v)^{-1/2}\iota(s)\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\}dG(s\mid\boldsymbol{\tau})dH(v\mid\boldsymbol{\nu}), \tag{11}$$

which we denote by $\mathbf{Y}\sim\mathcal{SHSMSN}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},HG,\kappa,\iota)$.

II. **Shape mixtures of SSMN (hereinafter SHSSMN)**: If $\eta(v,s)=\kappa(v)^{1/2}\iota(s)$, then, for all $\mathbf{y}\in\mathbb{R}^p$,

$$f(\mathbf{y}\mid\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\boldsymbol{\nu},\boldsymbol{\tau})=2\int_{\mathcal{S}(H)}\phi_p\{\mathbf{y}\mid\boldsymbol{\mu},\kappa(v)\boldsymbol{\Sigma}\}dH(v\mid\boldsymbol{\nu})\int_{\mathcal{S}(G)}\Phi_1\{\iota(s)\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\}dG(s\mid\boldsymbol{\tau}), \tag{12}$$

which we denote by $\mathbf{Y}\sim\mathcal{SHSSMN}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},HG,\kappa,\kappa^{1/2}\iota)$.

In the construction of the SHSMSN and SHSSMN subfamilies, a relevant aspect that we must consider is that if the distribution of $\iota(S)$ is symmetric (with respect to the origin), then (11) and (12) become symmetric SMN pdfs because $\mathrm{E}[\Phi_1\{\iota(S)x\}]=1/2$ for all $x\in\mathbb{R}$, a fact that was proved in [7] in a univariate context. For instance, this fact occurs when the function $\iota(\cdot)$ is odd and $G(\cdot|\boldsymbol{\tau})$ is symmetric (with respect to the origin). Obviously, to avoid this undesirable result for $\boldsymbol{\lambda}\neq\mathbf{0}$, it is sufficient to impose the condition that $\iota(S)$ is not symmetrically distributed around zero. In particular, when $\iota(\cdot)$ is odd it is sufficient to assume that $G(\cdot|\boldsymbol{\tau})$ is not symmetric around zero. Moreover, a convenient choice for the mixing functions is $\kappa(v)=1/v$ with $\eta(v,s)=s$ for the SHSMSN class given by (11), and with $\eta(v,s)=v^{-1/2}s$ for the SHSSMN class given by (12), while discarding in both cases the symmetry (around the origin) for the distribution of S. Under such specifications, several members of the SHSMSN and SHSSMN classes defined above in I and II, respectively, can be obtained by combining the $\mathcal{G}(\nu/2,\nu/2)$ and $\mathcal{B}(\nu,1)$ distributions for V with the distributions $\mathcal{N}_1(1,1)$ and $\mathcal{HN}_1(0,1)$ for S. In fact, as examples in the SHSSMN class we find:

IIa. Modified skew-t-normal (MSTN) distribution: $V\sim\mathcal{G}(\nu/2,\nu/2)$ and $S\sim\mathcal{N}(1,1)$, with pdf given, for all $\mathbf{y}\in\mathbb{R}^p$, by

$$f(\mathbf{y}\mid\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\boldsymbol{\nu},\boldsymbol{\tau})=2t_p(\mathbf{y}\mid\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)\Phi_1\left[\frac{\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})}{\sqrt{1+\{\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\}^2}}\right].$$

IIb. Modified skew-slash-normal (MSSLN) distribution: $V\sim\mathcal{B}(\nu,1)$ and $S\sim\mathcal{N}(1,1)$, with pdf given, for all $\mathbf{y}\in\mathbb{R}^p$, by

$$f(\mathbf{y}\mid\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\boldsymbol{\nu},\boldsymbol{\tau})=2sl_p(\mathbf{y}\mid\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)\Phi_1\left[\frac{\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})}{\sqrt{1+\{\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\}^2}}\right].$$

IIc. Skew-t-Cauchy (STC) distribution: $V\sim\mathcal{G}(\nu/2,\nu/2)$ and $S\sim\mathcal{HN}(0,1)$, with pdf given, for all $\mathbf{y}\in\mathbb{R}^p$, by

$$f(\mathbf{y}\mid\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\boldsymbol{\nu},\boldsymbol{\tau})=2t_p(\mathbf{y}\mid\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)C_1\{\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\}.$$

IId. Skew-slash-Cauchy (SSLC) distribution: $V\sim\mathcal{B}(\nu,1)$ and $S\sim\mathcal{HN}(0,1)$, with pdf given, for all $\mathbf{y}\in\mathbb{R}^p$, by

$$f(\mathbf{y}\mid\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\boldsymbol{\nu},\boldsymbol{\tau})=2sl_p(\mathbf{y}\mid\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)C_1\{\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\}.$$

Here $t_p(\cdot\mid\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$ and $sl_p(\cdot\mid\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$ denote, respectively, the p-variate pdfs of the Student t and the slash distributions with location $\boldsymbol{\mu}$, dispersion $\boldsymbol{\Sigma}$ and ν degrees of freedom, and $C_1(x)=1/2+(1/\pi)\arctan(x)$ is the cdf of the unit Cauchy distribution. Note, however, that the above examples can also be obtained without the assumption of independence between V and S by considering $\kappa(v)=1/v$ and $\eta(v,s)=s/v$, with: IIa. $V\sim\mathcal{G}(\nu/2,\nu/2)$ and $S\mid V=v\sim\mathcal{N}_1(0,v)$; IIb. $V\sim\mathcal{B}(\nu,1)$ and $S\mid V=v\sim\mathcal{N}_1(0,v)$; IIc. $V\sim\mathcal{G}(\nu/2,\nu/2)$ and $S\mid V=v\sim\mathcal{HN}_1(0,v)$; IId. $V\sim\mathcal{B}(\nu,1)$ and $S\mid V=v\sim\mathcal{HN}_1(0,v)$.

3. Main probabilistic properties

In this section, we study some of the main properties of the SSMSN family given in Definition 1. In particular, we show that the SSMSN family is closed under nonsingular linear transformations, which is a highly desirable property in families of parametric distributions.

3.1. Linear transformations

Proposition 3. *If $\mathbf{Y}\sim\mathcal{SSMSN}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},Q,\kappa,\eta)$, then for any $q\times p$ matrix $\mathbf{A}$ of rank q and $q\times 1$ vector $\mathbf{b}$, $\mathbf{b}+\mathbf{AY}\sim\mathcal{SSMSN}_q(\mathbf{b}+\mathbf{A}\boldsymbol{\mu},\boldsymbol{\Sigma}_A,\boldsymbol{\lambda}_A,Q,\kappa,\eta_A)$, where $\boldsymbol{\Sigma}_A=\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top$, $\boldsymbol{\lambda}_A=\boldsymbol{\Sigma}_A^{-1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}$ and $\eta_A(v,s)=\eta(v,s)/\{1+\eta(v,s)^2(\boldsymbol{\lambda}^\top\boldsymbol{\lambda}-\boldsymbol{\lambda}_A^\top\boldsymbol{\lambda}_A)\}^{1/2}$.*

Proof. Let $\mathbf{Y}_A = \mathbf{b} + \mathbf{A}\mathbf{Y}$, where by definition $\mathbf{Y} \mid V = v, S = s \sim \mathcal{SN}_p[\boldsymbol{\mu}, \kappa(v)\boldsymbol{\Sigma}, \eta(v,s)\boldsymbol{\lambda}]$. Then by adapting Eq. (11) in [13] to our parametrization, we have $\mathbf{Y}_A \mid V = v, S = s \sim \mathcal{SN}_q[\boldsymbol{\mu}_A, \kappa(v)\boldsymbol{\Sigma}_A, \eta_A(v,s)\boldsymbol{\lambda}_A]$. Thus the result is proved. □

It follows from Proposition 3 that a linear transformation of a random vector with SSMSN distribution changes, in general, both the skewness parameter from $\boldsymbol{\lambda}$ to $\boldsymbol{\lambda}_A$ and the shape function from $\eta(v,s)$ to $\eta_A(v,s)$. An exception occurs when $p = q$, which corresponds to a nonsingular linear transformation, in which $\boldsymbol{\lambda}_A = \boldsymbol{\Sigma}_A^{1/2}\mathbf{A}^{-\top}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda}$ and so $\eta_A(v,s) = \eta(v,s)$. The next corollary follows straightforwardly from Proposition 3 and considers only a nonsingular linear transformation.

