



On Multivariate Skewness and Kurtosis

Sreenivasa Rao Jammalamadaka
University of California, Santa Barbara, USA

Emanuele Taufer 
University of Trento, Trento, Italy

Gyorgy H. Terdik
University of Debrecen, Debrecen, Hungary

Abstract

A unified treatment of all currently available cumulant-based indexes of multivariate skewness and kurtosis is provided here, expressing them in terms of the third and fourth-order cumulant vectors respectively. Such a treatment helps reveal many subtle features and inter-connections among the existing indexes as well as some deficiencies, which are hitherto unknown. Computational formulae for obtaining these measures are provided for spherical and elliptically-symmetric, as well as skew-symmetric families of multivariate distributions, yielding several new results and a systematic exposition of many known results.

AMS (2000) subject classification. 62H05, 60E05, 60E10.

Keywords and phrases. Multivariate skewness, Multivariate kurtosis, Cumulant vectors, Interconnections between different measures, Symmetric and skew multivariate distributions

1 Introduction

Using the standard normal distribution as the yardstick, statisticians have defined notions of skewness (asymmetry) and kurtosis (peakedness) in the univariate case. Since all the odd central moments (when they exist) are zero for a symmetric distribution on the real line, a first attempt at measuring asymmetry is to ask how different the 3rd central moment is from zero, although in principle, one could use any other odd central moment, or even a combination of them. Several alternate indexes of asymmetry, using the mode or the quantiles are also available –see for instance (Arnold and Groeneveld, 1995; Averous and Meste, 1997; Ekström and Jammalamadaka, 2012) for recent work and review. Similarly for kurtosis, taking again the

standard normal distribution as the yardstick, a coefficient of kurtosis has been developed using the 4th central moment.

When dealing with multivariate distributions, the notions of symmetry, measurement of skewness, as well as of kurtosis, are not uniquely defined. For example, the mode-based approach in Arnold and Groeneveld (1995), although very popular, does not seem extendable to the multivariate case. Also, in interpreting the cumulant-based measures of skewness and kurtosis, one has to pause – especially for the kurtosis if there are multiple peaks.

Our focus in this paper is on multivariate distributions and one may remark that the two fundamental papers by Rao (1948a, b) take us back to the early days of such multivariate analysis. A good starting point is the monograph by Fang et al. (2017) and a more recent review article by Serfling (2004). We consider symmetry of a d -dimensional random vector \mathbf{X} , around a given point, which we will assume without loss of generality, to be the origin. Such an \mathbf{X} is said to be *spherically symmetric* or rotationally symmetric if for all $(d \times d)$ orthogonal matrices \mathbf{A} , \mathbf{X} has the same distribution as \mathbf{AX} . One may generalize this to *elliptical or ellipsoidal symmetry* in an obvious way. A more common and practical notion of symmetry in multi-dimensions, is the *reflective or antipodal symmetry*, and we say \mathbf{X} has this property if it has the same distribution as $-\mathbf{X}$. For measuring departures from such symmetry, various notions of skewness have been proposed by different authors, and the list includes (Mardia, 1970; Malkovich and Afifi, 1973; Isogai, 1982; Srivastava, 1984; Song, 2001) and Móri et al. (1994); see also Sumikawa et al. (2013) for an extension of Mardia's multivariate skewness index to the high-dimensional case. In a broad discussion and analysis of multivariate cumulants, their properties and their use in inference, Jammalamadaka et al. (2006) proposed using the full vector of third and fourth order cumulants, as vectorial measures for multivariate skewness and kurtosis respectively. Such measures based on the cumulant vectors, were further discussed by Balakrishnan et al. (2007) and Kollo (2008). A systematic treatment of asymptotic distributions of skewness and kurtosis indexes can be found in Baringhaus and Henze (1991a, b, 1992), and Klar (2002).

In this paper, one of our primary goals is to look at these disparate looking definitions of skewness and kurtosis based on cumulants which have been proposed in the literature, and to assess, and relate them from a unified perspective in terms of the cumulant vectors discussed in Jammalamadaka et al. (2006). Such a unified treatment helps reveal several relationships and features of the many existing proposals. For example it will be shown: (i) that Mardia (1970) and Malkovich and Afifi (1973) skewness measures can be equivalent; (ii) that Balakrishnan et al. (2007) and Móri et al. (1994)

skewness vectors are just proportional to each other; (iii) that Kollo (2008) vectorial measure can be a null vector even for some asymmetric distributions, and (iv) that Srivastava (1984) index is not affine invariant.

In Section 4, we also introduce alternate measures for skewness and Kurtosis based just on the *distinct* cumulants, and evaluate their performance with some examples in Section 7.

Another significant contribution of the paper is to provide clear and easy formulae for computing cumulants up to the fourth order for spherically symmetric, elliptically symmetric, and skew elliptical families along with several specific examples. These, comprehensive and mostly new, results allow for a straightforward computation of all the indexes of skewness and kurtosis discussed here for several important multivariate distributions.

The analysis presented here is based on the *cumulant vectors* of the third and fourth order, defined below. In our derivations we utilize an elegant and powerful tool—the so-called *T*-derivative, which we now describe. Let $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d)^\top$ be a d -dimensional vector of constants and let $\boldsymbol{\phi}(\boldsymbol{\lambda}) = (\phi_1(\boldsymbol{\lambda}), \dots, \phi_m(\boldsymbol{\lambda}))^\top$ denote a m -dimensional vector valued function ($m \in \mathbb{N}$), which is differentiable in all its arguments. The Jacobian matrix of $\boldsymbol{\phi}$ is defined by

$$D_{\boldsymbol{\lambda}}\boldsymbol{\phi}(\boldsymbol{\lambda}) = \frac{\partial\boldsymbol{\phi}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}^\top} = \left[\frac{\partial\phi_i(\boldsymbol{\lambda})}{\partial\lambda_j} \right]_{i=1,\dots,m; j=1,\dots,d}.$$

Then the operator $D_{\boldsymbol{\lambda}}^{\otimes}$, which we refer to as the *T*-derivative, is defined as

$$D_{\boldsymbol{\lambda}}^{\otimes}\boldsymbol{\phi}(\boldsymbol{\lambda}) = \text{vec} \left(\frac{\partial\boldsymbol{\phi}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}^\top} \right)^\top = \boldsymbol{\phi}(\boldsymbol{\lambda}) \otimes \frac{\partial}{\partial\boldsymbol{\lambda}}, \tag{1.1}$$

where the symbol \otimes here and everywhere else in the paper, denotes the Kronecker (tensor) product. Assuming $\boldsymbol{\phi}$ is k times differentiable, the k -th *T*-derivative is given by

$$D_{\boldsymbol{\lambda}}^{\otimes k}\boldsymbol{\phi}(\boldsymbol{\lambda}) = D_{\boldsymbol{\lambda}}^{\otimes} \left(D_{\boldsymbol{\lambda}}^{\otimes k-1}\boldsymbol{\phi}(\boldsymbol{\lambda}) \right), \tag{1.2}$$

which is a vector of order $m \times d^k$ containing all possible partial derivatives of entries of $\boldsymbol{\phi}(\boldsymbol{\lambda})$ according to the tensor product $\left(\frac{\partial}{\partial\lambda_1}, \dots, \frac{\partial}{\partial\lambda_d} \right)^{\top \otimes k}$. Refer to Jammalamadaka et al. (2006) for further details, properties, examples, and applications of the operator $D_{\boldsymbol{\lambda}}^{\otimes}$ (see also Terdik 2002).

We note that if $\phi_{\mathbf{X}}(\boldsymbol{\lambda})$ denotes the characteristic function of a d -dimensional random vector \mathbf{X} , then the operator $D_{\boldsymbol{\lambda}}^{\otimes k}$ applied to ϕ , and $\log \phi$ will provide the vector of moments of order k , and cumulants of order k respectively.

One may also refer to MacRae (1974) for a similar definition of a matrix derivative using the tensor product and a differential operator; indeed, the T-derivative we obtain by vectorizing the transposed Jacobian (1.1), can be seen as closely related to that. While these results can also be obtained by using tensor-calculus, as is done in McCullagh (1987) and Speed (1990), our approach is more straightforward and simpler, requiring only the knowledge of calculus of several variables. We believe that using only the tensor products of vectors leads to an intuitive and natural way to deal with higher order moments and cumulants for multivariate distributions, as it will be demonstrated in the paper. Another comprehensive reference on matrix derivatives is the book by Mathai (1997).

The paper is organized as follows: Sections 2 and 3 introduce respectively the skewness and kurtosis vectors, and treat several existing measures of skewness and kurtosis based on these cumulant vectors. Section 4 discusses a linear transformation on the skewness and kurtosis vectors, which helps remove from them redundant/duplicate information, and proposes skewness and kurtosis indexes based on distinct elements of the corresponding vectors. Sections 5 and 6 provide computational formulae for the skewness and kurtosis vectors for spherical, elliptical and asymmetric/skew multivariate distributions. Section 7 provides some examples while Section 8 provides some final considerations. To improve readability of the paper, more technical details and proofs are placed in an Appendix. A word about the notations: bold uppercase letters are used for random vectors and matrices while bold lowercase letters denote their specific values.

2 Multivariate Skewness

In this section and the next one, it will be shown that all cumulant-based measures of skewness and kurtosis that appear in the literature can be expressed in terms of the third and fourth cumulant vectors respectively. Also, several hitherto unnoticed relationships between different indexes will be brought out.

Let \mathbf{X} be a d -dimensional random vector whose first four moments exist. We will denote $E \mathbf{X} = \boldsymbol{\mu}$, with a positive-definite variance-covariance matrix $\text{Var } \mathbf{X} = \boldsymbol{\Sigma}$. Consider the standardized vector

$$\mathbf{Y} = \boldsymbol{\Sigma}^{-1/2} (\mathbf{X} - \boldsymbol{\mu}) \quad (2.1)$$

with zero means and identity matrix for its variance-covariance. A complete picture about the skewness, is contained in the “skewness vector” of \mathbf{X} or

\mathbf{Y} , defined as $\boldsymbol{\kappa}_3 = \text{Cum}_3(\mathbf{Y})$. Since the third order cumulants are the same as the third order central moments, we may write

$$\boldsymbol{\kappa}_3 = \text{E } \mathbf{Y}^{\otimes 3}. \tag{2.2}$$

Note that for a d -dimensional vector $\mathbf{Y} = (Y_1, \dots, Y_d)^\top$, $\boldsymbol{\kappa}_3$ has length d^3 and it contains all terms of the form $\text{E } Y_r^3, \text{E } Y_r^2 Y_s, \text{E } Y_r Y_s Y_t$ for $1 \leq r, s, t \leq d$. Among these d^3 elements, only $d(d+1)(d+2)/6$ are distinct elements, and in Section 4 we discuss a linear transformation which allows one to get the distinct elements of $\boldsymbol{\kappa}_3$. In the examples below, we denote the unit matrix of dimension k by \mathbf{I}_k , and the k -vector of ones by $\mathbf{1}_k$.

The following 6 examples reveal several relationships among various indexes of skewness which appear in the literature, and their connection to the third-order cumulant vector $\boldsymbol{\kappa}_3$, which can actually be seen as the common denominator.

Example 1. Mardia (1970) suggested the square norm of the vector $\boldsymbol{\kappa}_3$ as a measure of departure from symmetry, viz.

$$\kappa_3 = \|\boldsymbol{\kappa}_3\|^2,$$

which is denoted by $\beta_{1,d}$. If \mathbf{Y}_1 and \mathbf{Y}_2 are two independent copies of \mathbf{Y} , then we may write

$$\kappa_3 = \|\text{E } \mathbf{Y}^{\otimes 3}\|^2 = \text{E } \mathbf{Y}_1^{\top \otimes 3} \text{E } \mathbf{Y}_2^{\otimes 3} = \text{E } \left(\mathbf{Y}_1^\top \mathbf{Y}_2 \right)^3.$$

Example 2. Móri et al. (1994), after observing that

$$(\text{vec } \mathbf{I}_d)^\top \mathbf{Y}^{\otimes 2} = \mathbf{Y}^{\top \otimes 2} \text{vec } \mathbf{I}_d = \text{vec } \mathbf{Y}^\top \mathbf{Y} = \sum_{i=1}^d Y_i^2$$

define a “skewness vector”

$$s(\mathbf{Y}) = \text{E } \left(\mathbf{Y}^\top \mathbf{Y} \right) \mathbf{Y} = \left((\text{vec } \mathbf{I}_d)^\top \otimes \mathbf{I}_d \right) \text{E } \mathbf{Y}^{\otimes 3} = \left((\text{vec } \mathbf{I}_d)^\top \otimes \mathbf{I}_d \right) \boldsymbol{\kappa}_3. \tag{2.3}$$

Note that $(\text{vec } \mathbf{I}_d)^\top \otimes \mathbf{I}_d$ is a matrix of dimension $(d \times d^3)$ which contains d unit values per-row whereas all the others are 0; as a consequence, this measure does not take into account the contribution of cumulants of the type $\text{E}(Y_r Y_s Y_t)$, where (r, s, t) are all different.

Example 3. Kollo (2008), noting the fact that not all third-order mixed moments appear in $s(\mathbf{Y})$, proposes an alternate skewness vector $b(\mathbf{Y})$ which can again be expressed in terms of $\boldsymbol{\kappa}_3$ as follows:

$$b(\mathbf{Y}) = E \left[\mathbf{1}_{d^2}^\top (\mathbf{Y} \otimes \mathbf{Y}) \right] \otimes \mathbf{Y} = \left(\mathbf{1}_{d^2}^\top \otimes \mathbf{I}_d \right) E \mathbf{Y}^{\otimes 3} = \left(\mathbf{1}_{d^2}^\top \otimes \mathbf{I}_d \right) \boldsymbol{\kappa}_3. \quad (2.4)$$

Comparing $b(\mathbf{Y})$ in Eq. 2.4 to $s(\mathbf{Y})$ in Eq. 2.3, we see that the difference between the two expressions comes from the fact that $b(\mathbf{Y})$ has the term $\mathbf{1}_{d^2}^\top$, compared to $(\text{vec } \mathbf{I}_d)^\top$ in $s(\mathbf{Y})$. This results in the Kollo measure summing the elements of d consecutive groups of size d^2 in $\boldsymbol{\kappa}_3$.

Following this line of reasoning, note that if $d = 2$ and $\boldsymbol{\kappa}_3 = [1, -1, -1, 1, -1, 1, 1, -1]$, then $b(\mathbf{Y}) = 0$ even for an asymmetric distribution, making it not a valid measure of skewness. Section 7 gives specific examples where this actually happens.

