This version of the article has been accepted for publication. The version of record of this article, published in TEST, is available at Publisher's website: https://doi.org/10.1007/s11749-018-0583-5

A stochastic ordering based on the canonical transformation of skew-normal vectors

Jorge M. Arevalillo · **Hilario Navarro**

Abstract In this paper we define a new skewness ordering that enables stochastic comparisons for vectors that follow a multivariate skew-normal distribution. The new ordering is based on the canonical transformation associated with the multivariate skew-normal distribution and on the well-known convex transform order applied to the only skewed component of such canonical transformation. We examine the connection between the proposed ordering and the multivariate convex transform order studied by Belzunce et al (2015). Several standard skewness measures like Mardia's and Malkovich-Afifi's indices are revisited and interpreted in connection with the new ordering; we also study its relationship with the J-divergence between skew-normal and normal random vectors and with the *Negentropy*. Some artificial data are used in simulation experiments to illustrate the theoretical discussion; a real data application is provided as well.

Keywords Skew-normal distribution · Canonical transformation · Convex transform order

Mathematics Subject Classification (2000) 60E05 · 62H05

1 Introduction

Although the normal distribution has appealing properties for modeling multivariate observations, it has practical limitations since many of the data collected in real life applications do not fit to the normal model. Thus, more

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

flexible models that account for the non normality of the data have been developed in order to get rid of the normality corset; an alternative that has become increasingly popular for handling departures from normality is the multivariate skew-normal (SN) distribution defined in Azzalini and Dalla Valle (1996). The SN model has received attention in many applications where the collected data at hand exhibit asymmetry departures from normality; they include finance and risk management (Adcock 2004; Vernic 2006; Adcock et al 2015; Taniguchi et al 2015), genetics (Hardin and Wilson 2009), remote sensing (Zadkarami and Rowhani 2010), and psychiatry and psychology (Counsell et al 2011), just to name a few representative instances.

In this work we revisit the main properties concerned with skewness for the family of multivariate SN distributions; we examine the concept in a multivariate fashion by introducing a new ordering that allows to deal with multivariate stochastic skewness comparisons within this family. The new stochastic ordering is defined by application of the well-known Van Zwet's convex transform ordering (Van Zwet 1964) to the only skewed variate obtained from the canonical transformation of the SN vector (Loperfido 2010). The idea of using Van Zwet's convex transform ordering on scalar variables to define a multivariate stochastic ordering is not new; it has been utilized previously to establish a kurtosis ordering for the family of multivariate elliptically contoured distributions (Wang 2009; Arevalillo and Navarro 2012). In this paper we adopt the general idea and bring it to the context of asymmetric multivariate models like the SN family.

Once the stochastic ordering has been introduced, some standard measures of multivariate skewness like Mardia's and Malkovich-Afifi's indices (Mardia 1970; Malkovich and Afifi 1973) are revisited in order to show their role as indicators compatible with the proposed ordering; we also study how the new order is related to the J-divergence measure between SN and normal vectors and to the *Negentropy*, as a non-normality measure.

The paper is organized as follows: in the next section we review some theoretical background regarding the convex transform stochastic ordering between scalar variables as well as the underpinnings of the multivariate SN distribution, with an emphasis on specific issues that will be used along the paper. Section 3 addresses the main results of the paper: a new stochastic ordering is proposed to enable skewness comparisons between vectors that follow a multivariate SN distribution; in this section we also deal with the relationship between the new ordering and some standard skewness indices. Section 4 contains two simulation experiments that shed light on the theoretical discussion and an application to real data. We finish the paper giving some concluding remarks and suggestions for future research advances.

2 Background

2.1 Some background on skewness and the convex transform ordering

Skewness arises in probability theory as a concept to assess departures from symmetry. An approach for measuring skewness is the use of indices that quantify it; a great deal of ongoing research with proposals for new skewness indices has emerged since the pioneer work by Pearson (1895). In spite of the great deal of measures that allow to quantify the "departure from symmetry" through indices, this is a rather complex concept that needs complex stochastic ideas to become quantified. Thus, a more flexible alternative, based on the stochastic comparison of distributions, has been proposed in order to describe the concept in a more precise manner (Van Zwet 1964; Oja 1981; MacGillivray 1986); so the larger departures should correspond to the more skewed distributions in the stochastic order.

A skewness stochastic ordering between scalar variables can be defined in a rather intuitive way on the basis of the following simple argument: given a random variable X with distribution function F, we can inject skewness into X by application of an increasing convex transformation $Y = \psi(X)$; this skewing mechanism leads to a more skewed variable Y. Actually, the transformation that maps X on Y is given by $\psi = G^{-1}F$, where G denotes the distribution function of Y (Marshall and Olkin 2007); whenever $G^{-1}F$ is convex we can guarantee that F is less skewed to the right than G. This handy idea has inspired the following definition for a skewness stochastic order (Van Zwet 1964; Oja 1981).

Definition 1 (Convex transform order) Let X and Y be two random variables with distribution functions F and G respectively. Then we say that F is less skewed to the right than G, and denote it by $F \leq_c G$ or by $X \leq_c Y$, if $G^{-1}(F(x))$ is a convex function for all x belonging to the support of F.

According to Oja (1981), any index $S(\cdot)$ should be compatible with the previous ordering in order to be considered a measure of skewness. It should meet the following properties:

1. $S(F) = 0$ for any symmetric distribution F.
2. $S(a + bF) = sign(b)S(F)$ for every $a, b \in \mathbb{R}$.
3. If F and G are two distribution functions such that $F \leq_c G$ then $S(F) \leq S(G)$.

The first two properties are natural to the intuition of skewness; the third one states that $S(\cdot)$ must be compatible with the $\leq_c$ ordering. Some indicators like

$$E\left[\frac{X - E(X)}{\sqrt{Var(X)}}\right]^{2n+1} : n = 1, 2, \ldots,$$

which can be defined as long as the expectations exist, or the measure defined in Arnold and Groeneveld (1995) meet these properties. Some others, like Pearson measure of skewness

$$\frac{E(X) - Md(X)}{\sqrt{Var(X)}},$$

are not compatible with the convex transform skewness ordering.

Recently, the convex transform order for scalar variables has been extended to the multivariate setting in order to enable stochastic comparisons between random vectors (Belzunce et al 2015). The rationale behind this extension can be summarized as follows.

Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be p-dimensional vectors having distribution functions $\boldsymbol{F}$ and $\boldsymbol{G}$ and let us consider the vectors $\boldsymbol{x} = (x_1, \ldots, x_p)$ and $\boldsymbol{u} = (u_1, \ldots, u_p) \in [0.1]^p$. The approach combines the multivariate quantile transformation, defined by $\boldsymbol{Q_Y}(\boldsymbol{u}) = \big(\boldsymbol{Q}_{\boldsymbol{Y},1}(u_1), \ldots, \boldsymbol{Q}_{\boldsymbol{Y},p}(\boldsymbol{u})\big)$ where $\boldsymbol{Q}_{\boldsymbol{Y},1}(u_1) = G^{-1}_{Y_1}(u_1)$ and $\boldsymbol{Q}_{\boldsymbol{Y},k}(u_1, \ldots, u_k) = G^{-1}_{\left(Y_k | \bigcap_{j=1}^{k-1} Y_j = \boldsymbol{Q}_{\boldsymbol{Y},j}(u_1,\ldots,u_j)\right)}(u_k)$ for $2 \leq k \leq p$, and the multivariate distributional transformation, which is given by $\boldsymbol{D_X}(\boldsymbol{x}) = (\boldsymbol{D}_{\boldsymbol{X},1}(x_1), \ldots, \boldsymbol{D}_{\boldsymbol{X},p}(\boldsymbol{x}))$ with $\boldsymbol{D}_{\boldsymbol{X},1}(x_1) = F_{X_1}(x_1)$ and $\boldsymbol{D}_{\boldsymbol{X},k}(x_1, \ldots, x_k) = F_{\left(X_k | \bigcap_{j=1}^{k-1} X_j = x_j\right)}(x_k)$ for $2 \leq k \leq p$. The multivariate convex transform order is defined as follows.

Definition 2 (Multivariate convex transform order) We say that vector $\boldsymbol{X}$ is less skewed than vector $\boldsymbol{Y}$, and we denote it by $\boldsymbol{X} \leq_{mc} \boldsymbol{Y}$, if and only if all the components $\boldsymbol{\Phi}_i : i = 1, 2, \ldots, p$ of the transformation $\boldsymbol{\Phi}(\boldsymbol{x}) = \boldsymbol{Q_Y}(\boldsymbol{D_X}(\boldsymbol{x}))$ are convex functions in their support.

Note that when $p = 1$ the multivariate order reduces to the univariate convex transform order. On the other hand, if the components of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent then the multivariate convex transform order is equivalent to the scalar convex transform order between their components. This is an interesting fact which reveals enlightening implications for the SN distribution, as any SN vector can be transformed into a new one of independent components through its canonical transformation (Azzalini and Capitanio 1999; Loperfido 2010; Capitanio 2012). This issue is addressed in detail in the next section.