Corollary 1. *If $\mathbf{Y} \sim \mathcal{SSMSN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, Q, \kappa, \eta)$, then for any nonsingular matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$ and vector $\mathbf{b} \in \mathbb{R}^p$, we have $\mathbf{b} + \mathbf{A}\mathbf{Y} \sim \mathcal{SSMSN}_p(\mathbf{b} + \mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top, \boldsymbol{\lambda}_A, Q, \kappa, \eta)$.*

In particular, if $\mathbf{b} = -\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}$ and $\mathbf{A} = \boldsymbol{\Sigma}^{-1/2}$ then the "standardized" type of random vector is $\mathbf{X} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu}) \sim \mathcal{SSMSN}_p(\mathbf{0}, \mathbf{I}_p, \boldsymbol{\lambda}, Q, \kappa, \eta) \equiv \mathcal{SSMSN}_p(\boldsymbol{\lambda}, Q, \kappa, \eta)$. This yields the following corollary.

Corollary 2. *If $\mathbf{X} \sim \mathcal{SSMSN}_p(\mathbf{0}, \mathbf{I}_p, \boldsymbol{\lambda}, Q, \kappa, \eta)$, then $\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{X} \sim \mathcal{SSMSN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, Q, \kappa, \eta)$.*

According to Corollary 2, most properties of the SSMSN family can be studied from the "standardized" random vector $\mathbf{X} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu}) \sim \mathcal{SSMSN}_p(\mathbf{0}, \mathbf{I}_p, \boldsymbol{\lambda}, Q, \kappa, \eta)$. For instance, as shown by the following corollary, there exist linear transformations whose distribution is canonical SSMSN.

Corollary 3. *Let $\mathbf{X}_* = \boldsymbol{\Gamma}\mathbf{X}$, where $\mathbf{X} \sim \mathcal{SSMSN}_p(\mathbf{0}, \mathbf{I}_p, \boldsymbol{\lambda}, Q, \kappa, \eta)$ and $\boldsymbol{\Gamma} \in \mathbb{R}^{p\times p}$ is an orthogonal matrix such that $\boldsymbol{\Gamma}\boldsymbol{\lambda} = \lambda_*\mathbf{e}_1$, with $\lambda_* = (\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}$ and $\mathbf{e}_1$ being the first unit vector of $\mathbb{R}^p$. Then, $\mathbf{X}_* \sim \mathcal{SSMSN}_p(\mathbf{0}, \mathbf{I}_p, \lambda_*\mathbf{e}_1, Q, \kappa, \eta)$.*

An important consequence of Corollary 3 is that conditionally on V and S, the components $X_{*1}, \ldots, X_{*p}$ of $\mathbf{X}_*$ are independent random variables, with $X_{*1} \mid V = v, S = s \sim \mathcal{SN}_1(0, \kappa(v), \eta(v,s)\lambda_*)$ and the random variables $X_{*2}, \ldots, X_{*p}$ are independent and identically distributed (iid) as follows: $X_{*i} \mid V = v \overset{iid}{\sim} \mathcal{N}_1[0, \kappa(v)]$ for $i \in \{2, \ldots, p\}$. This means that $X_{*1} \sim \mathcal{SSMSN}_1(0, 1, \lambda_*, Q, \kappa, \eta)$ and $X_{*i} \overset{iid}{\sim} \mathcal{SMN}_1(0, 1, G, \kappa)$ for $i \in \{2, \ldots, p\}$, thus extending a similar result proved by [13] for the multivariate SN distribution. As $\mathbf{X} = \boldsymbol{\Gamma}^\top\mathbf{X}_*$, the result of Corollary 3 also allows to substantially simplify the calculation of moments of higher order and other related quantities such as Mardia's measures of multivariate skewness and kurtosis.

The next corollary shows that in the linear transformation $\boldsymbol{\lambda}^\top\mathbf{X}$ the shape function is also not altered. Its proof is a direct consequence from Proposition 3 by letting $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \mathbf{I}_p$, $\mathbf{b} = \mathbf{0}$ and $\mathbf{A} = \boldsymbol{\lambda}$ $(q = 1)$.

Corollary 4. *If $\mathbf{X} \sim \mathcal{SSMSN}_p(\mathbf{0}, \mathbf{I}_p, \boldsymbol{\lambda}, Q, \kappa, \eta)$, then $\boldsymbol{\lambda}^\top\mathbf{X} \sim \mathcal{SSMSN}_1(0, \lambda_*^2, \lambda_*, Q, \kappa, \eta)$, where $\lambda_* = (\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}$.*

3.2. Marginal distributions

Given a partition $\mathbf{Y}_1 \in \mathbb{R}^q$ and $\mathbf{Y}_2 \in \mathbb{R}^{p-q}$ of $\mathbf{Y} \sim \mathcal{SSMSN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, Q, \kappa, \eta)$, suppose we want to obtain the marginal pdf of $\mathbf{Y}_1$ and the conditional pdf of $\mathbf{Y}_2$ given $\mathbf{Y}_1 = \mathbf{y}_1$. Denote the induced partitions for $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\lambda}$ and $\bar{\boldsymbol{\lambda}} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}$ by

$$\boldsymbol{\mu} = \begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}, \quad \boldsymbol{\lambda} = \begin{pmatrix}\boldsymbol{\lambda}_1\\ \boldsymbol{\lambda}_2\end{pmatrix} \quad \text{and} \quad \bar{\boldsymbol{\lambda}} = \begin{pmatrix}\bar{\boldsymbol{\lambda}}_1\\ \bar{\boldsymbol{\lambda}}_2\end{pmatrix}.$$

Also, consider the standard notation $\boldsymbol{\mu}_{2\cdot1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$ and $\boldsymbol{\Sigma}_{22\cdot1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$.

Proposition 4. *If $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top \sim \mathcal{SSMSN}_{q+p-q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, Q, \kappa, \eta)$, then $\mathbf{Y}_1 \sim \mathcal{SSMSN}_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\lambda}_{(1)}, Q, \kappa, \eta_{(1)})$, where $\boldsymbol{\lambda}_{(1)} = \boldsymbol{\Sigma}_{11}^{-1/2}\bar{\boldsymbol{\lambda}}_1 = \boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}$, with $\mathbf{A} = (\mathbf{I}_q, \mathbf{0})$, $\eta_{(1)}(v,s) = \eta(v,s)/\{1 + \eta(v,s)^2\bar{\boldsymbol{\lambda}}_{2\cdot1}^\top\boldsymbol{\Sigma}_{22\cdot1}^{-1}\bar{\boldsymbol{\lambda}}_{2\cdot1}\}^{1/2}$, with $\bar{\boldsymbol{\lambda}}_{2\cdot1} = -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\bar{\boldsymbol{\lambda}}_1 + \bar{\boldsymbol{\lambda}}_2 = \mathbf{B}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}$ and $\mathbf{B} = (-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}, \mathbf{I}_{p-q})$.*

Proof. It is immediate from Proposition 3 with $\mathbf{A} = (\mathbf{I}_q, \mathbf{0})$. □

By using similar arguments as in Proposition 1, the marginal pdf of the random vector $\mathbf{Y}_1 \sim \mathcal{SSMSN}_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, Q, \kappa, \eta_{(1)})$ can also be computed as

$$f(\mathbf{y}_1 \mid \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\nu}, \boldsymbol{\tau}) = 2f_q(\mathbf{y}_1 \mid \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\nu})\mathrm{E}[F_{Z_{0\mathbf{z}_1}|W_{1\mathbf{z}_1}}\{W_{1\mathbf{z}_1}A_1(\mathbf{z}_1)\}],$$

where $f_q(\mathbf{y}_1 \mid \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\nu})$ is the respective SMN marginal pdf,

$$\mathbf{z}_1 = \boldsymbol{\Sigma}_{11}^{-1/2}(\mathbf{y}_1 - \boldsymbol{\mu}_1), \quad A_1(\mathbf{z}_1) = \boldsymbol{\lambda}_{(1)}^\top\mathbf{z}_1, \quad Z_{0\mathbf{z}_1} \overset{d}{=} \{Z_0 \mid \mathbf{Z}_{01} = \mathbf{z}_1\}, \quad W_{1\mathbf{z}_1} = \eta_{(1)}(V_{\mathbf{z}_1}, S_{\mathbf{z}_1}) \overset{d}{=} \{\eta_{(1)}(V,S) \mid \mathbf{Z}_{01} = \mathbf{z}_1\},$$

and $(Z_0, \mathbf{Z}_{01}) = \kappa(V)^{1/2}(U_0, \mathbf{U}_{01})$, with $U_0 \sim \mathcal{N}_1(0,1)$, $\mathbf{U}_{01} \sim \mathcal{N}_q(\mathbf{0}, \mathbf{I}_q)$ and (V,S) being independent.