Example 4. Malkovich and Afifi (1973) (see also Balakrishnan and Scarpa 2012) consider the following approach to measuring skewness: let \mathbb{S}_{d-1} be the $(d - 1)$ dimensional unit sphere in \mathbb{R}^d . First, for $\mathbf{u} \in \mathbb{S}_{d-1}$, note that

$$\underline{\text{Cum}}_3 \left(\mathbf{u}^\top \mathbf{Y} \right) = \left(\mathbf{u}^\top \right)^{\otimes 3} \underline{\text{Cum}}_3 (\mathbf{Y}) = \mathbf{u}^{\top \otimes 3} E \mathbf{Y}^{\otimes 3} = \mathbf{u}^{\top \otimes 3} \boldsymbol{\kappa}_3. \quad (2.5)$$

Malkovich–Afifi define their measure as

$$b_1^* (\mathbf{Y}) = \sup_{\mathbf{u}} \left(\left(\mathbf{u}^{\top \otimes 3} \boldsymbol{\kappa}_3 \right)^2 \right).$$

Consider

$$\left(\mathbf{u}^{\top \otimes 3} \boldsymbol{\kappa}_3 \right)^2 = \left\| \mathbf{u}^{\top \otimes 3} \right\|^2 \left\| \boldsymbol{\kappa}_3 \right\|^2 \cos^2 \left(\mathbf{u}^{\top \otimes 3}, \boldsymbol{\kappa}_3 \right),$$

where $\cos(\mathbf{a}, \mathbf{b})$ indicates the cosine of the angle between the vectors \mathbf{a} and \mathbf{b} ; next note that

$$\left\| \mathbf{u}^{\top \otimes 3} \right\|^2 = \mathbf{u}^{\top \otimes 3} \mathbf{u}^{\otimes 3} = \left(\mathbf{u}^\top \mathbf{u} \right)^{\otimes 3} = 1,$$

and $\sup_{\mathbf{u}} \cos \left(\mathbf{u}^{\top \otimes 3}, \boldsymbol{\kappa}_3 \right)$ could be 1 when there would exist a \mathbf{u}_0 such that $\mathbf{u}_0^{\top \otimes 3} = \boldsymbol{\kappa}_3 / \left\| \boldsymbol{\kappa}_3 \right\|$. This can happen only when the normed $\boldsymbol{\kappa}_3 / \left\| \boldsymbol{\kappa}_3 \right\|$ has the same form as $\mathbf{u}^{\top \otimes 3}$. It follows that

$$b_1^* (\mathbf{Y}) \leq \left\| \boldsymbol{\kappa}_3 \right\|^2.$$

Example 5. Balakrishnan et al. (2007) discuss a multivariate extension of Malkovich–Afifi measure. Denoting $\Omega(d\mathbf{u})$ as the normalized Lebesgue element of surface area on \mathbb{S}_{d-1} , they suggest

$$\mathbf{T} = \int_{\mathbb{S}_{d-1}} \mathbf{u} \left(\mathbf{u}^{\top \otimes 3} \boldsymbol{\kappa}_3 \right) \Omega(d\mathbf{u}) = \int_{\mathbb{S}_{d-1}} \mathbf{u} \otimes \mathbf{u}^{\top \otimes 3} \Omega(d\mathbf{u}) \boldsymbol{\kappa}_3 \tag{2.6}$$

and we see that this extension is a constant times (a matrix-multiple of) the skewness vector $\boldsymbol{\kappa}_3$. Indeed, one can use Theorem 3.3 of Fang et al (1990) (Fang et al., 2017) to show that this matrix-multiple reduces to

$$\int_{\mathbb{S}_{d-1}} \mathbf{u} \otimes \mathbf{u}^{\top \otimes 3} \Omega(d\mathbf{u}) = \frac{3}{d(d+2)} \left((\text{vec } \mathbf{I}_d)^{\top} \otimes \mathbf{I}_d \right).$$

Therefore \mathbf{T} defined in Eq. 2.6 becomes a scalar-multiple, $3/d(d+2)$ times $s(\mathbf{Y})$ defined in Eq. 2.3 by Móri, Székely and Rohatgi (1984). In particular when $d = 3$, we have $\mathbf{T} = \frac{3}{15}s(\mathbf{Y})$. It follows that, as in Móri et al. (1994), the vector \mathbf{T} does not take into account the contribution of cumulants of the type $E(Y_r Y_s Y_t)$, where r, s, t are all different.

Example 6. Srivastava (1984). If \mathbf{X} is a d -dimensional random vector with variance matrix $\boldsymbol{\Sigma}$, then consider $\Gamma^{\top} \boldsymbol{\Sigma} \Gamma = \text{Diag}(\lambda_1, \dots, \lambda_d) = \mathbf{D}_{\lambda}$, with orthogonal matrix Γ . The skewness measure defined by Srivastava can be written as

$$b_1^2(\mathbf{Y}) = \frac{1}{d} \sum_{j=1}^d \left(E \tilde{Y}_j^3 \right)^2.$$

where $\tilde{\mathbf{Y}} = \mathbf{D}_{\lambda}^{-1} \Gamma (\mathbf{X} - E\mathbf{X})$. We have $\tilde{\mathbf{Y}} = \mathbf{D}_{\lambda}^{-1} \Gamma \boldsymbol{\Sigma}^{1/2} \mathbf{Y} = \mathbf{D}_{\lambda}^{-1/2} \Gamma \mathbf{Y}$, and the i^{th} coordinate of $\tilde{\mathbf{Y}}$ is $\tilde{Y}_i = \mathbf{e}_i^{\top} \tilde{\mathbf{Y}}$, where \mathbf{e}_i is the i^{th} coordinate axis. Since

$E\tilde{\mathbf{Y}} = 0$, and $\text{Var}(\tilde{\mathbf{Y}}) = \mathbf{D}_{\lambda}^{-1}$, Srivastava measure can be re-expressed as

$$E \tilde{Y}_i^3 = \underline{\text{Cum}}_3(\tilde{Y}_i) = \underline{\text{Cum}}_3\left(\mathbf{e}_i^{\top} \mathbf{D}_{\lambda}^{-1/2} \Gamma \mathbf{Y}\right) = \mathbf{e}_i^{\top \otimes 3} \left(\mathbf{D}_{\lambda}^{-1/2}\right)^{\otimes 3} \Gamma^{\otimes 3} \boldsymbol{\kappa}_3,$$

and

$$b_1^2(\mathbf{Y}) = \frac{1}{d} \sum_{j=1}^d \left(\mathbf{e}_j^{\top \otimes 3} \left(\mathbf{D}_{\lambda}^{-1/2}\right)^{\otimes 3} \Gamma^{\otimes 3} \boldsymbol{\kappa}_3 \right)^2.$$

One notices that $\mathbf{e}_i^{\top \otimes 3}$ is a unit axis vector in the Euclidean space \mathbb{R}^{d^3} , so that the measure $b_1^2(\mathbf{Y})$ is the norm square of the projection of $\boldsymbol{\kappa}_3$ to the subspace of \mathbb{R}^{d^3} , and it does not contain all the information contained in the vector $\boldsymbol{\kappa}_3$. Note that this index is NOT affine invariant (nor the corresponding kurtosis index).

3 Multivariate Kurtosis

The kurtosis of \mathbf{Y} is measured by the 4th order cumulant vector denoted by $\boldsymbol{\kappa}_4$, and is computed as

$$\boldsymbol{\kappa}_4 = \text{Cum}_4(\mathbf{Y}) = \mathbb{E} \mathbf{Y}^{\otimes 4} - \mathbf{K}_{2,2} \text{Cum}_2(\mathbf{Y})^{\otimes 2} = \mathbb{E} \mathbf{Y}^{\otimes 4} - \mathbf{K}_{2,2} [\text{vec } \mathbf{I}_d]^{\otimes 2}, \tag{3.1}$$

where $\mathbf{K}_{2,2}$ denotes the commutator matrix (1.2) (see Appendix 1 for details). Note that $\boldsymbol{\kappa}_4$ turns out to be the zero vector for multivariate Gaussian distributions, and may be used as the “standard”.

This kurtosis vector $\boldsymbol{\kappa}_4$ forms the basis for all multivariate measures of kurtosis proposed in the literature, as our next 6 examples demonstrate. For instance, one may define its square norm as a scalar index of kurtosis, called the *total kurtosis* defined by

$$\kappa_4 = \|\boldsymbol{\kappa}_4\|^2 \tag{3.2}$$

which is one of the measures. We now connect the kurtosis vector $\boldsymbol{\kappa}_4$ and the total kurtosis κ_4 to various other indexes discussed in the literature.

Example 7. Mardia (1970), defined an index of kurtosis as $\beta_{2,d} = \mathbb{E} (\mathbf{Y}^\top \mathbf{Y})^2$. Note that

$$\mathbb{E} \text{vec} \left(\mathbf{Y}^\top \mathbf{Y} \right)^2 = \mathbb{E} [\mathbf{Y}]^{\top \otimes 2} [\mathbf{Y}]^{\otimes 2} = \mathbb{E} [\mathbf{Y}]^{\top \otimes 4} \text{vec } \mathbf{I}_{d^2}$$

and this is related to the kurtosis vector $\boldsymbol{\kappa}_4$ as follows:

$$\beta_{2,d} = (\text{vec } \mathbf{I}_{d^2})^\top \boldsymbol{\kappa}_4 + (\text{vec } \mathbf{I}_{d^2})^\top \mathbf{K}_{2,2} [\text{vec } \mathbf{I}_d]^{\otimes 2}.$$

In particular for the standard Gaussian vector \mathbf{Y} , we have $\boldsymbol{\kappa}_4 = 0$, so that $\beta_{2,d} = \mathbb{E} (\mathbf{Y}^\top \mathbf{Y})^2 = d(d+2)$. As a consequence, for such a \mathbf{Y} we have the equation

$$(\text{vec } \mathbf{I}_{d^2})^\top \mathbf{K}_{2,2} [\text{vec } \mathbf{I}_d]^{\otimes 2} = d(d+2).$$

Note that $\text{vec } \mathbf{I}_{d^2}$ contains only d^2 ones and hence Mardia’s measure does not take into account all the entries of $\boldsymbol{\kappa}_4$; it includes only some of the entries of $\mathbb{E} \mathbf{Y}^{\otimes 4}$ namely

$$\beta_{2,d} = \sum_{r=1}^d \mathbb{E} Y_r^4 + \sum_{r \neq s} \mathbb{E} Y_r^2 Y_s^2.$$

Example 8. Koziol (1989) considered the following index of kurtosis. Let $\tilde{\mathbf{Y}}$ be an independent copy of \mathbf{Y} , then

$$\mathbb{E} \left(\tilde{\mathbf{Y}}^\top \mathbf{Y} \right)^4 = \mathbb{E} \tilde{\mathbf{Y}}^{\top \otimes 4} \mathbf{Y}^{\otimes 4} = \|\mathbb{E} \mathbf{Y}^{\otimes 4}\|^2,$$

is the next higher degree analogue of Mardia’s skewness index $\beta_{1,d}$. Specifically

$$\begin{aligned} \|\mathbb{E} \mathbf{Y}^{\otimes 4}\|^2 &= \left\| \boldsymbol{\kappa}_4 + \mathbf{K}_{2,2} [\text{vec } \mathbf{I}_d]^{\otimes 2} \right\|^2 \\ &= \kappa_4 + 2\boldsymbol{\kappa}_4^\top \mathbf{K}_{2,2} (\text{vec } \mathbf{I}_d)^{\otimes 2} + d^2 \\ &= \kappa_4 + 6\beta_{2,d} - d^2 \end{aligned}$$

where $\beta_{2,d}$ is Mardia (1970) index of kurtosis.

Example 9. Móri et al. (1994) define kurtosis of \mathbf{Y} as

$$K(\mathbf{Y}) = \mathbb{E} \left(\mathbf{Y}\mathbf{Y}^\top \mathbf{Y}\mathbf{Y}^\top \right) - (d+2) \mathbf{I}_d = \mathbb{E} \left(\mathbf{Y}^\top \mathbf{Y} \right) \mathbf{Y}\mathbf{Y}^\top - (d+2) \mathbf{I}_d.$$

Then $\text{vec } K(\mathbf{Y}) = \left(\mathbf{I}_{d^2} \otimes (\text{vec } \mathbf{I}_d)^\top \right) \mathbb{E} \mathbf{Y}^{\otimes 4} - (d+2) \text{vec } \mathbf{I}_d$ which can be expressed in terms of $\boldsymbol{\kappa}_4$ as $\text{vec } K(\mathbf{Y}) = \left(\mathbf{I}_{d^2} \otimes (\text{vec } \mathbf{I}_d)^\top \right) \boldsymbol{\kappa}_4$.

As in the case of their skewness measure, this measure does not take into account the contribution of cumulants of the type $\mathbb{E} (Y_r Y_s Y_t Y_u)$ where r, s, t, u are all different.

Example 10. Malkovich and Afifi (1973). Similar to the discussion regarding skewness, a measure proposed by Malkovich and Afifi simply provides a different derivation of the total kurtosis in the form

$$b_2^*(\mathbf{Y}) = \sup_{\mathbf{u}} \left(\left(\mathbf{u}^{\top \otimes 4} \boldsymbol{\kappa}_4 \right)^2 \right) \leq \|\boldsymbol{\kappa}_4\|^2,$$

because of the fact that

$$\left(\mathbf{u}^{\top \otimes 4} \boldsymbol{\kappa}_4 \right)^2 = \left\| \mathbf{u}^{\top \otimes 4} \right\|^2 \|\boldsymbol{\kappa}_4\|^2 \cos^2 \left(\mathbf{u}^{\top \otimes 4}, \boldsymbol{\kappa}_4 \right) = \|\boldsymbol{\kappa}_4\|^2 \cos^2 \left(\mathbf{u}^{\top \otimes 4}, \boldsymbol{\kappa}_4 \right).$$

Remark 1. It may be noted that the idea used in our (2.6), namely integrating $\mathbf{u} \left(\mathbf{u}^{\top \otimes 4} \boldsymbol{\kappa}_Y \right)$ over the unit sphere, will not work for the kurtosis since it can be verified that this will result in a zero vector.

Example 11. Kollo (2008) introduces the kurtosis matrix $\mathbf{B}(\mathbf{Y})$ as

$$\begin{aligned} \mathbf{B}(\mathbf{Y}) &= \sum_{i,j=1}^d \mathbb{E} Y_i Y_j \mathbf{Y}\mathbf{Y}^\top = \mathbb{E} \sum_{i,j=1}^d Y_i Y_j \mathbf{Y}\mathbf{Y}^\top = \mathbb{E} \left(\sum_{j=1}^d Y_j \right)^2 \mathbf{Y}\mathbf{Y}^\top \\ &= \mathbb{E} \left[\mathbf{1}_{d^2}^\top (\mathbf{Y} \otimes \mathbf{Y}) \right] \mathbf{Y}\mathbf{Y}^\top. \end{aligned}$$

Then $\text{vec } \mathbf{B}(\mathbf{Y}) = \text{E} \left(\sum_{j=1}^d Y_j \right)^2 (\mathbf{Y} \otimes \mathbf{Y})$, which can be written as

$$\begin{aligned} \text{vec } \mathbf{B}(\mathbf{Y}) &= \text{E} \left[\mathbf{1}_{d^2}^\top (\mathbf{Y} \otimes \mathbf{Y}) \right] \text{vec } \mathbf{Y} \mathbf{Y}^\top = \text{E} \mathbf{Y}^{\otimes 2} \left[\mathbf{1}_{d^2}^\top (\mathbf{Y} \otimes \mathbf{Y}) \right] \\ &= \left(\mathbf{I}_{d^2} \otimes \mathbf{1}_{d^2}^\top \right) \text{E} \mathbf{Y}^{\otimes 4} = \left(\mathbf{I}_{d^2} \otimes \mathbf{1}_{d^2}^\top \right) \left(\boldsymbol{\kappa}_4 + \mathbf{K}_{2,2} (\text{vec } \mathbf{I}_d)^{\otimes 2} \right). \end{aligned}$$

4 Alternative Measures Based on Distinct Elements of the Cumulant Vectors

The skewness $\boldsymbol{\kappa}_3$ and the kurtosis $\boldsymbol{\kappa}_4$ vectors contain d^3 and d^4 elements respectively, which are not all distinct. Just as the covariance matrix of a d -dimensional vector contains only $d_2 = d(d+1)/2$ distinct elements, a simple computation shows that $\boldsymbol{\kappa}_3$ contains $d_3 = d(d+1)(d+2)/6$ distinct elements, while $\boldsymbol{\kappa}_4$ contains $d_4 = d(d+1)(d+2)(d+3)/24$ distinct elements.