2.2 Some background on the skew-normal distribution

The multivariate SN distribution was introduced by Azzalini and Dalla Valle (1996) to regulate departures from normality; it has become a widely used model to deal with skewness in multivariate data. The study of its theoretical properties and applications have originated a fruitful research topic (Azzalini and Capitanio 1999; Capitanio et al 2003; Azzalini 2005; Contreras-Reyes and Arellano-Valle 2012; Balakrishnan and Scarpa 2012; Balakrishnan et al 2014).

2.2.1 Definition

Here 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 variate with location vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_p)'$ and scale matrix $\boldsymbol{\Omega}$ as follows:

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

In the expression above $\phi_p(\cdot; \boldsymbol{\Omega})$ denotes the normal density function of a p-dimensional normal variable 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 $\bar{\boldsymbol{\Omega}} = \boldsymbol{\omega}^{-1}\boldsymbol{\Omega}\boldsymbol{\omega}^{-1}$ is a correlation matrix and $\boldsymbol{\alpha}$ is a p-dimensional shape vector. We will write $\boldsymbol{X} \sim SN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ to denote that $\boldsymbol{X}$ follows a p-dimensional SN distribution with density function (1).

The shape vector $\boldsymbol{\alpha}$ regulates the asymmetry of the model in a directional fashion; it describes a perturbation of the normal model in which skewness is injected across a direction. Note that when $\boldsymbol{\alpha} = \boldsymbol{0}$ we come back to the multivariate normal distribution with mean $\boldsymbol{\xi}$ and covariance matrix $\boldsymbol{\Omega}$.

As an illustration we can see in Figure 1 the plots of the density function for the normal and three SN distributions with different shape vectors; the contour plots for each density are also depicted. We can observe how they become deformed as the asymmetry is injected into the normal model across different directions; the deformation depends of the amount of asymmetry injected as we can see by simple comparison of Figures 1b and 1c, for which $||\alpha|| = \sqrt{2}$ and $||\alpha|| = 2$ respectively.

We can observe that $\boldsymbol{X}$ can be rewritten as $\boldsymbol{X} = \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{Z}$, where $\boldsymbol{Z}$ is a SN vector with density function

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

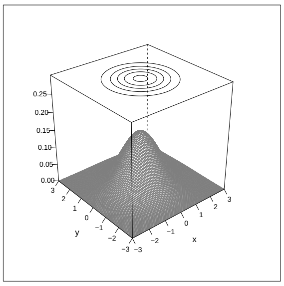
The expression in equation (2) gives the density function of a “normalized” version of the multivariate SN variable $\boldsymbol{X}$ (Azzalini and Capitanio 2014); we will use the normalized version of $\boldsymbol{X}$ to define the new stochastic ordering between SN vectors.

2.2.2 The canonical transformation of the SN vector

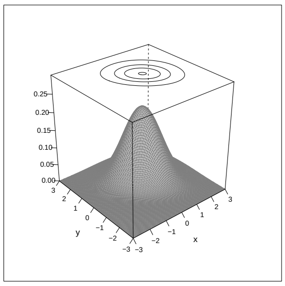
An interesting property of the SN model is concerned with the existence of a linear combination of the components of $\boldsymbol{X}$ that captures all the asymmetry of the model. This property is stated by the following proposition.

Proposition 1 (Azzalini and Capitanio (1999)) *Let $\boldsymbol{X}$ be a vector such that $\boldsymbol{X} \sim SN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$, with $\boldsymbol{Z} = \boldsymbol{\omega}^{-1}(\boldsymbol{X} - \boldsymbol{\xi})$ its normalized vector. There exist a linear transform $\boldsymbol{U} = \boldsymbol{A}^*\boldsymbol{Z}$ such that $\boldsymbol{U} \sim SN_p(\boldsymbol{I}_p, \boldsymbol{\alpha}^*)$, where $\boldsymbol{I}_p$ is the p-dimensional identity matrix and $\boldsymbol{\alpha}^*$ is a skewness vector with at most one non-zero component. The density function of $\boldsymbol{U}$ is given by*

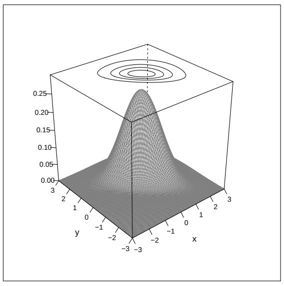
$$f(\boldsymbol{u}; \boldsymbol{\alpha}^*, \boldsymbol{I}_p) = 2\prod_{i=1}^{p} \phi(u_i)\Phi(\alpha_1^* u_1), \tag{3}$$



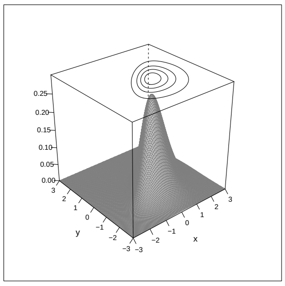
(a) $\boldsymbol{\alpha}' = (0, 0)$



(b) $\boldsymbol{\alpha}' = (1, 1)$



(c) $\boldsymbol{\alpha}' = (2, 2)$



(d) $\boldsymbol{\alpha}' = (2, -2)$

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

where $\alpha_1^* = (\boldsymbol{\alpha}'\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha})^{1/2}$ *is the only non-zero component of* $\boldsymbol{\alpha}^*$ *and* ϕ *denotes the density function of a* $N(0, 1)$ *random variable.*

The matrix $\boldsymbol{A}^*$ is given by $\boldsymbol{A}^* = (\boldsymbol{C}^{-1}\boldsymbol{P})'$ where $\boldsymbol{C}$ is a matrix verifying that $\boldsymbol{C}'\boldsymbol{C} = \bar{\boldsymbol{\Omega}}$, and $\boldsymbol{P}$ is an orthogonal matrix whose first column lies in the direction of $\boldsymbol{C}\boldsymbol{\alpha}$ (Azzalini and Capitanio 1999). The transformed vector $\boldsymbol{U}$ is known as the canonical transformation of the SN model; it can be found using the constructive procedure introduced by Capitanio (2012) as follows.

Proposition 2 (Capitanio (2012)) *Let* $\boldsymbol{X}$ *be a* p*-dimensional vector such that* $\boldsymbol{X} \sim SN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ *and let us define* $\boldsymbol{M} = \boldsymbol{\Omega}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Omega}^{-1/2}$, *with* $\boldsymbol{\Omega}^{1/2}$ *the unique square root of* $\boldsymbol{\Omega}$ *and* $\boldsymbol{\Sigma}$ *the covariance matrix of* $\boldsymbol{X}$. *If* $\boldsymbol{Q}$ *and* $\boldsymbol{\Lambda}$ *denote the orthogonal and diagonal matrices that provide the spectral decomposition of* $\boldsymbol{M}$ *then the transformation* $\boldsymbol{U} = \boldsymbol{H}'(\boldsymbol{X} - \boldsymbol{\xi})$, *with* $\boldsymbol{H} = \boldsymbol{\Omega}^{-1/2}\boldsymbol{Q}$, *gives the canonical form of* $\boldsymbol{X}$.

The canonical transformation of Proposition 2 maps the original vector $\boldsymbol{X}$ into a vector $\boldsymbol{U}$ whose components are mutually independent scalar variables with one variable, say the first one U_1, absorbing all the skewness of $\boldsymbol{X}$ into the quantity α_1^*. Hence $\boldsymbol{U}$ can be interpreted as a representation of the SN vector that isolates the multivariate asymmetry into a single component. Moreover, by trivial implication, we see that the linear combination $U_1 = \boldsymbol{a}_1' \boldsymbol{Z}$, where $\boldsymbol{a}_1'$ denotes the first row of the matrix $\boldsymbol{H}$, is a random variable such that $U_1 \sim SN_1(0, 1, \alpha_1^*)$.

The implications of the canonical form are illustrated in Figure 2 for a vector $\boldsymbol{X} \sim SN_2(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ with $\boldsymbol{\xi} = (0, 0)$, $\boldsymbol{\alpha}' = (3, 3)$ and $\boldsymbol{\Omega} = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$. The contour plots of the original SN vector and its canonical form show how all the multivariate skewness gets summarized into the skewness of the first canonical variate, whereas the second canonical variate follows a standard normal distribution.

The canonical representation has highlighted an appealing property of the SN family, which has implications in model-based projection pursuit: it turns out that the projection yielding the maximum skewness lies on the direction of vector $\boldsymbol{\omega}^{-1}\boldsymbol{\alpha}$ (Loperfido 2010), which is also proportional to the vector $\boldsymbol{a}_1$ that yields the skewed component of the canonical form (Capitanio 2012). This finding and its connection with Mardia's and Malkovich-Afifi's skewness has allowed to derive their analytical expressions for the family of SN distributions (Loperfido 2010). Some related work pointing out to the same issue for the multivariate skew-t family, a model that incorporates both skewness and kurtosis parameters to handle non-normality, has been launched more recently by Arevalillo and Navarro (2015).

3 Main contribution

The canonical representation of Proposition 1 has received increasingly attention in the literature due to its appealing theoretical and applied implications (Loperfido 2010; Capitanio 2012; Balakrishnan et al 2014; Arevalillo and Navarro 2015). In this paper we use it to propose a new skewness stochastic ordering between SN vectors.

3.1 A new skewness ordering for SN vectors

A new skewness stochastic ordering is proposed on the basis of the following simple idea: the result established by Proposition 1 is utilized to define an ordering between SN vectors through the convex transform ordering (Van Zwet 1964) of the skewed components in their canonical representations; this idea is quite natural and obeys to the fact that all the skewness of the SN vector is absorbed by a single component of the canonical transformation.

