This version of the article has been accepted for publication. The version of record of this article, published in Advances in Data Analysis and Classification, is available at Publisher's website: https://doi.org/10.1007/s11634-020-00388-6

Data projections by skewness maximization under scale mixtures of skew-normal vectors

Jorge M Arevalillo · Hilario Navarro

Abstract Multivariate scale mixtures of skew-normal distributions are flexible models that account for the non-normality of data by means of a tail weight parameter and a shape vector representing the asymmetry of the model in a directional fashion. Its stochastic representation involves a skew-normal vector and a non negative mixing scalar variable, independent of the skew-normal vector, that injects tail weight behavior into the model. In this paper we look into the problem of finding the projection that maximizes skewness for vectors that follow a scale mixture of skew-normal distribution; when a simple condition on the moments of the mixing variable is fulfilled, it can be shown that the direction yielding the maximal skewness is proportional to the shape vector. This finding stresses the directional nature of the shape vector to regulate the asymmetry; it also provides the theoretical foundations motivating the skewness based projection pursuit problem in this class of distributions. Some examples that illustrate the application of our results are also given; they include a simulation experiment with artificial data, which sheds light on the usefulness and implications of our results, and the application to real data.

Keywords Skew-normal · Scale mixtures of Skew-normal distributions · Maximal skewness projection

Jorge M Arevalillo E-mail: jmartin@ccia.uned.es · Hilario Navarro E-mail: hnavarro@ccia.uned.es

Department of Statistics, Operational Research and Numerical Analysis, University Nacional Educación a Distancia (UNED), Paseo Senda del Rey 9, 28040, Madrid, Spain
Tel.: +34-91-3987264, Fax: +34-91-3987260

1 Introduction

1.1 Skew-normal and scale mixtures of skew-normal distributions

The multivariate skew-normal (SN) distribution is a widely used distribution due to its flexibility to regulate asymmetry departures from normality. The study of its theoretical properties and applications has originated a great deal of research (Azzalini and Capitanio 1999; Capitanio et al 2003; Azzalini 2005; Contreras-Reyes and Arellano-Valle 2012; Balakrishnan and Scarpa 2012; Balakrishnan et al 2014; Azzalini and Capitanio 2014). In this paper we adopt the notation of the seminal works by Azzalini and Dalla Valle (1996) and Azzalini and Capitanio (1999) to define the density function of a p-dimensional SN vector with location vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_p)'$ and scale matrix $\boldsymbol{\Omega}$ as follows:

$$f(\boldsymbol{x}; \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\Omega}) = 2\phi_p(\boldsymbol{x} - \boldsymbol{\xi}; \boldsymbol{\Omega})\Phi(\boldsymbol{\alpha}'\boldsymbol{\omega}^{-1}(\boldsymbol{x} - \boldsymbol{\xi})) \quad : \quad \boldsymbol{x} \in \mathbb{R}^p, \tag{1}$$

where $\phi_p(\cdot; \boldsymbol{\Omega})$ denotes the p-dimensional normal density function with zero mean and covariance matrix $\boldsymbol{\Omega}$, Φ is the distribution function of a standard $N(0,1)$ variable, $\boldsymbol{\omega} = diag(\omega_1, \ldots, \omega_p)$ is a scale diagonal matrix with non negative entries such that $\overline{\boldsymbol{\Omega}} = \boldsymbol{\omega}^{-1}\boldsymbol{\Omega}\boldsymbol{\omega}^{-1}$ is a correlation matrix and $\boldsymbol{\alpha}$ is a p-dimensional shape vector that accounts for the asymmetry of the multivariate model. Note that the scale matrix $\boldsymbol{\omega}$ can be written as $\boldsymbol{\omega} = (\boldsymbol{\Omega} \odot \boldsymbol{I}_p)^{1/2}$, where the symbol $\odot$ denotes the entry-wise matrix product.

We put $\boldsymbol{X} \sim SN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ to denote that $\boldsymbol{X}$ follows a p-dimensional SN distribution whose density function is given by (1); if $\boldsymbol{\alpha} = \boldsymbol{0}$ then we have $\boldsymbol{X} \sim N_p(\boldsymbol{\xi}, \boldsymbol{\Omega})$. We can also observe that $\boldsymbol{X} = \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{Z}$, where $\boldsymbol{Z}$ is a normalized multivariate SN variable with density function given by

$$f(\boldsymbol{z}; \boldsymbol{0}, \boldsymbol{\alpha}, \boldsymbol{\Omega}) = 2\phi_p(\boldsymbol{z}; \overline{\boldsymbol{\Omega}})\Phi(\boldsymbol{\alpha}'\boldsymbol{z}). \tag{2}$$

The multivariate scale mixture of skew-normal (SMSN) distribution is an extension of the SN model that incorporates an additional parameter to regulate tail weight behavior. The SMSN family was introduced as a subclass of the more general class of skew-elliptical distributions (Branco and Dey 2001); it has become an increasingly popular multivariate model because it defines a wide class of distributions for handling skewness and tail weight simultaneously.

The SMSN family contains popular models used in the statistical practice like the skew-t or the double exponential multivariate distributions. The family is essentially characterized by the product of a SN vector and an independent non negative scalar variable; while the former controls the non-normality of the multivariate distribution in terms of asymmetry, the later injects tail weight into the multivariate model. Thus, the family has become a flexible model to handle departures from normality in multivariate data analysis. Theoretically insights and properties of the SMSN class have been described in the literature to show its richness and flexibility (Kim 2008; Lachos et al 2010b; Capitanio 2012; Kim and Kim 2017).

In this paper we use the notation adopted by Capitanio (2012) to define the family of multivariate SMSN distributions as follows.

Definition 1 Let $\boldsymbol{Z}$ be a random vector such that $\boldsymbol{Z} \sim SN_p(\boldsymbol{0}, \overline{\boldsymbol{\Omega}}, \boldsymbol{\alpha})$, with density function (2), and let U be a non negative scalar variable, independent of $\boldsymbol{Z}$. The random vector $\boldsymbol{X} = \boldsymbol{\xi} + \boldsymbol{\omega} U^{-1/2} \boldsymbol{Z} = \boldsymbol{\xi} + \boldsymbol{\omega} S \boldsymbol{Z}$, with $S = U^{-1/2}$ and $\boldsymbol{\omega}$ a scale diagonal matrix, is said to follow a multivariate SMSN distribution.

The scalar variable S is known as the mixing variable of the SMSN model, which accounts for the tail weight behavior of the multivariate distribution. The density function of $\boldsymbol{X}$ can be obtained by integrating out its conditional distribution given that $U = u$. Taking into account (1), we obtain as a result the density function:

$$\psi_p(\boldsymbol{x}; \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\Omega}, H) = \int_0^\infty 2\phi_p(\boldsymbol{x} - \boldsymbol{\xi}; u^{-1}\boldsymbol{\Omega}) \Phi(u^{1/2} \boldsymbol{\alpha}' \boldsymbol{\omega}^{-1} (\boldsymbol{x} - \boldsymbol{\xi})) dH(u) \quad (3)$$

where H is the distribution function of the scalar variable U. Here, H may be an absolutely continuous distribution or a discrete distribution.

We will write $\boldsymbol{X} \sim SMSN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, H)$ to indicate that $\boldsymbol{X}$ follows a multivariate SMSN distribution. If $\boldsymbol{\alpha} = \boldsymbol{0}$ then $\boldsymbol{X}$ becomes a scale mixture of normal distributions, a subclass of the multivariate elliptical family. Note that, when H is degenerate at $U = 1$ we have $\boldsymbol{X} \sim SN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$.

1.2 Skewness based projection pursuit

Projection pursuit is an exploratory data analysis tool for finding interesting data projections that may reveal the existence of outliers or some kind of structure in multivariate data; this task is accomplished by the maximization of a projection index quantifying the interestingness of a data projection (Friedman and Tukey 1974; Huber 1985; Friedman 1987; Jones and Sibson 1987; Cook et al 1993; Caussinus and Ruiz-Gazen 2010). The well-known principal component analysis (PCA) can be considered as a projection pursuit method which focuses on second order moments and uses a scale measure as a projection index. Actually, the PCA approach captures the whole structure of dependence when the underlying model follows a multivariate normal distribution; however, in non-normal scenarios with skewed multivariate data, PCA entails obvious limitations since multivariate associations will involve higher order moments. Hence, recent works resort to the use of skewness based projections as a byproduct for addressing the projection pursuit problem when the multivariate data at hand exhibit asymmetries (Loperfido 2018; 2019).

Although projection pursuit was born as an exploratory data analysis technique, the need for advances towards inferential results motivated the study of multivariate models whose parameters are related to the maximization of a non-normality projection index (Loperfido 2010; Arevalillo and Navarro 2015;

Loperfido 2018). This paper contributes to the problem by addressing skewness based projection pursuit for a wide and flexible family of multivariate distributions, the class of SMSN distributions: Specifically, when the input vector is such that $\boldsymbol{X} \sim SMSN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, H)$, the direction yielding the maximal skewness projection for $\boldsymbol{X}$ is found; its connection with the shape vector $\boldsymbol{\eta}' = \boldsymbol{\alpha}'\boldsymbol{\omega}^{-1}$ that regulates the asymmetry of the multivariate distribution is also discussed. The same problem was addressed for SN vectors by Loperfido (2010), who also suggested its extension to a more general framework. In this paper we revisit the problem and establish its generalization for SMSN vectors. A second order condition on the moments of the mixing variable suffices to derive an analytical solution to the skewness maximization problem for SNSM vectors; the results stress the interpretation of the shape vector as a parameter for assessing the asymmetry of the multivariate model in a directional fashion. Several examples illustrating the theoretical results are also given in order to elaborate on the computational and applied facets of the theoretical findings through a simulation experiment and applications to real data.

The paper is organized as follows: In Section 2 we address the problem of finding the maximal skewness projection for vectors that follow a SMSN distribution; this is the main theoretical contribution of the paper. The applications shed light on the theoretical findings and are presented in Section 3. The last section gives some summarizing and concluding remarks.

2 Skewness maximization

Let $\boldsymbol{X}$ be a vector such that $\boldsymbol{X} \sim SMSN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, H)$ and let us denote by $\boldsymbol{U} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{X}-\boldsymbol{\xi})$ its scaled version, with $\boldsymbol{\Sigma}$ the covariance matrix of $\boldsymbol{X}$. From Definition 1 we obtain the stochastic representation: $\boldsymbol{X} = \boldsymbol{\xi}+\boldsymbol{\omega}S\boldsymbol{Z} = \boldsymbol{\xi}+S\boldsymbol{Z}^*$, with $\boldsymbol{Z}^* \sim SN_p(0, \boldsymbol{\Omega}, \boldsymbol{\eta})$ where $\boldsymbol{\eta} = \boldsymbol{\omega}^{-1}\boldsymbol{\alpha}$ and $\boldsymbol{\Omega} = \boldsymbol{\omega}\overline{\boldsymbol{\Omega}}\boldsymbol{\omega}$.