Similar to the fact that there are many applications as well as measures which consider only the distinct elements of a covariance matrix, it is quite sensible and reasonable to follow this approach and define skewness and kurtosis measures based on just the distinct elements of the corresponding cumulant vectors. For example, in estimating the “total skewness” index $\|\boldsymbol{\kappa}_3\|^2$ discussed in Mardia (1970), one may use the “elimination matrix” since the terms in the summation are symmetric like in a covariance matrix.

Selection of the distinct elements from the vectors $\boldsymbol{\kappa}_3$ and $\boldsymbol{\kappa}_4$ can be accomplished via linear transformations. This approach can be traced back to Magnus and Neudecker (1980) who introduce two transformation matrices, \mathbf{L} and \mathbf{D} , which consist of zeros and ones. For any (n, n) arbitrary matrix \mathbf{A} , \mathbf{L} eliminates from $\text{vec}(\mathbf{A})$ the supra-diagonal elements of \mathbf{A} , while \mathbf{D} performs the reverse transformation for a symmetric \mathbf{A} .

In the case of a covariance matrix $\boldsymbol{\Sigma}$ of a vector \mathbf{X} , the elimination matrix \mathbf{L} above is a matrix acting on $\text{vec}(\boldsymbol{\Sigma})$ and in the approach defined in this paper it holds that $\text{vec}(\boldsymbol{\Sigma}) = \text{Cum}_2(\mathbf{X}, \mathbf{X})$, i.e. the distinct elements of $\text{vec}(\boldsymbol{\Sigma})$ correspond to the distinct elements of a tensor product $\mathbf{X} \otimes \mathbf{X}$. In this way the elimination matrix \mathbf{L} can be generalized to tensor products of higher orders in a simple way.

We shall use elimination matrices $\mathbf{G}(3, d)$ and $\mathbf{G}(4, d)$, for shortening the skewness and kurtosis vectors respectively and keeping just the distinct entries in them. See Meijer (2005) for details where the notations \mathbf{T}_3^+ and \mathbf{T}_4^+ are used respectively, which are actually the Moore–Penrose inverses of triplication and quadruplication matrices. This gives cumulant vectors of distinct elements as

$$\boldsymbol{\kappa}_{3,D} = \mathbf{G}(3, d) \boldsymbol{\kappa}_3 \text{ and } \boldsymbol{\kappa}_{4,D} = \mathbf{G}(4, d) \boldsymbol{\kappa}_4.$$

In such a case, the distinct element vector has dimension $\dim(\boldsymbol{\kappa}_{3,D}) = d(d+1)(d+2)/6$. For instance, when $d = 2$, $\dim(\boldsymbol{\kappa}_{3,D})$ is 50% of the $\dim(\boldsymbol{\kappa}_3)$, and in general, the percentage of distinct elements decreases in the proportion $(1 + 1/d)(1 + 2/d)/6$, getting close to $1/6$ for large d . Similarly, the fraction of distinct elements in $\boldsymbol{\kappa}_{4,D}$ relative to $\boldsymbol{\kappa}_4$ approaches $1/24$ for large d — a significant reduction.

Following the discussion of this section one could define indexes of total skewness and total kurtosis exploiting the square norms of the skewness and kurtosis vectors containing only the distinct elements, i.e.

$$\kappa_{3,D} = \|\boldsymbol{\kappa}_{3,D}\|^2 = \|\mathbf{G}(3, d) \boldsymbol{\kappa}_3\|^2 \quad \text{and} \quad \kappa_{4,D} = \|\boldsymbol{\kappa}_{4,D}\|^2 = \|\mathbf{G}(4, d) \boldsymbol{\kappa}_4\|^2. \tag{4.1}$$

In Section 7 some numerical evidence about the performance of these indexes, which eliminate duplication of information, will be given.

5 Multivariate Symmetric Distributions

Two important classes of symmetric distributions are the spherically symmetric distributions and the elliptically symmetric distributions. In this section, we discuss them in turn and derive their cumulants. Some related results and discussion may be found in Fang et al. (2017).

5.1. *Multivariate Spherically Symmetric Distributions* A d -vector $\mathbf{W} = (W_1, \dots, W_d)^\top$ has a *spherically* symmetric distribution if that distribution is invariant under the group of rotations in \mathbb{R}^d . This is equivalent to saying that \mathbf{W} has the stochastic representation

$$\mathbf{W} = R\mathbf{U}, \tag{5.1}$$

where R is a non negative random variable, $\mathbf{U} = (U_1, \dots, U_d)^\top$ is uniform on the sphere \mathbb{S}_{d-1} , and R and \mathbf{U} are *independent* (see e.g. Fang et al., 2017, Theorem 2.5). The moments of the components of \mathbf{W} , when they exist, can be expressed in terms of a one-dimensional integral, Fang et al., (2017 Theorem 2.8, p. 34) and the characteristic function has the form

$$\phi_{\mathbf{W}}(\boldsymbol{\lambda}) = g(\boldsymbol{\lambda}^\top \boldsymbol{\lambda}), \tag{5.2}$$

where g is called the “characteristic generator” and R is the generating variate, with say, a generating distribution F . The relationship between the distribution F of R and g is given through the characteristic function of the uniform distribution on the sphere (see Fang et al., 2017, p. 30).

The marginal distributions of any such spherically or elliptically symmetric distributions have zero skewness, and the same kurtosis value given by the “kurtosis parameter” of the form (see Muirhead, 2009, p. 41)

$$\kappa_0 = \frac{g''(0) - g'(0)^2}{g'(0)^2}. \tag{5.3}$$

The next lemma provides the moments of the uniform distribution on \mathbb{S}_{d-1} , and is proved in the [Appendix](#).

Lemma 1. *Let \mathbf{U} be uniform on sphere \mathbb{S}_{d-1} . Then*

1. *For odd-order moments*

$$\mathbb{E} \mathbf{U}^{\otimes(2k+1)} = 0, \quad \underline{\text{Cum}}_{2k+1}(\mathbf{U}) = 0,$$

while for even-order moments

$$\mathbb{E} \prod_{i=1}^d U_i^{2k_i} = \frac{1}{(d/2)_k} \prod_{i=1}^d \frac{(2k_i)!}{2^{2k_i} k_i!}. \tag{5.4}$$

For T-products we have

$$\mathbb{E} \mathbf{U}^{\otimes 4} = \frac{1}{d(d+2)} (3\mathbf{e}(d^3, d) + \mathbf{e}(d^4)), \tag{5.5}$$

where zero-one vectors $\mathbf{e}(d^3, d)$ and $\mathbf{e}(d^4)$ are given by Eq. 1.8 and 1.9 respectively, more over the sum of the entries is

$$\sum \mathbb{E} \mathbf{U}^{\otimes 4} = \frac{3d}{d+2}.$$

2. *The odd moments of the modulus*

$$\mathbb{E} \prod_{i=1}^d |U_i|^{k_i} = \sqrt{\frac{1}{\pi^{d_1}}} \frac{1}{G_k} \prod_{i=1}^{d_1} \Gamma((k_i + 1)/2) \tag{5.6}$$

where $k = \sum k_i$, see Eq. 6.3 for G_k , and d_1 is the number of nonzero k_i , in particular

$$\mathbb{E} |U_i|^{2k+1} = \sqrt{\frac{1}{\pi}} \frac{k!}{G_{2k+1}}.$$

For T-products we have

$$\mathbb{E} |\mathbf{U}|^{\otimes 2} = \frac{1}{d} \text{vec } \mathbf{I}_d + \frac{1}{\pi} \frac{1}{G_2} (\mathbf{1}_{d^2} - \text{vec } \mathbf{I}_d)$$

and

$$E|\mathbf{U}|^{\otimes 3} = \sqrt{\frac{1}{\pi}} \frac{1}{G_3} \left(\mathbf{1}_{d^3}(3) + \frac{1}{2} \mathbf{1}_{d^3}(1,2) + \frac{1}{\pi} \mathbf{1}_{d^3}(1,1,1) \right),$$

where the zero-one vectors are given in Eq. 1.10 and

$$\begin{aligned} E|\mathbf{U}|^{\otimes 4} &= \frac{3}{d(d+2)} \mathbf{1}_{d^4}(4) + \frac{1}{4G_4} \mathbf{1}_{d^4}(2,2) \\ &\quad + \frac{1}{\pi} \frac{1}{G_4} \left(\mathbf{1}_{d^4}(1,3) + \sqrt{\frac{1}{\pi}} \mathbf{1}_{d^4}(2,1,1) + \frac{1}{\pi} \mathbf{1}_{d^4}(1,1,1,1) \right) \end{aligned}$$

where the zero-one vectors are given in Eq. 1.11.

Remark 2. Observe that there are only three distinct elements of the third order moments, and four distinct elements of the fourth order moments!

The next lemma provides the cumulants of \mathbf{W} (proof in the [Appendix](#)).

Lemma 2. Let \mathbf{W} be spherically symmetric with the representation $\mathbf{W} = R\mathbf{U}$ and characteristic generator g having second derivative at 0; then $\underline{\text{Cum}}_1(\mathbf{W}) = \underline{\text{Cum}}_3(\mathbf{W}) = 0$, $\underline{\text{Cum}}_2(\mathbf{W}) = -2g'(0) \text{vec } \mathbf{I}_d$, and

$$\underline{\text{Cum}}_4(\mathbf{W}) = 4 \left(g''(0) - g'(0)^2 \right) \mathbf{K}_{2,2}(\text{vec } \mathbf{I}_d)^{\otimes 2}.$$

In terms of kurtosis parameter κ_0 , we have $\kappa_4 = 3d(d+2)\kappa_0^2$, $\kappa_{4,D} = \kappa_0^2(9d + d(d-1)/2)$ and Mardia's $\beta_{2,d} = d(d+2)(\kappa_0 + 1)$.

From the representation of \mathbf{W} given in Eq. 5.1, it is easy to see that $\underline{\text{Cum}}_{2\ell+1}(\mathbf{W}) = 0$, $\ell = 0, 1, \dots$, while the second and fourth order cumulants are calculated directly using the Lemmas 1 and 2 above, so that we have the following

Theorem 1. If \mathbf{W} is spherically distributed with the representation $\mathbf{W} = R\mathbf{U}$ and $E(R^4) < \infty$, then $\underline{\text{Cum}}_{2\ell+1}(\mathbf{W}) = 0$, $\underline{\text{Cum}}_2(\mathbf{W}) = \frac{ER^2}{d} \text{vec } \mathbf{I}_d$ and

$$\underline{\text{Cum}}_4(\mathbf{W}) = ER^4 E\mathbf{U}^{\otimes 4} - 3 \left(\frac{ER^2}{d} \right)^2 (\text{vec } \mathbf{I}_d)^{\otimes 2}.$$

In terms of these moments, the kurtosis parameter becomes

$$\kappa_0 = \left(\frac{d}{d+2} \frac{ER^4}{(ER^2)^2} - 1 \right).$$

5.2. *Multivariate elliptically symmetric distributions* A d -vector \mathbf{X} has an elliptically symmetric distribution if it has the representation

$$\mathbf{X} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{W}$$

where $\boldsymbol{\mu} \in \mathbb{R}^d$, $\boldsymbol{\Sigma}$ is a variance-covariance matrix and \mathbf{W} is spherically distributed. Hence the cumulants of \mathbf{X} are just constant times the cumulants of \mathbf{W} except for the mean i.e.

$$\underline{\text{Cum}}_m(\mathbf{X}) = \left(\boldsymbol{\Sigma}^{1/2}\right)^{\otimes m} \underline{\text{Cum}}_m(\mathbf{W}),$$

from which one gets $\underline{\text{Cum}}_1(\mathbf{X}) = \boldsymbol{\mu}$, $\underline{\text{Cum}}_{2\ell+1}(\mathbf{X}) = 0$, $\ell = 1, \dots$ and

$$\underline{\text{Cum}}_{2\ell}(\mathbf{X}) = \left(\boldsymbol{\Sigma}^{1/2}\right)^{\otimes 2\ell} \underline{\text{Cum}}_{2\ell}(\mathbf{W}).$$

Moments of elliptically symmetric distributions have also been discussed by Berkane and Bentler (1986). As a special case, we now discuss

5.2.1. *Multivariate t -distribution.* Multivariate t -distribution is spherically symmetric (see Example 2.5 Fang et al., 2017, p.32). Consult also the monograph by Kotz and Nadarajah (2004) for further details. Let

$$\mathbf{W} = \sqrt{\frac{m}{S^2}}\mathbf{Z}$$

where $\mathbf{Z} \in \mathcal{N}_d(0, \mathbf{I}_d)$ standard normal, and S^2 is χ^2 distributed with m degrees of freedom. $\mathbf{W} \in Mt_d(m, 0, \mathbf{I}_d)$, we have

$$\mathbf{W} = \sqrt{m} \frac{\|\mathbf{Z}\|}{S} \mathbf{Z} / \|\mathbf{Z}\| = R^* \mathbf{U}, \tag{5.7}$$

where $R^* = \sqrt{m} \|\mathbf{Z}\| / S$, and R^{*2}/d has an F-distribution with d and m degrees of freedom. Let $\boldsymbol{\mu} \in \mathbb{R}^d$, and \mathbf{A} is an $d \times d$ matrix and $\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}^\top \mathbf{W}$ then $\mathbf{X} \in Mt_d(m, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \mathbf{A}^\top \mathbf{A}$, hence \mathbf{X} is an elliptically symmetric random variable. Since the characteristic function is quite complicated (see Fang et al., 2017 Section 3.3.6 p. 85), we utilize the stochastic representation given in Eq. 5.7 for deriving the kurtosis (note that the skewness is zero). The proof of the next Lemma is in the [Appendix](#).

Lemma 3. *Let \mathbf{W} be a multivariate t -vector with dimension d and degrees of freedom m , with $m > 4$, then $E\mathbf{W} = 0$, $\underline{\text{Cum}}_2(\mathbf{W}) = \frac{m}{m-2} \text{vec } \mathbf{I}_d$ and $\underline{\text{Cum}}_3(\mathbf{W}) = 0$. From Theorem 1, the kurtosis parameter in Eq. 5.3 becomes*

$\kappa_0 = \frac{2}{m-4}$, and the kurtosis $\kappa_4 = 3d^2\kappa_0$. Moreover, if $\mathbf{X} \in Mt_d(m, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \mathbf{A}^\top \mathbf{A}$, then $E \mathbf{X} = \boldsymbol{\mu}$,

$$\underline{\text{Cum}}_2(\mathbf{X}) = \frac{m}{m-2} \text{vec } \boldsymbol{\Sigma}, \quad \underline{\text{Cum}}_3(\mathbf{X}) = 0,$$

and kurtosis parameter κ_0 and kurtosis κ_4 are the same as for W .

6. Multivariate Skew Distributions

Starting with Azzalini and Dalla Valle (1996) who suggest methods for obtaining multivariate skew-normal distributions, several authors discuss different approaches for obtaining asymmetric multivariate distributions, by skewing a spherically or an elliptically symmetric distribution. We mention here (Branco and Dey, 2001) who extend the work in Azzalini and Dalla Valle (1996) to multivariate skew-elliptical distribution, Arnold and Beaver (2002) who use a conditioning approach on a d -dimensional random vector to an elliptically contoured distribution (although this conditioning is not strictly necessary), Sahu et al. (2003) who use transformation and conditioning techniques, Dey and Liu (2005) who discuss an approach based on linear constraints, Genton and Loperfido (2005) who introduce a general class of multivariate skew-elliptical distributions– the so-called multivariate generalized skew-elliptical (GSE) distribution. See also Genton (2004) and the references therein.

Here we provide a systematic treatment of several skew-multivariate distributions by providing general formulae for cumulant vectors up to the fourth order, which are needed in deriving the corresponding skewness and kurtosis measures discussed in Sections 2 and 3.