The next statement formalizes the definition of the skewness stochastic ordering for vectors that follow a SN distribution.

Definition 3 (Canonical convex transform order) Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two p-dimensional random vectors, with distribution functions $F_{\boldsymbol{X}}$ and $F_{\boldsymbol{Y}}$, such that $\boldsymbol{X} \sim SN_p(\boldsymbol{\xi_X}, \boldsymbol{\Omega_X}, \boldsymbol{\alpha_X})$ and $\boldsymbol{Y} \sim SN_p(\boldsymbol{\xi_Y}, \boldsymbol{\Omega_Y}, \boldsymbol{\alpha_Y})$ with normalized random vectors: $\boldsymbol{Z_X} = \boldsymbol{\omega_X}^{-1}(\boldsymbol{X} - \boldsymbol{\xi_X})$ and $\boldsymbol{Z_Y} = \boldsymbol{\omega_Y}^{-1}(\boldsymbol{Y} - \boldsymbol{\xi_Y})$. Let us denote

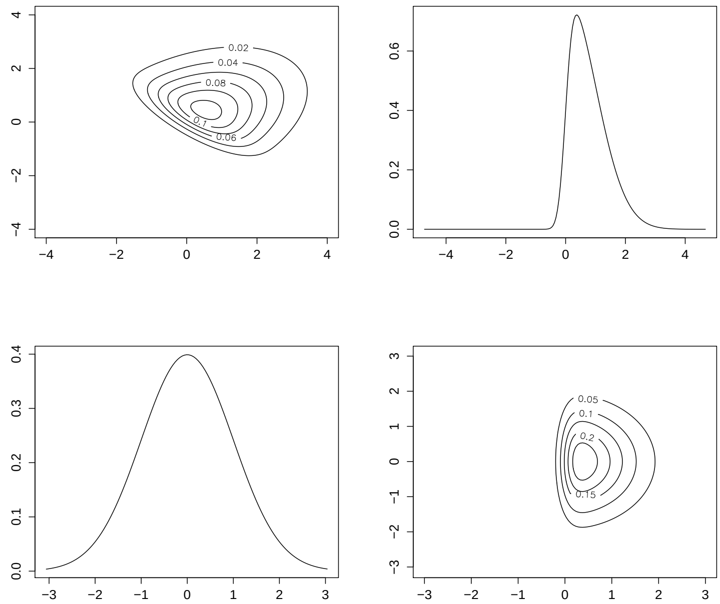


Fig. 2: Contour plots for the bivariate SN distribution (top left) and for the canonical transform (bottom right), as well as density functions for the first (top right) and second (bottom left) canonical variates.

by $U_{1\boldsymbol{X}} = \boldsymbol{a}'_{1\boldsymbol{X}}\boldsymbol{Z}_{\boldsymbol{X}}$ and $U_{1\boldsymbol{Y}} = \boldsymbol{a}'_{1\boldsymbol{Y}}\boldsymbol{Z}_{\boldsymbol{Y}}$ the skewed components of the canonical representations (3) of $\boldsymbol{Z}_{\boldsymbol{X}}$ and $\boldsymbol{Z}_{\boldsymbol{Y}}$, having distribution functions $F_{U_{1\boldsymbol{X}}}$ and $F_{U_{1\boldsymbol{Y}}}$. We say that $F_{\boldsymbol{X}}$ is less skewed to the right than $F_{\boldsymbol{Y}}$ in the canonical convex transform order, and we denote it by $F_{\boldsymbol{X}} \leq_{cc} F_{\boldsymbol{Y}}$ or by $\boldsymbol{X} \leq_{cc} \boldsymbol{Y}$, when $F_{U_{1\boldsymbol{X}}} \leq_c F_{U_{1\boldsymbol{Y}}}$.

Figure 3 displays the contour plots for bidimensional SN vectors with common location $\boldsymbol{\xi} = (0, 0)$ and scale matrix $\boldsymbol{\Omega} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$; their shape vectors are given by $\boldsymbol{\alpha}' = (0, 0)$, $\boldsymbol{\alpha}' = (-1, 1)$ and $\boldsymbol{\alpha}' = (1, 1)$ respectively. It is worthwhile noting that the distribution functions of their canonical variates are stochastically dominated with respect to the slant parameter $\alpha_1^* = (\boldsymbol{\alpha}'\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha})^{1/2}$ of their scalar SN canonical variates. This is not an unexpected phenomenon; it arises within the general framework of distributions generated by a perturbation scheme of a symmetric density function when certain mild conditions are fulfilled for the perturbation (Azzalini and Regoli 2012; Propositions 4-6). Although the dominance between distribution functions provides a simple way to introduce a stochastic ordering, our approach focuses on skewness comparisons which resort to the convex transform order of Definition 3.

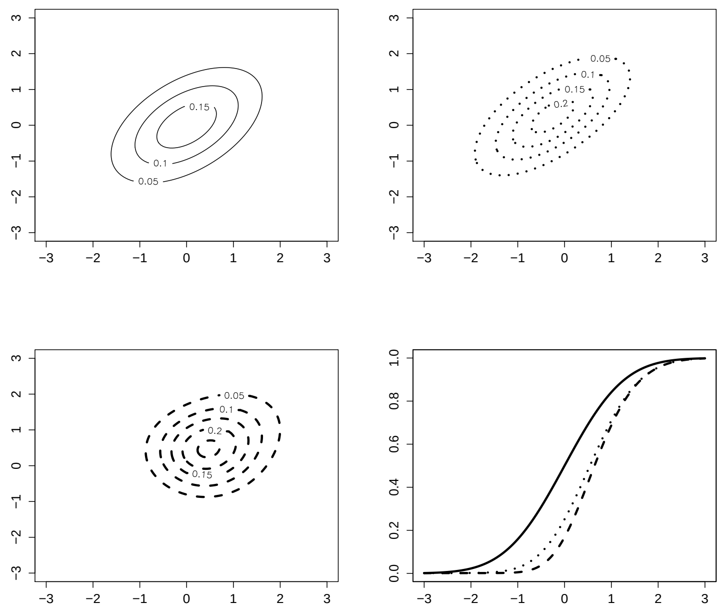


Fig. 3: Contour plots for three bidimensional SN distributions with shape parameters $\boldsymbol{\alpha}' = (0, 0)$ (solid line), $\boldsymbol{\alpha}' = (-1, 1)$ (dotted line) and $\boldsymbol{\alpha}' = (1, 1)$ (dashed line), along with the cumulative distribution functions of their first canonical variates (bottom right plots).

On the other hand, we can easily see the close connection between the canonical convex transform order $\leq_{cc}$ and the multivariate convex transform order $\leq_{mc}$ of the canonical forms derived from the SN vectors. The following proposition states it explicitly.

Proposition 3 *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two p-dimensional random vectors such that $\boldsymbol{X} \sim SN_p(\boldsymbol{\xi_X}, \boldsymbol{\Omega_X}, \boldsymbol{\alpha_X})$ and $\boldsymbol{Y} \sim SN_p(\boldsymbol{\xi_Y}, \boldsymbol{\Omega_Y}, \boldsymbol{\alpha_Y})$ with normalized random vectors $\boldsymbol{Z_X} = \boldsymbol{\omega_X}^{-1}(\boldsymbol{X} - \boldsymbol{\xi_X})$ and $\boldsymbol{Z_Y} = \boldsymbol{\omega_Y}^{-1}(\boldsymbol{Y} - \boldsymbol{\xi_Y})$. Let us denote by $\boldsymbol{U_X} = \boldsymbol{A}^*_{\boldsymbol{X}} \boldsymbol{Z_X}$ and $\boldsymbol{U_Y} = \boldsymbol{A}^*_{\boldsymbol{Y}} \boldsymbol{Z_Y}$ their respective canonical transformations. Then $\boldsymbol{X} \leq_{cc} \boldsymbol{Y}$ if and only if $\boldsymbol{U_X} \leq_{mc} \boldsymbol{U_Y}$.*

Proof.

It is straightforward from Definition 2, the independence of the components of vectors $\boldsymbol{U_X}$ and $\boldsymbol{U_Y}$, and the result of Proposition 1. □

Note that the multivariate convex transform order shown in Proposition 3 follows from the fact that both canonical forms share the copula of independence (Belzunce et al 2015). Also note that the result provides a characterization for the comparison of skewness of two SN vectors in terms of the *mc* order. Moreover, in accordance to Definition 3, a natural way to make sense of the idea “being an indicator of multivariate skewness” can be described as follows: since the skewed component of the canonical form absorbs all the asymmetry of the SN vector, it then stands to reason to say that a functional $S(\cdot)$ is an indicator of multivariate skewness if it is consistent with the canonical convex transform order. This fact can be formalized as follows.

Definition 4 (Indicators of multivariate skewness) Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be p-dimensional SN random vectors with distribution functions $F_{\boldsymbol{X}}$ and $F_{\boldsymbol{Y}}$. We say that a measure $S(\cdot)$ is an indicator of multivariate skewness when it fulfills that if $F_{\boldsymbol{X}} \leq_{cc} F_{\boldsymbol{Y}}$ then $S(F_{\boldsymbol{X}}) \leq S(F_{\boldsymbol{Y}})$.