We address the problem of finding the direction $\boldsymbol{c}$ for which the scalar variable $Y = \boldsymbol{c'U}$ attains the maximum skewness. Our goal is to solve the optimization problem: $\max_{\boldsymbol{c}\in\mathbb{R}_0^p} \gamma_1(\boldsymbol{c'U})$, with γ_1 the skewness index defined by $\gamma_1(Y) = E^2\left(\dfrac{Y-\mu_Y}{\sigma_Y}\right)^3$ and $\mathbb{R}_0^p$ the set of all non-null p-dimensional vectors.

Since γ_1 is scale invariant, we can confine to vectors such that $\boldsymbol{c}'\boldsymbol{c} = 1$; hence, the problem of finding the directional skewness can be described as

$$\gamma_{1,p}^D(\boldsymbol{X}) = \max_{\boldsymbol{c}\in\mathbb{S}_p} \gamma_1(\boldsymbol{c'U}) \tag{4}$$

where $\mathbb{S}_p = \{\boldsymbol{c} \in \mathbb{R}^p : \boldsymbol{c}'\boldsymbol{c} = 1\}$. Alternatively, it can be stated by the following equivalent formulation:

$$\gamma_{1,p}^D(\boldsymbol{X}) = \max_{\boldsymbol{d}\in\mathbb{S}_p^*} \gamma_1(\boldsymbol{d'X}) \tag{5}$$

where $\boldsymbol{d} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{c}$ and $\mathbb{S}_p^* = \{\boldsymbol{d} \in \mathbb{R}^p : \boldsymbol{d}'\boldsymbol{\Sigma}\boldsymbol{d} = 1\}$.

The solutions of any of the previous equivalent problems are given by

$$\boldsymbol{\lambda_X} = \arg\max_{\boldsymbol{d}\in\mathbb{S}_p^*} \gamma_1(\boldsymbol{d'X}) \ , \ \boldsymbol{\lambda_U} = \arg\max_{\boldsymbol{c}\in\mathbb{S}_p} \gamma_1(\boldsymbol{c'U}) \tag{6}$$

Both solutions satisfy that $\boldsymbol{\lambda_X} \propto \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda_U}$.

2.1 Main contribution

First of all we need to prove the next auxiliary lemma.

Lemma 1 *Let $\boldsymbol{X}$ be a vector such that $\boldsymbol{X} \sim SMSN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, H)$, whose mixing variable has finite second order moment. If $\boldsymbol{\Sigma}$ is the covariance matrix of $\boldsymbol{X}$ then $\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}$, where $\boldsymbol{\gamma} = \dfrac{\boldsymbol{\Omega\eta}}{\sqrt{1+\boldsymbol{\eta'\Omega\eta}}}$, is proportional to $\boldsymbol{\eta} = \boldsymbol{\omega}^{-1}\boldsymbol{\alpha}$.*

Proof. The covariance matrix for SMSN multivariate distributions is given by

$$\boldsymbol{\Sigma} = E(S^2)\boldsymbol{\Omega} - \frac{2}{\pi}E^2(S)\boldsymbol{\gamma\gamma'} = E(S^2)\left(\boldsymbol{\Omega} - \frac{2}{\pi}\frac{E^2(S)}{E(S^2)}\boldsymbol{\gamma\gamma'}\right)$$

(Capitanio 2012; Azzalini and Capitanio 2014). In order to calculate $\boldsymbol{\Sigma}^{-1}$, we use the well-known Sherman-Morrison formula which is given by

$$(\boldsymbol{A} + \boldsymbol{uv'})^{-1} = \boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{uv'}\boldsymbol{A}^{-1}}{1+\boldsymbol{v'A}^{-1}\boldsymbol{u}}. \tag{7}$$

Taking $\boldsymbol{A} = \boldsymbol{\Omega}$, $\boldsymbol{u} = -\dfrac{2}{\pi}\dfrac{E^2(S)}{E(S^2)}\boldsymbol{\gamma}$ and $\boldsymbol{v} = \boldsymbol{\gamma}$ we obtain that

$$E(S^2)\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Omega}^{-1} - \frac{\boldsymbol{\Omega}^{-1}\boldsymbol{\gamma}\left(-\frac{2}{\pi}\frac{E^2(S)}{E(S^2)}\right)\boldsymbol{\gamma'\Omega}^{-1}}{1+\left(-\frac{2}{\pi}\frac{E^2(S)}{E(S^2)}\right)\boldsymbol{\gamma'\Omega}^{-1}\boldsymbol{\gamma}} = \boldsymbol{\Omega}^{-1} + \frac{\boldsymbol{\Omega}^{-1}\boldsymbol{\gamma\gamma'\Omega}^{-1}}{\frac{\pi}{2}\frac{E(S^2)}{E^2(S)} - \boldsymbol{\gamma'\Omega}^{-1}\boldsymbol{\gamma}}$$

from which we get $\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma} = \dfrac{\boldsymbol{\Omega}^{-1}\boldsymbol{\gamma}}{E(S^2) - \frac{2}{\pi}E^2(S)\boldsymbol{\gamma'\Omega}^{-1}\boldsymbol{\gamma}}$ after some matrix calculations. Consequently, $\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}$ is proportional to $\boldsymbol{\Omega}^{-1}\boldsymbol{\gamma} = \dfrac{\boldsymbol{\eta}}{\sqrt{1+\boldsymbol{\eta'\Omega\eta}}}$, which implies the assertion of the statement. □

The quantity $\gamma_{1,p}^D$ in (4) is a multivariate skewness index that captures the directional nature of the asymmetry (Malkovich and Afifi 1973). Although it depends on the form of the stochastic representation of the SMSN vector, specifically on the distribution of the mixing variable S, we now show that the vector yielding the maximum skewness projection lies on the direction of $\boldsymbol{\eta}$. This is true when it holds a simple condition on the moments of S, as stated by the next theorem.

Theorem 1 *Let $\boldsymbol{X}$ be a random vector such that $\boldsymbol{X} \sim SMSN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, H)$ with mixing variable having finite first and second order moments satisfying that $\frac{4}{\pi}E^2(S) \geq E(S^2)$. Then the maximum skewness in (5) is attained at the direction of the vector $\boldsymbol{\eta}' = \boldsymbol{\alpha}'\boldsymbol{\omega}^{-1}$.*

Proof. Taking into account the following equivalent restrictions: $\boldsymbol{c} \in \mathbb{S}_p$ or $\boldsymbol{d} \in \mathbb{S}_p^*$ from (4) or (5), we get

$$\gamma_1(Y) = \gamma_1(\boldsymbol{c}'\boldsymbol{U}) = E^2[\boldsymbol{c}'(\boldsymbol{U} - E(\boldsymbol{U}))]^3 = E^2\left[\boldsymbol{d}'\left(\boldsymbol{X} - \boldsymbol{\xi} - E(S)\sqrt{\frac{2}{\pi}}\boldsymbol{\gamma}\right)\right]^3,$$

with $\boldsymbol{d} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{c}$ and $\boldsymbol{\gamma} = \dfrac{\boldsymbol{\omega}\overline{\boldsymbol{\Omega}}\boldsymbol{\alpha}}{\sqrt{1+\boldsymbol{\alpha}'\overline{\boldsymbol{\Omega}}\boldsymbol{\alpha}}} = \dfrac{\boldsymbol{\Omega}\boldsymbol{\eta}}{\sqrt{1+\boldsymbol{\eta}'\boldsymbol{\Omega}\boldsymbol{\eta}}}$.

The previous expression for γ_1 admits the following reformulations:

$$\gamma_1(Y) = \gamma_1(\boldsymbol{c}'\boldsymbol{U}) = \gamma_1(\boldsymbol{d}'\boldsymbol{X}) = \gamma_1(\boldsymbol{d}'(\boldsymbol{X} - \boldsymbol{\xi})) \tag{8}$$

where $\boldsymbol{d}'(\boldsymbol{X} - \boldsymbol{\xi}) = SZ_0$ and $Z_0 = \boldsymbol{d}'\boldsymbol{Z}^*$ is a scalar random variable such that $Z_0 \sim SN_1(0, \omega_{\boldsymbol{d}}, \alpha_{\boldsymbol{d}})$, with $\alpha_{\boldsymbol{d}} = \dfrac{\boldsymbol{d}'\boldsymbol{\gamma}}{\sqrt{\omega_{\boldsymbol{d}} - (\boldsymbol{d}'\boldsymbol{\gamma})^2}}$ and $\omega_{\boldsymbol{d}} = \boldsymbol{d}'\boldsymbol{\Omega}\boldsymbol{d}$ (Azzalini and Capitanio 2014; formula (5.44)). Alternatively, Z_0 can be represented by $Z_0 = \omega_{\boldsymbol{d}}^{1/2}U_0$ with $U_0 \sim SN_1(0, 1, \alpha_{\boldsymbol{d}})$.

In order to find an analytical expression for the quantity in (8), we need the moments of U_0 up to the third one. They are given by

$$E(U_0) = \sqrt{\frac{2}{\pi}}\delta_0 \ , \ E(U_0^2) = 1 \text{ and } E(U_0^3) = \sqrt{\frac{2}{\pi}}(3\delta_0 - \delta_0^3),$$

where $\delta_0^2 = \dfrac{\alpha_{\boldsymbol{d}}^2}{1+\alpha_{\boldsymbol{d}}^2} = \omega_{\boldsymbol{d}}^{-1}t$, with $t = (\boldsymbol{d}'\boldsymbol{\gamma})^2$, is a quantity such that $0 \leq \delta_0^2 = \omega_{\boldsymbol{d}}^{-1}t \leq 1$. Inserting these moments into (8), we get

$$\gamma_1(\boldsymbol{d}'(\boldsymbol{X} - \boldsymbol{\xi})) = E^2[(SZ_0 - E(SZ_0))^3] = \omega_{\boldsymbol{d}}^3 E^2[(SU_0 - E(SU_0))^3]$$

$$= \omega_{\boldsymbol{d}}^3[E(S^3)E(U_0^3) - 3E(S^2)E(S)E(U_0^2)E(U_0) + 2E^3(S)E^3(U_0)]^2$$

$$= \omega_{\boldsymbol{d}}^3\left[E(S^3)\sqrt{\frac{2}{\pi}}(3\delta_0 - \delta_0^3) - 3E(S^2)E(S)\sqrt{\frac{2}{\pi}}\delta_0 + 2E^3(S)\sqrt{\frac{2}{\pi}}\frac{2}{\pi}\delta_0^3\right]^2$$

$$= \frac{2}{\pi}\omega_{\boldsymbol{d}}^3\delta_0^2[a\delta_0^2 - 3b]^2 = \frac{2}{\pi}t[at - 3b\omega_{\boldsymbol{d}}]^2,$$

where $a = \frac{4}{\pi}E^3(S) - E(S^3)$ and $b = E(S)E(S^2) - E(S^3)$.