On the other hand, the conditions $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}^\top = \mathbf{0}$ and $\boldsymbol{\lambda}_2 = \mathbf{0}$ allow us to recover the skewness function for the marginal distribution of $\mathbf{Y}_1$, but they remove the skewness from the marginal distribution of $\mathbf{Y}_2$. Moreover, they are conditionally independent given $V=v$. In fact, under these conditions it easily follows that the conditional joint pdf of $\mathbf{Y}_1$ and $\mathbf{Y}_2$ given $(V,S)=(v,s)$ reduces to

$$f(\mathbf{y}_1,\mathbf{y}_2\mid v,s)=2\phi_q\{\mathbf{y}_1\mid\boldsymbol{\mu}_1,\kappa(v)\Sigma_{11}\}\phi_{p-q}\{\mathbf{y}_2\mid\boldsymbol{\mu}_2,\kappa(v)\Sigma_{22}\}\Phi_1\{\eta(v,s)\kappa(v)^{-1/2}\bar{\boldsymbol{\lambda}}_1^\top\Sigma_{11}^{-1}(\mathbf{y}_1-\boldsymbol{\mu}_1)\}.$$

These facts are summarized next.

Corollary 5. *If in Proposition 4, $\Sigma_{12}=\mathbf{0}$ and $\boldsymbol{\lambda}_2=\mathbf{0}$ then $\mathbf{Y}_1\sim\mathcal{SSMSN}_q(\boldsymbol{\mu}_1,\Sigma_{11},\boldsymbol{\lambda}_{(1)},Q,\kappa,\eta)$, with $\boldsymbol{\lambda}_{(1)}$ as before, and $\mathbf{Y}_2$ is symmetrically distributed as $\mathbf{Y}_2\sim\mathcal{SMN}_{p-q}(\boldsymbol{\mu}_2,\Sigma_{22},G,\kappa)$. Also, they are conditionally independent given V, with $\mathbf{Y}_1\mid V=v\sim\mathcal{SHMSN}_q[\boldsymbol{\mu}_1,\kappa(v)\Sigma_{11},\boldsymbol{\lambda}_{(1)},H,\eta]$ and $\mathbf{Y}_2\mid V=v\sim\mathcal{N}_{p-q}[\boldsymbol{\mu}_2,\kappa(v)\Sigma_{22}]$.*

3.3. Selection representation

As indicated in the next proposition, the SSMSN distributions are also selection distributions as those defined by [6,11].

Proposition 5. *Let $X_0=\kappa(V)^{1/2}\{\eta(V,S)\boldsymbol{\lambda}^\top\mathbf{Z}_1-Z_0\}$ and $\mathbf{X}_1=\kappa(V)^{1/2}\mathbf{Z}_1$, where $Z_0\sim\mathcal{N}(0,1)$, $\mathbf{Z}_1\sim\mathcal{N}_p(\mathbf{0},\mathbf{I}_p)$, $(V,S)\sim Q(v,s\mid\boldsymbol{\nu},\boldsymbol{\tau})$, which are mutually independent. Then, the selection random vector defined by*

$$\mathbf{X}\stackrel{d}{=}(\mathbf{X}_1\mid X_0>0)=\kappa(V)^{1/2}(\mathbf{Z}_1\mid Z_0<\eta(V,S)\boldsymbol{\lambda}^\top\mathbf{Z}_1),$$

is $\mathcal{SSMSN}_p(\mathbf{0},\mathbf{I}_p,\boldsymbol{\lambda},HG,\kappa,\eta)$ distributed.

Proof. First note that conditionally on $(V,S)=(v,s)$, the random variables X_0 and $\mathbf{X}_1$ have a multivariate normal joint distribution given by

$$\begin{pmatrix}X_0\\ \mathbf{X}_1\end{pmatrix}\Bigg|V=v,S=s\sim\mathcal{N}_{1+p}\left[\begin{pmatrix}0\\ \mathbf{0}\end{pmatrix},\kappa(v)\begin{pmatrix}1+\eta(v,s)^2\boldsymbol{\lambda}^\top\boldsymbol{\lambda} & \eta(v,s)\boldsymbol{\lambda}^\top\\ \eta(v,s)\boldsymbol{\lambda} & \mathbf{I}_p\end{pmatrix}\right].$$

Thus using Eq. (13) in [6], we find that $\mathbf{X}\mid V=v,S=s$ has a $\mathcal{SN}_p(0,\kappa(v)\mathbf{I}_p,\eta(v,s)\boldsymbol{\lambda})$ conditional pdf: $f(\mathbf{x}\mid v,s)=2\phi_p\{\mathbf{x};\mathbf{0},\kappa(v)\Sigma\}\Phi_1\{\kappa(v)^{-1/2}\eta(v,s)\boldsymbol{\lambda}^\top\mathbf{x}\}$, which concludes the proof. □

Proposition 5 provides an alternative way to define the SSMSN class as:

$$\mathbf{Y}=\boldsymbol{\mu}+\Sigma^{1/2}\mathbf{X},\quad\text{with }\mathbf{X}=\kappa(V)^{1/2}\mathbf{Z},\tag{13}$$

where $\mathbf{Z}\stackrel{d}{=}(\mathbf{Z}_1\mid Z_0<\eta(V,S)\boldsymbol{\lambda}^\top\mathbf{Z}_1)$ and is such that $\mathbf{Z}\mid V=v,S=s\sim\mathcal{SN}_p[\mathbf{0},\mathbf{I}_p,\eta(v,s)]$, since $(V,S)\sim Q(v,s\mid\boldsymbol{\nu},\boldsymbol{\tau})$, $Z_0\sim\mathcal{N}_1(0,1)$ and $\mathbf{Z}_1\sim\mathcal{N}_p(\mathbf{0},\mathbf{I}_p)$ are mutually independent.

When the mixing variables V and S are independent, we can also identify some subfamilies of SSMSN distributions such as the SHSMSN and SHSSMN classes given by (11) and (12), respectively, according to the specification of $\eta(v,s)$ as indicated below.

In fact, for the SHSMSN class in (11), V and S are independent and $\eta(v,s)=\iota(s)$. Hence, in (13), $\mathbf{Z}\mid S=s\sim\mathcal{SN}_p(\mathbf{0},\mathbf{I}_p,\iota(s)\boldsymbol{\lambda})$ and is independent of V. This means that $\mathbf{Z}$ has a SHMSN distribution and is independent of V, and therefore the distribution of $\mathbf{X}=\kappa(V)^{1/2}\mathbf{Z}$ becomes also a scale mixture of SHMSN distributions. In this case, the random vector $\mathbf{X}$ can be represented in terms of the SMN spherical random vector $(U_0,\mathbf{U}_1)=\kappa(V)^{1/2}(Z_0,\mathbf{Z}_1)$ as $\mathbf{X}\stackrel{d}{=}(\mathbf{U}_1\mid U_0<\iota(S)\boldsymbol{\lambda}^\top\mathbf{U}_1)$. Therefore, since S and V are independent, the conditional pdf of $\mathbf{X}\mid S=s$ belongs to the class of skew-spherical pdfs of the form $2f^{(p)}(\|\mathbf{x}\|^2)F^{(1)}_{\|\mathbf{x}\|^2}\{\iota(s)\boldsymbol{\lambda}^\top\mathbf{x}\}$, where $f^{(p)}$ is a SMN p-dimensional spherical density generator and $F_a^{(1)}(x)$ is the cdf induced by the univariate conditional density generator $f_a^{(1)}(u)=f^{(p+1)}(u+a)/f^{(p)}(a)$. This means that the location-scale SHSMSN distributions are shape mixtures of SMSN distributions.

Similarly, in the SHSSMN class defined in (12), in which V and S are also independent and $\eta(v,s)=\kappa(v)^{1/2}\iota(s)$, by letting $\iota(s)=\tilde{\kappa}(s)^{-1/2}$ we have that $\mathbf{X}\stackrel{d}{=}(\mathbf{U}_1\mid U_0<\boldsymbol{\lambda}^\top\mathbf{U}_1)$, but in this case $U_0=\tilde{\kappa}(S)^{1/2}Z_0$ and $\mathbf{U}_1=\kappa(V)^{1/2}\mathbf{Z}_1$ are independent SMN spherical random variables. This yields a class of multivariate skew-symmetric distributions with pdf of the form $2f^{(p)}(\|\mathbf{x}\|^2)\tilde{F}^{(1)}(\boldsymbol{\lambda}^\top\mathbf{x})$, where $f^{(p)}$ is a p-dimensional spherical density generator and $\tilde{F}^{(1)}$ is a univariate cdf induced by a univariate spherical density generator $\tilde{f}^{(1)}$. The corresponding location-scale distributions obtained under such conditions are only a subclass of (12). The univariate SSMSN class presented recently by [24] is obtained following this alternative route for $p=1$.

3.4. Stochastic representation

Next, we extend the stochastic representation of the multivariate SN distribution, initially discussed in [8,11,14], to the whole SSMSN family.