For a $SN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ vector $\boldsymbol{X}$ it is straightforward to show that the quantity $\alpha_1^* = (\boldsymbol{\alpha}' \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha})^{1/2}$ is an indicator of multivariate skewness in the sense of Definition 4. This assertion can be proved using a simple argument: we know that α_1^* is the shape parameter of the only skewed component in the canonical form of $\boldsymbol{X}$, which is a $SN_1(0, 1, \alpha_1^*)$ scalar variable; it can be shown that α_1^* is a one to one function of its standardized third order central moment, which in turn preserves the convex transform order (Van Zwet 1964; Oja 1981; MacGillivray 1986).

We can also prove the reciprocal, which implies that the ordering $\leq_{cc}$ is a total multivariate stochastic order for the family of SN multivariate distributions. The following proposition is essential for the proof.

Proposition 4 *Let X and Y be two SN scalar variables having distribution functions F and G such that $X \sim SN_1(0, 1, \alpha_1)$ and $Y \sim SN_1(0, 1, \alpha_2)$ with $0 < \alpha_1 < \alpha_2$. Then it is satisfied that $F \leq_c G$.*

Proof.

Since F and G are the distribution functions of $SN_1(0, 1, \alpha_1)$ and $SN_1(0, 1, \alpha_2)$ scalar variables, we know that

$$F(x) = \Phi(x) - 2T(x, \alpha_1) \quad \text{and} \quad G(x) = \Phi(x) - 2T(x, \alpha_2),$$

where $T(x, \lambda) = \phi(x) \displaystyle\int_0^\lambda \frac{\phi(xz)}{1 + z^2}\, dz$ is the Owen's T function.

We must prove that $H(x) = G^{-1}(F(x))$ is a convex function. This is equivalent to prove the condition that, for any straight line $a + bx$, the function $H(x) - (a + bx)$ has at most two zeros and whenever it has two zeros it is negative between them. This condition is straightforward for $b \leq 0$ because H is a non decreasing function such that $\lim_{x \to -\infty} H(x) = -\infty$ and $\lim_{x \to \infty} H(x) = \infty$.

In order to prove the previous condition when $b > 0$, we will define the function $D(x) = \Phi(x) - 2T(x, \alpha_1) - \Phi(a + bx) + 2T(a + bx, \alpha_2)$ which has the same number of zeros as $H(x) - (a + bx)$. We know that

$$D'(x) = 2b\phi(a + bx)\Phi(\alpha_2(a + bx)) \left[e^{m(x)} - 1 \right],$$

with $m(x)$ the function defined in Lemma 1. Then the number of zeros of m will give the monotone behavior of D which in turn will allow to determine its sign pattern, just taking into account the limits:

$$\lim_{x\to-\infty} D(x) = \lim_{x\to\infty} D(x) = 0. \tag{4}$$

Now, we distinguish the following cases:

(i) $0 < b < \sqrt{\frac{\alpha_1^2+1}{\alpha_2^2+1}}$. Lemma 1 shows that $D'(x)$ has two zeros and sign pattern $-+-$; so function D has a decreasing - increasing - decreasing monotone behavior. In addition, taking into account that $0 < \alpha_1 < \alpha_2$ we can assert that $D(x) > 0$ when $x > \dfrac{a}{1-b}$. This fact, together with the monotone behavior of D and the limits in (4), will necessarily imply that D has only one zero and sign pattern $-+$.

(ii) $\sqrt{\frac{\alpha_1^2+1}{\alpha_2^2+1}} < b < 1$. Once again we can observe that $D(x) > 0$ when $x > \dfrac{a}{1-b}$. In this case we must distinguish two situations: firstly, if m has three zeros —see item (ii) in Lemma 1— D' will follow the sign pattern $+-+-$ and consequently D has an increasing - decreasing- increasing - decreasing monotone behavior which, taking into account (4), leads to either a non-negative D or to a $+-+$ sign pattern so that D has two zeros and it is negative between them. Secondly, in case m has only one zero —see once again item (ii) in Lemma 1— then D will follow and increasing - decreasing monotone behavior pattern and, taking into account (4), we obtain a non-negative D.

(iii) $b > 1$. In this case we can see that $D(x) > 0$ when $x < \dfrac{a}{1-b}$, and from item (iii) in the previous lemma, we can state that D has an increasing - decreasing - increasing pattern. This monotone behavior, the limits in (4) and the fact that $D(x) > 0$ when $x < \dfrac{a}{1-b}$ imply that D has one zero and sign pattern $+-$.

(iv) $b = 1, a > 0$. From item (iv) in Lemma 1 we can conclude that D has an increasing - decreasing - increasing monotone behavior which, together with the limits in (4), will imply that it has only one zero and follows the sign pattern $+-$.

(v) $b = 1, a < 0$. In this case, taking into account item (v) in Lemma 1, we see that D' has only one zero; therefore, D follows an increasing - decreasing pattern. This monotone behavior and (4) imply that D is non-negative.

(vi) $b = \sqrt{\frac{\alpha_1^2+1}{\alpha_2^2+1}}, a > 0$. Taking into account item (vi) of Lemma 1, we obtain a decreasing - increasing - decreasing monotone behavior pattern for the function D, with $D(x) > 0$ when $x > \dfrac{a}{1-\sqrt{\frac{\alpha_1^2+1}{\alpha_2^2+1}}}$. Both facts, along with the limits in (4), yield a function D with one zero and sign pattern $-+$.

(vii) $b=\sqrt{\frac{\alpha_1^2+1}{\alpha_2^2+1}}$, $a<0$. In this case, taking into account (vii) of Lemma 1 as well as the argument used to prove (ii) of this lemma, we can conclude that either D is non-negative or it has two zeros, being negative between them.

All the alternatives from (i) to (vii) have resulted in a function D with at most two zeros, and when it has two zeros D is negative between them. This fact implies the convexity of the function H and proves that $0<\alpha_1<\alpha_2$ is a sufficient condition for the convex ordering. □

The next theorem states that the canonical convex transform order is a total order for the family of multivariate SN distributions.

Theorem 1 *Let us consider $\boldsymbol{X}$ and $\boldsymbol{Y}$ vectors, with distribution functions $F_{\boldsymbol{X}}$ and $F_{\boldsymbol{Y}}$, such that $\boldsymbol{X}\sim SN_p(\boldsymbol{\xi_X},\boldsymbol{\Omega_X},\boldsymbol{\alpha_X})$ and $\boldsymbol{Y}\sim SN_p(\boldsymbol{\xi_Y},\boldsymbol{\Omega_Y},\boldsymbol{\alpha_Y})$, and let $\boldsymbol{Z_X}=\boldsymbol{\omega_X}^{-1}(\boldsymbol{X}-\boldsymbol{\xi_X})$ and $\boldsymbol{Z_Y}=\boldsymbol{\omega_Y}^{-1}(\boldsymbol{Y}-\boldsymbol{\xi_Y})$ be their standardized versions with $\bar{\boldsymbol{\Omega}}_{\boldsymbol{X}}=\boldsymbol{\omega_X}^{-1}\boldsymbol{\Omega_X}\boldsymbol{\omega_X}^{-1}$ and $\bar{\boldsymbol{\Omega}}_{\boldsymbol{Y}}=\boldsymbol{\omega_Y}^{-1}\boldsymbol{\Omega_Y}\boldsymbol{\omega_Y}^{-1}$ correlation matrices. Then it is satisfied that $F_{\boldsymbol{X}}\le_{cc}F_{\boldsymbol{Y}}$ if and only if $S(F_{\boldsymbol{X}})\le S(F_{\boldsymbol{Y}})$, where $S(F_{\boldsymbol{X}})=(\boldsymbol{\alpha}'_{\boldsymbol{X}}\bar{\boldsymbol{\Omega}}_{\boldsymbol{X}}\boldsymbol{\alpha_X})^{1/2}$ and $S(F_{\boldsymbol{Y}})=(\boldsymbol{\alpha}'_{\boldsymbol{Y}}\bar{\boldsymbol{\Omega}}_{\boldsymbol{Y}}\boldsymbol{\alpha_Y})^{1/2}$.*

Proof.

As we have just mentioned the implication $F_{\boldsymbol{X}}\le_{cc}F_{\boldsymbol{Y}}\Rightarrow S(F_{\boldsymbol{X}})\le S(F_{\boldsymbol{Y}})$ is straightforward. For the reciprocal implication, taking into account Propositions 1 and 4, we conclude that $F_{U_{1\boldsymbol{X}}}\le_c F_{U_{1\boldsymbol{Y}}}$, where F_{U_1} and F_{U_2} denote the distribution functions of the SN scalar variates of the corresponding canonical representations; consequently, $F_{\boldsymbol{X}}\le_{cc}F_{\boldsymbol{Y}}$ as we aimed to prove. □

Next, we are going to review several indices of non-normality in the SN model and will show how they can be interpreted as indicators of skewness in accordance to the result established by Theorem 1.