The previous calculations show that $\gamma_1(Y) = \gamma_1(\boldsymbol{c}'\boldsymbol{U}) = \gamma_1(\boldsymbol{d}'(\boldsymbol{X} - \boldsymbol{\xi})$ is a function of the quantity $t = (\boldsymbol{d}'\boldsymbol{\gamma})^2$, specifically

$$\gamma_1(Y) = h(t) = \frac{2}{\pi}t[at - 3b\omega_{\boldsymbol{d}}]^2 = \frac{2}{\pi}\delta_0^2\omega_{\boldsymbol{d}}^3[a\delta_0^2 - 3b]^2 \tag{9}$$

where as before $\boldsymbol{\gamma} = \dfrac{\boldsymbol{\Omega\eta}}{\sqrt{1 + \boldsymbol{\eta}'\boldsymbol{\Omega\eta}}}$.

Firstly, we prove the non-decreasing behaviour of $h(t)$, whose first derivative is given by

$$h'(t) = \frac{6\omega_{\boldsymbol{d}}^2}{\pi}\left(\frac{at}{\omega_{\boldsymbol{d}}} - 3b\right)\left(\frac{at}{\omega_{\boldsymbol{d}}} - b - \frac{2bct}{\omega_{\boldsymbol{d}}}\right) : 0 \leq \omega_{\boldsymbol{d}}^{-1}t \leq 1,$$

with a and b as previously defined, and the quantity c given by $c = \frac{2}{\pi}\frac{E^2(S)}{E(S^2)}$.

The well-known moment inequality $E(S^3) \geq E(S)E(S^2)$ implies that $b \leq 0$, so we are going to distinguish two cases: if $a \geq 0$ then $h'(t) > 0$ and $h(t)$ is a non-decreasing function. On the other hand, when $a < 0$ the condition on the moments of S assumed in the statement implies that $b \leq a$, which in turn gives $\frac{b}{a} \geq 1$. Taking into account that $0 \leq \omega_{\boldsymbol{d}}^{-1}t \leq 1$ we can assert that $\frac{at}{\omega_{\boldsymbol{d}}} - b > 0$ from which we obtain that $h'(t) > 0$ and once again we conclude that $h(t)$ is a non-decreasing function.

Since $h(t)$ is non-decreasing, its maximum is attained at the maximum value of $t = (\boldsymbol{d}'\boldsymbol{\gamma})^2$; so our problem in (6) is simplified to finding the direction that maximizes $(\boldsymbol{d}'\boldsymbol{\gamma})^2$. We know that

$$(\boldsymbol{d}'\boldsymbol{\gamma})^2 = (\boldsymbol{c}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\gamma})^2 \leq (\boldsymbol{c}'\boldsymbol{c})(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\gamma})'(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\gamma}) = \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma},$$

with the maximum of the previous scalar product attained when $\boldsymbol{c}$ is proportional to $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\gamma}$. Hence, we can conclude that $\boldsymbol{\lambda}_U \propto \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\gamma}$ which, taking into account that $\boldsymbol{\lambda}_X \propto \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda}_U$ from (6), implies that $\boldsymbol{\lambda}_X \propto \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}$. This finding, together with the result of Lemma 1, proves the statement. □

Observe that the condition $a = \frac{4}{\pi}E^3(S) - E(S^3) \geq 0$ also ensures the validity of Theorem 1 as already shown by the arguments employed for proving the theorem. In fact, taking into account that $b = E(S)E(S^2) - E(S^3) \leq 0$, this third order moment condition implies the condition $\frac{4}{\pi}E^2(S) - E(S^2) \geq 0$ given by the statement of Theorem 1. However, this third moment condition is not satisfied by some popular subfamilies within the SMSN family.

As a result of Theorem 1, we could also calculate the Malkovich-Afifi's skewness index (5) for multivariate SMSN distributions as follows: We just have to take for $\boldsymbol{d}$ in (9) the vector giving the maximal skewness direction

$\boldsymbol{\lambda_X} = \dfrac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}{\sqrt{\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}}$. When $\boldsymbol{d} = \boldsymbol{\lambda_X}$, some calculations show that the quantities involved in expression (9) become

$$\omega_{\boldsymbol{d}} = \frac{\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}{\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}} = \frac{1}{E(S^2) - \frac{2}{\pi}E^2(S)\boldsymbol{\gamma}'\boldsymbol{\Omega}^{-1}\boldsymbol{\gamma}}$$

$$t = (\boldsymbol{\lambda}_{\boldsymbol{X}}'\boldsymbol{\gamma})^2 = \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma} = \frac{\boldsymbol{\gamma}'\boldsymbol{\Omega}^{-1}\boldsymbol{\gamma}}{E(S^2) - \frac{2}{\pi}E^2(S)\boldsymbol{\gamma}'\boldsymbol{\Omega}^{-1}\boldsymbol{\gamma}} = \omega_{\boldsymbol{d}}(\boldsymbol{\gamma}'\boldsymbol{\Omega}^{-1}\boldsymbol{\gamma}),$$

and $\delta_0^2 = \boldsymbol{\gamma}'\boldsymbol{\Omega}^{-1}\boldsymbol{\gamma} = \dfrac{\boldsymbol{\gamma}'\boldsymbol{\eta}}{\sqrt{1+\boldsymbol{\eta}'\boldsymbol{\Omega}\boldsymbol{\eta}}} = \dfrac{\boldsymbol{\eta}'\boldsymbol{\Omega}\boldsymbol{\eta}}{1+\boldsymbol{\eta}'\boldsymbol{\Omega}\boldsymbol{\eta}} = \dfrac{\lambda^2}{1+\lambda^2}$, with $\lambda^2 = \boldsymbol{\eta}'\boldsymbol{\Omega}\boldsymbol{\eta}$.

Therefore, we get $\gamma_1(Y) = \dfrac{2}{\pi}\dfrac{\lambda^2}{1+\lambda^2}\dfrac{\left[a\frac{\lambda^2}{1+\lambda^2} - 3b\right]^2}{\left[E(S^2) - \frac{2}{\pi}E^2(S)\frac{\lambda^2}{1+\lambda^2}\right]^3}$. Taking into account the expressions for a and b, as used in the proof of Theorem 1, we obtain that the theoretic value of the maximal skewness given by

$$\gamma_{1,p}^D(\boldsymbol{X}) = \frac{2\lambda^2\left[(\frac{4}{\pi}E^3(S) - 3E(S)E(S^2) + 2E(S^3))\lambda^2 - 3(E(S)E(S^2) - E(S^3))\right]^2}{\pi\left[E(S^2) + \left(E(S^2) - \frac{2}{\pi}E^2(S)\right)\lambda^2\right]^3} \tag{10}$$

2.2 Examples

In this section we present some particular examples within the SMSN family for which the result of Theorem 1 is valid. They are all well-established and used multivariate models.

2.2.1 The multivariate SN distribution

The SN multivariate model is obtained when the mixing variable is degenerate at $S = 1$. In this case $a \geq 0$ and the result of Theorem 1 is satisfied; this finding was previously derived by Loperfido (2010). It is also worthwhile noting that in this case $E(S) = E(S^2) = E(S^3) = 1$ and the maximal skewness in (10) reduces to

$$\gamma_{1,p}^D(\boldsymbol{X}) = 2(4-\pi)^2\left\{\frac{\lambda^2}{\pi + (\pi-2)\lambda^2}\right\}^3 \tag{11}$$

which agrees with the Malkovich-Afifi's skewness measure for the multivariate SN distribution (Loperfido 2010; Azzalini and Capitanio 2014).

2.2.2 The multivariate skew-t distribution

The multivariate skew-t (ST) distribution arises when the mixing variable of the SMSN vector is $S = U^{-1/2}$ with $U \sim \chi^2_\nu/\nu$. In this case, as stated by Azzalini and Capitanio (2003), we obtain that the density function (3) becomes

$$f(\boldsymbol{x};\boldsymbol{\xi},\boldsymbol{\Omega},\boldsymbol{\alpha},\nu) = 2\, t_p(\boldsymbol{x};\nu) T_1\left(\boldsymbol{\alpha}'\boldsymbol{\omega}^{-1}(\boldsymbol{x}-\boldsymbol{\xi})\left(\frac{\nu+p}{Q_{\boldsymbol{x}}+\nu}\right)^{1/2};\nu+p\right) : \boldsymbol{x}\in\mathbb{R}^p \tag{12}$$

where $t_p(\boldsymbol{x};\nu)$ is the density function of a p-dimensional t variable with ν degrees of freedom given by $t_p(\boldsymbol{x};\nu) = \dfrac{\Gamma((\nu+p)/2)}{|\boldsymbol{\Omega}|^{1/2}(\pi\nu)^{p/2}\Gamma(\nu/2)}\left(1+\dfrac{Q_{\boldsymbol{x}}}{\nu}\right)^{-(\nu+p)/2}$, $T_1(y;\nu+p)$ is the distribution function of a t scalar variable with $\nu+p$ degrees of freedom and $Q_{\boldsymbol{x}} = (\boldsymbol{x}-\boldsymbol{\xi})'\boldsymbol{\Omega}^{-1}(\boldsymbol{x}-\boldsymbol{\xi})$.

We write $\boldsymbol{X} \sim ST_p(\boldsymbol{\xi},\boldsymbol{\Omega},\boldsymbol{\alpha},\nu)$ to denote that $\boldsymbol{X}$ follows a p-dimensional ST distribution with density function (12). Figure 1 shows how the vector $\boldsymbol{\alpha}$ deforms the symmetry of the t distribution when the asymmetry is injected into the multivariate model across different directions; the contoured plots for each density function are also depicted. It is worthwhile noting that when $\nu \to \infty$, the ST becomes the SN distribution, i.e. $\boldsymbol{X} \sim SN_p(\boldsymbol{\xi},\boldsymbol{\Omega},\boldsymbol{\alpha})$.

In this case the moments of S are $E(S^k) = \dfrac{(\nu/2)^{k/2}\Gamma\left(\frac{\nu-k}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}$ provided that $\nu > k : k \geq 1$. From this expression we obtain that for $\nu > 3$

$$a = E(S)\left(\frac{4}{\pi}E^2(S) - \frac{\nu}{\nu-3}\right) = \nu E(S)\left(\frac{2}{\pi}\frac{\Gamma^2\left(\frac{\nu-1}{2}\right)}{\Gamma^2\left(\frac{\nu}{2}\right)} - \frac{1}{\nu-3}\right),$$

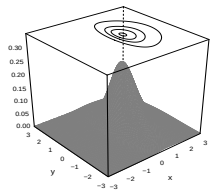
with $a < 0$ when $\nu < 9$ and $a > 0$ when $\nu \geq 9$. This assertion can be proved if a is rewritten as $a = \nu E(S)\dfrac{1}{\nu-3}\dfrac{2}{\pi}\left(h(\nu) - \dfrac{\pi}{2}\right)$, with $h(\nu) = \dfrac{(\nu-3)\Gamma^2\left(\frac{\nu-1}{2}\right)}{\Gamma^2\left(\frac{\nu}{2}\right)}$, taking into account that h is a non-decreasing function.