6.1. *Multivariate skew spherical distributions* Let $\mathbf{Z} = [\mathbf{Z}_1^\top, \mathbf{Z}_2^\top]^\top$ be spherically symmetric distributed in dimension $(m+d)$. Define the canonical fundamental skew-spherical (CFSS) distribution (Arellano-Valle and Genton 2005, Prop. 3.3) by

$$\mathbf{X} = \boldsymbol{\Delta} |\mathbf{Z}_1| + \left(\mathbf{I}_d - \boldsymbol{\Delta} \boldsymbol{\Delta}^\top \right)^{1/2} \mathbf{Z}_2, \tag{6.1}$$

where the modulus is taken element-wise, and $\boldsymbol{\Delta}$ is the $d \times m$ skewness matrix, and let R be the generating variate. A simple construction of $\boldsymbol{\Delta}$ is given by $\boldsymbol{\Delta} = \boldsymbol{\Lambda} (\mathbf{I}_m + \boldsymbol{\Lambda}^\top \boldsymbol{\Lambda})^{-1/2}$ with some real matrix $\boldsymbol{\Lambda}$ of dimension $d \times m$. If $\mathbf{Z} = R\mathbf{U}$, then $\mathbf{Z}_1 = p_1 R \mathbf{U}_1 \in \mathbb{R}^m$ and $\mathbf{Z}_2 = p_2 R \mathbf{U}_2 \in \mathbb{R}^d$, then p_1^2 is $Beta(m/2, d/2)$, $p_2^2 = 1 - p_1^2$, which is $Beta(d/2, m/2)$. The variables R, p_1^2, \mathbf{U}_1 and \mathbf{U}_2 are independent by Theorem 2.6 in Fang et al. (2017).

Then we have $E\mathbf{Z} = 0$ and $\text{Cov}(\mathbf{Z}) = ER^2\mathbf{I}/d$. Introduce the function

$$G_{\ell_1, \ell_2}(m, d) = \frac{B(m/2 + \ell_1, d/2 + \ell_2)}{B(m/2, d/2)}. \tag{6.2}$$

and,

$$G_k(d) = \frac{\Gamma((d+k)/2)}{\Gamma(d/2)} \tag{6.3}$$

written as G_k for short. We may express the joint moments of p_1, p_2 as

$$E p_1^{\ell_1} p_2^{\ell_2} = \frac{1}{B(m/2, d/2)} \int_0^1 x^{m/2+\ell_1-1} (1-x)^{d/2+\ell_2-1} dx = G_{\ell_1, \ell_2}(m, d). \tag{6.4}$$

In particular $E p_1 = G_{1,0}(m, d)$, $E p_1 p_2 = G_{1,1}(m, d)$. The cumulants of \mathbf{X} can be obtained from these moments. Using Eq. 6.4 above, and Lemma (1) for the moments of p_k and $|\mathbf{U}_j|$, we obtain the following result (see the Appendix for the proof).

Theorem 2. *Let the d -vector \mathbf{X} have a CFSS distribution as defined in Eq. 6.1 and denote by $\mathbf{V}_2 = (\mathbf{I} - \Delta\Delta^\top)^{1/2} \mathbf{Z}_2$ for ease of notation, with*

$$E \mathbf{V}_2^{\otimes 2} = \frac{1}{d} (\text{vec } \mathbf{I}_d - \Delta^{\otimes 2} \text{vec } \mathbf{I}_m).$$

The moments of \mathbf{X} , assuming $E(R^4) < \infty$, are given by:

$$E \mathbf{X} = \Delta E p_1 E R \mathbf{1}_m,$$

$$E \mathbf{X}^{\otimes 2} = E p_1^2 E R^2 \Delta^{\otimes 2} E |\mathbf{U}_1|^{\otimes 2} + E (p_2 R)^2 E \mathbf{V}_2^{\otimes 2},$$

$$E \mathbf{X}^{\otimes 3} = E p_1^3 E R^3 E |\mathbf{U}_1|^{\otimes 3} + E p_1 p_2^2 E R^3 \mathbf{K}_{2,1} (\mathbf{I}_{d^2} \otimes \Delta) (E \mathbf{V}_2^{\otimes 2} \otimes E |\mathbf{U}_1|),$$

$$E \mathbf{X}^{\otimes 4} = E p_1^4 E R^4 \Delta^{\otimes 4} E |\mathbf{U}_1|^{\otimes 4} + E p_1^2 p_2^2 E R^4 \mathbf{K}_{2,1,1} (\Delta^{\otimes 2} \otimes \mathbf{I}_{m^2}) \times \\ \times (E |\mathbf{U}_1|^{\otimes 2} \otimes E \mathbf{V}_2^{\otimes 2}) + (E p_2^4 E R^4) \Delta_1^{\otimes 4} E \mathbf{U}_2^{\otimes 4},$$

where $\Delta_1 = (\mathbf{I}_d - \Delta\Delta^\top)^{1/2}$, with the commutators as given in Eqs. 1.1 and 1.2. The cumulants of \mathbf{X} are obtained by using the relations in Section 1.2.

Remark 3. *It can be seen that the cumulants depend on Δ , and the moments of the generating variate R . For instance*

$$\text{Cum}_2(\mathbf{X}) = E R^2 \mathbf{D}_2(m, d, \Delta) - (ER)^2 (E p_1)^2 (E |\mathbf{U}_1|)^{\otimes 2}$$

where

$$\begin{aligned} \mathbf{D}_2(m, d, \mathbf{\Delta}) &= \frac{G_{2,0}}{m} \mathbf{\Delta}^{\otimes 2} \left(\text{vec } \mathbf{I}_m + \frac{1}{\pi} \frac{1}{G_2(m)} (\mathbf{1}_{m^2} - \text{vec } \mathbf{I}_m) \right) \\ &\quad + \frac{G_{0,2}}{d} (\mathbf{I}_{d^2} - \mathbf{\Delta}^{\otimes 2}) \text{vec } \mathbf{I}_m. \end{aligned}$$

6.2. *Multivariate Skew-t Distribution* This distribution goes back to Azzalini and Capitanio (2003); consult also Kim and Mallick (2003) for a derivation of moments up to the 4th order and moments of quadratic forms of the multivariate skew t-distribution. Let \mathbf{X} be a d -dimensional vector having multivariate skew-normal distribution, $SN_d(0, \mathbf{\Omega}, \boldsymbol{\alpha})$ (see Section 6.3 below for details), and S^2 be a random variable which follows a χ^2 distribution with m degrees of freedom. Then the random vector

$$\mathbf{V} = \boldsymbol{\mu} + \frac{\sqrt{m}}{S} \mathbf{X}$$

has a multivariate skew t-distribution denoted by $St_d(\boldsymbol{\mu}, \mathbf{\Omega}, \boldsymbol{\alpha}, m)$. Derivation of the cumulants of \mathbf{V} up to the 4th order are provided in the next theorem, with detailed proof given in the Appendix.

Theorem 3. *Let $\mathbf{V} \in St_d(\boldsymbol{\mu}, \mathbf{\Omega}, \boldsymbol{\alpha}, m)$ with $m > 4$, then its first four cumulants are:*

$$\underline{\text{Cum}}_1(\mathbf{V}) = \mathbf{E} \mathbf{V} = \boldsymbol{\mu} + G_1(m) \sqrt{\frac{m}{\pi}} \boldsymbol{\delta},$$

$$\underline{\text{Cum}}_2(\mathbf{V}) = \frac{m}{m-2} \text{vec } \mathbf{\Omega} - \frac{m}{\pi} (G_1(m))^2 \boldsymbol{\delta}^{\otimes 2},$$

$$\begin{aligned} \underline{\text{Cum}}_3(\mathbf{V}) &= \sqrt{\frac{2}{\pi}} \left(\frac{m}{2}\right)^{3/2} G_1(m) \times \\ &\quad \left(\left(\frac{4}{\pi} G_1(m)^2 - \frac{2}{m-3}\right) \boldsymbol{\delta}^{\otimes 3} + \frac{2}{(m-3)(m-2)} \mathbf{K}_{2,1}(\text{vec } \mathbf{\Omega} \otimes \boldsymbol{\delta}) \right), \end{aligned}$$

$$\begin{aligned} \underline{\text{Cum}}_4(\mathbf{V}) &= \frac{4}{\pi} \left(\frac{m}{2}\right)^2 G_{-1}(m)^2 \left(\frac{4}{m-3} - \frac{6}{\pi} G_{-1}(m)^2\right) \boldsymbol{\delta}^{\otimes 4} \\ &\quad + \left(\frac{m}{2}\right)^2 \frac{8}{(m-4)(m-2)^2} \mathbf{K}_{2,2}((\text{vec } \mathbf{\Omega})^{\otimes 2}) \\ &\quad - \frac{8}{\pi} \left(\frac{m}{2}\right)^2 G_{-1}(m)^2 \frac{1}{(m-3)(m-2)} \mathbf{K}_{2,1,1}(\text{vec } \mathbf{\Omega} \otimes \boldsymbol{\delta}^{\otimes 2}). \end{aligned}$$

From Theorem 3 the skewness and kurtosis vectors are $\kappa_3 = (\Sigma_V^{-1/2})^{\otimes 3} \underline{\text{Cum}}_3(\mathbf{V})$, and $\kappa_4 = (\Sigma_V^{-1/2})^{\otimes 4} \underline{\text{Cum}}_4(\mathbf{V})$, where

$$\Sigma_V = \text{Var } \mathbf{V} = \frac{m}{m-2} \mathbf{\Omega} - \frac{m}{\pi} (G_1(m))^2 \boldsymbol{\delta} \boldsymbol{\delta}^\top.$$

6.3. Multivariate Skew-Normal Distribution Consider the multivariate skew-normal distribution introduced by Azzalini and Dalla Valle (1996), whose marginal densities are scalar skew-normals. A d -dimensional random vector \mathbf{X} is said to have a multivariate skew-normal distribution, $\text{SN}_d(\boldsymbol{\mu}, \mathbf{\Omega}, \boldsymbol{\alpha})$ with shape parameter $\boldsymbol{\alpha}$ if it has the density function

$$2\varphi(\mathbf{X}; \boldsymbol{\mu}, \mathbf{\Omega}) \Phi\left(\boldsymbol{\alpha}^\top (\mathbf{X} - \boldsymbol{\mu})\right), \quad \mathbf{X} \in \mathbb{R}^d.$$

where $\varphi(\mathbf{X}; \boldsymbol{\mu}, \mathbf{\Omega})$ is the d -dimensional normal density with mean $\boldsymbol{\mu}$ and correlation matrix $\mathbf{\Omega}$; here φ and Φ denote the univariate standard normal density and the cdf. The cumulant function of $\text{SN}_d(0, \mathbf{\Omega}, \boldsymbol{\alpha})$ is given by

$$C_{\mathbf{X}}(\boldsymbol{\lambda}) = \log 2 - \frac{1}{2} \boldsymbol{\lambda}^\top \mathbf{\Omega} \boldsymbol{\lambda} + \log \Phi\left(i \boldsymbol{\delta}^\top \boldsymbol{\lambda}\right),$$

where

$$\boldsymbol{\delta} = \frac{1}{(1 + \boldsymbol{\alpha}^\top \mathbf{\Sigma} \boldsymbol{\alpha})^{1/2}} \mathbf{\Omega} \boldsymbol{\alpha} \quad \text{and} \quad \boldsymbol{\alpha} = \frac{1}{(1 - \boldsymbol{\delta}^\top \mathbf{\Sigma}^{-1} \boldsymbol{\delta})^{1/2}} \mathbf{\Omega}^{-1} \boldsymbol{\delta}. \quad (6.5)$$

Note that the cumulants of order higher than 2 do not depend on $\mathbf{\Omega}$ but only on $\boldsymbol{\delta}$. Here we use the approach discussed in this paper to get explicit expressions for the cumulants of the multivariate SN distribution. See also Genton et al. (2001) and Kollo et al. (2018) for moments of SN and its quadratic forms. The proof of the next Lemma is similar to that of Lemma 5 that is coming later, and is omitted.

Lemma 4. *The cumulants of the multivariate skew-normal distribution, $\text{SN}_d(\boldsymbol{\mu}, \mathbf{\Omega}, \boldsymbol{\alpha})$ are the following:*

$$\underline{\text{Cum}}_1(\mathbf{X}) = \text{E } \mathbf{X} = \sqrt{\frac{2}{\pi}} \boldsymbol{\delta}, \quad \text{and} \quad \underline{\text{Cum}}_2(\mathbf{X}) = \text{vec } \mathbf{\Omega} - \frac{2}{\pi} \boldsymbol{\delta}^{\otimes 2},$$

while for $k = 3, 4, \dots$, $\text{Cum}_k(\mathbf{X}) = c_k \boldsymbol{\delta}^{\otimes k}$. In particular

$$c_3 = 2 \left(\sqrt{\frac{2}{\pi}} \right)^3 - \sqrt{\frac{2}{\pi}}, \quad c_4 = -6 \left(\sqrt{\frac{2}{\pi}} \right)^4 + 4 \left(\sqrt{\frac{2}{\pi}} \right)^2,$$

$$c_5 = 24 \left(\sqrt{\frac{2}{\pi}} \right)^5 - 20 \left(\sqrt{\frac{2}{\pi}} \right)^3 + 3\sqrt{\frac{2}{\pi}}.$$

From this Lemma one observes that $\boldsymbol{\delta} = \sqrt{\pi/2} \boldsymbol{\mu}_{\mathbf{X}}$, from which $\text{Cum}_3(\mathbf{X}) = (2 - \frac{\pi}{2}) \boldsymbol{\mu}_{\mathbf{X}}^{\otimes 3}$. Hence Mardia’s skewness measure becomes

$$\beta_{1,d} = \left(\frac{4 - \pi}{2} \right)^2 \left\| \boldsymbol{\mu}_{\mathbf{X}}^{\otimes 3} \right\|^2 = \left(\frac{4 - \pi}{2} \right)^2 \left\| \boldsymbol{\mu}_{\mathbf{X}} \right\|^6$$

and the kurtosis measure $\beta_{2,d} = (2\pi - 6)^2 (\text{vec } \mathbf{I}_{d^2})^\top \boldsymbol{\mu}_{\mathbf{X}}^{\otimes 4} + d(d + 2)$.