3.2 Non-normality measures and the canonical convex ordering

A classical skewness measure is Mardia's index (Mardia 1970). It is given by

$$\gamma_{1,p}^M=E[(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu})]^3, \tag{5}$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent and identically distributed random vectors with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

Another multivariate skewness measure capturing the directional nature of the concept was proposed by Malkovich and Afifi (1973). It is defined by

$$\gamma_{1,p}^D=\sup_{\boldsymbol{c}\in\mathbb{R}_0^p}\gamma_1(\boldsymbol{c}'\boldsymbol{X}), \tag{6}$$

where $\mathbb{R}_0^p$ is the set of all non-null p-dimensional vectors, with γ_1 the standard skewness index for scalar variables given by $\gamma_1(Y)=E^2\left(\frac{Y-\mu_Y}{\sigma_Y}\right)^3$.

We show that both indices, $\gamma_{1,p}^M$ and $\gamma_{1,p}^D$, are skewness measures compatible with the canonical convex ordering for SN vectors.

Corollary 1 *Let* $\boldsymbol{X}$ *be a vector such that* $\boldsymbol{X} \sim SN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$. *Then* $\gamma_{1,p}^M$ *and* $\gamma_{1,p}^D$ *are indicators of multivariate skewness in the sense of Definition 4.*

Proof.

The result for Mardia's measure is straightforward because

$$\gamma_{1,p}^M = 2(4-\pi)^2 \left[\frac{\boldsymbol{\alpha}' \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha}}{\pi + (\pi - 2)\boldsymbol{\alpha}' \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha}} \right]^3$$

agrees with the standard skewness index of the skewed scalar component in the canonical form of $\boldsymbol{X}$ (Azzalini and Capitanio 1999; Capitanio 2012). In order to prove the statement of the corollary for $\gamma_{1,p}^D$, it will suffice to recall that $\gamma_{1,p}^D = \gamma_{1,p}^M$ for the SN multivariate distribution (Loperfido 2010). □

Another way to assess departures from normality is resorting to the use of measures based on distances between distributions. A well- known distance between vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ is the J-divergence, which is defined by

$$J(\boldsymbol{X}, \boldsymbol{Y}) = D_{KL}(\boldsymbol{X}, \boldsymbol{Y}) + D_{KL}(\boldsymbol{Y}, \boldsymbol{X}), \tag{7}$$

where D_{KL} denotes the so called Kullback-Leibler divergence:

$$D_{KL}(\boldsymbol{X}, \boldsymbol{Y}) = \int_{\mathcal{X}} f(w) \ln \frac{f(w)}{g(w)} dw \tag{8}$$

with f and g the probability density functions of $\boldsymbol{X}$ and $\boldsymbol{Y}$.

The next statement shows how to it can be interpreted as an indicator of multivariate skewness compatible with the canonical convex ordering.

Corollary 2 *Let* $\boldsymbol{X}$ *and* $\boldsymbol{X}_0$ *be random vectors such that* $\boldsymbol{X} \sim SN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ *and* $\boldsymbol{X}_0 \sim N_p(\boldsymbol{\xi}, \boldsymbol{\Omega})$. *Then the J-divergence,* $J(\boldsymbol{X}, \boldsymbol{X}_0)$, *between* $\boldsymbol{X}$ *and* $\boldsymbol{X}_0$ *is an indicator of multivariate skewness in the sense of Definition 4.*

Proof.

From Contreras-Reyes and Arellano-Valle (2012) we know that

$$J(\boldsymbol{X}, \boldsymbol{X}_0) = E[\log \Phi(\alpha_1^* W)] - E[\log \Phi(\alpha_1^* W_0)],$$

where W is a scalar variable such that $W \sim SN_1(0, 1, \alpha_1^*)$ and W_0 is another scalar variable such that $W_0 \sim N(0, 1)$. Now, some simple calculus leads to

$$J(\boldsymbol{X}, \boldsymbol{X}_0) = \int_0^{\infty} [2\Phi(\alpha_1^* w) - 1] \log \left(\frac{1 - \Phi(\alpha_1^* w)}{\Phi(\alpha_1^* w)} \right) \phi(w)\, dw.$$

Taking into account that the quantity above is non decreasing with respect to α_1^* (Contreras-Reyes and Arellano-Valle 2012), we can conclude the result of the statement. □

Another measure that quantifies the non-normality of a multivariate random vector $\boldsymbol{X}$ is the *Negentropy* (Contreras-Reyes and Arellano-Valle 2012), which is defined by

$$H_N(\boldsymbol{X}) = H(\boldsymbol{X}_0) - H(\boldsymbol{X}), \tag{9}$$

with $H(\boldsymbol{X})$ the entropy of $\boldsymbol{X}$ and $H(\boldsymbol{X}_0)$ the entropy of a normal random vector having the same covariance matrix as $\boldsymbol{X}$.

In the next corollary we show that the *Negentropy* is a non-normality measure compatible with the proposed canonical convex order. The key issue for proving the corollary is found in writing the *Negentropy* H_N as a function of the slant parameter α_1^*.

Corollary 3 *Let $\boldsymbol{X}$ be a vector such that $\boldsymbol{X} \sim SN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$. The Negentropy $H_N(\boldsymbol{X})$ is also an indicator of multivariate skewness.*

Proof.

From Azzalini and Capitanio (1999; 2014) we know that the covariance matrix of $\boldsymbol{X}$ is given by $\boldsymbol{\Sigma} = \mathrm{var}(\boldsymbol{X}) = \boldsymbol{\Omega} - \frac{2}{\pi}\boldsymbol{\omega}\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\omega}$, where $\boldsymbol{\delta} = \frac{\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha}}{(1+\boldsymbol{\alpha}'\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha})^{1/2}}$. Then some simple calculus shows that $H(\boldsymbol{X}_0)$ in (9) is

$$H(\boldsymbol{X}_0) = \frac{1}{2}\log\{(2\pi e)^p|\boldsymbol{\Sigma}|\} = \frac{1}{2}\log\{(2\pi e)^p|\boldsymbol{\Omega}|\} + \frac{1}{2}\log\left(1 - \frac{2}{\pi}\frac{\alpha_1^{*2}}{1+\alpha_1^{*2}}\right).$$

On the other hand, using the expression of the density function (1) of the SN distribution, the entropy of vector $\boldsymbol{X}$ can be calculated as follows:

$$H(\boldsymbol{X}) = -E\{f(\boldsymbol{X};\boldsymbol{\xi},\boldsymbol{\alpha},\boldsymbol{\Omega})\} = \frac{1}{2}[\log\{(2\pi)^p|\boldsymbol{\Omega}|\} + E\{(\boldsymbol{X}-\boldsymbol{\xi})'\boldsymbol{\Omega}^{-1}(\boldsymbol{X}-\boldsymbol{\xi})\}]$$

$$-E\{\log\{2\Phi(Y)\}\} = \frac{1}{2}\log\{(2\pi e)^p|\boldsymbol{\Omega}|\} - E\{\log\{2\Phi(Y)\}\},$$

with $Y = \boldsymbol{\alpha}'\boldsymbol{\omega}^{-1}(\boldsymbol{X}-\boldsymbol{\xi})$ a scalar variable. Taking into account Azzalini and Capitanio (2014; Section 5.1.6) we can assert that $Y \sim SN_1(0, \alpha_1^{*2}, \alpha_1^*)$ or equivalently $Y = \alpha_1^* W$ with $W \sim SN_1(0, 1, \alpha_1^*)$; consequently

$$H(\boldsymbol{X}) = \frac{1}{2}\log\{(2\pi e)^p|\boldsymbol{\Omega}|\} - E\{\log\{2\Phi(\alpha_1^* W)\}\},$$

from which we get

$$H_N(\boldsymbol{X}) = H(\boldsymbol{X}_0) - H(\boldsymbol{X}) = \frac{1}{2}\log\left(1 - \frac{2}{\pi}\frac{\alpha_1^{*2}}{1+\alpha_1^{*2}}\right) + E\{\log\{2\Phi(\alpha_1^* W)\}\}.$$

This is a non decreasing function of α_1^* (see Figure 4 obtained by simulation), which implies the assertion of the statement. □

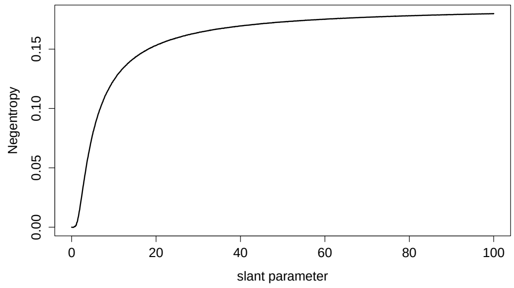


Fig. 4: Plot of $H_N(X)$ against the slant parameter for $\alpha_1^* \in [0, 100]$

4 Examples

Our findings have shown that the slant parameter $\alpha_1^{*2} = \boldsymbol{\alpha}'\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha}$ is an indicator that can be used to make stochastic skewness comparisons between vectors that follow a SN distribution. This section provides some examples, using both artificial and real data, that illustrate the theoretical results.