Proposition 6. *Suppose that* $\mathbf{X} = \kappa(V)^{1/2}\mathbf{Z}$*, where*

$$\mathbf{Z} = \frac{\eta(V,S)\boldsymbol{\lambda}}{\sqrt{1+\eta(V,S)^2\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}}|Z_0| + \left\{\mathbf{I}_p - \frac{\eta(V,S)^2\boldsymbol{\lambda}\boldsymbol{\lambda}^\top}{1+\eta(V,S)^2\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}\right\}^{1/2}\mathbf{Z}_1, \tag{14}$$

with $Z_0 \sim \mathcal{N}(0,1)$, $\mathbf{Z}_1 \sim \mathcal{N}_p(\mathbf{0},\mathbf{I}_p)$ *and* $(V,S) \sim Q(v,s \mid \boldsymbol{\nu},\boldsymbol{\tau})$ *being mutually independent. Then* $\mathbf{Y} = \boldsymbol{\mu} + \Sigma^{1/2}\mathbf{X} \sim \mathcal{SSMSN}_p(\boldsymbol{\mu},\Sigma,\boldsymbol{\lambda},Q,\kappa,\eta)$*.*

Proof. By (14), $\kappa(V)^{1/2}\mathbf{Z} \mid V = v, S = s \sim \mathcal{SN}_p[\mathbf{0},\kappa(v)\mathbf{I}_p,\eta(v,s)\boldsymbol{\lambda}]$, which is equal to the conditional distribution of $\mathbf{X}$ given $(V,S) = (v,s)$. This proves by definition that $\mathbf{X} \sim \mathcal{SSMSN}_p(\mathbf{0},\mathbf{I}_p,\boldsymbol{\lambda},Q,\kappa,\eta)$. Thus, the proof follows by Corollary 2. □

As a consequence of Proposition 6, we have the following corollary.

Corollary 6. *Let* $R^2 = (\mathbf{Y}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{Y}-\boldsymbol{\mu})$*, where* $\mathbf{Y} \sim \mathcal{SSMSN}_p(\boldsymbol{\mu},\Sigma,\boldsymbol{\lambda},Q,\kappa,\iota)$*. Then,* $R^2 \stackrel{d}{=} \kappa(V)R_0^2$*, where* $R_0^2 \sim \chi_p^2$ *and is independent of* $V \sim H$*.*

Proof. Let $\mathbf{X} = \Sigma^{-1/2}(\mathbf{Y}-\boldsymbol{\mu})$. Since $\mathbf{X} \stackrel{d}{=} \kappa(V)^{1/2}\mathbf{Z}$, with $\mathbf{Z}$ given by (14), then $R^2 = \mathbf{X}^\top\mathbf{X} \stackrel{d}{=} \kappa(V)\mathbf{Z}^\top\mathbf{Z}$. Since, $\mathbf{Z} \mid V = v, S = s \sim \mathcal{SN}_p[\mathbf{0},\mathbf{I}_p,\eta(v,s)\boldsymbol{\lambda}]$, by Proposition 7 in [13] it follows that $(\mathbf{Z}^\top\mathbf{Z} \mid V = v, S = s) \stackrel{d}{=} \mathbf{Z}_0^\top\mathbf{Z}_0$ for all (v,s), where $\mathbf{Z}_0 \sim \mathcal{N}_p(\mathbf{0},\mathbf{I}_p)$ which is independent of (V,S) and hence of V. Since $\mathbf{Z}_0^\top\mathbf{Z}_0$ is also independent of V, then $\kappa(V)\mathbf{Z}^\top\mathbf{Z} \stackrel{d}{=} \kappa(V)\mathbf{Z}_0^\top\mathbf{Z}_0$ which completes the proof. □

3.5. Hierarchical representation

Considering the stochastic representation in (14), we can represent hierarchically the SSMSN distributions in terms of the multivariate normal distribution. Such representation is established in the following proposition and has a key role for classical and Bayesian statistical analysis of the SSMSN models. In order to simplify the notation, we let

$$\bar{\boldsymbol{\lambda}} = \Sigma^{1/2}\boldsymbol{\lambda}, \quad \Psi(v,s) = \Sigma - \gamma(v,s)^2\bar{\boldsymbol{\lambda}}\bar{\boldsymbol{\lambda}}^\top \quad \text{and} \quad \gamma(v,s) = \frac{\eta(v,s)}{\sqrt{1+\eta(v,s)^2\bar{\boldsymbol{\lambda}}^\top\Sigma^{-1}\bar{\boldsymbol{\lambda}}}}.$$

Proposition 7. *If* $\mathbf{Y} \sim \mathcal{SSMSN}_p(\boldsymbol{\mu},\Sigma,\boldsymbol{\lambda},Q,\kappa,\eta)$*, then*

$$\mathbf{Y} \mid U = u, V = v, S = s \sim \mathcal{N}_p[\boldsymbol{\mu} + \kappa(v)^{1/2}\gamma(v,s)\bar{\boldsymbol{\lambda}}u, \kappa(v)\Psi(v,s)],$$

where $U = |Z_0| \sim \mathcal{HN}(0,1)$ *and* $(V,S) \sim Q(v,s \mid \boldsymbol{\nu},\boldsymbol{\tau})$ *are independent.*

The hierarchical representation given in Proposition 7 is very useful to implement the MCMC and/or EM methodologies for obtaining Bayesian and/or classical inference under the SSMSN models. A useful consequence of the hierarchical representation in Proposition 7 is considered next. To this end, the following lemma will be used.

Lemma 1. *If* $\Psi = \Sigma - \gamma^2\bar{\boldsymbol{\lambda}}\bar{\boldsymbol{\lambda}}^\top$ *and* $\gamma = \eta/(1+\eta^2\bar{\boldsymbol{\lambda}}^\top\Sigma^{-1}\bar{\boldsymbol{\lambda}})^{1/2}$*, then*

$$\phi_p(\mathbf{y} \mid \boldsymbol{\mu} + \kappa^{1/2}\gamma\bar{\boldsymbol{\lambda}}u, \kappa\Psi)\phi_1(u \mid 0,1) = \phi_p(\mathbf{y} \mid \boldsymbol{\mu},\kappa\Sigma)\phi_1\{u \mid \kappa^{-1/2}\gamma\bar{\boldsymbol{\lambda}}^\top\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu}), 1-\gamma^2\bar{\boldsymbol{\lambda}}^\top\Sigma^{-1}\bar{\boldsymbol{\lambda}}\}.$$

Let $q(v,s \mid \boldsymbol{\nu},\boldsymbol{\tau})$ denote the joint pdf (or probability mass function) of V and S. By Proposition 7, Lemma 1 and the change of variable $W = \kappa(V)^{1/2}\gamma(v,s)U$, the joint distribution of $(\mathbf{Y},V,S,W)$ can be represented, for all $\mathbf{y} \in \mathbb{R}^p$, $v \in \mathcal{S}(H)$, $s \in \mathcal{S}(G)$, $w > 0$, as

$$f(\mathbf{y},v,s,w \mid \boldsymbol{\mu},\Sigma,\boldsymbol{\lambda},\boldsymbol{\nu},\boldsymbol{\tau}) = 2\phi_p\{\mathbf{y} \mid \boldsymbol{\mu},\kappa(v)\Sigma\}\phi_1\{w \mid \eta(v,s)\bar{\boldsymbol{\lambda}}^\top\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu}),\kappa(v)\}q(v,s \mid \boldsymbol{\nu},\boldsymbol{\tau}). \tag{15}$$

Eq. (15) is very useful to simplify the implementation of MCMC and/or EM methodologies for obtaining Bayesian and/or classical inference under the SSMSN models. In this sense, the following conditional *posterior* distributions and moments related to the latent random quantities (V,S,W) are sometimes necessary. To establish this result, it is assumed that the scale and shape mixing variables V and S are independent, with marginal pdfs (or probability functions) given, respectively, by $h(v \mid \boldsymbol{\nu})$ and $g(s \mid \boldsymbol{\tau})$. The proof of this result is straightforward from (15), the Bayes formula and some basic probabilistic properties.