Using Lemma 1 from Arevalillo and Navarro (2015), we can conclude the validity of the second order moment condition: $\dfrac{4}{\pi}E^2(S) \geq E(S^2)$.

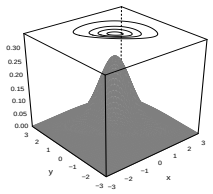
2.2.3 The multivariate skew double exponential distribution

We say that a p-dimensional vector follows a double exponential (DE) distribution with location $\boldsymbol{\xi}$ and full rank scale matrix $\boldsymbol{\Omega}$ if it has a density function given by

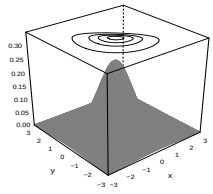
$$f(\boldsymbol{x};\boldsymbol{\xi},\boldsymbol{\Omega}) = \frac{\Gamma(\frac{p}{2})}{\pi^{p/2}\Gamma(p)2^{1+p}}|\boldsymbol{\Omega}|^{-1/2}\exp\left\{-\frac{1}{2}\left[(\boldsymbol{x}-\boldsymbol{\xi})'\boldsymbol{\Omega}^{-1}(\boldsymbol{x}-\boldsymbol{\xi})\right]^{1/2}\right\}, \tag{13}$$



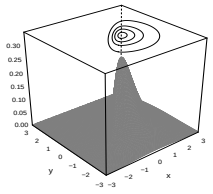
(a) $\boldsymbol{\alpha}' = (3, 0)$



(b) $\boldsymbol{\alpha}' = (3, 3)$



(c) $\boldsymbol{\alpha}' = (-3, -3)$



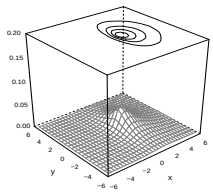
(d) $\boldsymbol{\alpha}' = (3, -3)$

Fig. 1: Density functions of the bivariate ST variable, with location $\boldsymbol{\xi} = (0, 0)$, scale matrix $\boldsymbol{\Omega} = \boldsymbol{I}_2$ and $\nu = 4$, for different shape vectors.

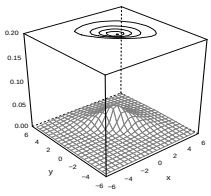
The multivariate DE distribution is a scale mixture of normal variables with mixing variate $S = U^{-1/2}$, where $U = W^{-1}$ and $W \sim Gamma\left(\frac{p+1}{2}, \frac{1}{8}\right)$ (Gómez-Sánchez-Manzano et al 2008); equivalently, we could say that the scalar variable U follows an inverse gamma distribution. When we put this mixing variable in Definition 1, we get the multivariate skew double exponential (SDE) distribution. We write $\boldsymbol{X} \sim SDE_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ to indicate that the random vector $\boldsymbol{X}$ follows a p-dimensional SDE distribution with location $\boldsymbol{\xi}$, scale matrix $\boldsymbol{\Omega} = \boldsymbol{\omega}\overline{\boldsymbol{\Omega}}\boldsymbol{\omega}$ and shape vector $\boldsymbol{\alpha}$. If the distribution H in (3) is replaced by the inverse gamma distribution then we obtain the density function of the multivariate SDE model, which is given by

$$\psi_p(\boldsymbol{x}; \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\Omega}) = \frac{2}{\Gamma\left(\frac{p+1}{2}\right) 8^{\frac{p+1}{2}}} \cdot \int_0^{\infty} \phi_p(\boldsymbol{x} - \boldsymbol{\xi}; u^{-1}\boldsymbol{\Omega}) \Phi(u^{1/2}\boldsymbol{\alpha}'\boldsymbol{\omega}^{-1}(\boldsymbol{x} - \boldsymbol{\xi}) u^{-\frac{p+1}{2}-1} e^{-1/8u} du \quad (14)$$

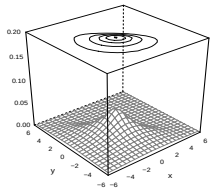
The plots of the bivariate SDE densities are depicted by Figure 2; note the effect of parameter $\boldsymbol{\alpha}$ on the shape of the densities and their contoured plots after injection of asymmetry across different directions. A simple comparison with Figure 1 gives an idea of the differences between both subclasses.



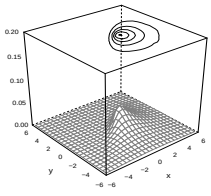
(a) $\boldsymbol{\alpha}' = (3, 0)$



(b) $\boldsymbol{\alpha}' = (3, 3)$



(c) $\boldsymbol{\alpha}' = (-3, -3)$



(d) $\boldsymbol{\alpha}' = (3, -3)$

Fig. 2: Density functions of the bivariate skew double exponential with $\boldsymbol{\xi} = (0, 0)$ and scale matrix $\boldsymbol{\Omega} = \boldsymbol{I}_2$, for different shape vectors.

In this case the moments of the mixing variable are given by the general formula: $E(S^k) = \dfrac{8^{k/2}\Gamma\left(\frac{p+k+1}{2}\right)}{\Gamma\left(\frac{p+1}{2}\right)} = \dfrac{2^{k/2}\Gamma\left(\frac{p}{2}\right)\Gamma(p+k)}{\Gamma(p)\Gamma\left(\frac{p+k}{2}\right)} : k \geq 1$, where the last equality follows from the well-known property: $\Gamma\left(\dfrac{n}{2}\right) = \dfrac{2^{1-n}(n-1)!\sqrt{\pi}}{\left(\frac{n-1}{2}\right)!}$. Taking in this general formula $k = 1, 2$, we obtain that

$$a = E(S)\left(\frac{4}{\pi}E^2(S) - 4(p+2)\right) = 4E(S)\left(\frac{2}{\pi}\frac{p^2\Gamma^2\left(\frac{p}{2}\right)}{\Gamma^2\left(\frac{p+1}{2}\right)} - (p+2)\right),$$

which gives negative values when $p < 5$ and positive values when $p \geq 5$.

In order to check if the moment condition $\dfrac{4}{\pi}E^2(S) \geq E(S^2)$ is held, we define the function:

$$g(p) = \frac{1}{p+1}\left[\frac{p\Gamma\left(\frac{p}{2}\right)}{\Gamma\left(\frac{p+1}{2}\right)}\right]^2 \quad : \quad p \geq 1.$$

After taking logarithms, we can see that its first derivative is given by

$$g'(p) = g(p)\left[\frac{2}{p} + \psi\left(\frac{p}{2}\right) - \psi\left(\frac{p+1}{2}\right) - \frac{1}{p+1}\right],$$

where $\psi(x) = \dfrac{\Gamma'(x)}{\Gamma(x)}$ is the digamma function.

Taking into account the well-known property: $\psi(x+1) = \dfrac{1}{x} + \psi(x)$ and the following inequalities for the digamma function: $\log\left(x - \dfrac{1}{2}\right) < \psi(x) < \log(x) - \dfrac{1}{2x}$ when $x > \dfrac{1}{2}$ (Merkle 1998), we get

$$g'(p) = g(p)\left[\psi\left(\frac{p+2}{2}\right) - \psi\left(\frac{p+1}{2}\right) - \frac{1}{p+1}\right]$$

$$> g(p)\left[\log\left(\frac{p+1}{2}\right) - \log\left(\frac{p+1}{2}\right) + \frac{1}{p+1} - \frac{1}{p+1}\right] = 0,$$

which implies that $g(p)$ is a non-decreasing function for $p \geq 1$. Consequently, $g(p) \geq g(1) = \dfrac{\pi}{2}$ or equivalently $\dfrac{2}{\pi}\left[\dfrac{p\Gamma\left(\frac{p}{2}\right)}{\Gamma\left(\frac{p+1}{2}\right)}\right]^2 \geq p+1$. This inequality leads to the moment condition $\dfrac{4}{\pi}E^2(S) \geq E(S^2)$, so the result of Theorem 1 is also valid for the multivariate SDE distribution.

2.2.4 The multivariate skew-slash distribution

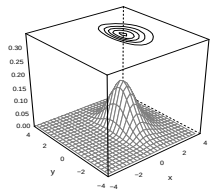
Another flexible model that combines both asymmetry and tail weight behavior is the skew-slash (SSL) multivariate distribution (Wang and Genton 2006). It corresponds to the mixing variable $S = U^{-1/2}$ with $U \sim Beta(q/2, 1)$, where q is a tail weight parameter such that $q > 0$. We put $\boldsymbol{X} \sim SSL_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, q)$ to indicate that $\boldsymbol{X}$ follows a p-dimensional SSL distribution with location $\boldsymbol{\xi}$, scale matrix $\boldsymbol{\Omega} = \boldsymbol{\omega}\overline{\boldsymbol{\Omega}}\boldsymbol{\omega}$, shape vector $\boldsymbol{\alpha}$ and tail weight parameter $q > 0$. The density function of a SSL vector can be easily derived from (3) taking into account that H is the distribution of a $Beta(q/2, 1)$ random variable; it is given by

$$\psi_p(\boldsymbol{x}; \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\Omega}, q) = q\int_0^\infty \phi_p(\boldsymbol{x} - \boldsymbol{\xi}; u^{-1}\boldsymbol{\Omega})\Phi(u^{1/2}\boldsymbol{\alpha}'\boldsymbol{\omega}^{-1}(\boldsymbol{x} - \boldsymbol{\xi}))u^{q/2-1}du \quad (15)$$

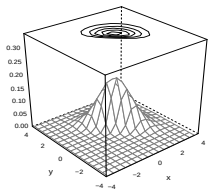
Note that when $q \to \infty$, the SSL becomes a SN distribution. Figure 3 shows the bivariate SSL density functions for several directions of asymmetry; the contoured curves are also displayed for each case.

In this case we know that the moments of the mixing variable are given by $E(S^k) = E(U^{-k/2}) = \dfrac{q}{q-k}$ for $q > k$. From now on, we assume that $q > 3$.

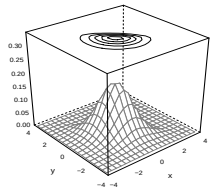
Let us define the function: $g(q) = \dfrac{E(S^2)}{E(S)^2} = \dfrac{(q-1)^2}{q(q-2)}$. Since g is a decreasing function, we have $g(q) \leq g(4) \leq \dfrac{4}{\pi}$; consequently, the second order moment condition of Theorem 1 also holds.



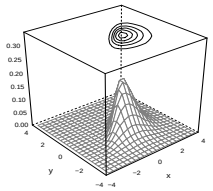
(a) $\boldsymbol{\alpha}' = (3, 0)$



(b) $\boldsymbol{\alpha}' = (3, 3)$



(c) $\boldsymbol{\alpha}' = (-3, -3)$



(d) $\boldsymbol{\alpha}' = (3, -3)$

Fig. 3: Density functions of the bivariate skew-slash variable, with location $\boldsymbol{\xi} = (0, 0)$, scale matrix $\boldsymbol{\Omega} = \boldsymbol{I}_2$ and tail weight parameter $q = 5$, for different shape vectors.