6.4. Canonical Fundamental Skew-Normal (CFUSN) Distribution

Arellano-Valle and Genton (2005) introduced the CFUSN distribution (cf. their Proposition 2.3), to include all existing definitions of SN distributions. The marginal stochastic representation of \mathbf{X} with distribution $\text{CFUSN}_{d,m}(\boldsymbol{\Delta})$ is given by

$$\mathbf{X} = \boldsymbol{\Delta} |\mathbf{Z}_1| + \left(\mathbf{I}_d - \boldsymbol{\Delta} \boldsymbol{\Delta}^\top \right)^{1/2} \mathbf{Z}_2 \tag{6.6}$$

where $\boldsymbol{\Delta}$, is the $d \times m$ skewness matrix such that $\|\boldsymbol{\Delta} \mathbf{a}\| < 1$, for all $\|\mathbf{a}\| = 1$, and $\mathbf{Z}_1 \in \mathcal{N}(0, \mathbf{I}_m)$ and $\mathbf{Z}_2 \in \mathcal{N}(0, \mathbf{I}_d)$ are independent (Proposition 2.2. (Arellano-Valle and Genton, 2005)). A simple construction of $\boldsymbol{\Delta}$ is $\boldsymbol{\Delta} = \boldsymbol{\Lambda} (\mathbf{I}_m + \boldsymbol{\Lambda}^\top \boldsymbol{\Lambda})^{-1/2}$ with some real matrix $\boldsymbol{\Lambda}$ with $d \times m$. The $\text{CFUSN}_{d,m}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$ can be defined via the linear transformation $\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2} \mathbf{X}$. We then have the following

$$\mathbf{E} \mathbf{X} = \boldsymbol{\Delta} \mathbf{E} |\mathbf{Z}_1| = \sqrt{\frac{2}{\pi}} \boldsymbol{\Delta} \mathbf{1}_m,$$

$$\text{Var } \mathbf{X} = \boldsymbol{\Delta} \text{Var} |\mathbf{Z}_1| \boldsymbol{\Delta}^\top + \left(\mathbf{I}_d - \boldsymbol{\Delta} \boldsymbol{\Delta}^\top \right)^{1/2} \left(\mathbf{I}_d - \boldsymbol{\Delta} \boldsymbol{\Delta}^\top \right)^{1/2},$$

$$\text{Var} |\mathbf{Z}_1| = \mathbf{I}_m - \frac{2}{\pi} \mathbf{I}_m, \quad \text{Var } \mathbf{X} = \mathbf{I}_d - \frac{2}{\pi} \boldsymbol{\Delta} \boldsymbol{\Delta}^\top.$$

The cumulant function of a $\text{CFUSN}_{d,m}(0, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$ is

$$C_{\mathbf{X}}(\boldsymbol{\lambda}) = m \log 2 - \frac{1}{2} \boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda} + \log \Phi_m(i \boldsymbol{\Delta}^\top \boldsymbol{\lambda}),$$

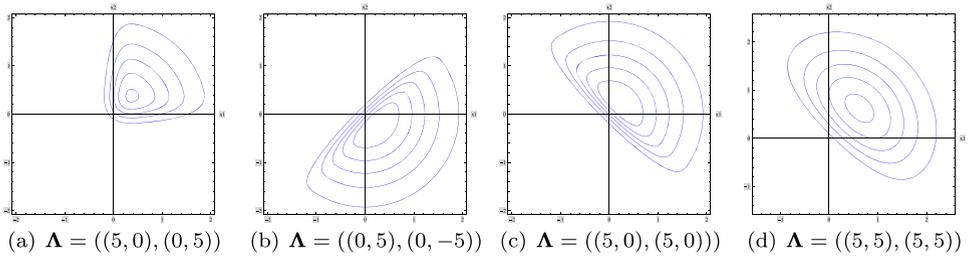


Figure 1: Contour plots of the densities of $CFUSN_{2,2}(0, \mathbf{I}_2, \mathbf{\Delta})$ distribution with generating $\mathbf{\Delta}$ as indicated above

where Φ_m denotes the standard normal distribution function of dimension m . Note again that the cumulants of order higher than 2 do not depend on $\mathbf{\Sigma}$ but just on $\mathbf{\Delta}$. The proof of the next Lemma is given in the [Appendix](#).

Lemma 5. *The cumulants of $CFUSN_{d,m}(0, \mathbf{\Sigma}, \mathbf{\Delta})$ are given by*

$$\underline{\text{Cum}}_1(\mathbf{X}) = E\mathbf{X} = \sqrt{\frac{2}{\pi}}\mathbf{\Delta}\mathbf{1}_m, \tag{6.7}$$

$$\underline{\text{Cum}}_2(\mathbf{X}) = \text{vec}\mathbf{\Sigma} - \frac{2}{\pi}\mathbf{\Delta}^{\otimes 2}\text{vec}\mathbf{I}_m, \tag{6.8}$$

Table 1: Skewness and kurtosis measures for the bivariate CFUSN distribution of Fig. 1

| | | Figure 1a | Figure 1b | Figure 1c | Figure 1d | | | | |
|------|---------------------|-----------|-----------|-----------|-----------|------|------|------|------|
| Skew | κ_3 - Mardia | 1.45 | 0.84 | 0.84 | 0.46 | | | | |
| | $\kappa_{3,D}$ | 1.45 | 0.42 | 0.42 | 0.23 | | | | |
| | MSR | 0.85 | 0.85 | 0.65 | -0.65 | 0.65 | 0.65 | 0.48 | 0.48 |
| | Kollo | 0.85 | 0.85 | 0.00 | 0.00 | 1.30 | 1.30 | 0.96 | 0.96 |
| Kurt | κ_4 | 0.99 | 0.61 | 0.61 | 0.17 | | | | |
| | $\kappa_{4,D}$ | 0.99 | 0.19 | 0.19 | 0.05 | | | | |
| | Mardia | 9.41 | 8.78 | 8.78 | 8.41 | | | | |
| | Koziol | 15.8 | 14.2 | 14.2 | 13.0 | | | | |
| | MSR | 0.71 | 0.00 | 0.39 | -0.39 | 0.39 | 0.39 | 0.21 | 0.21 |
| | Kollo | 4.71 | 2.00 | 4.00 | 2.00 | 4.78 | 2.78 | 4.41 | 2.41 |

and for $k = 3, 4 \dots$

$$\underline{\text{Cum}}_k(\mathbf{X}) = c_k \mathbf{\Delta}^{\otimes k} \text{vec} \left[\mathbf{e}_p^{\otimes k-1} \right]_{p=1:m} . \tag{6.9}$$

where \mathbf{e}_k denotes the k^{th} unit vector in \mathbb{R}^m . Expressions for c_k are provided in Lemma 4.

7 Some Illustrative Examples

Figure 1 provides the contour plots of the density of a CFUSN $_{2,2}(\underline{0}, \mathbf{I}_2, \mathbf{\Delta})$ for selected choices of the generating $\mathbf{\Lambda}$ as described under each figure.

Table 1 reports the values of skewness and kurtosis indexes computed with the help of Lemma 5. These results have been further verified by simulations. The index of Malkovich and Afifi and that of Balakrishnan et al. are not reported as they are equivalent to Mardia and MSR measures respectively.

Among the skewness indexes, note that Kollo’s vector measure is not able to capture the presence of skewness (although it is quite strong) in case b). As far as the kurtosis indexes are concerned, note that the total kurtosis κ_4 and $\kappa_{4,D}$ are quite effective in showing differences among the four cases.

Figure 2 reports the contour plots of the density of St $_2(\underline{0}, \mathbf{I}_2, \underline{\alpha}, 10)$ for some choices of $\underline{\alpha}$ as given under the figure. Table 2 reports the corresponding values of skewness and kurtosis measures

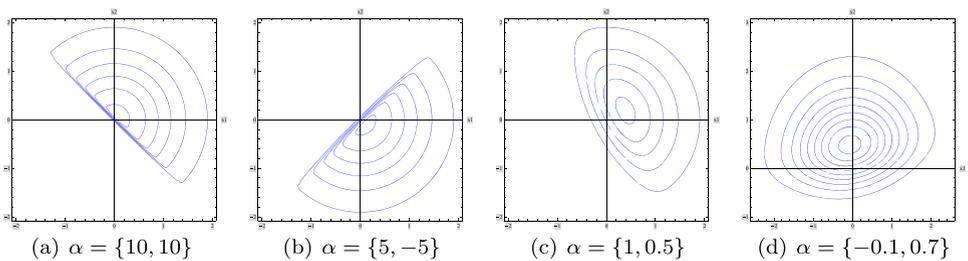


Figure 2: Contour plots of the densities of Skew t distribution with unit covariance matrix and $\underline{\alpha}$ as indicated above

Table 2: Skewness and kurtosis measures for the bivariate Skew-t distribution of Fig. 2

| Skew | κ_3 - Mardia | Figure 2a | Figure 2b | Figure 2c | Figure 2d |
|------|---------------------|-----------|-----------|-----------|-----------|
| | $\kappa_{3,D}$ | 2.17 | 2.00 | 0.26 | 0.09 |
| | MSR | 1.37 | 1.27 | 0.21 | 0.08 |
| | Kollo | 1.14 | 1.10 | -1.10 | -0.05 |
| | κ_4 | 2.04 | 0.24 | 0.74 | 0.09 |
| | $\kappa_{4,D}$ | 13.7 | 12.7 | 3.65 | 2.99 |
| | Mardia | 7.41 | 6.97 | 2.83 | 2.39 |
| | Koziol | 13.2 | 13.1 | 11.1 | 10.8 |
| | MSR | 36.2 | 34.8 | 21.8 | 20.6 |
| | Kollo | 2.62 | 2.54 | -1.14 | 1.35 |
| | | 7.83 | 5.40 | 2.60 | -0.02 |
| | | | | 5.86 | 2.67 |
| | | | | 2.94 | 5.34 |
| | | | | 5.53 | 1.47 |
| | | | | | 5.45 |

8 Concluding Remarks

In this paper we have taken a vectorial approach to express information about skewness and kurtosis of multivariate distributions. This can be achieved by applying a vector derivative operator that we call the T-derivative, to the cumulant generating function. Although some of our methods may appear as similar to some existing results, we demonstrate that they lead to a direct and natural way of expressing higher order cumulants and moments in the multivariate case. This approach can also be easily extended to obtain moments and cumulants beyond the fourth order.

Our careful analysis of existing measures of skewness and kurtosis via the third and fourth order cumulant vectors, reveals some hidden features and relationships among them.

Explicit formulae for cumulant vectors for several distributions have been obtained. Several results are new, such as those in Lemmas 1 and 3, while others complement and complete existing results available in the literature.

The availability of explicit formulae for κ_3 and κ_4 together with available computing formulae for commutators provides a systematic treatment of higher order moments and cumulants for general classes of symmetric and asymmetric multivariate distributions, which are needed in applications, estimation and testing.

Acknowledgments. The work of the third author is partially supported by the Project EFOP3.6.2-16-2017-00015 of the European Union and co-financed by the European Social Fund.

Funding Information. Open access funding provided by Università degli Studi di Trento within the CRUI-CARE Agreement.

Open Access. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

1.1. * Commutation matrices

In order to define the skewness and kurtosis in terms of the moments, some preliminary discussion of what are known as “commutation matrices” is helpful. We use the notation $1 : d$ to denote $1, 2, \dots, d$ and $E_{i,j}$ to represent the $(d \times d)$ elementary matrix, i.e. one whose entries are all zero except for just the i, j^{th} element which is 1.

Note that the vector $\text{vec } E_{i,j}^\top = \text{vec } E_{j,i}$, is the $((i - 1)d + j)^{th}$, $i = 1 : d$, $j = 1 : d$, unit vector of the identity matrix \mathbf{I}_{d^2} . We now define the matrix $\mathbf{K}_{(1,2)}$ by stacking these column vectors $\text{vec } E_{i,j}^\top$ into a matrix according to the order defined by the operator vec , i.e. we follow the column-wise ordering of entries $\text{vec } E_{1,1}^\top$ followed by $\text{vec } E_{1,2}^\top$ etc. into a matrix of dimension $d^2 \times d^2$, (cf. Graham 2081). Then

$$\mathbf{K}_{(1,2)} (\underline{a}_1 \otimes \underline{a}_2) = \underline{a}_2 \otimes \underline{a}_1,$$

which is the reason why it is called a commutation matrix (see e.g. Graham 2081, Magnus and Neudecker 1999). By changing the neighboring elements of a tensor product, we can reach any permutation of them. For instance, if $\mathbf{p} = (i_1, i_2, i_3, i_4)$ is a permutation of the numbers 1,2,3,4, we can introduce the commutator matrix $\mathbf{K}_{\mathbf{p}}$ for changing the order of a tensor product, namely

$$\mathbf{K}_{\mathbf{p}} (\underline{a}_1 \otimes \underline{a}_2 \otimes \underline{a}_3 \otimes \underline{a}_4) = \underline{a}_{i_1} \otimes \underline{a}_{i_2} \otimes \underline{a}_{i_3} \otimes \underline{a}_{i_4}.$$

In particular if we set $\mathbf{p}_1 = (1, 3, 2, 4)$

$$\begin{aligned} & \mathbf{K}_{\mathbf{p}_1} (\underline{a}_1 \otimes \underline{a}_2 \otimes \underline{a}_3 \otimes \underline{a}_4) \\ &= (\mathbf{I}_{d_1} \otimes \mathbf{K}_{(1,2)} \otimes \mathbf{I}_{d_4}) (\underline{a}_1 \otimes \underline{a}_2 \otimes \underline{a}_3 \otimes \underline{a}_4) = \underline{a}_1 \otimes \underline{a}_3 \otimes \underline{a}_2 \otimes \underline{a}_4. \end{aligned}$$

It is worth noting that $\mathbf{K}_{(1,2)}$, and in general each commutator matrix, depends on the dimensions of the vectors under consideration. In our example, the dimension of $\mathbf{K}_{(1,2)}$ is $d^2 \times d^2$, while the dimension of $\mathbf{K}_{\mathbf{p}_1}$ is $d^4 \times d^4$.

Commutator matrices are very useful for expressing the formulae for moments in terms of cumulants and vice versa, as shown in the next Section 1.2.

1.2. * Conversion of moments and cumulants

Since we will be dealing with the cumulants of different variables, from now on, we will index the cumulants by the variable as well and write $\kappa_{X,k} = \text{Cum}_k(\mathbf{X})$. In terms of the commutator matrices, we can write

$$E \mathbf{X}^{\otimes 3} = \kappa_{X,3} + \mathbf{K}_{2,1} (\kappa_{X,2} \otimes \underline{\kappa}_{X,1}) + \kappa_{X,1}^{\otimes 3},$$

where

$$\mathbf{K}_{2,1} = \mathbf{I}_{d^3} + \mathbf{K}_{(1,3,2)}^{-1} (d_{1:3}) + \mathbf{K}_{(3,1,2)}^{-1} (d_{1:3}), \tag{1.1}$$

while

$$E \mathbf{X}^{\otimes 4} = \kappa_{X,4} + \mathbf{K}_{3,1} (\kappa_{X,3} \otimes \underline{\kappa}_{X,1}) + \mathbf{K}_{2,2} \kappa_{X,2}^{\otimes 2} + \mathbf{K}_{2,1,1} (\kappa_{X,2} \otimes \underline{\kappa}_{X,1}^{\otimes 2}) + \kappa_{X,1}^{\otimes 4},$$

where

$$\begin{aligned} \mathbf{K}_{3,1} &= \mathbf{I}_{d^4} + \mathbf{K}_{(1,2,4,3)}^{-1} + \mathbf{K}_{(1,4,2,3)}^{-1} + \mathbf{K}_{(4,1,2,3)}^{-1}, \\ \mathbf{K}_{2,2} &= \mathbf{I}_{d^4} + \mathbf{K}_{(1,3,2,4)}^{-1} + \mathbf{K}_{(1,4,2,3)}^{-1}, \\ \mathbf{K}_{2,1,1} &= \mathbf{I}_{d^4} + \mathbf{K}_{(1,3,2,4)}^{-1} + \mathbf{K}_{(3,1,2,4)}^{-1} + \mathbf{K}_{(3,1,4,2)}^{-1} + \mathbf{K}_{(1,3,4,2)}^{-1} + \mathbf{K}_{(3,4,1,2)}^{-1}. \end{aligned} \tag{1.2}$$

For cumulants we have (see e.g. Jammalamadaka et al., 2006)

$$\begin{aligned} \kappa_{X,2} &= E \mathbf{X}^{\otimes 2} - (E \mathbf{X})^{\otimes 2}, \\ \kappa_{X,3} &= E \mathbf{X}^{\otimes 3} - \mathbf{K}_{2,1} (E \mathbf{X}^{\otimes 2} \otimes E \mathbf{X}) + (E \mathbf{X})^{\otimes 3}, \\ \kappa_{X,4} &= E \mathbf{X}^{\otimes 4} - \mathbf{K}_{3,1} (E \mathbf{X}^{\otimes 3} \otimes E \mathbf{X}) - \mathbf{K}_{2,2} (E \mathbf{X}^{\otimes 2})^{\otimes 2} + \\ &\quad 2\mathbf{K}_{2,1,1} (E \mathbf{X}^{\otimes 2} \otimes (E \mathbf{X})^{\otimes 2}) - 6(E \mathbf{X})^{\otimes 4} \end{aligned}$$

where the commutators are as defined in Eqs. 1.1 and 1.2.