Proposition 8. *Under the joint pdf in (15), with* $q(v,s \mid \boldsymbol{\nu},\boldsymbol{\tau}) = h(v \mid \boldsymbol{\nu})g(s \mid \boldsymbol{\tau})$ *for all* (v,s)*, the conditional distribution of* (V,S,W) *given* $\mathbf{Y} = \mathbf{y}$ *is such that* $f(v,s,w \mid \mathbf{y}) = f(w \mid v,s,\mathbf{y})f(s \mid v,\mathbf{y})f(v \mid \mathbf{y})$*, where*

$$f(w \mid v,s,\mathbf{y}) = \frac{1}{\Phi_1\{\kappa(v)^{-1/2}\eta(v,s)\bar{\boldsymbol{\lambda}}^\top\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})\}}\phi_1\{w \mid \eta(v,s)\bar{\boldsymbol{\lambda}}^\top\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu}),\kappa(v)\}, \quad w > 0,$$

$$f(s \mid v, \mathbf{y}) = \frac{1}{\Psi_1\{\bar{\boldsymbol{\lambda}}^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu}) \mid v\}} \Phi_1\{\kappa(v)^{-1/2}\eta(v,s)\bar{\boldsymbol{\lambda}}^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\} g(s \mid \boldsymbol{\tau}), \quad s \in \mathcal{S}(G),$$
$$f(v \mid \mathbf{y}) = \frac{2}{f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \bar{\boldsymbol{\lambda}}, \boldsymbol{\nu}, \boldsymbol{\tau})} \phi_p\{\mathbf{y} \mid \boldsymbol{\mu}, \kappa(v)\boldsymbol{\Sigma}\}\, h(v \mid \boldsymbol{\nu}) \Psi_1\{\bar{\boldsymbol{\lambda}}^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu}) \mid v\}, \quad v \in \mathcal{S}(H),$$

where $\Psi_1(x \mid v) = \int_{\mathcal{S}(G)} \Phi_1\{\kappa(v)^{-1/2}\eta(v,s)x\} g(s \mid \boldsymbol{\tau}) ds$. *Also, if* $\mu_w = \eta(v,s)\bar{\boldsymbol{\lambda}}^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})$ *and* $\sigma_w^2 = \kappa(v)$, *then*

$$\mathrm{E}(W \mid v, s, \mathbf{y}) = \mu_w + \sigma_w \frac{\phi_1(\mu_w/\sigma_w)}{\Phi_1(\mu_w/\sigma_w)}, \quad \mathrm{E}(W^2 \mid v, s, \mathbf{y}) = \mu_w^2 + \sigma_w^2 + \mu_w\sigma_w \frac{\phi_1(\mu_w/\sigma_w)}{\Phi_1(\mu_w/\sigma_w)}.$$

4. Maximum likelihood estimation for SSMSN

4.1. EM-algorithm

In this section, we describe how to use the EM-type algorithm for maximum likelihood (ML) estimation of the SSMSN model parameters. To this end, we use the ECME, a fast extension of the original EM algorithm proposed by [28]. Given a random sample $\mathbf{y}_1, \ldots, \mathbf{y}_n$ from the $\mathcal{SSMSN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\tau})$ distribution, the corresponding log-likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\tau})$ is

$$\ell(\boldsymbol{\theta} \mid \mathbf{y}) = \sum_{i=1}^n \ln \left[\int_{\mathcal{S}(G)} \int_{\mathcal{S}(H)} 2\phi_p\{\mathbf{y}_i \mid \boldsymbol{\mu}, \kappa(v_i)\boldsymbol{\Sigma}\} \Phi_1\{\kappa(v_i)^{-1/2}\eta(v_i, s_i)\boldsymbol{\lambda}^\top \mathbf{z}_i\} q(v_i, s_i \mid \boldsymbol{\nu}, \boldsymbol{\tau}) ds_i dv_i \right], \tag{16}$$

where $\mathbf{z}_i = \boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu})$, and we assume that $q(v, s \mid \boldsymbol{\nu}, \boldsymbol{\tau}) = g(v \mid \boldsymbol{\nu}) h(s \mid \boldsymbol{\tau})$, $\forall\, (v, s)$, in which, except for the unknown parameters (scalar or vectors) $\boldsymbol{\nu}$ and $\boldsymbol{\tau}$, $g(\cdot \mid \boldsymbol{\nu})$ and $h(\cdot \mid \boldsymbol{\tau})$ are known pdfs or probability mass functions. In general, it is difficult to maximize (16) directly with respect to $\boldsymbol{\theta}$. Considering the hierarchical representation described in Section 3.5 for the SSMSN family, we develop in this section an EM algorithm to search for the ML estimator, $\widehat{\boldsymbol{\theta}}$, of $\boldsymbol{\theta}$.

We start by assuming the scale and shape functions $\kappa(\cdot)$ and $\eta(\cdot)$ to be known, and for the ith sample unit we denote the complete data arising from (15) by $\mathbf{d}_i = (\mathbf{y}_i^\top, v_i, s_i, w_i)^\top$ for all $i \in \{1, \ldots, n\}$, where $\mathbf{y}_1, \ldots, \mathbf{y}_n$ are observed data and (v_i, s_i, w_i) for all $i \in \{1, \ldots, n\}$, are missing data. Also, we let $\mathbf{d} = (\mathbf{y}^\top, \mathbf{v}^\top, \mathbf{s}^\top, \mathbf{w}^\top)^\top$, where $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{v} = (v_1, \ldots, v_n)^\top$, $\mathbf{s} = (s_1, \ldots, s_n)^\top$ and $\mathbf{w} = (w_1, \ldots, w_n)^\top$. The corresponding random vector is denoted by $\mathbf{D} = (\mathbf{Y}^\top, \mathbf{V}^\top, \mathbf{S}^\top, \mathbf{W}^\top)^\top$. By (15) the complete log-likelihood function is

$$\begin{aligned}\ell_c(\boldsymbol{\theta} \mid \mathbf{d}) = {} & \sum_{i=1}^n \ln \phi_p\{\mathbf{y}_i \mid \boldsymbol{\mu}, \kappa(v_i)\boldsymbol{\Sigma}\} + \sum_{i=1}^n \ln \phi_1\{w_i \mid \eta(v_i, s_i)\bar{\boldsymbol{\lambda}}^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}), \kappa(v_i)\} + \sum_{i=1}^n \ln g(v_i \mid \boldsymbol{\nu}) \\ & + \sum_{i=1}^n \ln h(s_i \mid \boldsymbol{\tau}) + n \ln 2.\end{aligned} \tag{17}$$

The last three terms of $\ell_c(\boldsymbol{\theta} \mid \mathbf{d})$ are not relevant to the estimation of $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, therefore they can be ignored in the implementation of the ECM step of the ECME algorithm described below.

To facilitate the estimation process, consider the reparameterization, $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda}$. The E-Step on the $(k+1)$th iteration of the ECME algorithm requires the calculation of $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) = \mathrm{E}_{\boldsymbol{\theta}^{(k)}}\{\ell_c(\boldsymbol{\theta} \mid \mathbf{D}) \mid \mathbf{Y} = \mathbf{y}\}$, where $\boldsymbol{\theta}^{(k)}$ is the estimated value of $\boldsymbol{\theta}$ in the kth algorithm step. To do this, we need to calculate different conditional expectations as described below. In fact, from (17) it follows that

$$\begin{aligned}Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) = {} & c(\boldsymbol{\theta}^{(k)}|\mathbf{y}) - \frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^n \widehat{\kappa_i^{-1}}^{(k)} (\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}) - \frac{1}{2}\boldsymbol{\Delta}^\top \sum_{i=1}^n \widehat{\kappa_i^{-1}\eta_i^2}^{(k)} (\mathbf{y}_i - \boldsymbol{\mu})(\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Delta} \\ & + \boldsymbol{\Delta}^\top \sum_{i=1}^n \widehat{\kappa_i^{-1}\eta_i w_i}(\mathbf{y}_i - \boldsymbol{\mu}) + \sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}^{(k)}}\{\ln g(V_i \mid \boldsymbol{\nu}) \mid \mathbf{Y}_i = \mathbf{y}_i\} + \sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}^{(k)}}\{\ln h(S_i \mid \boldsymbol{\tau}) \mid \mathbf{Y}_i = \mathbf{y}_i\},\end{aligned} \tag{18}$$

in which

$$c(\boldsymbol{\theta}^{(k)}|\mathbf{y}) = n\left\{\ln 2 - \frac{p+1}{2}\ln(2\pi)\right\} + \frac{p+1}{2}\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}^{(k)}}\{\ln \kappa(V_i) \mid \mathbf{Y}_i = \mathbf{y}_i\} - \frac{1}{2}\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}^{(k)}}\{\kappa(V_i)^{-1}W_i^2 \mid \mathbf{Y}_i = \mathbf{y}_i\},$$

and, for all $i \in \{1, \ldots, n\}$,

$$\begin{aligned}\widehat{\kappa_i^{-1}}^{(k)} &= \mathrm{E}_{\boldsymbol{\theta}^{(k)}}\{\kappa(V_i)^{-1} \mid \mathbf{Y}_i = \mathbf{y}_i\}, \\ \widehat{\kappa_i^{-1}\eta_i^2}^{(k)} &= \mathrm{E}_{\boldsymbol{\theta}^{(k)}}\{\kappa(V_i)^{-1}\eta(V_i, S_i)^2 \mid \mathbf{Y}_i = \mathbf{y}_i\}, \\ \widehat{\kappa_i^{-1}\eta_i w_i}^{(k)} &= \mathrm{E}_{\boldsymbol{\theta}^{(k)}}\{\kappa(V_i)^{-1}\eta(V_i, S_i)W_i \mid \mathbf{Y}_i = \mathbf{y}_i\}.\end{aligned}$$