2.2.5 The multivariate skew-contaminated normal distribution

The family of skew-contaminated normal distributions arises when the mixing variable of the multivariate SMSN model is $S = U^{-1/2}$ with U a scalar variable following a discrete distribution such that

$$P(U = \gamma) = \epsilon \quad , \quad P(U = 1) = 1 - \epsilon$$

where $0 \leq \epsilon \leq 1$ is a parameter quantifying the amount of contamination and $0 < \gamma \leq 1$ is a scale inflation factor (Lachos et al 2010a). In this case the density function of the input vector $\boldsymbol{X}$ is given by

$$f(\boldsymbol{x}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \gamma, \epsilon) = 2 \Big[\epsilon \phi_p(\boldsymbol{x} - \boldsymbol{\xi}; \gamma^{-1}\boldsymbol{\Omega}) \Phi(\gamma^{1/2} \boldsymbol{\alpha}' \boldsymbol{\omega}^{-1} (\boldsymbol{x} - \boldsymbol{\xi})) + (1 - \epsilon) \phi_p(\boldsymbol{x} - \boldsymbol{\xi}; \boldsymbol{\Omega}) \Phi(\boldsymbol{\alpha}' \boldsymbol{\omega}^{-1} (\boldsymbol{x} - \boldsymbol{\xi})) \Big] \quad (16)$$

We will write $\boldsymbol{X} \sim SCN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \gamma, \epsilon)$ to denote that the vector $\boldsymbol{X}$ follows a p-dimensional skew-contaminated normal distribution. It can be observed that if we take $\epsilon = 0$ or $\gamma = 1$ then the model reduces the multivariate SN distribution.

In order to check the validity of the moment condition $\frac{4}{\pi}E^2(S) \geq E(S^2)$ from Theorem 1, we must take into account that $E(S^k) = \gamma^{-k/2}\epsilon + 1 - \epsilon$. The expression $\frac{4}{\pi}E^2(S) - E(S^2)$ can be written as a function of $\gamma^{-1/2}$ as follows: $\frac{4}{\pi}E^2(S) - E(S^2) = h(\gamma^{-1/2})$, where h is a function defined by the following second order degree polynomial:

$$h(x) = \frac{4}{\pi}\left[1 + \epsilon(x-1)\right]^2 - \left[1 + \epsilon(x^2-1)\right]$$

$$= \epsilon\left(\frac{4\epsilon}{\pi} - 1\right)x^2 + \frac{8}{\pi}\epsilon(1-\epsilon)x + \frac{4}{\pi}(\epsilon-1)^2 - 1 + \epsilon,$$

whose roots are given by the values

$$x_1 = \frac{-\frac{8}{\pi}\epsilon(1-\epsilon) - \sqrt{4\epsilon(1-\epsilon)\left(\frac{4}{\pi}-1\right)}}{2\epsilon\left(\frac{4\epsilon}{\pi}-1\right)} \qquad \text{and}$$

$$x_2 = \frac{-\frac{8}{\pi}\epsilon(1-\epsilon) + \sqrt{4\epsilon(1-\epsilon)\left(\frac{4}{\pi}-1\right)}}{2\epsilon\left(\frac{4\epsilon}{\pi}-1\right)}.$$

We can see that h is a convex function provided that $\frac{\pi}{4} \leq \epsilon \leq 1$. Therefore, if this inequality holds then we can assert that $x_1 < x_2$ and we arrive to the following implications: The first implication is that the function h is negative when $x \in (x_1, x_2)$ and positive when $x \notin (x_1, x_2)$. The second implication is that, as a result of the condition $\frac{\pi}{4} \leq \epsilon \leq 1$, the inequality $\sqrt{4\epsilon(1-\epsilon)\,(4/\pi - 1)} < 2\epsilon(4/\pi - 1)$ holds and consequently $x_2 < 1$. Therefore, we can conclude that the quantity $\frac{4}{\pi}E^2(S) - E(S^2) = h(\gamma^{-1/2}) \geq 0$.

On the other hand, if $0 \leq \epsilon < \frac{\pi}{4}$ then we can easily see that $x_2 < x_1$ and h is a concave function such that $h(x) \geq 0$ when $x \in (x_2, x_1)$ and $h(x) \leq 0$ when $x \notin (x_2, x_1)$. Moreover, once again it holds that $x_2 < 1$. Consequently, we can state that $\frac{4}{\pi}E^2(S) - E(S^2) = h(\gamma^{-1/2}) \geq 0$ as long as $\gamma^{-1/2} < x_1$; however, the sufficient condition of Theorem 1 is not satisfied when $\gamma^{-1/2} \geq x_1$.

This example has provided a situation for which the condition on the moments of the mixing variable doesn't hold; so we cannot guarantee the result established by Theorem 1 when $0 \leq \epsilon < \frac{\pi}{4}$ and $\gamma^{-1/2} \geq x_1$. This fact doesn't mean that the shape vector $\boldsymbol{\eta}' = \boldsymbol{\alpha}'\boldsymbol{\omega}^{-1}$ will fail at yielding the maximal skewness projection since the moment condition of the theorem is a sufficient condition.

3 Simulations and application

In this section the theoretical results are illustrated by means of a simulation experiment for artificial data and a couple of applications to real data, which are aimed at illustrating the usefulness of the *MaxSkew* principle for feature construction and pattern recognition in the analysis of genomic data, and also for class discovery in data. We will focus on the computational side for the estimation of $\boldsymbol{\eta}$ using the functionalities implemented in the MaxSkew R package (Franceschini and Loperfido 2016).

3.1 Simulation experiment

This experiment comprises several simulation scenarios for some of the SMSN subfamilies studied in Section 2.2; they include the multivariate skew-t, skew-slash and skew-contaminated normal distributions. The simulation study is carried out for different dimensions of the input vector, $p = 2, 10$, and sample sizes $n = 100, 500$. In order to establish the common framework that will parameterize all the multivariate distributions, we take the following settings for location and scale parameters: $\boldsymbol{\xi} = \mathbf{0}$ and $\bar{\boldsymbol{\Omega}}$ a correlation Toeplitz matrix defined by $\bar{\boldsymbol{\Omega}} = (\omega_{i,j})_{1 \leq i,j \leq p}$, with $\omega_{i,j} = \rho^{|i-j|} : 1 \leq i \leq j \leq p$, where for the values of ρ we take: $\rho = -0.3, 0.3, -0.8, 0.8$, and $\boldsymbol{\omega}$ defined by a diagonal matrix whose entries are generated at random from the set of integers $\{1, 2 \ldots, 100\}$. The shape vector of the distributions is set on the basis of the idea of injecting skewness into the multivariate model through a singular direction such as the first principal component of the scale matrix; another direction like the last component of the scale matrix is also considered for a comparative purpose.

For each one of these simulation scenarios, 5000 replicates of the experiment are carried out by drawing samples from the corresponding skewed input vector $\boldsymbol{X}$. The maximal skewness direction is estimated using the third order moment matrix, on the basis of the implementation provided by the MaxSkew R package (Franceschini and Loperfido 2016) —from now on *MaxSkew* estimation— and the mean square error (MSE) is computed by comparing the unit length vectors corresponding to the direction estimated by the maximal skewness principle and the exact theoretical direction.

3.1.1 Results for the multivariate skew-t distribution

In this scenario we consider a skew-t input vector $\boldsymbol{X}$ whose tail weight is regulated by the degrees of freedom ν. In this case, for each combination of the triple (n, p, ρ), three simulation trials are carried out for $\nu = 4, 8, 50$.

The MSEs of the *MaxSkew* estimation when the shape vector lies on the direction of the first principal component of the scale matrix are shown in Table 1: Observe that the errors are quite similar for opposite values of ρ; overall slight lower errors are obtained for the lower $|\rho|$. On the other hand, as expected, the smaller errors appear for the larger sample size. However,

Table 1: MSEs obtained from the skew-t distribution with shape vector lying on the direction of the first principal component of the scale matrix.

	$n \diagdown \nu$	4	8	50	4	8	50
		$\rho = -0.8$			$\rho = 0.8$		
$p = 2$	100	1.3170	1.3571	1.5563	1.3071	1.3546	1.5463
	500	0.9991	0.8809	1.0047	0.9997	0.8550	1.0077
$p = 10$	100	1.8749	1.8755	1.9093	1.8713	1.8728	1.9057
	500	1.7624	1.6091	1.4369	1.7633	1.6102	1.4182
		$\rho = -0.3$			$\rho = 0.3$		
$p = 2$	100	1.2054	1.2935	1.5518	1.1953	1.2718	1.5501
	500	0.8501	0.7541	1.0097	0.8590	0.7375	1.0026
$p = 10$	100	1.7634	1.7962	1.9084	1.7711	1.7988	1.8948
	500	1.6023	1.5229	1.6771	1.6079	1.5182	1.6518

we don't observe a clear pattern of variability of the MSE with respect to ν. Finally, we can see that the accuracy deteriorates for the larger p.

For comparison we also study the case of a shape vector lying on the direction of the last principal component of the scale matrix. The results are shown in Table 2 only for $\rho = 0.3, 0.8$, as they provide similar outputs as the negative values of ρ. Note that in nearly all the cases the errors are smaller than in Table 1 with more noticeable differences in the models which are furthest from the SN. Moreover, the increasing trend of the MSE with respect to ρ is more remarkable in this case.

Table 2: MSEs obtained from the skew-t distribution with shape vector lying on the direction of the last principal component of the scale matrix.

	$n \diagdown \nu$	4	8	50	4	8	50
		$\rho = 0.3$			$\rho = 0.8$		
$p = 2$	100	0.6495	0.8115	1.3418	0.9737	1.1358	1.5733
	500	0.2335	0.1991	0.6670	0.5362	0.4885	1.0033
$p = 10$	100	1.4027	1.5105	1.8312	1.6563	1.7149	1.9065
	500	1.0188	1.0132	1.5572	1.4262	1.4192	1.7587

3.1.2 Results for the multivariate skew-slash distribution

Now we assume that the input vector follows a skew-slash distribution with the previously defined settings for location and scale parameters. In this case the tail weight is controlled by the parameter q; so each simulation experiment is carried out for the following values of the tail weight: $q = 5, 25, 50$.

The results are provided by Table 3 which displays the MSEs when the shape vector lies on the direction of the principal eigenvector of the scale

matrix: It is observed that the impact of ρ on the accuracy of the estimations is more remarkable when the dimension increases. We can also see that the MSE increases with p and, to a lesser extend, with the tail weight q.

Table 3: MSEs obtained from the skew-slash distribution with shape vector lying on the direction of the first principal component of the scale matrix.