1.3. * Proofs

We start with the proof of Theorem 3, which is one of the central results of the paper. Proofs of the remaining results follow.

PROOF OF THEOREM 3. In order to derive the cumulants of \mathbf{V} , set $S_m = \frac{\sqrt{m}}{S}$ and let

$$\mathbf{W} = \frac{\sqrt{m}}{S} \mathbf{X} = S_m \mathbf{X}.$$

Note that the cumulants of \mathbf{W} are the same as those of \mathbf{V} except for the mean, and that (cf. Johnson et al., 1994, p. 452)

$$\begin{aligned} E \left(\sqrt{S^2/m} \right)^{-k} &= E S_m^k = \frac{(m/2)^{k/2} \Gamma((m-k)/2)}{\Gamma(m/2)}, \\ &= \mu_k(S_m) = \left(\frac{m}{2}\right)^{k/2} G_{-k}(m) \end{aligned} \tag{1.3}$$

where $G_k(m)$ is as given in Eq. 6.3. We have

$$G_{-2}(m) = \frac{2}{m-2}, G_{-3}(m) = \frac{2}{m-3} G_{-1}(m), G_{-4}(m) = \frac{4}{(m-4)(m-2)}.$$

Consider first the computation of mean and variance.

$$\kappa_{V,1} = E \mathbf{V} = \boldsymbol{\mu} + G_{-1}(m) \sqrt{\frac{m}{\pi}} \boldsymbol{\delta},$$

with δ as also α given in Eq. 6.5. From the formula for cumulants of products (see Terdik 2002) we need to compute:

$$\begin{aligned} \kappa_{W,2} &= \text{Cum}_2(S_m \mathbf{X}) \\ &= C_2(S_m) \kappa_{X,2} + C_2(S_m) \kappa_{X,1}^{\otimes 2} + C_1(S_m)^2 \kappa_{X,2}, \end{aligned}$$

where, from Lemma 4,

$$\kappa_{X,1} = E \mathbf{X} = \sqrt{\frac{2}{\pi}} \delta, \kappa_{X,2} = \text{vec } \Sigma - \frac{2}{\pi} \delta^{\otimes 2},$$

and

$$\begin{aligned} C_1(S_m) &= E \left(\sqrt{S^2/m} \right)^{-1} = \sqrt{\frac{m}{2}} G_{-1}(m), \\ C_2(S_m) &= \frac{m}{2} \left(G_{-2}(m) - (G_{-1}(m))^2 \right) = \frac{m}{2} \left(\frac{2}{m-2} - (G_{-1}(m))^2 \right). \end{aligned}$$

For $m > 2$, using the results above

$$\begin{aligned} \kappa_{W,2} &= \left(\frac{m}{m-2} - \frac{m}{2} (G_{-1}(m))^2 \right) \left(\text{vec } \Sigma - \frac{2}{\pi} \delta^{\otimes 2} + \frac{2}{\pi} \delta^{\otimes 2} \right) \\ &\quad + \frac{m}{2} (G_{-1}(m))^2 \left(\text{vec } \Sigma - \frac{2}{\pi} \delta^{\otimes 2} \right) \\ &= \frac{m}{m-2} \text{vec } \Sigma - \frac{m}{\pi} (G_{-1}(m))^2 \delta^{\otimes 2}. \end{aligned}$$

We note that Sahu et al., (2003, p.137) provide a different approach and expression for the second cumulant. Consider now the third order cumulant. Again, from the formula for cumulants of products (see Terdik 2002) we need to compute:

$$\begin{aligned} \kappa_{W,3} &= C_3(S_m) \left(\kappa_{X,3} + \mathbf{K}_{2,1}(\kappa_{X,2} \otimes \kappa_{X,1}) + \kappa_{X,1}^{\otimes 3} \right) \\ &\quad + C_1(S_m) C_2(S_m) (3\kappa_{X,3} + 2\mathbf{K}_{2,1}(\kappa_{X,2} \otimes \kappa_{X,1})) \\ &\quad + C_1(S_m)^3 \kappa_{X,3}. \end{aligned} \tag{1.4}$$

Notice that \mathbf{W} is a scalar multiple of a vector valued variate, making the general formula for vector valued case somewhat simplified. Now, we consider separately the terms in Eq. 1.4: first those depending on \mathbf{X} , namely

$$\begin{aligned} & \kappa_{X,3} + \mathbf{K}_{2,1}(\kappa_{X,2} \otimes \kappa_{X,1}) + \kappa_{X,1}^{\otimes 3} \\ = & \left(2 \left(\sqrt{\frac{2}{\pi}} \right)^3 - \sqrt{\frac{2}{\pi}} \right) \boldsymbol{\delta}^{\otimes 3} + \sqrt{\frac{2}{\pi}} \mathbf{K}_{2,1}(\text{vec } \boldsymbol{\Sigma} \otimes \boldsymbol{\delta}) \\ & - 3 \left(\sqrt{\frac{2}{\pi}} \right)^3 \boldsymbol{\delta}^{\otimes 3} + \left(\sqrt{\frac{2}{\pi}} \right)^3 \boldsymbol{\delta}^{\otimes 3} \\ = & -\sqrt{\frac{2}{\pi}} \boldsymbol{\delta}^{\otimes 3} + \sqrt{\frac{2}{\pi}} \mathbf{K}_{2,1}(\text{vec } \boldsymbol{\Sigma} \otimes \boldsymbol{\delta}). \end{aligned}$$

Here, as well as in the next set of calculations, we use Lemma 4 which expresses the cumulants of a multivariate skew normal variable in terms of its parameters. As a side result we have

$$E \mathbf{X}^{\otimes 3} = -\sqrt{\frac{2}{\pi}} \boldsymbol{\delta}^{\otimes 3} + \sqrt{\frac{2}{\pi}} \mathbf{K}_{2,1}(\text{vec } \boldsymbol{\Sigma} \otimes \boldsymbol{\delta}), \tag{1.5}$$

since the third moment

$$E \mathbf{X}^{\otimes 3} = \kappa_{X,3} + \mathbf{K}_{2,1} \kappa_{X,2} \otimes \kappa_{X,1} + \kappa_{X,1}^{\otimes 3},$$

cf. Genton et al. (2001) where the mean is zero. The next term in Eq. 1.4 contains

$$\begin{aligned} 3\kappa_{X,3} + 2\mathbf{K}_{2,1}(\kappa_{X,1} \otimes \kappa_{X,2}) &= \left(6 \left(\sqrt{\frac{2}{\pi}} \right)^3 - 3\sqrt{\frac{2}{\pi}} \right) \boldsymbol{\delta}^{\otimes 3} \\ &+ 2\sqrt{\frac{2}{\pi}} \mathbf{K}_{2,1}(\text{vec } \boldsymbol{\Sigma} \otimes \boldsymbol{\delta}) - 6 \left(\sqrt{\frac{2}{\pi}} \right)^3 \boldsymbol{\delta}^{\otimes 3} \\ &= -3\sqrt{\frac{2}{\pi}} \boldsymbol{\delta}^{\otimes 3} + 2\sqrt{\frac{2}{\pi}} \mathbf{K}_{2,1}(\text{vec } \boldsymbol{\Sigma} \otimes \boldsymbol{\delta}). \end{aligned}$$

Plugging these results into Eq. 1.4, we obtain

$$\begin{aligned} \kappa_{W,3} &= C_3(S_m) \left(-\sqrt{\frac{2}{\pi}} \boldsymbol{\delta}^{\otimes 3} + \sqrt{\frac{2}{\pi}} (3 \text{vec } \boldsymbol{\Sigma} \otimes \boldsymbol{\delta}) \right) \\ &+ C_1(S_m) C_2(S_m) \left(-3\sqrt{\frac{2}{\pi}} \boldsymbol{\delta}^{\otimes 3} + 2\sqrt{\frac{2}{\pi}} \mathbf{K}_{2,1} (\text{vec } \boldsymbol{\Sigma} \otimes \boldsymbol{\delta}) \right) \\ &+ C_1(S_m)^3 \left(2 \left(\sqrt{\frac{2}{\pi}} \right)^3 - \sqrt{\frac{2}{\pi}} \right) \boldsymbol{\delta}^{\otimes 3}. \end{aligned} \tag{1.6}$$

The cumulants of S_m that are needed, come from the expressions for cumulants in terms of moments and the moment formula (1.3). They are as follows

$$\begin{aligned} C_3(S_m) &= \mu_3(S_m) - 3\mu_1(S_m) \mu_2(S_m) + 2\mu_1(S_m)^3 \\ &= \left(\frac{m}{2}\right)^{3/2} G_{-1}(m) \left(-\frac{4m-14}{(m-3)(m-2)} + 2(G_{-1}(m))^2 \right) \end{aligned}$$

and

$$\begin{aligned} C_1(S_m) C_2(S_m) &= \mu_1(S_m) \mu_2(S_m) - \mu_1(S_m)^3 \\ &= \left(\frac{m}{2}\right)^{3/2} G_{-1}(m) \left(\frac{2}{m-2} - G_{-1}(m)^2 \right) \end{aligned}$$

$C_1(S_m)^3$ is simply the third power of the expected value. Finally we collect the coefficients in Eq. 1.6 of different terms. Starting with $\mathbf{K}_{2,1} (\boldsymbol{\kappa}_{X,2} \otimes \boldsymbol{\kappa}_{X,1})$, we have

$$\begin{aligned} &\sqrt{\frac{2}{\pi}} \left(\frac{m}{2}\right)^{3/2} G_{-1}(m) \left(-\frac{4m-14}{(m-3)(m-2)} + 2(G_{-1}(m))^2 \right) \\ &+ 2\sqrt{\frac{2}{\pi}} \left(\frac{m}{2}\right)^{3/2} G_{-1}(m) \left(\frac{2}{m-2} - G_{-1}(m)^2 \right) \\ &= \sqrt{\frac{2}{\pi}} \left(\frac{m}{2}\right)^{3/2} G_{-1}(m) \frac{2}{(m-3)(m-2)}. \end{aligned}$$

Then collecting the coefficients of $\delta^{\otimes 3}$ gives

$$\begin{aligned} & -\sqrt{\frac{2}{\pi}} \left(\frac{m}{2}\right)^{3/2} G_{-1}(m) \left(-\frac{4m-14}{(m-3)(m-2)} + 2(G_{-1}(m))^2\right) \\ & -3\sqrt{\frac{2}{\pi}} \left(\frac{m}{2}\right)^{3/2} G_{-1}(m) \left(\frac{2}{m-2} - G_{-1}(m)^2\right) \\ & + \left(\frac{m}{2}\right)^{3/2} G_{-1}(m)^3 \left(2\left(\sqrt{\frac{2}{\pi}}\right)^3 - \sqrt{\frac{2}{\pi}}\right) \\ = & -\sqrt{\frac{2}{\pi}} \left(\frac{m}{2}\right)^{3/2} G_{-1}(m) \left(\frac{2}{m-3} - \frac{4}{\pi}G_{-1}(m)^2\right). \end{aligned}$$

Putting these results together, we obtain

$$\begin{aligned} \kappa_{W,3} = & -\sqrt{\frac{2}{\pi}} \left(\frac{m}{2}\right)^{3/2} G_{-1}(m) \left(\frac{2}{m-3} - \frac{4}{\pi}G_{-1}(m)^2\right) \delta^{\otimes 3} \\ & + \frac{2}{(m-3)(m-2)} \sqrt{\frac{2}{\pi}} \left(\frac{m}{2}\right)^{3/2} G_{-1}(m) \mathbf{K}_{2,1}(\delta \otimes \text{vec } \Sigma). \end{aligned}$$

For the fourth cumulant ($k = 4$), again expressing the cumulant of product in terms of product of cumulants (see Terdik et al., 2002), and writing $C_k(S_m) = \text{Cum}_k(S_m)$ for short, we have

$$\begin{aligned} \kappa_{W,4} = & C_4(S_m) (\kappa_{X,4} + \mathbf{K}_{3,1}(\kappa_{X,3} \otimes \kappa_{X,1}) + \mathbf{K}_{2,2}\kappa_{X,2}^{\otimes 2} + \mathbf{K}_{2,1,1}(\kappa_{X,2} \otimes \kappa_{X,1}^{\otimes 2}) + \kappa_{X,1}^{\otimes 4}) \quad (1.7) \\ & + C_3(S_m)C_1(S_m) (4\kappa_{X,4} + 3\mathbf{K}_{3,1}(\kappa_{X,3} \otimes \kappa_{X,1}) + 4\mathbf{K}_{2,2}\kappa_{X,2}^{\otimes 2} + 2\mathbf{K}_{2,1,1}(\kappa_{X,2} \otimes \kappa_{X,1}^{\otimes 2})) \\ & + C_2(S_m)^2 (3\kappa_{X,4} + 3\mathbf{K}_{3,1}(\kappa_{X,3} \otimes \kappa_{X,1}) + 2\mathbf{K}_{2,2}\kappa_{X,2}^{\otimes 2} + 2\mathbf{K}_{2,1,1}(\kappa_{X,2} \otimes \kappa_{X,1}^{\otimes 2})) \\ & + C_2(S_m)C_1(S_m)^2 (6\kappa_{X,4} + 3\mathbf{K}_{3,1}(\kappa_{X,3} \otimes \kappa_{X,1}) + 4\mathbf{K}_{2,2}\kappa_{X,2}^{\otimes 2}) + C_1(S_m)^4 \kappa_{X,4}. \end{aligned}$$

Remark 4. *In the most general case when one considers cumulants of products such as*

$\text{Cum}_4(X_1Y_1, X_2Y_2, X_3Y_3, X_4Y_4)$ *in terms of products of cumulants; there are 2465 terms according to the indecomposable partitions with respect to partition $\mathcal{L} = \{(1, 2), (3, 4), (5, 6), (7, 8)\}$ of 8 elements. In our case it simplifies because we consider four products of the same independent variables $\text{Cum}_4(XY)$. One can also check these formulae by expressing the 4th order cumulant in terms of expected values and using independence.*

In doing so, one should pay attention to the fact that the first order cumulants are not zero.

The cumulants contained in this expression are considered in Lemma 4. The cumulants of \mathbf{X} , except the second order one, depend on constant times tensor powers $\delta^{\otimes k}$, thus making the use of commutator matrices unnecessary. The products including $\kappa_{X,2} = \text{vec } \Sigma - 2/\pi \delta^{\otimes 2}$, needs to be considered

$$\begin{aligned} \mathbf{K}_{2,2} \kappa_{X,2}^{\otimes 2} &= \mathbf{K}_{2,2} \left(\left(\text{vec } \Omega - \frac{2}{\pi} \delta^{\otimes 2} \right) \otimes \left(\text{vec } \Omega - \frac{2}{\pi} \delta^{\otimes 2} \right) \right) \\ &= \mathbf{K}_{2,2} (\text{vec } \Omega)^{\otimes 2} - \frac{2}{\pi} \mathbf{K}_{2,2} (\text{vec } \Omega \otimes \delta^{\otimes 2} + \delta^{\otimes 2} \otimes \text{vec } \Omega) \\ &\quad + 3 \left(\frac{2}{\pi} \right)^2 \delta^{\otimes 4}. \end{aligned}$$

The term with $\text{vec } \Omega$ and $\delta^{\otimes 2}$ becomes

$$\mathbf{K}_{2,2} (\text{vec } \Omega \otimes \delta^{\otimes 2} + \delta^{\otimes 2} \otimes \text{vec } \Omega) = \mathbf{K}_{2,1,1} (\text{vec } \Omega \otimes \delta^{\otimes 2}).$$

Calculations for kurtosis are similar to those that have been used for skewness. The expression (1.7) contains five groups of cumulants, which are as follows:

1. It follows from Section 1.2

$$\kappa_{X,4} + \mathbf{K}_{3,1} (\kappa_{X,3} \otimes \kappa_{X,1}) + \mathbf{K}_{2,2} \kappa_{X,2}^{\otimes 2} + \mathbf{K}_{2,1,1} (\kappa_{X,2} \otimes \kappa_{X,1}^{\otimes 2}) + \kappa_{X,1}^{\otimes 4} = \mathbf{E} \mathbf{X}^{\otimes 4}$$

(cf. Genton et al., 2001). Observe that the coefficient of $\delta^{\otimes 4}$ is 0.