Analytical expressions of these conditional expectations can be explored from Proposition 5. Otherwise, they can be calculated numerically using, e.g., the MCMC procedure. The properties of the conditional expectation can help us to do this, since $\mathrm{E}\{\kappa(V)^{-1}\eta(V,S)^2 \mid \mathbf{Y}\} = \mathrm{E}[\kappa(V)^{-1}\mathrm{E}\{\eta(V,S)^2 \mid V, \mathbf{Y}\} \mid \mathbf{Y}]$ and $\mathrm{E}\{\kappa(V)^{-1}\eta(V,S)W \mid \mathbf{Y}\} = \mathrm{E}[\kappa(V)^{-1}\mathrm{E}\{\eta(V,S)\mathrm{E}(W \mid V,S,\mathbf{Y}) \mid V,\mathbf{Y}\} \mid \mathbf{Y}]$, where by Proposition 5,

$$\mathrm{E}(W \mid V=v, S=s, \mathbf{Y}=\mathbf{y}) = \eta(v,s)\boldsymbol{\Delta}^\top(\mathbf{y}-\boldsymbol{\mu}) + \kappa(v)^{1/2}\frac{\phi_1\{\kappa(v)^{-1/2}\eta(v,s)\boldsymbol{\Delta}^\top(\mathbf{y}-\boldsymbol{\mu})\}}{\Phi_1\{\kappa(v)^{-1/2}\eta(v,s)\boldsymbol{\Delta}^\top(\mathbf{y}-\boldsymbol{\mu})\}},$$

$$\mathrm{E}\{\eta(V,S)^2 \mid V=v, \mathbf{Y}=\mathbf{y}\} = \frac{1}{\Psi_1\{\boldsymbol{\Delta}^\top(\mathbf{y}-\boldsymbol{\mu}) \mid v\}}\int_{\mathcal{S}(G)} \eta(v,s)^2\Phi_1\{\kappa(v)^{-1/2}\eta(v,s)\boldsymbol{\Delta}^\top(\mathbf{y}-\boldsymbol{\mu})\}dG(s \mid \boldsymbol{\tau}),$$

$$\mathrm{E}\{t(V) \mid \mathbf{Y}=\mathbf{y}\} = \frac{2}{f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}, \boldsymbol{\nu}, \boldsymbol{\tau})}\int_{\mathcal{S}(H)} t(v)\phi_p\{\mathbf{y} \mid \boldsymbol{\mu}, \kappa(v)\boldsymbol{\Sigma}\}\Psi_1\{\boldsymbol{\Delta}^\top(\mathbf{y}-\boldsymbol{\mu}) \mid v\}dH(v \mid \boldsymbol{\nu}),$$

provided that $\eta(v,s)^2$ and $t(v)$ are integrable functions. Some simplifications occur for the SHSMSN and SHSSMN cases in which $\eta(v,s) = \iota(s)$ and $\eta(v,s) = \kappa(v)^{1/2}\iota(s)$, respectively.

E-step. Compute $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)})$ using (18), where $c(\boldsymbol{\theta}^{(k)}|\mathbf{y})$ can be ignored.

CM-steps. Proceed in the following steps as:

CM-Step 1. Update $\boldsymbol{\mu}^{(k)}$ by

$$\boldsymbol{\mu}^{(k+1)} = \left(\sum_{i=1}^n \widehat{\kappa_i^{-1}}^{(k)}(\boldsymbol{\Sigma}^{(k)})^{-1} + \sum_{i=1}^n \widehat{\kappa_i^{-1}\eta_i^2}^{(k)}\boldsymbol{\Delta}^{(k)}\widehat{\boldsymbol{\Delta}}^{(k)\top}\right)^{-1}$$
$$\times \left\{(\boldsymbol{\Sigma}^{(k)})^{-1}\sum_{i=1}^n \widehat{\kappa_i^{-1}}^{(k)}\mathbf{y}_i + \boldsymbol{\Delta}^{(k)}\widehat{\boldsymbol{\Delta}}^{(k)\top}\sum_{i=1}^n \widehat{\kappa_i^{-1}\eta_i^2}^{(k)}\mathbf{y}_i - \sum_{i=1}^n \widehat{\kappa_i^{-1}\eta_i w_i}^{(k)}\boldsymbol{\Delta}^{(k)}\right\}.$$

CM-step 2. Update $\boldsymbol{\Sigma}^{(k)}$ by

$$\boldsymbol{\Sigma}^{(k+1)} = \frac{1}{n}\sum_{i=1}^n \widehat{\kappa_i^{-1}}^{(k)}(\mathbf{y}_i - \boldsymbol{\mu}^{(k+1)})(\mathbf{y}_i - \boldsymbol{\mu}^{(k+1)})^\top.$$

CM-Step 3. Update $\boldsymbol{\Delta}^{(k)}$ by

$$\boldsymbol{\Delta}^{(k+1)} = \left\{\sum_{i=1}^n \widehat{\kappa_i^{-1}\eta_i^2}^{(k)}(\mathbf{y}_i - \boldsymbol{\mu}^{(k+1)})(\mathbf{y}_i - \boldsymbol{\mu}^{(k+1)})^\top\right\}^{-1}\sum_{i=1}^n \widehat{\kappa_i^{-1}\eta_i w_i}^{(k)}(\mathbf{y}_i - \boldsymbol{\mu}^{(k+1)}).$$

CM-Step 4. Update $\boldsymbol{\lambda}^{(k)}$ by $\boldsymbol{\lambda}^{(k+1)} = (\boldsymbol{\Sigma}^{(k+1)})^{1/2}\boldsymbol{\Delta}^{(k+1)}$.

CM-Step 5. Update $(\boldsymbol{\nu}^{(k)}, \boldsymbol{\tau}^{(k)})$ by

$$\boldsymbol{\nu}^{(k+1)} = \mathrm{argmax}_{\boldsymbol{\nu}}\left[\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}^{(k)}}\{\ln g(V_i \mid \boldsymbol{\nu}) \mid \mathbf{Y}_i = \mathbf{y}_i\}\right], \quad \boldsymbol{\tau}^{(k+1)} = \mathrm{argmax}_{\boldsymbol{\tau}}\left[\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}^{(k)}}\{\ln h(S_i \mid \boldsymbol{\tau}) \mid \mathbf{Y}_i = \mathbf{y}_i\}\right].$$

In those cases where these last expectations are difficult to evaluate, the CM-step 4 can be replaced by the following CML-step as suggested by [28]:

CML-Step. Update $(\boldsymbol{\nu}, \boldsymbol{\tau})$ by

$$(\boldsymbol{\nu}^{(k+1)}, \boldsymbol{\tau}^{(k+1)}) = \mathrm{argmax}_{(\boldsymbol{\nu},\boldsymbol{\tau})}\left\{\sum_{i=1}^n \ln f(\mathbf{y}_i \mid \boldsymbol{\mu}^{(k+1)}, \boldsymbol{\Sigma}^{(k+1)}, \boldsymbol{\lambda}^{(k+1)}, \boldsymbol{\nu}, \boldsymbol{\tau})\right\}.$$

The CML-step requires uni- or bi-dimensional searches, which can be easily accomplished by using, for example, the `optim` or `optimize` routines in `R` [32]. The iterations of the ECME algorithm continue until the difference between two successive log-likelihood values, $|\ell(\boldsymbol{\theta}^{(k+1)}|\mathbf{y}) - \ell(\boldsymbol{\theta}^{(k)}|\mathbf{y})|$, is sufficiently small, say 10^{-4}. As in [21], the initial values used in the EM algorithm are the vector of sample means for $\boldsymbol{\mu}$, the sample covariance matrix for $\boldsymbol{\Sigma}$ and the vector of sample skewnesses for $\boldsymbol{\lambda}$. An alternative to attain global maximization is to consider several starting values.

Expressions for the observed information matrix can also be derived directly from the observed likelihood function in (16). To assume the mixing distributions to be completely known facilitates substantially these derivations, since in this case we only need the derivatives corresponding to the SN model. Another way to obtain estimates of standard errors is by using the approximation from the derivatives of the Q-function as proposed by Louis [29].

Table 1
Descriptive statistics of the wind speed data.