	$n \diagdown q$	5	25	50	5	25	50
			$\rho = -0.8$			$\rho = 0.8$	
$p = 2$	100	1.4258	1.6016	1.6276	1.4413	1.5821	1.6251
	500	0.9892	1.0712	1.1035	1.0007	1.0706	1.1185
$p = 10$	100	1.8838	1.9075	1.9132	1.8786	1.9165	1.9166
	500	1.7157	1.4453	1.4436	1.7282	1.4463	1.4407
			$\rho = -0.3$			$\rho = 0.3$	
$p = 2$	100	1.3777	1.6167	1.6500	1.3660	1.5864	1.6367
	500	0.8895	1.1078	1.1509	0.9081	1.1224	1.1587
$p = 10$	100	1.8063	1.9320	1.9347	1.8170	1.9269	1.9482
	500	1.6529	1.7508	1.7925	1.6518	1.7737	1.7861

Once again we also study the case of a shape vector lying on the direction of the last principal component of the scale matrix. The results are summarized by Table 4 for $\rho = 0.3, 0.8$. Similarly to the skew-t, the MSE increases with the ρ parameter and the dimension. The outputs show a more accurate estimation than in Table 3 only when $q = 5$, which corresponds to the model furthest from the skew-normal.

Table 4: MSEs obtained from the skew-slash distribution with shape vector lying on the direction of the last principal component of the scale matrix.

	$n \diagdown q$	5	25	50	5	25	50
			$\rho = 0.3$			$\rho = 0.8$	
$p = 2$	100	1.0026	1.4891	1.5131	1.2724	1.6352	1.7102
	500	0.3532	0.9307	0.9677	0.6965	1.1921	1.2698
$p = 10$	100	1.5366	1.9213	1.9028	1.7142	1.9793	1.9602
	500	1.1654	1.8005	1.7949	1.5498	1.9101	1.9070

3.1.3 Results for the multivariate skew-contaminated normal distribution

Finally, we consider an input vector $\boldsymbol{X}$ following a skew-contaminated normal distribution. In this case we will fix the inflation factor to $\gamma = 0.5$ and we compute the MSEs, under the established settings for location and scale, varying the simulation trials for the following proportions of contamination: $\epsilon = 0.3, 0.5, 0.8$.

The results are displayed in Table 5. We observe the same pattern for opposite values of ρ; the impact of ρ in the errors is the same as in the skew-slash. Once again, the results reveal that the MSE increases for the larger dimension, which comes up as a quite natural finding for nearly all the trials under consideration. In this case an interesting finding is that the MSE decreases as we approach the model with equal contamination $\epsilon = 0.5$.

Table 5: MSEs obtained from the skew-contaminated normal with shape vector lying on the direction of the first principal component of the scale matrix.

	$n \diagdown \epsilon$	0.3	0.5	0.8	0.3	0.5	0.8
			$\rho = -0.8$			$\rho = 0.8$	
$p = 2$	100	1.4416	1.4166	1.5128	1.4563	1.4840	1.5014
	500	0.9026	0.8742	0.9580	0.9062	0.8674	0.9625
$p = 10$	100	1.8838	1.8813	1.8913	1.8860	1.8817	1.8969
	500	1.3923	1.3433	1.3694	1.3840	1.3479	1.3857
			$\rho = -0.3$			$\rho = 0.3$	
$p = 2$	100	1.4150	1.3947	1.4973	1.4376	1.4514	1.4931
	500	0.8403	0.8130	0.9441	0.8478	0.8095	0.9419
$p = 10$	100	1.8567	1.8640	1.8912	1.8641	1.8650	1.9048
	500	1.5166	1.5042	1.6267	1.5152	1.5105	1.6296

The previous experimental trials are now repeated taking a shape vector that lies on the direction of the last principal component; the resulting MSEs are shown in Table 6. In this case the MSE exhibits the same increasing pattern with respect to ρ as the one observed in Tables 2 and 4.

Table 6: MSEs obtained from the skew-contaminated normal with shape vector lying on the direction of the last principal component of the scale matrix.

	$n \diagdown \epsilon$	0.3	0.5	0.8	0.3	0.5	0.8
			$\rho = 0.3$			$\rho = 0.8$	
$p = 2$	100	1.1699	1.0891	1.3195	1.3510	1.4039	1.4683
	500	0.4014	0.3449	0.5902	0.7197	0.6948	0.9221
$p = 10$	100	1.6780	1.7306	1.8136	1.8620	1.8641	1.9066
	500	1.2028	1.1829	1.4982	1.5771	1.5427	1.7244

Overall, the simulations have revealed an accurate and quite homogeneous behavior of the *MaxSkew* approach when applied under several underlying models that belong to the wide family of SMSN distributions. Hence, we advocate the use of the maximal skewness principle for analyzing multivariate skewed data. The next sections present a couple of real data applications that illustrate its usefulness.

3.2 Application to genomic data analysis

Individuals diagnosed with triple negative breast cancer (TNBC) are clinically heterogeneous and have more adverse prognosis than patients affected by other breast cancer subtypes. A genomic data set containing gene expression measures of TNBC patients was obtained from Gene Expression Omnibus (GEO) repository through accession series GSE31519. The data contained gene expression measures for 13146 genes from 579 TNBC tumors; after removal of 85 patients treated with neoadjuvant chemotherapy we obtained a data set with 13146 gene expression measures for 495 TNBC tumors samples.

Recent studies on genomics, proteomics and microRNA data have revealed the usefulness of probabilistic graphical models to obtain association networks that provide insights about an underlying functional biological structure (Gamez-Pozo et al 2015; Zapater-Moros et al 2018; Prado-Vázquez et al 2019). In brief, the approach takes a set of variables and, using mutual information to assess variable association, it applies the Chow-Liu algorithm (Chow and Liu 1968) to search the closest tree to the underlying dependence structure. The method provides a structure of connected branches which can be analyzed separately (Edwards et al 2010); this may be an important initial step to undertake dimension reduction in high-dimensional data as it will help to identify groups of variables on the basis of their associations.

When applied to the TNBC data using the 2000 most variable genes as input variables it allowed to identify 26 functional biological groups of genes, similarly as in Prado-Vázquez et al (2019). In this example we propose to apply data projections, using the skewness maximization principle, in order carry out a dimension reduction that leads to a representative gene for each one of the functional groups (labelled by "Node.." in Figures 5 and 6). Due to the asymmetries observed in gene expression measures, we advocate the use of the maximal skewness principle for constructing and assessing metagenes (genes related to a specific biological function), as an alternative to standard averages (Rody et al 2011; Prado-Vázquez et al 2019); we call the maximal skewness genes the *MaxSkew* metagenes. Our ultimate goal is to use such skewness based projection pursuit approach, as a dimension reduction method, in order to highlight outstanding observations or to uncover hidden patterns in data on the basis of the application of the multidimensional scaling (MDS) technique to the expression measures provided by the *MaxSkew* metagenes. The results are compared with an exploratory data analysis that applies only MDS to all the genes and with an application of the MDS to the expression measures obtained from the first principal component derived on each functional group.

The results provide the following insights: when the MDS is applied to all the genes, the representation of the data on the first two MDS coordinates shows a noisy cloud of points (top left plot of Figure 4); the fact is observed when the first principal component is used to summarize each one of the functional groups and the MDS is applied accordingly (top right plot of Figure 4). Meanwhile, when MDS is applied to the expression measures of the *MaxSkew* metagenes, it reveals the presence of a compact bulk of points, another small

group of individuals surrounding the bulk and two isolated observations that could be labeled as outliers (see the plot at the bottom of Figure 4). These preliminary insights are corroborated by the scatter plot matrices corresponding to the top six most variable *MaxSkew* metagenes and the bottom six least variable *MaxSkew* metagenes depicted by Figure 6 which, once again, can be compared with the scatter matrices obtained from the first principal component metagenes (Figure 5). Once again, it is highlighted the existence of a few abnormal observations that were unnoticed by the first principal component.

There is another intriguing fact revealed by a simple exploration of the correlations between the metagenes. The plots of Figure 7 show that the *MaxSkew* method leads to much lower correlated metagenes than the first principal component method. This issue is relevant from the biological viewpoint since it might be desirable for interpretability to handle a few uncorrelated or nearly uncorrelated biological factors which may be useful to shed light on the molecular and genomic underpinnings of the disease; we argue that they can be summarized by the *MaxSkew* metagenes, obtained by the maximal skewness based projection pursuit approach, as they lead to lower correlated constructs.

Therefore, we have shown that the maximal skewness principle can be a useful tool to solve the relevant biological problem of data reduction in high-dimensional genomic data by defining *MaxSkew* metagenes from functional groups of genes. This example illustrates how the skewness based projection pursuit method contributes to summarize the gene expression of high-dimensional genomic data; it also provides a powerful approach to identify outliers in high-dimensional genomic data, to uncover potential hidden patterns and to identify outstanding individuals as well.

3.3 Application to the detection of classes in data

Here, we present a simple example that illustrates the usefulness of the maximal skewness principle as a data dimension reduction tool for unsupervised learning that allows to uncover hidden groups or classes in data. The example deals with a data set that consists of biomedical measures collected by the Australian Institute of Sport (AIS) for 202 athletes (Cook and Weisberg 2009); let us consider the input vector $\boldsymbol{X} = (RCC, WCC, Bfat, LBM, Ht, Wt)$ whose components quantify the following measures: red cell count, white cell count, percent body fat, lean body mass, height (cm) and weight (kg). Multivariate skewed distributions were reported as adequate models for fitting this data (Azzalini and Capitanio 1999). Lin (2010) showed that AIS data can be better modelled by mixtures of multivariate skew distributions, as later evidenced by Lin et al (2014) and Lee and McLachlan (2016).

From the scatter plots of Figure 8, it can be observed the lack of well-defined groups, perhaps with the exception of the scatter plot for the inputs $Bfat$ and LBM. Now, we apply the proposed method to the construction of new features and use them for group detection; an exhaustive analysis showed that the maximal skewness projections led to *MaxSkew* features derived from

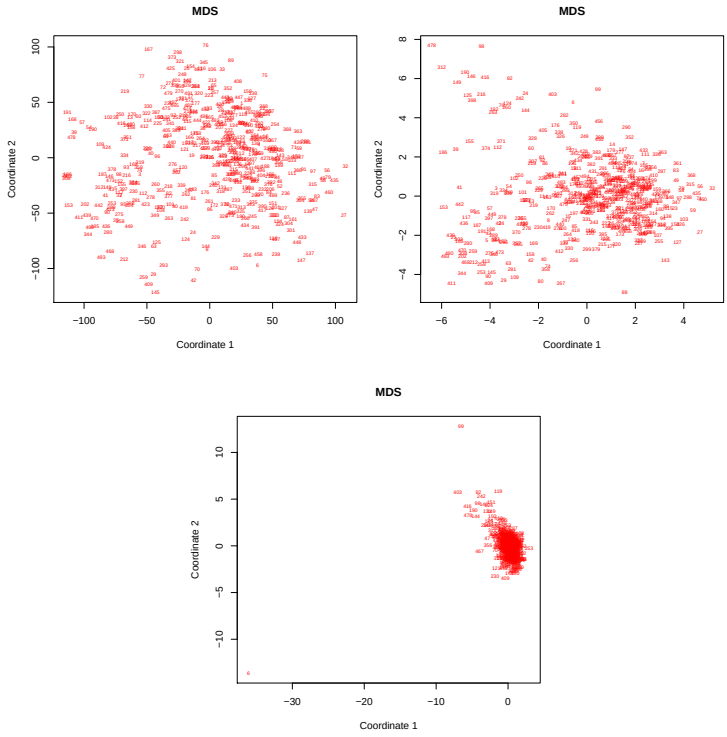


Fig. 4: Projection of data on the MDS coordinates obtained from all the genes (top left), on the MDS coordinates from the 26 metagenes given by the first principal component (top right) and on the MDS coordinates obtained from the 26 *MaxSkew* metagenes (bottom).

the triples $(WCC, Bfat, Wt)$, (WCC, LBM, Ht) and (WCC, Ht, Wt), which highlight two groups when used as the new features for two standard clustering methods like K-means and model based clustering with the BIC criterion for model selection —see the left plots of Figures 9 and 10 which depict the resulting groups. Actually, both findings are in close agreement with an underlying classification by sex, as shown by the right plots in both figures. Hence, the maximal skewness principle is also helpful as an auxiliary feature engineering tool for uncovering groups in data.