- 2.

$$\begin{aligned} &4\kappa_{X,4} + 3\mathbf{K}_{3,1} (\kappa_{X,3} \otimes \kappa_{X,1}) + 4\mathbf{K}_{2,2} \kappa_{X,2}^{\otimes 2} + 2\mathbf{K}_{2,1,1} (\kappa_{X,2} \otimes \kappa_{X,1}^{\otimes 2}) \\ &= 4 \frac{2}{\pi} \delta^{\otimes 4} + 4\mathbf{K}_{2,2} (\text{vec } \Sigma)^{\otimes 2} - 2 \frac{2}{\pi} \mathbf{K}_{2,1,1} (\text{vec } \Sigma \otimes \delta^{\otimes 2}) \end{aligned}$$

and the coefficient of $\delta^{\otimes 4}$ is $4 \frac{2}{\pi}$.

- 3.

$$\begin{aligned} &3\kappa_{X,4} + 3\mathbf{K}_{3,1} (\kappa_{X,3} \otimes \kappa_{X,1}) + 2\mathbf{K}_{2,2} \kappa_{X,2}^{\otimes 2} + 2\mathbf{K}_{2,1,1} \kappa_{X,2} \otimes \kappa_{X,1}^{\otimes 2} \\ &= 2 \left(\mathbf{K}_{2,2} (\text{vec } \Sigma)^{\otimes 2} \right) \end{aligned}$$

and the coefficient of $\delta^{\otimes 4}$ is 0.

4.

$$\begin{aligned}
 &6\boldsymbol{\kappa}_{X,4} + 3\mathbf{K}_{3,1}(\boldsymbol{\kappa}_{X,3} \otimes \boldsymbol{\kappa}_{X,1}) + 4\mathbf{K}_{2,2}\boldsymbol{\kappa}_{X,2}^{\otimes 2} \\
 &= 12\frac{2}{\pi}\boldsymbol{\delta}^{\otimes 4} + 4\mathbf{K}_{2,2}(\text{vec } \boldsymbol{\Sigma})^{\otimes 2} - 4\frac{2}{\pi}\mathbf{K}_{2,1,1}((\text{vec } \boldsymbol{\Sigma} \otimes \boldsymbol{\delta}^{\otimes 2}))
 \end{aligned}$$

and the coefficient of $\boldsymbol{\delta}^{\otimes 4}$ is $12\frac{2}{\pi}$

5. $\boldsymbol{\kappa}_{X,4} = \left(-6\left(\frac{2}{\pi}\right)^2 + 4\frac{2}{\pi}\right)\boldsymbol{\delta}^{\otimes 4}$,

Now we collect the coefficients for $\boldsymbol{\delta}^{\otimes 4}$

$$\begin{aligned}
 &\left(-6\left(\frac{2}{\pi}\right)^2 + 4\frac{2}{\pi}\right)C_1(S_m)^4 + 12\frac{2}{\pi}C_2(S_m)C_1(S_m)^2 + 4\frac{2}{\pi}C_3(S_m)C_1(S_m) \\
 &= 2\left(\frac{m}{2}\right)^2\frac{2}{\pi}G_{-1}(m)^2\left(\frac{4}{m-3} - 3\frac{2}{\pi}G_{-1}(m)^2\right),
 \end{aligned}$$

and for $\mathbf{K}_{2,2}(\text{vec } \boldsymbol{\Sigma})^{\otimes 2}$

$$\begin{aligned}
 &C_4(S_m) + 4C_3(S_m)C_1(S_m) + 2C_2(S_m)^2 + 4C_2(S_m)C_1(S_m)^2 \\
 &= \left(\frac{m}{2}\right)^2\frac{8}{(m-4)(m-2)^2},
 \end{aligned}$$

and for $\mathbf{K}_{2,1,1}(\text{vec } \boldsymbol{\Sigma} \otimes \boldsymbol{\delta}^{\otimes 2})$

$$\begin{aligned}
 &-4\frac{2}{\pi}C_2(S_m)C_1(S_m)^2 - 2\frac{2}{\pi}C_3(S_m)C_1(S_m) \\
 &= -\frac{2}{\pi}\left(\frac{m}{2}\right)^2\frac{4}{(m-3)(m-2)}G_{-1}(m)^2.
 \end{aligned}$$

Combining these five groups, one arrives at the proof of the Theorem.

PROOF OF LEMMA 1. This Lemma can be obtained essentially from the results in Fang et al. (2017), by rearranging them into a vector form, and using our notations.

Writing $d_1 = (d^3 - 1) / (d - 1)$, and defining the matrix $\mathbf{I}(d^3, d) = \mathbf{I}_{d^3}(:, [1, 1 + d_1, \dots, 1 + (d - 1)d_1])$, (i.e. d columns of the unit matrix \mathbf{I}_{d^3}), we will denote

$$\mathbf{e}(d^3, d) = \text{vec } \mathbf{I}(d^3, d). \tag{1.8}$$

Also let

$$\begin{aligned} \mathbf{e}(d^2, d) &= \sum_{k=1}^{d-1} \sum_{m=1}^{d-k} (\mathbf{I}_{d^2}(:, (k-1)d+k+m) \otimes \mathbf{I}_{d^2}(:, (k-1+m)d+k) \\ &\quad + \mathbf{I}_{d^2}(:, (k-1+m)d+k) \otimes \mathbf{I}_{d^2}(:, (k-1)d+k+m)), \end{aligned}$$

$$\mathbf{e}(d^4) = \text{vec}(\mathbf{I}_d \otimes \mathbf{I}_d) + (\text{vec} \mathbf{I}_d)^{\otimes 2} + \mathbf{e}(d^2, d) - 2\mathbf{e}(d^3, d). \tag{1.9}$$

Note that the number of ones in $\mathbf{e}(d^3, d)$ and $\mathbf{e}(d^2, d)$ are d and $(d-1)d$ respectively. Let $\mathbf{I}_d = [\mathbf{e}_k]_{k=1:d}$, i.e. \mathbf{e}_k is a unit vector in \mathbb{R}^d . The lemma follows with the following additional notations

$$\begin{aligned} \mathbf{1}_{d^3}(3) &= \sum_{k=1}^d \mathbf{e}_k^{\otimes 3}, \\ \mathbf{1}_{d^3}(2, 1) &= \mathbf{K}_{2,1} \sum_{k \neq j} \mathbf{e}_k^{\otimes 2} \otimes \mathbf{e}_j, \\ \mathbf{1}_{d^3}(1, 1, 1) &= \sum_{j \neq k \neq \ell} \mathbf{e}_j \otimes \mathbf{e}_k \otimes \mathbf{e}_\ell, \end{aligned} \tag{1.10}$$

and

$$\begin{aligned} \mathbf{1}_{d^4}(4) &= \sum_{k=1}^d \mathbf{e}_k^{\otimes 4}, \\ \mathbf{1}_{d^4}(3, 1) &= \mathbf{K}_{3,1} \sum_{j \neq k} \mathbf{e}_k^{\otimes 3} \otimes \mathbf{e}_j, \\ \mathbf{1}_{d^4}(2, 2) &= \mathbf{K}_{2,2} \sum_{j \neq k} \mathbf{e}_j^{\otimes 2} \otimes \mathbf{e}_k^{\otimes 2}, \\ \mathbf{1}_{d^4}(2, 1, 1) &= \mathbf{K}_{2,1,1} \sum_{j \neq k \neq \ell} \mathbf{e}_j^{\otimes 2} \otimes \mathbf{e}_k \otimes \mathbf{e}_\ell, \\ \mathbf{1}_{d^4}(1, 1, 1, 1) &= \sum_{j \neq k \neq \ell \neq n} \mathbf{e}_j \otimes \mathbf{e}_k \otimes \mathbf{e}_\ell \otimes \mathbf{e}_n. \end{aligned} \tag{1.11}$$

PROOF OF LEMMA 2. Recall $\phi_{\mathbf{W}}(\boldsymbol{\lambda})$ defined in Eq. 5.2. Applying $D_{\boldsymbol{\lambda}}^{\otimes}$ to the cumulant generator $\psi_{\mathbf{W}}(\boldsymbol{\lambda}) = \log \phi_{\mathbf{W}}(\boldsymbol{\lambda})$, we have

$$\underline{\text{Cum}}_n(\mathbf{W}) = (-i)^n D_{\boldsymbol{\lambda}}^{\otimes n} \psi_{\mathbf{W}}(\boldsymbol{\lambda})|_{\boldsymbol{\lambda}=0}.$$

Writing $f(\boldsymbol{\lambda}^\top \boldsymbol{\lambda}) = g'(\boldsymbol{\lambda}^\top \boldsymbol{\lambda}) / g(\boldsymbol{\lambda}^\top \boldsymbol{\lambda})$, we have

$$\begin{aligned} D_{\boldsymbol{\lambda}}^{\otimes 1} \psi_{\mathbf{W}}(\boldsymbol{\lambda}) &= \frac{2g'(\underline{\boldsymbol{\lambda}}^\top \boldsymbol{\lambda})}{g(\underline{\boldsymbol{\lambda}}^\top \boldsymbol{\lambda})} \boldsymbol{\lambda} = 2f(\boldsymbol{\lambda}^\top \boldsymbol{\lambda}) \underline{\boldsymbol{\lambda}}, \\ D_{\boldsymbol{\lambda}}^{\otimes 2} \psi_{\mathbf{W}}(\boldsymbol{\lambda}) &= 4f'(\underline{\boldsymbol{\lambda}}^\top \boldsymbol{\lambda}) \boldsymbol{\lambda}^{\otimes 2} + 2f(\boldsymbol{\lambda}^\top \boldsymbol{\lambda}) \text{vec } \mathbf{I}_d, \\ D_{\boldsymbol{\lambda}}^{\otimes 3} \psi_{\mathbf{W}}(\boldsymbol{\lambda}) &= 8f''(\underline{\boldsymbol{\lambda}}^\top \boldsymbol{\lambda}) \boldsymbol{\lambda}^{\otimes 3} + 4f'(\boldsymbol{\lambda}^\top \underline{\boldsymbol{\lambda}}) (2\mathbf{I}_{d^3} + \mathbf{K}_z \otimes \mathbf{I}_d) (\boldsymbol{\lambda} \otimes \text{vec } \mathbf{I}_d), \end{aligned}$$

from which we obtain $\underline{\text{Cum}}_1(\mathbf{W}) = 0$, $\underline{\text{Cum}}_2(\mathbf{W}) = -2g'(0) \text{vec } \mathbf{I}_d$, and $\underline{\text{Cum}}_3(\mathbf{W}) = 0$. For $\underline{\text{Cum}}_4(\mathbf{W})$ we have

$$D_{\boldsymbol{\lambda}}^{\otimes 4} \psi_{\mathbf{W}}(\boldsymbol{\lambda})|_{\boldsymbol{\lambda}=0} = 4\left(g''(0) - g'(0)^2\right) (\mathbf{I}_{d^4} + \mathbf{K}_{(3,1,2,4)} + \mathbf{K}_{(1,3,2,4)}) (\text{vec } \mathbf{I}_d)^{\otimes 2}.$$

Mardia’s measure can now be obtained by noting that

$$\begin{aligned} &(\text{vec } \mathbf{I}_{d^2})^\top \left(\boldsymbol{\Sigma}_W^{-1/2}\right)^{\otimes 4} \underline{\text{Cum}}_4(\mathbf{W}) \\ &= 4 \frac{g''(0) - g'(0)^2}{4g'(0)^2} (\text{vec } \mathbf{I}_{d^2})^\top (\mathbf{I}_{d^4} + \mathbf{K}_{(3,1,2,4)} + \mathbf{K}_{(1,3,2,4)}) (\text{vec } \mathbf{I}_d)^{\otimes 2} \\ &= d(d+2) \kappa_0. \end{aligned}$$

PROOF OF LEMMA 3. Since R^{*2}/d has an F-distribution with d and m degrees of freedom, we have for $m > 4$,

$$\begin{aligned} \mathbb{E} R^{*2} &= \frac{dm}{m-2}, & \mathbb{E} R^{*4} &= d^2 \left(\frac{m}{d}\right)^2 \frac{d(d+2)}{(m-2)(m-4)}, \\ \frac{\mathbb{E} R^{*4}}{(\mathbb{E} R^{*2})^2} &= \frac{d+2}{d} \frac{m-2}{m-4}, \end{aligned}$$

and using Theorem 1. One may use Fang et al., (2017 p. 88) to obtain formulae for the moments of components of \mathbf{W} , and get

$$\begin{aligned} \mathbb{E} W_1^2 &= \frac{m}{m-2} \\ \mathbb{E} W_1^4 &= m^2 \frac{\Gamma(m/2-2)}{2^2 \Gamma(m/2)} \frac{4!}{2^2 2!} = \frac{3m^2}{(m-2)(m-4)}, \end{aligned}$$

hence $\text{Cum}_4(W_1/\sigma_{W_1}) = 3\left(\frac{m-2}{m-4} - 1\right)$; Theorem 1 can be used to get κ_0 .