	Mean	Covariance			Mardia's skewness	Marginal skewness	Mardia's kurtosis	Marginal kurtosis
vs	17.0	185.3	–	–	3.5	−0.8	23.7	4.4
gh	12.7	126.9	177.8	–	–	−0.7	–	2.6
kw	14.0	148.2	110.6	297.2	–	−0.4	–	3.0

Table 2
Fit of various SSMSN distributions to the wind speed dataset. Best fit indicated by #1, second best by #2, and third best by #3.

Family	Particular distributions	$\ell(\widehat{\boldsymbol{\theta}}\vert\mathbf{y})$	AIC	BIC
SN [14]	skew-normal	−3229.2	6482.4	6525.9
SMSN [17]	skew-t	−3180.7	6387.5 #2	6434.6 #2
	skew-contaminated-normal	−3186.4	6400.7	6451.5
	skew-slash	−3186.9	6399.7	6446.9
SHMSN [7]	skew-generalized-normal	−3236.6	6499.2	6546.4
	skew-curved-normal	−3217.5	6459.0	6502.5
	skew-normal-Cauchy	−3227.5	6479.1	6522.6
SSMN [21]	skew-t-normal	−3180.9	6387.8 #3	6434.9 #3
	contaminated-skew-normal	−3183.0	6394.0	6444.8
	skew-slash-normal	−3184.9	6395.8	6443.0
SHSSMN (this paper)	Modified skew-t-normal	−3199.9	6425.9	6473.1
	skew-t-Cauchy	−3178.7	6383.4 #1	6430.5 #1
	Modified skew-slash-normal	−3218.1	6462.2	6509.4
	skew-slash-Cauchy	−3227.7	6481.4	6528.5

4.2. Observed information matrix

To compute the asymptotic covariance of the ML estimates of $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\boldsymbol{\alpha}}^\top, \widehat{\boldsymbol{\lambda}}^\top)^\top$, whit $\boldsymbol{\alpha} = \mathrm{vec}(\mathbf{B})$, where $\Sigma^{1/2} = \mathbf{B} = \mathbf{B}(\boldsymbol{\alpha})$, we employ the information-based method suggested by [29] and [30]. We use the notation $\dot{\mathbf{B}}_k = \partial\mathbf{B}(\boldsymbol{\alpha})/\partial\alpha_k$, with $k \in \{1, \ldots, p(p+1)/2\}$. The empirical information matrix is defined as

$$\mathbf{I}_e(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^n \widehat{\mathbf{s}}_i\widehat{\mathbf{s}}_i^\top,$$

where $\widehat{\mathbf{s}}_i = (\widehat{\mathbf{s}}_{i,\boldsymbol{\beta}}^\top, \widehat{\mathbf{s}}_{i,\boldsymbol{\alpha}}^\top, \widehat{\mathbf{s}}_{i,\boldsymbol{\lambda}}^\top)^\top$ are the estimates of individual scores $\mathbf{s}(\mathbf{y}_i|\boldsymbol{\theta}) = \partial Q(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. The explicit expressions for the elements of $\widehat{\mathbf{s}}_i$ are summarized below:

$$\begin{aligned}
\widehat{\mathbf{s}}_{i,\boldsymbol{\beta}} &= \widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}_i - \widehat{\boldsymbol{\mu}}) - \widehat{\kappa_i\eta_i w_i}\widehat{\boldsymbol{\Delta}} + \widehat{\kappa_i\eta_i^2}\widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\Delta}}^\top(\mathbf{y}_i - \widehat{\boldsymbol{\mu}});\\
\widehat{\mathbf{s}}_{i,\alpha_k} &= -\mathrm{tr}(\widehat{\mathbf{B}}^{-1}\widehat{\dot{\mathbf{B}}}_k) + \frac{1}{2}\widehat{\kappa_i^{-1}}(\mathbf{y}_i - \widehat{\boldsymbol{\mu}})^\top\widehat{\mathbf{B}}^{-1}(\widehat{\mathbf{B}}^{-1}\widehat{\dot{\mathbf{B}}}_k + \widehat{\dot{\mathbf{B}}}_k\widehat{\mathbf{B}}^{-1})\widehat{\mathbf{B}}^{-1}(\mathbf{y}_i - \widehat{\boldsymbol{\mu}})\\
&\quad - \widehat{\boldsymbol{\lambda}}^\top\widehat{\mathbf{B}}^{-1}\widehat{\dot{\mathbf{B}}}_k\widehat{\mathbf{B}}^{-1}(\mathbf{y}_i - \widehat{\boldsymbol{\mu}})\left[\widehat{\kappa_i\eta_i w_i} + \widehat{\kappa_i^{-1}\eta_i^2}\widehat{\boldsymbol{\lambda}}^\top\widehat{\mathbf{B}}^{-1}(\mathbf{y}_i - \widehat{\boldsymbol{\mu}})\right], \quad \text{for all } k \in \{1, \ldots, p(p+1)/2\};\\
\widehat{\mathbf{s}}_{i,\boldsymbol{\lambda}} &= \widehat{\boldsymbol{\Sigma}}^{-1/2}(\mathbf{y}_i - \widehat{\boldsymbol{\mu}})\left[\widehat{\kappa_i\eta_i w_i} - \widehat{\kappa_i^{-1}\eta_i^2}(\mathbf{y}_i - \widehat{\boldsymbol{\mu}})^\top\widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\boldsymbol{\lambda}}\right].
\end{aligned}$$

5. Wind data application

To illustrate the flexibility of the SSMSN distributions, we use a wind speed dataset that consists of hourly average wind speed in the Pacific North-West of the United States collected at three meteorological towers approximately located on a line and ordered from west to east: Goodnoe Hills (gh), Kennewick (kw), and Vansyce (vs). The data was collected from 25 February to 30 November 2003 recorded at midnight, a time when wind speeds tend to peak. More information about the data can be found in [15].

Denote by $\mathbf{Y}(t)$ the three-dimensional vector of wind speed at the towers (gh, kw and vs) recorded at time $t \in \{1, \ldots, 278\}$. Azzalini and Genton [15] applied a Ljung–Box test to the data that indicated some serial correlation at the Goodnoe Hills tower, but not at the other two towers. So, the authors proposed a skew-t model that modeled heavy tails and asymmetric behaviors. Descriptive statistics of the wind speed data are reported in Table 1, confirming the presence of skewness and kurtosis.

Next, we fit various SSMSN models described in this paper to the wind data. The results of the fit in terms of log-likelihood, AIC and BIC are provided in Table 2. We can see that the models with higher log-likelihood are the skew-t, skew-t-normal and skew-t-Cauchy. We also see that both AIC and BIC criteria favor the skew-t-Cauchy model, and then the skew-t closely followed by the skew-t-normal.