4.1 Simulation examples with artificial data

It is well-known that Mardia's and Malkovich-Afifi's measures agree in the SN model. Their relationship with the slant parameter $\alpha_1^{*2} = \boldsymbol{\alpha}'\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha}$, established in Corollary 1, points to two methods that would allow to carry out stochastic skewness comparisons: the first method estimates de maximal skewness projection and calculates Malkovich-Afifi's index accordingly. Alternatively, Mardia index (5) can be estimated and, once again, we can use it for stochastic comparisons of SN vectors. We call these procedures *MaxSkew* and *Mardia* methods. Both alternatives can be implemented using standard functions of the R packages MaxSkew (Franceschini and Loperfido 2016), which uses the computational approach of Loperfido (2013), and psych (Revelle 2016). The following examples with artificial data show the performance of both methods for two specific parameterizations of the SN model.

4.1.1 SN model with a Permutation Symmetric scale matrix

Let us consider a vector such that $\boldsymbol{X} \sim SN_p(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ with $\boldsymbol{\Omega}$ a Permutation Symmetric scale matrix given by $\boldsymbol{\Omega} = \bar{\boldsymbol{\Omega}} = (\omega_{i,j})_{1 \le i,j \le p}$, where $\omega_{i,j} = \omega_{j,i} = \rho$ for $1 \le i < j \le p$ and $\omega_{i,i} = 1$ for $1 \le i \le p$, with $\rho > -\frac{1}{p-1}$ in order to have a positive definite matrix. For simplicity we are taking $\rho > 0$.

The shape parameter $\boldsymbol{\alpha}$ leading to the most skewed SN multivariate distribution with respect to the ordering $\le_{cc}$ lies in the direction that maximizes the quantity $\alpha_1^{*2} = \boldsymbol{\alpha}'\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha}$. In this case it lies in the direction of the normalized eigenvector $\mathbf{e}=\frac{1}{\sqrt{p}}\mathbf{1}_{p\times 1}$, where $\mathbf{1}_{p\times 1}$ denotes a vector whose components are all equal to 1. Then the maximum for α_1^{*2} is given by the largest eigenvalue: $\mathbf{e}'\bar{\boldsymbol{\Omega}}\mathbf{e} = 1+(p-1)\rho$ (Main et al 2015), which in turn gives the following Mardia's and Malkovich-Afifi's skewness coefficients:

$$\gamma_{1,p}^D = \gamma_{1,p}^M = 2(4-\pi)^2\left[\frac{1+(p-1)\rho}{\pi+(\pi-2)(1+(p-1)\rho)}\right]^3. \qquad (10)$$

In order to compare *MaxSkew* and *Mardia* methods, we carry out the following simulation experiment: a total of 1000 samples are drawn from the a random vector $\boldsymbol{X} \sim SN_p(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{e})$ for sample sizes $n = 100, 250, 500$, dimensions $p = 2, 5, 10$ and $\rho = 0.2, 0.5, 0.9$. Then (10) is estimated by both procedures for each sample and the Mean Squared Error (MSE) is calculated over the one thousand samples. Note that, for our simulations, we have taken the most skewed vector among the family of SN distributions with a Permutation Symmetric scale matrix.

	$n = 100$			$n = 250$			$n = 500$		
p \ ρ	0.2	0.5	0.9	0.2	0.5	0.9	0.2	0.5	0.9
2	0.05	0.05	0.06	0.01	0.01	0.01	0.003	0.004	0.005
5	0.24	0.25	0.22	0.04	0.04	0.05	0.01	0.01	0.02
10	1.00	1.10	0.83	0.10	0.11	0.13	0.02	0.04	0.05

Table 1: MSE obtained by the *MaxSkew* method.

	$n = 100$			$n = 250$			$n = 500$		
p \ ρ	0.2	0.5	0.9	0.2	0.5	0.9	0.2	0.5	0.9
2	0.08	0.09	0.10	0.02	0.02	0.02	0.01	0.01	0.01
5	4.08	4.18	4.04	0.72	0.74	0.73	0.19	0.19	0.19
10	154.68	153.72	152.69	26.55	26.51	26.45	6.86	6.84	6.86

Table 2: MSE obtained by *Mardia* method.

4.1.2 SN model with a Toeplitz scale matrix

Assume that $\boldsymbol{X}$ follows a SN distribution, $\boldsymbol{X} \sim SN_p(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$, with scale matrix a Kac-Murdock-Szegö symmetric Toeplitz matrix defined as follows: $\boldsymbol{\Omega} = \bar{\boldsymbol{\Omega}} = (\omega_{i,j})_{1\leq i,j\leq p}$ where $\omega_{i,j} = \rho^{|i-j|} : 1 \leq i \leq j \leq p$ with ρ an autoregressive parameter, which we are going to assume to be non-negative.

In this example we are not aware of a closed form for the direction yielding the maximum of $\alpha_1^{*2} = \boldsymbol{\alpha}'\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha}$, but it can calculated numerically at each scenario of the following simulation experiment. We have drawn 1000 samples from a SN vector having a Toeplitz scale matrix for sample sizes $n = 100, 250, 500$, dimensions $p = 2, 5, 10$ and $\rho = 0.2, 0.5, 0.9$. For every p and ρ we find the vector $\boldsymbol{e}$ solving the problem: $\arg\max_{\boldsymbol{\alpha}:||\boldsymbol{\alpha}||=1}(\boldsymbol{\alpha}'\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha})$, so that the most skewed vector in the ordering $\leq_{cc}$ is found. Then we calculate $\gamma_{1,p}^D$ and $\gamma_{1,p}^M$ using Corollary 1.

We have applied both *MaxSkew* and *Mardia* methods to estimate $\gamma_{1,p}^D = \gamma_{1,p}^M$; the MSE of the estimations is calculated over the one thousand samples in all the scenarios under consideration. The results show that once again that *MaxSkew* method outperforms *Mardia*.

	$n = 100$			$n = 250$			$n = 500$		
p \ ρ	0.2	0.5	0.9	0.2	0.5	0.9	0.2	0.5	0.9
2	0.05	0.05	0.06	0.01	0.01	0.01	0.003	0.003	0.005
5	0.27	0.23	0.26	0.03	0.04	0.05	0.01	0.01	0.02
10	1.08	1.03	0.88	0.10	0.09	0.12	0.02	0.02	0.05

Table 3: MSE obtained by the *MaxSkew* method.

	$n = 100$			$n = 250$			$n = 500$		
p \ ρ	0.2	0.5	0.9	0.2	0.5	0.9	0.2	0.5	0.9
2	0.09	0.09	0.10	0.02	0.02	0.02	0.005	0.006	0.007
5	4.14	4.06	4.18	0.72	0.73	0.73	0.18	0.19	0.19
10	155.18	153.73	153.31	26.65	26.83	26.50	6.86	6.88	6.95

Table 4: MSE obtained by *Mardia* method.

4.2 An example with real data

The results of the simulation study have shown the superiority of the *MaxSkew* method. Here, we apply it to real data in order to estimate $\gamma_{1,p}^D = \gamma_{1,p}^M$. so that skewness differences in the ordering $\leq_{cc}$ may be uncovered.

Data with biomedical measures were collected for 202 atheltes, 102 male and 100 female, by the Australian Institute of Sport (AIS) (Cook and Weisberg 2009). Let us consider a vector $\boldsymbol{X} = (BMI, SSF, Bfat, LBM)$ whose components denote the body mass index, the sum of skin folds, the body fat percentage and the lean body mass. The SN distribution has been previously reported to be a good model for fitting this data (Azzalini and Capitanio 1999).

Figure 5 displays the bivariate scatterplots of the variables in $\boldsymbol{X}$ for males and females. The contour plots of the fitted bivariate SN models highlight differences in asymmetry between both groups; perhaps the most remarkable difference corresponds to the pair $(Bfat, LBM)$. We quantify the observed differences by application of the *MaxSkew* method to estimate $\gamma_{1,p}^{D} = \gamma_{1,p}^{M}$ for all the bivariate combinations in both groups. The results, summarized in Table 5, confirm the visual inspection. Overall, we can conclude that pairings of variables for males are more skewed with respect to $\leq_{cc}$ than for females.

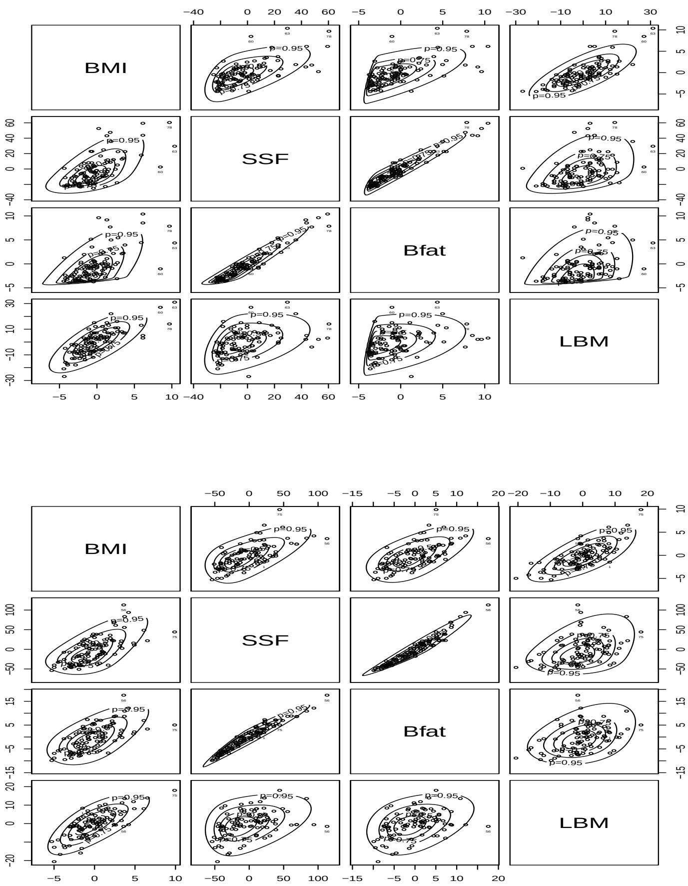


Fig. 5: Bivariate scatterplots of variables in $\boldsymbol{X}$ for males (top) and females (bottom) along with the contour levels of the fitted SN distribution.