Unsurprisingly, the previous patterns are also validated by a multidimensional scaling (MDS) representation of the maximal skewness features when projected onto the two-dimensional MDS space, as shown by Figure 11. Moreover, it is worthwhile noting that the cases 113, 160, 161, 163, 166 and 178 appear isolated; this finding is consistent with the results displayed by Figures 9 and 10 which also highlight the abnormality of these observations, a revealing issue that may deserve further investigation.

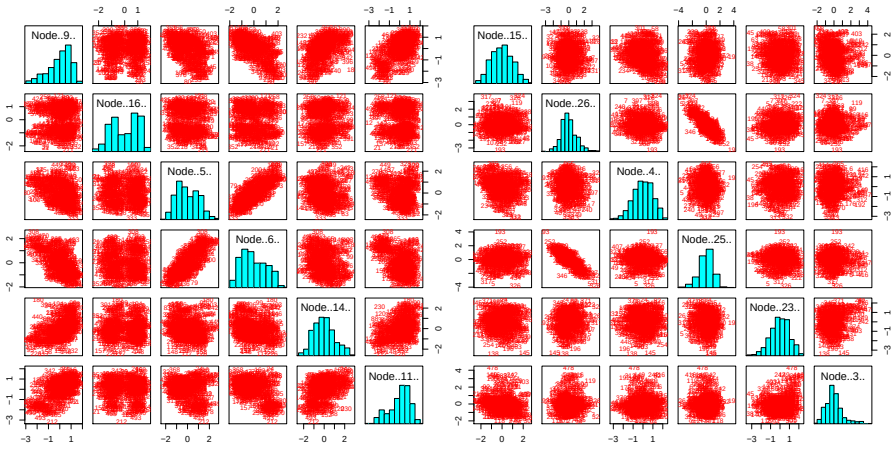


Fig. 5: Scatter plots for the six most variable metagenes generated by the first principal component (left) and the six least variable metagenes generated by the first principal component (right).

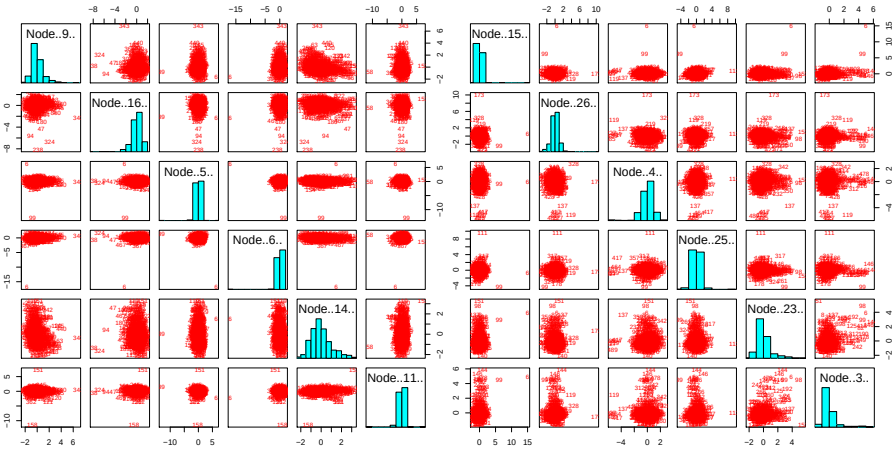


Fig. 6: Scatter plots for the six most variable *MaxSkew* metagenes (left) and the six least variable *MaxSkew* metagenes (right).

This example illustrates the maximal skewness principle providing an approach for dimension reduction that allows to summarize combos of inputs into the transformed *MaxSkew* features that convey a consistent representation of multivariate skewed data. When applied in combination with other exploratory techniques for group finding and pattern recognition, the proposed method reveals outstanding patterns and observations in multivariate data.

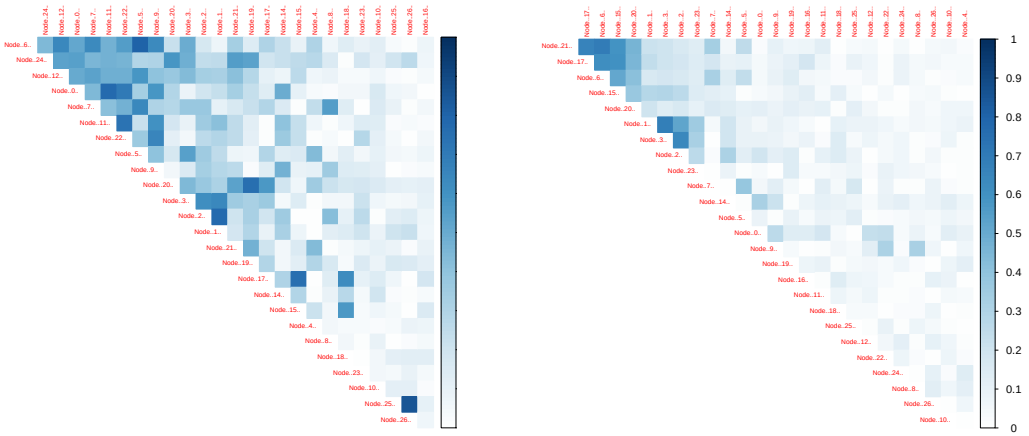


Fig. 7: Plots depicting the absolute value of correlations between first principal component metagenes (left) and *MaxSkew* metagenes (right).

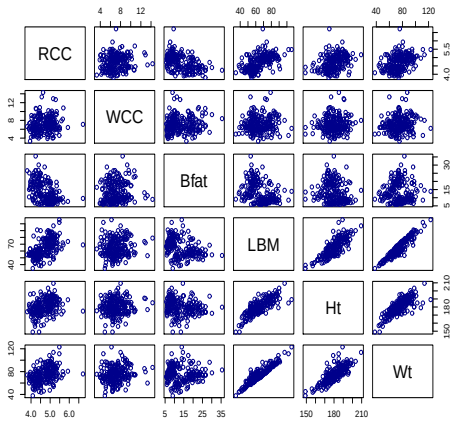


Fig. 8: Matrix of scatter plots for the AIS data set.

4 Summary and concluding remarks

In this paper we have addressed the problem of finding directions yielding maximal skewness projections for vectors that follow a SMSN distribution. We have found a simple condition on the moments of the mixing variable which guarantees that the direction yielding the maximal skewness projection is proportional to the shape vector $\boldsymbol{\eta}' = \boldsymbol{\alpha}'\boldsymbol{\omega}^{-1}$; this finding stresses the role of $\boldsymbol{\eta}'$ to regulate the asymmetry of the model in a directional fashion.

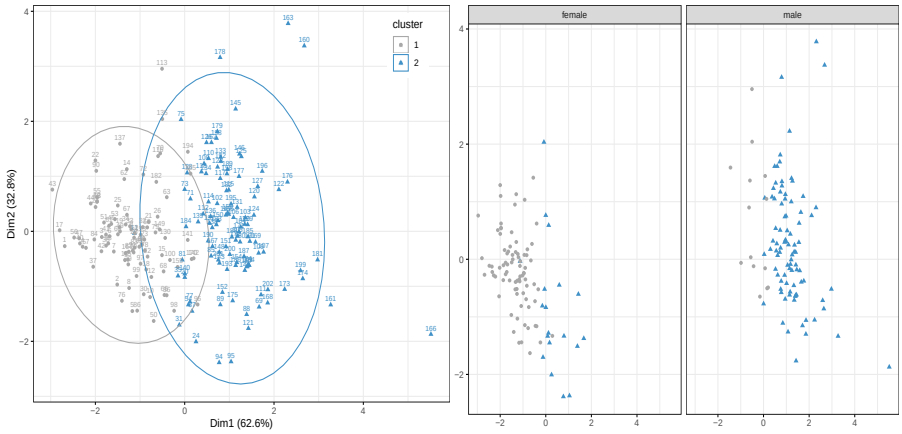


Fig. 9: K-means clustering with two groups derived from the *MaxSkew* features: scatter plot based on the first two principal components (% of variance), with the concentration ellipses at 0.95 level (left), and grouped by sex (right).

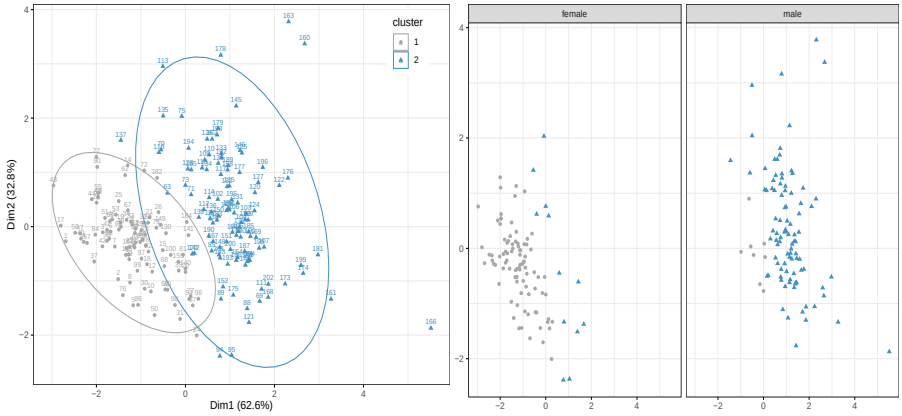


Fig. 10: Two component model based clustering derived from the *MaxSkew* features: scatter plot based on the first two principal components (% of variance), with the concentration ellipses at 0.95 level (left), and grouped by sex (right).