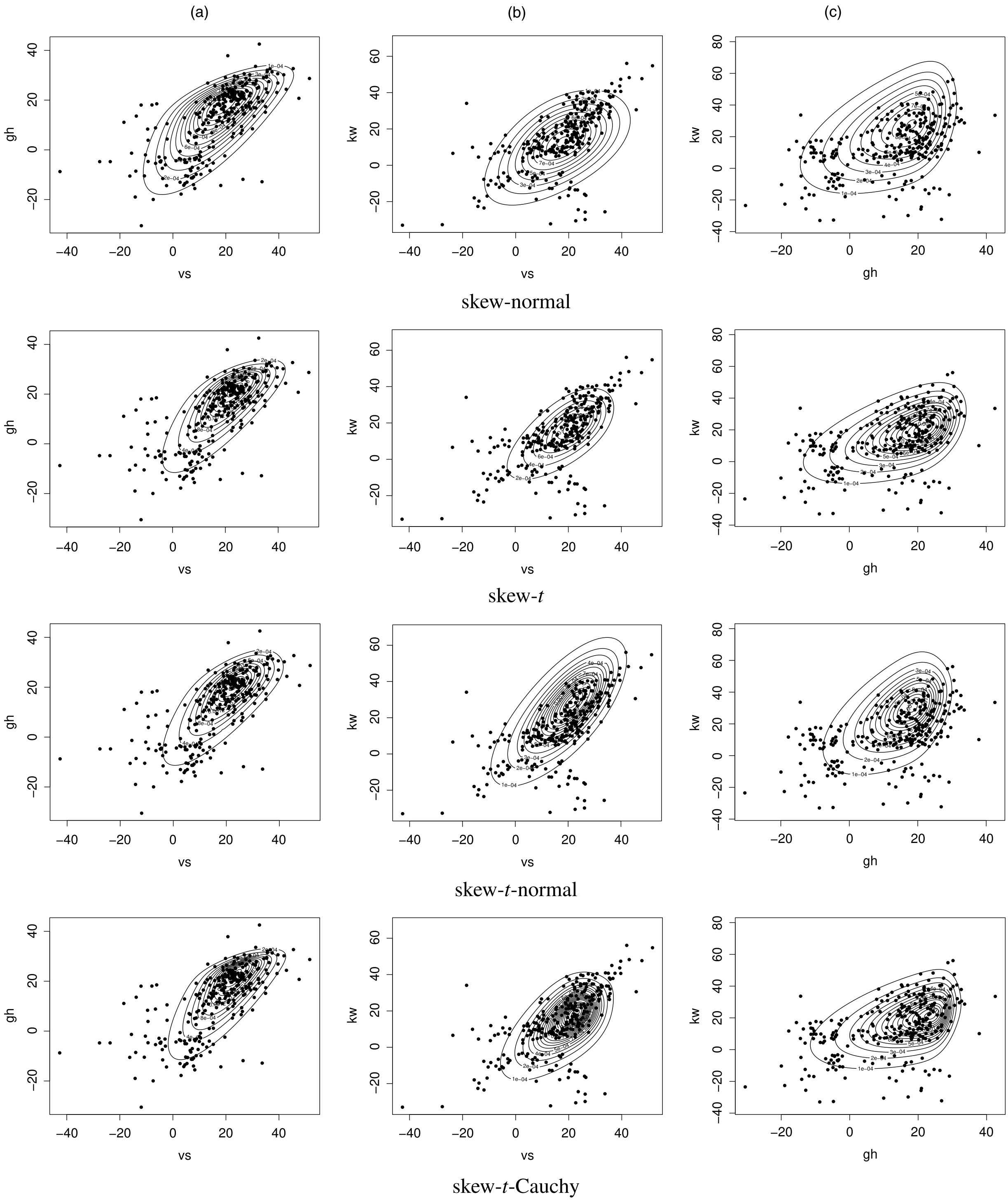

Fig. 1. Contours levels for SSMSN distributions fitted to wind speed data: (a) $vs \times gh$; (b) $vs \times kw$ and (c) $gh \times kw$.

According to fitted contour plots in Fig. 1, the skew-normal does not seem to fit all the data completely well, whereas the three models with smaller AIC and BIC, the skew-t, skew-t-normal and skew-t-Cauchy, seem to better fit the wind speed data.

6. Discussion

We proposed a broad and flexible class of multivariate SSMSN distributions which are obtained by both scale and shape mixtures of multivariate skew-normal distributions and allows to unify several subfamilies considered in the literature that can be seen as submodels of our proposal. We presented the main probabilistic properties of this family of distributions in detail and we gave the theoretical foundations for subsequent inference with this class of models. In particular, we showed that by specifying the shape mixing function as $\eta(v, s) = \iota(s)$ and $\eta(v, s) = \kappa(v)^{1/2}\iota(s)$ we obtain two different subclasses of SSMSN distributions that correspond to the shape (scale) mixtures of SMSN (SHMSN) distributions and the shape mixture of SSMN distributions. Further extensions of the SSMSN class considered here can be studied from the results in [9].

Acknowledgments

This research was supported by Fondecyt (Chile) 1120121 and 1150325, by FAPEMIG (Minas Gerais State Foundation for Research Development) grant CEX APQ 01944/17, and by the King Abdullah University of Science and Technology (KAUST). We thank the Editor, Associate Editor and four anonymous reviewers for comments that improved the paper. We also thank Prof. Adelchi Azzalini for suggesting Proposition 1 during a seminar presentation of this work at the University of Padova and Prof. Mauricio Castro for some initial discussions on the topic of this paper.

References

[1] D.F. Andrews, C.L. Mallows, Scale mixtures of normal distributions, J. R. Stat. Soc. Ser. B Stat. Methodol. 36 (1974) 99–102.
[2] R.B. Arellano-Valle, On some characterizations of spherical distributions, Statist. Probab. Lett. 54 (2001) 227–232.
[3] R.B. Arellano-Valle, A. Azzalini, On the unification of families of skew-normal distributions, Scand. J. Stat. 33 (2006) 561–574.
[4] R.B. Arellano-Valle, H. Bolfarine, On some characterizations of the *t* distribution, Stat. Probab. Lett. 25 (1995) 79–85.
[5] R.B. Arellano-Valle, H. Bolfarine, P.L. Iglesias, A predictivistic interpretation to the multivariate t distribution, TEST 3 (1994) 221–236.
[6] 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.
[7] 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, Bayesian Anal. 3 (2008) 513–540.
[8] R.B. Arellano-Valle, M.G. Genton, On fundamental skew distributions, J. Multivariate Anal. 96 (2005) 93–116.
[9] R.B. Arellano-Valle, M.G. Genton, R.H. Loschi, Shape mixtures of multivariate skew-normal distributions, J. Multivariate Anal. 100 (2009) 91–101.
[10] R.B. Arellano-Valle, H.W. Gomez, F.A. Quintana, A new class of skew-normal distributions, Comm. Statist. Theory Methods 33 (2004) 1465–1480.
[11] R.B. Arellano-Valle, R.B. del Pino, E.S. Martín, Definitions and probabilistic properties of skew distributions, Statist. Probab. Lett. 58 (2002) 111–121.
[12] O. Arslan, Variance-mean mixture of the multivariate skew-normal distribution, Statist. Papers 56 (2015) 353–378.
[13] A. Azzalini, A. Capitanio, Statistical applications of the multivariate skew-normal distribution, J. R. Stat. Soc. Ser. B Stat. Methodol. 61 (1999) 579–602.
[14] A. Azzalini, A. Dalla Valle, The multivariate skew-normal distribution, Biometrika 83 (1996) 715–726.
[15] A. Azzalini, M.G. Genton, Robust likelihood methods based on the skew-*t* and related distributions, Internat. Statist. Rev. 76 (2008) 106–129.
[16] A. Azzalini, A. Capitanio, The Skew-Normal and Related Families, Cambridge University Press, London, 2014.
[17] M.D. Branco, D.K. Dey, A general class of multivariate skew-elliptical distributions, J. Multivariate Anal. 79 (2001) 99–113.
[18] S. Cambanis, S. Huang, G. Simons, On the theory of elliptical contoured distributions, J. Multivariate Anal. 11 (1981) 368–385.
[19] K.T. Fang, S. Kotz, K.W. Ng, Symmetric Multivariate and Related Distributions, Chapman & Hall, London, 1990.
[20] C.S. Ferreira, H. Bolfarine, V.H. Lachos, Skew scale mixtures of normal distributions: Properties and estimation, Stat. Methodol. 8 (2011) 154–171.
[21] C.S. Ferreira, V.H. Lachos, H. Bolfarine, Likelihood-based inference for multivariate skew scale mixtures of normal distributions, AStA Adv. Stat. Anal. 100 (2016) 421–441.
[22] M.G. Genton, Skew-elliptical distributions and their applications: A journey beyond normality, Chapman & Hall / CRC, Boca Raton, FL, 2004.
[23] M.G. Genton, N.M.R. Loperfido, Generalized skew-elliptical distributions and their quadratic forms, Ann. Inst. Statist. Math. 57 (2005) 389–401.
[24] A. Jamalizadeh, T.-I. Lin, A general class of scale-shape mixtures of skew-normal distributions: Properties and estimation, Comput. Statist. 32 (2017) 451–474.
[25] F. Kahrari, R.B. Arellano-Valle, M. Rezaei, F. Yousefzadeh, Scale mixtures of skew-normal-Cauchy distributions, Statist. Probab. Lett. 126 (2017) 1–6.
[26] F. Kahrari, M. Rezaei, F. Yousefzadeh, R.B. Arellano-Valle, On the multivariate skew-normal-Cauchy distribution, Statist. Probab. Lett. 117 (2016) 80–88.
[27] K.L. Lange, J.S. Sinsheimer, Normal/independent distributions and their applications in robust regression, J. Comput. Graph. Statist. 2 (1993) 175–198.
[28] C. Liu, D.B. Rubin, The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence, Biometrika 80 (1994) 267–278.
[29] T.A. Louis, Finding the observed information matrix when using the EM algorithm, J. R. Stat. Soc. Ser. B Stat. Methodol. 44 (1982) 226–233.
[30] I. Meilijson, A fast improvement to the EM algorithm to its own terms, J. R. Stat. Soc. Ser. B Stat. Methodol. 51 (1989) 127–138.
[31] N. Nematollahi, R. Farnoosh, Z. Rahnamaei, Location-scale mixture of skew-elliptical distributions: Looking at the robust modelling, Stat. Methodol. 32 (2016) 131–146.
[32] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2016. https://www.R-project.org/.
[33] F. Vilca, N. Balakrishnan, C.B. Zeller, Multivariate skew-normal generalized hyperbolic distribution and its properties, J. Multivariate Anal. 128 (2014) 73–85.
[34] J. Wang, J. Boyer, M.G. Genton, A skew-symmetric representation of multivariate distributions, Statist. Sinica 14 (2004) 1259–1270.