	BMI SFF	BMI $Bfat$	BMI LBM	SSF $Bfat$	SSF LBM	$Bfat$ LBM
Males	1.96	2.54	2.59	2.46	2.26	2.64
Females	1.07	1.35	0.63	1.67	1.06	0.47

Table 5: Estimates of $\gamma^D_{1,p} = \gamma^M_{1,p}$ obtained by the *MaxSkew* method.

5 Summary and concluding remarks

In this paper we have addressed the problem of the skewness stochastic comparison between vectors that follow a multivariate SN distribution. A new skewness stochastic ordering has been proposed and its relationship with some classical measures that quantify departures from normality in the SN model has been studied. The new ordering is defined using the well-known convex transform ordering between the corresponding skewed scalar components in the canonical form of the vectors; this is a quite natural way for defining it, since all the multivariate skewness of the SN model is absorbed by such a skewed component. An extension of the convex transform order to the multivariate setting has been recently introduced by Belzunce et al (2015); we revisited it in the context of the canonical transformation of SN vectors and studied its connection with our skewness stochastic ordering.

We have also proved that the canonical convex transform order is a total order for the family of multivariate SN distributions. In addition, for a specific parametrization of the SN model, having location vector $\boldsymbol{\xi}$ and scale matrix $\boldsymbol{\Omega}$, the most skewed vector with respect to the ordering $\leq_{cc}$ can be obtained theoretically by finding the shape vector $\boldsymbol{\alpha}$ that maximizes the slant parameter, $\boldsymbol{\alpha}'\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha}$, of the only skewed component in the canonical representation.

Since the proposed stochastic ordering relies on the canonical representation of the SN vector and this canonical form has already been established for scale mixtures of skew-normal (SMSN) distributions (Capitanio 2012), future research would include the extension of the proposed ordering to the more general class of SMSN vectors. We believe that a program for such extension should begin by exploring the SMSN multivariate family due to the appealing form of its stochastic representation: for a vector $\boldsymbol{X}$ following a SMSN distribution it is given by

$$\boldsymbol{X} = \boldsymbol{\xi} + \boldsymbol{\omega} S \boldsymbol{Z},$$

where $\boldsymbol{\xi}$ is a location vector, $\boldsymbol{\omega}$ is a scale diagonal matrix with non-negative entries, $\boldsymbol{Z}$ is a SN vector with density function (2), and S is a non-negative scalar variable, independent of $\boldsymbol{Z}$, that injects an extra tail-weight parameter into the multivariate model (Capitanio 2012; Azzalini and Capitanio 2014).

The extension would involve an analogous approach, following the path of Definitions 3 and 4, that may ensure the validity of the results established in this paper for SN vectors. Deep understanding of the distributional form of the SMSN family is needed in order to address the issue.

Another alternative to generalize the results even more would explore wider families of multivariate skewed distributions such as the skew-elliptical family (Branco and Dey 2001; Genton 2004; Azzalini and Capitanio 2014) or different forms of skewed elliptical-based multivariate distributions (Azzalini and Capitanio 2003; Genton and Loperfido 2005; Arellano-valle and Genton 2010; Landsman et al 2017). To the best of our knowledge, a well-established canonical representation for multivariate distributions within these families is an open issue that should be addressed before moving forward.

Finally, we want to stress the potential use of Theorem 1 for making inference. Since it establishes that the slant parameter of the first canonical variate is an indicator to assess the *cc* ordering, we argue that any inferential procedure should rely on the estimation of such quantity. Currently the MaxSkew (Franceschini and Loperfido 2016) and sn (Azzalini 2016) R packages provide computational alternatives to address the issue in future research advances.

Appendix

Lemma 1 (Sign patterns) *Let us consider α_1 and α_2 non-negative parameters such that $0 < \alpha_1 < \alpha_2$. The function m defined by*

$$m(x) = -\frac{x^2}{2} + \frac{(a+bx)^2}{2} + \log \Phi(\alpha_1 x) - \log \Phi(\alpha_2(a+bx)) - \log b \; : x \in \mathbb{R},$$

with $b > 0$, has one, two or three zeros.

Proof.

We know that $e^{m(x)}$ is the quotient of the densities of two scalar random variables S and Z such that $S \sim SN_1(0,1,\alpha_1)$ and $Z = \dfrac{T-a}{b}$ with $T \sim SN_1(0,1,\alpha_2)$. Therefore, we can assert that the number of zeros of the function m is at least one. Since m can be rewritten as $m(x) = m_1(x) - m_2(x)$, with $m_1(x) = -\dfrac{x^2}{2} + \log \Phi(\alpha_1 x)$ and $m_2(x) = -\dfrac{(a+bx)^2}{2} + \log \Phi(\alpha_2(a+bx)) + \log b$, the number of intersections between m_1 and m_2 gives the number of zeros of m. We now examine the behavior of both functions.

$$\lim_{x \to -\infty} m_1(x) = \lim_{x \to \infty} m_1(x) = -\infty \text{ and } m_1(x) < 0 \;,\; x \in \mathbb{R}. \tag{11}$$

The first derivative of m_1 is given by

$$m_1'(x) = -x + \frac{\alpha_1 \phi(\alpha_1 x)}{\Phi(\alpha_1 x)} = -x + \frac{\alpha_1}{\frac{1}{\phi(\alpha_1 x)} - r(\alpha_1 x)} \;,\; \text{ with } r(t) = \frac{1-\Phi(t)}{\phi(t)}$$

for $t > 0$ the well-known Mills' ratio, which is a convex and strictly decreasing function Baricz (2008). Therefore, when $x > 0$ we get

$$m_1''(x) = -1 - \frac{\alpha_1}{(\frac{1}{\phi(\alpha_1 x)} - r(\alpha_1 x))^2} \left(\frac{\alpha_1^2 x}{\phi(\alpha_1 x)} - \alpha_1 r'(\alpha_1 x) \right) < 0.$$

When $x < 0$, taking into account that $m_1'(x) = -x + \dfrac{\alpha_1 \phi(\alpha_1 x)}{1 - \Phi(-\alpha_1 x)} = -x + \dfrac{\alpha_1}{r(-\alpha_1 x)}$ we obtain once again that

$$m_1''(x) = -1 + \frac{\alpha_1^2}{r(-\alpha_1 x)^2} r'(-\alpha_1 x) < 0.$$

Therefore, m_1 is negative concave function having a unique maximum.

On the other hand, for m_2 it is straightforward to show that, as long $b \leq 1$, we have

$$\lim_{x \to -\infty} m_2(x) = \lim_{x \to \infty} m_2(x) = -\infty \text{ with } m_2(x) < 0 \text{ , } x \in \mathbb{R}. \tag{12}$$

Now, using an analogous argument as previously for m_1 we conclude that m_2 is a negative concave function with a unique maximum provided that $b \leq 1$. Similarly, when $b > 1$ we can see that m_2 is a concave function with two zeros, being non-negative between the zeros.

Next, we study the asymptotic relative positions of m_1 and m_2, as well as the limit behavior of the function m.

Firstly, we calculate the limit

$$\lim_{x \to \infty} m(x) = \frac{a^2}{2} - \log b + \lim_{x \to \infty} \left(\frac{(b^2 - 1)x^2}{2} + abx \right) =$$

$$= \begin{cases} -\infty & \text{if } b < 1 \text{ or if } b = 1 \text{ and } a < 0 \\ \infty & \text{if } b > 1 \text{ or if } b = 1 \text{ and } a > 0 \end{cases}. \tag{13}$$

Secondly, we need the limit

$$\lim_{x \to -\infty} m(x) = \frac{a^2}{2} - \log b + \lim_{x \to -\infty} x^2 \left[\frac{b^2 - 1}{2} + \frac{ab}{x} + \frac{1}{x^2} \log \left(\frac{\Phi(\alpha_1 x)}{\Phi(\alpha_2 (a + bx))} \right) \right].$$