It is worthwhile noting that if the condition on the moments of the mixing variable is not satisfied then we cannot ensure that the shape vector lies on the direction maximizing skewness: This is the case for the skew-contaminated normal distribution when $0 \leq \epsilon < \pi/4$ and $\gamma^{-1/2} \geq x_1$; we also guess a similar situation comes up for the multivariate skewed exponential power distribution. However, in such cases we cannot assert that the shape vector fails at defining

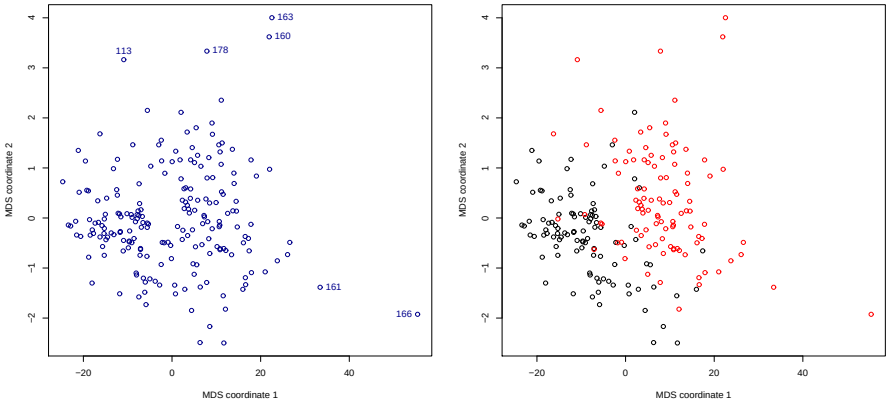


Fig. 11: Multidimensional scaling obtained from the *MaxSkew* measures (left). Sex label superimposed on the multidimensional scaling projections obtained from the *MaxSkew* features with the red color corresponding to males and the black to females (right).

the projection maximizing skewness since the moment inequality established by Theorem 1 defines a sufficient condition. Anyway, the condition is met by the most popular multivariate distributions within the SMSN family; they include the skew-normal, skew-t, skew double exponential and skew-slash distributions. The paper contributes to the field by extending previous work for the skew-normal and extended skew-normal families (Loperfido 2010; Franceschini and Loperfido 2014). In fact, it would allow to pursue research in the skewness based projection pursuit problem both from the theoretical and applied viewpoints under the general SMSN family.

The extension of the results established in this paper opens the road for future investigation: The SNSM family models skewness and kurtosis simultaneously; on the other hand, a family like the generalized skew-normal distributions (Loperfido 2004) models multimodality, although their skewness and kurtosis ranges are limited; hence, scales mixtures of generalized skew-normal distributions may define a broader and flexible family for modeling skewness, kurtosis and multimodality; we wonder if the results established in this paper can be generalized to this wider family. Another open problem would consist of looking into the connection between model based projection pursuit under SMSN vectors and multivariate kurtosis and skewness stochastic orderings, as established by previews works on the issue, along the lines pursued by Wang (2009) or by Arevalillo and Navarro (2012; 2019).

The theoretical findings established by this work have been illustrated through a simulation experiment that includes the skew-t, skew-slash and skew-contaminated normal distribution as the multivariate underlying non-normal model. The results have shown the accuracy of the *MaxSkew* method by

Franceschini and Loperfido (2016) to compute maximal skewness data projections. On the other hand, the simulations have also shown that the performance of skewness-based projection pursuit deteriorates as the number of variables increases; this result supports the theoretical findings in Bickel et al (2018) and might be motivating sparse projection pursuit. Finally, when applied to gene expression measures from a genomic cancer experiment, the *MaxSkew* method has proved its usefulness at providing a strategy for multivariate data reduction in genomic data, which may also motivate an alternative approach to define metagenes from biologically well-established functional groups of genes. These findings may highlight relevant insights for biologists and oncologists.

Acknowledgements The authors are grateful to two anonymous reviewers and the associate editor for helpful comments that improved the paper.

References

Arevalillo JM, Navarro H (2012) A study of the effect of kurtosis on discriminant analysis under elliptical populations. Journal of Multivariate Analysis 107:53 – 63

Arevalillo JM, Navarro H (2015) A note on the direction maximizing skewness in multivariate skew-t vectors. Statistics & Probability Letters 96:328–332

Arevalillo JM, Navarro H (2019) A stochastic ordering based on the canonical transformation of skew-normal vectors. TEST 28(2):475–498

Azzalini A (2005) The skew-normal distribution and related multivariate families. Scandinavian Journal of Statistics 32(2):159–188

Azzalini A, Capitanio A (1999) Statistical applications of the multivariate skew normal distribution. Journal of the Royal Statistical Society Series B 61(3):579–602

Azzalini A, Capitanio A (2003) Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. Journal of the Royal Statistical Society Series B 65(2):367–389

Azzalini A, Capitanio A (2014) The Skew-Normal and Related Families. IMS monographs. Cambridge University Press

Azzalini A, Dalla Valle A (1996) The multivariate skew-normal distribution. Biometrika 83(4):715–726

Balakrishnan N, Scarpa B (2012) Multivariate measures of skewness for the skew-normal distribution. Journal of Multivariate Analysis 104(1):73–87

Balakrishnan N, Capitanio A, Scarpa B (2014) A test for multivariate skew-normality based on its canonical form. Journal of Multivariate Analysis 128:19–32

Bickel PJ, Kur G, Nadler B (2018) Projection pursuit in high dimensions. Proceedings of the National Academy of Sciences 115(37):9151–9156

Branco MD, Dey DK (2001) A general class of multivariate skew-elliptical distributions. Journal of Multivariate Analysis 79(1):99 – 113

Capitanio A (2012) On the canonical form of scale mixtures of skew-normal distributions. arXiv/12070797

Capitanio A, Azzalini A, Stanghellini E (2003) Graphical models for skew-normal variates. Scandinavian Journal of Statistics 30(1):129–144

Caussinus H, Ruiz-Gazen A (2010) Exploratory Projection Pursuit, John Wiley & Sons, Ltd, chap 3, pp 67–92

Chow CK, Liu CN (1968) Approximating discrete probability distributions with dependence trees. IEEE Transactions on Information Theory 14(3):462–467

Contreras-Reyes JE, Arellano-Valle RB (2012) Kullback-Leibler divergence measure for multivariate skew-normal distributions. Entropy 14(9):1606–1626

Cook D, Buja A, Cabrera J (1993) Projection pursuit indexes based on orthonormal function expansions. Journal of Computational and Graphical Statistics 2(3):225–250

Cook R, Weisberg S (2009) An Introduction to Regression Graphics. Wiley Series in Probability and Statistics, Wiley

Edwards D, de Abreu GC, Labouriau R (2010) Selecting high-dimensional mixed graphical models using minimal AIC or BIC forests. BMC Bioinformatics 11(1):18

Franceschini C, Loperfido N (2014) Testing for Normality When the Sampled Distribution Is Extended Skew-Normal, Springer International Publishing, pp 159–169

Franceschini C, Loperfido N (2016) MaxSkew: Orthogonal Data Projections with Maximal Skewness. URL https://CRAN.R-project.org/package=MaxSkew, R package version 1.0

Friedman JH (1987) Exploratory projection pursuit. Journal of the American Statistical Association 82(397):249–266

Friedman JH, Tukey JW (1974) A projection pursuit algorithm for exploratory data analysis. IEEE Trans Comput 23(9):881–890

Gamez-Pozo A, Berges-Soria J, Arevalillo JM, Nanni P, Lopez-Vacas R, Navarro H, Grossmann J, Castaneda CA, Main P, Diaz-Almiron M, Espinosa E, Ciruelos E, Vara JAF (2015) Combined label-free quantitative proteomics and microRNA expression analysis of breast cancer unravel molecular differences with clinical implications. Cancer Research 75(11):2243–2253

Gómez-Sánchez-Manzano E, Gómez-Villegas M, Marín J (2008) Multivariate exponential power distributions as mixtures of normal distributions with bayesian applications. Communications in Statistics - Theory and Methods 37(6):972–985

Huber PJ (1985) Projection pursuit. The Annals of Statistics 13(2):435–475

Jones MC, Sibson R (1987) What is projection pursuit? Journal of the Royal Statistical Society Series A (General) 150(1):1–37

Kim HM (2008) A note on scale mixtures of skew normal distribution. Statistics & Probability Letters 78(13):1694 – 1701

Kim HM, Kim C (2017) Moments of scale mixtures of skew-normal distributions and their quadratic forms. Communications in Statistics - Theory and

Methods 46(3):1117–1126
Lachos VH, Ghosh P, Arellano-Valle RB (2010a) Likelihood based inference for skew-normal independent linear mixed models. Statistica Sinica 20(1):303–322
Lachos VH, Labra FV, Bolfarine H, Ghosh P (2010b) Multivariate measurement error models based on scale mixtures of the skew-normal distribution. Statistics 44(6):541–556
Lee SX, McLachlan GJ (2016) Finite mixtures of canonical fundamental skew t-distributions: the unification of the restricted and unrestricted skew t-mixture models. Statistics and Computing 26:573–589
Lin TI (2010) Robust mixture modeling using multivariate skew t distributions. Statistics and Computing 20:343–356
Lin TI, Ho HJ, Lee CR (2014) Flexible mixture modelling using the multivariate skew-t-normal distribution. Statistics and Computing 24:531–546
Loperfido N (2004) Generalized Skew-Normal Distributions, CRC/Chapman & Hall, chap 4, pp 65–80
Loperfido N (2010) Canonical transformations of skew-normal variates. TEST 19(1):146–165
Loperfido N (2018) Skewness-based projection pursuit: A computational approach. Computational Statistics & Data Analysis 120:42–57
Loperfido N (2019) Finite mixtures, projection pursuit and tensor rank: a triangulation. Advances in Data Analysis and Classification 13(1):145–173
Malkovich JF, Afifi AA (1973) On tests for multivariate normality. Journal of the American Statistical Association 68(341):176–179
Merkle M (1998) Conditions for convexity of a derivative and some applications to the gamma function. Aequationes Mathematicae 55(3):273–280
Prado-Vázquez G, Gámez-Pozo A, Trilla-Fuertes L, Arevalillo JM, Zapater-Moros A, Ferrer-Gómez M, Díaz-Almirón M, López-Vacas R, Navarro H, Maín P, Feliú J, Zamora P, Espinosa E, Fresno Vara JÁ (2019) A novel approach to triple-negative breast cancer molecular classification reveals a luminal immune-positive subgroup with good prognoses. Scientific Reports 9(1):1538
Rody A, Karn T, Liedtke C, Pusztai L, Ruckhaeberle E, Hanker L, Gaetje R, Solbach C, Ahr A, Metzler D, Schmidt M, Müller V, Holtrich U, Kaufmann M (2011) A clinically relevant gene signature in triple negative and basal-like breast cancer. Breast Cancer Research 13(5):R97
Wang J (2009) A family of kurtosis orderings for multivariate distributions. Journal of Multivariate Analysis 100(3):509–517
Wang J, Genton MG (2006) The multivariate skew-slash distribution. Journal of Statistical Planning and Inference 136(1):209–220
Zapater-Moros A, Gámez-Pozo A, Prado-Vázquez G, Trilla-Fuertes L, Arevalillo JM, Díaz-Almirón M, Navarro H, Maín P, Feliú J, Zamora P, Espinosa E, Fresno Vara JÁ (2018) Probabilistic graphical models relate immune status with response to neoadjuvant chemotherapy in breast cancer. Oncotarget 9(45):27,586–27,594