The calculus of this limit requires looking into some details about the asymptotic behavior of the last summand in the expression above. We sketch the computations as follows: taking into account that

$$\lim_{x \to -\infty} \frac{\Phi(\alpha_1 x)}{\Phi(\alpha_2 (a + bx))} = \frac{\alpha_1}{\alpha_2 b} \lim_{x \to -\infty} \frac{\phi(\alpha_1 x)}{\phi(\alpha_2 (a + bx))}$$

$$= \frac{\alpha_1}{\alpha_2 b} \exp \left\{ \lim_{x \to -\infty} \left(-\frac{\alpha_1^2 x^2}{2} + \frac{\alpha_2^2 (a + bx)^2}{2} \right) \right\} = 0 \text{ or } \infty,$$

depending on whether the limit of the exponential equals $-\infty$ or ∞, we would obtain that $\lim_{x\to-\infty} \log\left(\frac{\Phi(\alpha_1 x)}{\Phi(\alpha_2(a+bx))}\right) = \pm\infty$. It then stands to reason the application of L'hopital's rule in order to get

$$\lim_{x\to-\infty} \frac{1}{x^2}\log\left(\frac{\Phi(\alpha_1 x)}{\Phi(\alpha_2(a+bx))}\right) = \lim_{x\to-\infty} \frac{1}{2x}\left[\frac{\alpha_1\phi(\alpha_1 x)}{\Phi(\alpha_1 x)} - \frac{b\alpha_2\phi(\alpha_2(a+bx))}{\Phi(\alpha_2(a+bx))}\right]$$

$$= \lim_{x\to-\infty} \frac{1}{2x}\left[b\alpha_2^2(a+bx) - \alpha_1^2 x\right] = \frac{b^2\alpha_2^2 - \alpha_1^2}{2},$$

where the two step in the second line above has been carried out using the relation between the quotients in the square brackets and the Mills' ratio together with the asymptotic expansion of the Mills' ratio from Abramowitz and Stegun (1964; (26.2.12)). Consequently,

$$\lim_{x\to-\infty} m(x) = \frac{a^2}{2} - \log b + \lim_{x\to-\infty} x^2\left(\frac{b^2-1}{2} + \frac{ab}{x} + \frac{b^2\alpha_2^2 - \alpha_1^2}{2}\right)$$

whose result, that depends on the signs of the difference $b^2(\alpha_2^2+1) - (\alpha_1^2+1)$ and a, can be established as follows:

$$\lim_{x\to-\infty} m(x) = \begin{cases} -\infty & \text{if } b < \sqrt{\frac{\alpha_1^2+1}{\alpha_2^2+1}} \text{ or if } b = \sqrt{\frac{\alpha_1^2+1}{\alpha_2^2+1}} \text{ and } a > 0 \\ \infty & \text{if } b > \sqrt{\frac{\alpha_1^2+1}{\alpha_2^2+1}} \text{ or if } b = \sqrt{\frac{\alpha_1^2+1}{\alpha_2^2+1}} \text{ and } a < 0 \end{cases} \tag{14}$$

The limits (13) and (14) along with the fact $\sqrt{\frac{\alpha_1^2+1}{\alpha_2^2+1}} < 1$, resulting from the condition of the statement, are suggesting the cases to be considered in items (i)–(vii). For all of them we will take into account that both m_1 and m_2 are concave functions with a unique maximum.

Before studying the sign pattern of the function m in all these cases, let us introduce some notation: the symbol $-$ is used to denote negativity and the symbol $+$ is for positivity; for example, with the pattern $-+$ we denote that a function is changing from negative to positive.

(i) *$0 < b < \sqrt{\frac{\alpha_1^2+1}{\alpha_2^2+1}}$.* In this case, taking into account the conclusions derived in (13) and (14), we obtain: $\lim_{x\to-\infty} m(x) = \lim_{x\to\infty} m(x) = -\infty$, which implies that m_1 and m_2 have two intersection points and the function m has two zeros with sign pattern $-+-$.

(ii) *$\sqrt{\frac{\alpha_1^2+1}{\alpha_2^2+1}} < b < 1$.* Now, the results obtained in (13) and (14) lead to the following limits: $\lim_{x\to-\infty} m(x) = \infty$ and $\lim_{x\to\infty} m(x) = -\infty$, from which we can deduce that m_1 and m_2 may intersect once or three times, depending on the values of a. Hence, m may have one or three zeros with respective sign patterns $+-$ and $+-+-$.

(iii) $b > 1$. In this case from (13) and (14) we obtain: $\lim_{x\to-\infty} m(x) = \lim_{x\to\infty} m(x) = \infty$, which implies that m_1 and m_2 will intersect one each other twice and m has two zeros with sign pattern $+-+$.

(iv) $b = 1$, $a > 0$. For this case once again $\lim_{x\to-\infty} m(x) = \lim_{x\to\infty} m(x) = \infty$ and m will have two zeros with the same sign pattern as in (iii).

(v) $b = 1$, $a < 0$. This case leads leads to the same limits as those in item (ii): $\lim_{x\to-\infty} m(x) = \infty$ and $\lim_{x\to\infty} m(x) = -\infty$. However, in this case m has only one zero, with $+-$ sign pattern, due to the stochastic dominance of the distribution functions of variables S and $Z = T - a$.

(vi) $b = \sqrt{\frac{\alpha_1^2+1}{\alpha_2^2+1}}$, $a > 0$. In this case m has the same limit behavior as in item (i), so that it has two zeros with sign pattern $-+-$.

(vii) $b = \sqrt{\frac{\alpha_1^2+1}{\alpha_2^2+1}}$, $a < 0$. This case leads to the same limits for m as in item (ii) so the function m may have one or three zeros which in turn yield the sign patterns $+-$ and $+-+-$ respectively.

The conclusions derived in all these previous cases prove the statement of the lemma. □

References

Abramowitz M, Stegun IA (eds) (1964) Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables (Dover Books on Mathematics), 1st edn. Dover books on mathematics, Dover Publications

Adcock C (2004) Capital asset pricing for UK stocks under the multivariate skew-normal distribution, Chapman and Hall/CRC, pp 191–204

Adcock C, Eling M, Loperfido N (2015) Skewed distributions in finance and actuarial science: a review. The European Journal of Finance 21(13-14):1253–1285

Arellano-valle RB, Genton MG (2010) Multivariate unified skew-elliptical distributions. Chilean Journal of Statistics 1(1):17–33

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

Arnold BC, Groeneveld RA (1995) Measuring skewness with respect to the mode. The American Statistician 49(1)

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

Azzalini A (2016) The R package `sn`: The Skew-Normal and Skew-t distributions (version 1.4-0). Università di Padova, Italia, URL http://azzalini.stat.unipd.it/SN

Azzalini A, Capitanio A (1999) Statistical applications of the multivariate skew normal distribution. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 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

Azzalini A, Regoli G (2012) Some properties of skew-symmetric distributions. Annals of the Institute of Statistical Mathematics 64(4):857–879

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

Baricz A (2008) Mills' ratio: Monotonicity patterns and functional inequalities. Journal of Mathematical Analysis and Applications 340(2):1362–1370

Belzunce F, Mulero J, Ruíz MJ, Suárez-Llorens A (2015) On relative skewness for multivariate distributions. TEST 24(4):813–834

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

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

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

Counsell N, Cortina-Borja M, Lehtonen A, Stein A (2011) Modelling psychiatric measures using skew-normal distributions. European psychiatry : the journal of the Association of European Psychiatrists 26(2):112–114

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

Genton M (2004) Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality. CRC Press

Genton M, Loperfido N (2005) Generalized skew-elliptical distributions and their quadratic forms. Annals of the Institute of Statistical Mathematics 57(2):389–401

Hardin J, Wilson J (2009) A note on oligonucleotide expression values not being normally distributed. Biostatistics 10(3):446–450

Landsman Z, Makov U, Shushi T (2017) Extended generalized skew-elliptical distributions and their moments. Sankhya A: The Indian Journal of Statistics 79(1):76–100

Loperfido N (2010) Canonical transformations of skew-normal variates. TEST 19(1):146–165

Loperfido N (2013) Skewness and the linear discriminant function. Statistics & Probability Letters 83(1):93–99

MacGillivray HL (1986) Skewness and asymmetry: Measures and orderings. Ann Statist 14(3):994–1011, DOI 10.1214/aos/1176350046

Main P, Arevalillo JM, Navarro H (2015) Local effect of asymmetry deviations from gaussianity using information-based measures. In: Proceedings of the 2nd Int. Electron. Conf. Entropy Appl., vol 2, pp 15–30

Malkovich JF, Afifi AA (1973) On tests for multivariate normality. Journal of the American Statistical Association 68(341):176–179

Mardia KV (1970) Measures of multivariate skewness and kurtosis with applications. Biometrika 57:519–530

Marshall A, Olkin I (2007) Life Distributions: Structure of Nonparametric, Semiparametric, and Parametric Families. Springer Series in Statistics, Springer New York

Oja H (1981) On location, scale, skewness and kurtosis of univariate distributions. Scandinavian Journal of Statistics 8(3):154–168

Pearson K (1895) Contributions to the mathematical theory of evolution. ii. skew variation in homogeneous material. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 186:343–414

Revelle W (2016) psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois, URL https://CRAN.R-project.org/package=psych, R package version 1.6.9

Taniguchi M, Petkovic A, Kase T, DiCiccio T, Monti AC (2015) Robust portfolio estimation under skew-normal return processes. The European Journal of Finance 21:1091–1112

Van Zwet WR (1964) Convex Transformations of Random Variables. Mathematish Centrum, Amsterdam

Vernic R (2006) Multivariate skew-normal distributions with applications in insurance. Insurance Mathematics & Economics 38:413–426

Wang J (2009) A family of kurtosis orderings for multivariate distributions. Journal of Multivariate Analysis 100(3):509–517

Zadkarami MR, Rowhani M (2010) Application of skew-normal in classification of satellite image. Journal of Data Science 8:597–606