PROOF OF THEOREM 2. Recall (6.4) and Lemma 1 for the moments of p_k and $|\mathbf{U}_j|$. Set $\mathbf{V}_1 = \Delta |\mathbf{Z}_1|$, $\mathbf{V}_2 = (\mathbf{I} - \Delta \Delta^\top)^{1/2} \mathbf{Z}_2$ for simplifying notations. The commutation matrices are given in Eqs. 1.1 and 1.2.

$$\mathbf{E} \mathbf{X} = \Delta \sqrt{\frac{1}{\pi}} \frac{G_{1,0}}{G_1(m)} \mathbf{E} R \mathbf{1}_m,$$

$$\begin{aligned} \mathbf{E} \mathbf{X}^{\otimes 2} &= \mathbf{E} \mathbf{V}_1^{\otimes 2} + \mathbf{E} \mathbf{V}_2^{\otimes 2} \\ &= \mathbf{E} p_1^2 \mathbf{E} R^2 \Delta^{\otimes 2} \mathbf{E} |\mathbf{U}_1|^{\otimes 2} + \mathbf{E} p_2^2 \mathbf{E} R^2 \mathbf{E} \mathbf{V}_2^{\otimes 2}, \end{aligned}$$

$$\mathbf{E} \mathbf{X}^{\otimes 3} = \mathbf{E} (\mathbf{V}_1 + \mathbf{V}_2)^{\otimes 3} = \mathbf{E} \mathbf{V}_1^{\otimes 3} + \mathbf{K}_{2,1} (\mathbf{E} \mathbf{V}_2^{\otimes 2} \otimes \mathbf{V}_1).$$

where $\mathbf{E} \mathbf{V}_1^{\otimes 3} = \Delta^{\otimes 3} \mathbf{E} |\mathbf{Z}_1|^{\otimes 3} = \mathbf{E} p_1^3 \mathbf{E} R^3 \mathbf{E} |\mathbf{U}_1|^{\otimes 3}$ and

$$\mathbf{K}_{2,1} \mathbf{E} \mathbf{V}_2^{\otimes 2} \otimes \mathbf{V}_1 = \mathbf{E} p_1 p_2^2 \mathbf{E} R^3 \mathbf{K}_{2,1} (\mathbf{I}_{d^2} \otimes \Delta) (\mathbf{E} \mathbf{V}_2^{\otimes 2} \otimes \mathbf{E} |\mathbf{U}_1|).$$

$$\mathbf{E} \mathbf{X}^{\otimes 4} = \mathbf{E} \mathbf{V}_1^{\otimes 4} + \mathbf{K}_{2,1,1} (\mathbf{E} \mathbf{V}_1^{\otimes 2} \otimes \mathbf{V}_2^{\otimes 2}) + \mathbf{E} \mathbf{V}_2^{\otimes 4},$$

where $\mathbf{E} \mathbf{V}_1^{\otimes 4} = \mathbf{E} p_1^4 \mathbf{E} R^4 \Delta^{\otimes 4} \mathbf{E} |\mathbf{U}_1|^{\otimes 4}$ and

$$\mathbf{E} \mathbf{V}_2^{\otimes 4} = (\mathbf{E} p_2^4 \mathbf{E} R^4) \left((\mathbf{I} - \Delta \Delta^\top)^{1/2} \right)^{\otimes 4} \mathbf{E} \mathbf{U}_2^{\otimes 4},$$

$$\mathbf{E} \mathbf{V}_1^{\otimes 2} \otimes \mathbf{V}_2^{\otimes 2} = \mathbf{E} p_1^2 p_2^2 \mathbf{E} R^4 (\Delta^{\otimes 2} \otimes \mathbf{I}_{m^2}) (\mathbf{E} |\mathbf{U}_1|^{\otimes 2} \otimes \mathbf{E} \mathbf{V}_2^{\otimes 2}).$$

PROOF OF LEMMA 5. In the proof, the operator D_λ^\otimes is used repeatedly; let $C_{\mathbf{X}}$ denote the cumulant function of \mathbf{X} and note that

$$D_\lambda^\otimes \log \Phi_m(i \Delta^\top \lambda) = i \frac{1}{\Phi_m(i \Delta^\top \lambda)} \Delta \left[\Phi_m(\mathbf{Y}) \frac{\partial}{\partial \mathbf{Y}^\top} \Big|_{\mathbf{Y}=i \Delta^\top \lambda} \right]^\top,$$

put \mathbf{e}_k for the k^{th} unit vector, since Φ_m is the distribution function of the m -variate standard normal distribution which is simply the product of univariate standard normal distributions; hence

$$\begin{aligned} \Phi_m(\mathbf{Y}) \frac{\partial}{\partial \mathbf{Y}^\top} \Big|_{\mathbf{Y}=i \Delta^\top \lambda} &= \sum_k \Phi_{m-1}(\mathbf{Y}) \varphi(y_k) \mathbf{e}_k^\top \Delta^\top \Big|_{\mathbf{Y}=i \Delta^\top \lambda} \\ &= \Phi_m(\mathbf{Y}) \left[\frac{\varphi(y_k)}{\Phi(y_k)} \right]_{k=1:m} \Delta^\top \Big|_{\mathbf{Y}=i \Delta^\top \lambda}, \end{aligned}$$

and

$$D_{\lambda}^{\otimes} \log \Phi_m \left(i \Delta^{\top} \lambda \right) = -i \Delta \left[\frac{\varphi(y_k)}{\Phi(y_k)} \right]_{k=1:m}^{\top} \Big|_{\mathbf{Y}=i \Delta^{\top} \lambda}.$$

The first two cumulants are obtained, by evaluating the derivatives at $\lambda = 0$. We have:

$$D_{\lambda}^{\otimes} C_{\mathbf{X}} = -\Sigma_{\Delta} + D_{\lambda}^{\otimes} \log \Phi_m \left(i \Delta^{\top} \lambda \right).$$

Consider next

$$\begin{aligned} \Delta \left[\frac{\varphi(y_k)}{\Phi(y_k)} \right]_{k=1:m} \Big|_{\mathbf{Y}=i \Delta^{\top} \lambda} \frac{\partial}{\partial \lambda^{\top}} &= \Delta \left[\frac{\varphi(y_k)}{\Phi(y_k)} \left(y_k - \frac{\varphi(y_k)}{\Phi(y_k)} \right) \mathbf{e}_k^{\top} \Delta^{\top} \right]_{k=1:m} \\ &= \Delta \left[\frac{\varphi(y_k)}{\Phi(y_k)} \left(y_k - \frac{\varphi(y_k)}{\Phi(y_k)} \right) \mathbf{e}_k^{\top} \right]_{k=1:m} \Delta^{\top}, \end{aligned}$$

to obtain

$$\begin{aligned} D_{\lambda}^{\otimes 2} C_{\mathbf{X}} &= -\text{vec } \Sigma + i \text{vec } \Delta \left[\frac{\varphi(y_k)}{\Phi(y_k)} \left(y_k - \frac{\varphi(y_k)}{\Phi(y_k)} \right) \mathbf{e}_k^{\top} \right]_{k=1:m} \Delta^{\top} \\ &= -\text{vec } \Sigma + i \Delta^{\otimes 2} \text{vec} \left(\text{Diag} \left[\frac{\varphi(y_k)}{\Phi(y_k)} \left(y_k - \frac{\varphi(y_k)}{\Phi(y_k)} \right) \right] \right). \end{aligned}$$

For the third order cumulant, we have

$$\begin{aligned} &\left(\text{Diag} \left[\frac{\varphi(y_k)}{\Phi(y_k)} \left(y_k - \frac{\varphi(y_k)}{\Phi(y_k)} \right) \right] \right) \frac{\partial}{\partial \lambda^{\top}} \\ &= \text{vec} \left(\left[\mathbf{e}_k^{\top} \frac{\varphi(y_k)}{\Phi(y_k)} \left(y_k - \frac{\varphi(y_k)}{\Phi(y_k)} \right) \right] \right) \frac{\partial}{\partial \lambda^{\top}} \\ &= \left[\mathbf{e}_k \frac{\partial}{\partial y_k} \frac{\varphi(y_k)}{\Phi(y_k)} \left(y_k - \frac{\varphi(y_k)}{\Phi(y_k)} \right) \mathbf{e}_k^{\top} \Delta^{\top} \right], \end{aligned}$$

and

$$\begin{aligned} &\text{vec} \left(\Delta \left[\frac{\partial}{\partial y_1} \frac{\varphi(y_1)}{\Phi(y_1)} \mathbf{e}_1 \mathbf{e}_1^{\top}, \dots, \frac{\partial}{\partial y_m} \frac{\varphi(y_m)}{\Phi(y_m)} \mathbf{e}_m \mathbf{e}_m^{\top} \right] \Delta^{\top \otimes 2} \right) \\ &= \Delta^{\top \otimes 3} \text{vec} \left[\frac{\partial}{\partial y_1} \frac{\varphi(y_1)}{\Phi(y_1)} \mathbf{e}_1 \mathbf{e}_1^{\top}, \dots, \frac{\partial}{\partial y_m} \frac{\varphi(y_m)}{\Phi(y_m)} \mathbf{e}_m \mathbf{e}_m^{\top} \right] \\ &= \Delta^{\top \otimes 3} \text{vec} \left[\frac{\partial}{\partial y_1} \frac{\varphi(y_1)}{\Phi(y_1)} \mathbf{e}_1^{\otimes 2}, \dots, \frac{\partial}{\partial y_m} \frac{\varphi(y_m)}{\Phi(y_m)} \mathbf{e}_m^{\otimes 2} \right], \end{aligned}$$

hence

$$\text{Cum}_3(\mathbf{X}) = \sqrt{\frac{2}{\pi}} \left(\frac{4}{\pi} - 1 \right) \Delta^{\otimes 3} \text{vec} \left[\mathbf{e}_k^{\otimes 2} \right]_{k=1:m}.$$

Here $[\mathbf{e}_k^{\otimes 2}]_{k=1:m}$ is a matrix of vectors of $\mathbf{e}_k^{\otimes 2}$ with dimension $m^2 \times m$. Higher order cumulants come out as constant times $\Delta^{\otimes k} \text{vec} [\mathbf{e}_p^{\otimes k-1}]_{p=1:m}$, $k = 3, 4, \dots$

Note that c_3 and c_4 coincide with those given in Appendix A.2 of Azzalini and Capitanio (1999), and we omit details about the other constants c_k .

References

- ARELLANO-VALLE, R B and GENTON, M G (2005). On fundamental skew distributions. *Journal of Multivariate Analysis* **96**, 1, 93–116.
- ARNOLD, B C and GROENEVELD, R A (1995). Measuring skewness with respect to the mode. *The American Statistician* **49**, 1, 34–38.
- ARNOLD, B C and BEAVER, R J (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Test* **11**, 1, 7–54.
- AVEROUS, J and MESTE, M (1997). Skewness for multivariate distributions: two approaches. *The Annals of Statistics* **25**, 5, 1984–1997.
- AZZALINI, A and DALLA VALLE, A (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 4, 715–726.
- AZZALINI, A and 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 and 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 (Statistical Methodology)* **65**, 2, 367–389.
- BALAKRISHNAN, N, BRITO, M R and QUIROZ, A J (2007). A vectorial notion of skewness and its use in testing for multivariate symmetry. *Communications in Statistics. Theory and Methods* **36**, 9, 1757–1767.
- BALAKRISHNAN, N and SCARPA, B (2012). Multivariate measures of skewness for the skew-normal distribution. *Journal of Multivariate Analysis* **104**, 1, 73–87.
- BARINGHAUS, L (1991a). Testing for spherical symmetry of a multivariate distribution. *The Annals of Statistics* **19**, 2, 899–917.
- BARINGHAUS, L and HENZE, N (1991b). Limit distributions for measures of multivariate skewness and kurtosis based on projections. *Journal of Multivariate Analysis* **38**, 1, 51–69.
- BARINGHAUS, L and HENZE, N (1992). Limit distributions for mardia's measure of multivariate skewness. *Ann. Stat.* **20**, 4, 1889–1902.
- BERKANE, M and BENTLER, P M (1986). Moments of elliptically distributed random variates. *Statistics & Probability Letters* **4**, 6, 333–335.
- BRANCO, M D and DEY, D K (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis* **79**, 1, 99–113.
- DEY, D K and LIU, J (2005). A new construction for skew multivariate distributions. *Journal of Multivariate Analysis* **95**, 2, 323–344.
- EKSTRÖM, M and JAMMALAMADAKA, S R (2012). A general measure of skewness. *Statistics & Probability Letters* **82**, 8, 1559–1568.
- FANG, K W, KOTZ, S and NG, KW (2017). *Symmetric Multivariate and Related Distributions*. Chapman and Hall/CRC, London.

- GENTON, M G, HE, L and LIU, X (2001). Moments of skew-normal random vectors and their quadratic forms. *Statistics & Probability Letters* **51**, 4, 319–325.
- GENTON, MG (2004). *Skew-elliptical distributions and their applications: a journey beyond normality*. CRC Press, London.
- GENTON, M G and LOPERFIDO, N M R (2005). Generalized skew-elliptical distributions and their quadratic forms. *Annals of the Institute of Statistical Mathematics* **57**, 2, 389–401.
- GRAHAM, A (2081). *Kronecker products and matrix calculus with applications*. Courier Dover Publications, USA.
- ISOGAI, T (1982). On a measure of multivariate skewness and a test for multivariate normality. *Annals of the Institute of Statistical Mathematics* **34**, 1, 531–541.
- JAMMALAMADAKA, S R, SUBBA RAO, T and TERDIK, G (2006). Higher order cumulants of random vectors and applications to statistical inference and time series. *Sankhya (Ser A, Methodology)* **68**, 326–356.
- JOHNSON, N, KOTZ, S and BALAKRISHNAN, N (1994). *Continuous Univariate Distributions, 1*, 2nd edn. Houghton Mifflin, Boston.
- KIM, HJ and MALLICK, BK (2003). Moments of random vectors with skew t distribution and their quadratic forms. *Statistics & Probability Letters* **63**, 4, 417–423. doi: [10.1016/S0167-7152\(03\)00121-4](https://doi.org/10.1016/S0167-7152(03)00121-4), <http://www.sciencedirect.com/science/article/pii/S0167715203001214>.
- KLAR, B (2002). A treatment of multivariate skewness, kurtosis, and related statistics. *Journal of Multivariate Analysis* **83**, 1, 141–165.
- KOLLO, T (2008). Multivariate skewness and kurtosis measures with an application in ICA. *Journal of Multivariate Analysis* **99**, 10, 2328–2338.
- KOLLO, T, KÄÄRIK, M and SELART, A (2018). Asymptotic normality of estimators for parameters of a multivariate skew-normal distribution. *Communications in Statistics-Theory and Methods* **47**, 15, 3640–3655.
- KOTZ, S and NADARAJAH, S (2004). *Multivariate t-distributions and their applications*. Cambridge University Press, Cambridge.
- KOZIOL, JA (1989). A note on measures of multivariate kurtosis. *Biometrical Journal* **31**, 5, 619–624. doi: [10.1002/bimj.4710310517](https://doi.org/10.1002/bimj.4710310517).
- MACRAE, E C (1974). Matrix derivatives with an application to an adaptive linear decision problem. *The Annals of Statistics* **2**, 2, 337–346.
- MAGNUS, JR and NEUDECKER, H (1980). The elimination matrix: Some lemmas and applications. *SIAM Journal on Algebraic Discrete Methods* **1**, 4, 422–449. doi: [10.1137/0601049](https://doi.org/10.1137/0601049).
- MAGNUS, JR and NEUDECKER, H (1999). *Matrix differential calculus with applications in statistics and econometrics*. Wiley, Chichester, revised reprint of the 1988 original.
- MALKOVICH, J F and AFIFI, A A (1973). On tests for multivariate normality. *Journal of the American Statistical Association* **68**, 341, 176–179.
- MARDIA, K V (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57**, 519–530.
- MATHAI, AM (1997). *Jacobians of matrix transformations and functions of matrix arguments*. World Scientific Publishing Company.
- MCCULLAGH, P (1987). *Tensor methods in statistics Monographs on Statistics and Applied Probability*. Chapman & Hall, London.
- MEIJER, E (2005). Matrix algebra for higher order moments. *Linear Algebra and its Applications* **410**, 112–134. doi: [10.1016/j.laa.2005.02.040](https://doi.org/10.1016/j.laa.2005.02.040).

- MÓRI, T F, ROHATGI, V K and SZÉKELY, G J (1994). On multivariate skewness and kurtosis. *Theory of Probability & Its Applications* **38**, 3, 547–551.
- MUIRHEAD, R J (2009). *Aspects of multivariate statistical theory*, 197. Wiley, New York.
- RAO, CR (1948a). Tests of significance in multivariate analysis. *Biometrika* **35**, 1/2, 58–79.
- RAO, CR (1948b). The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society Series B (Methodological)* **10**, 2, 159–203.
- SAHU, S K, DEY, D K and BRANCO, M D (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *Canadian Journal of Statistics* **31**, 2, 129–150.
- SERFLING, R J (2004). Multivariate symmetry and asymmetry. *Encyclopedia of statistical sciences* **8**, 5338–5345.
- SONG, K S (2001). Rényi information, loglikelihood and an intrinsic distribution measure. *Journal of Statistical Planning and Inference* **93**, 1-2, 51–69.
- SPEED, T P (1990). Invariant moments and cumulants. In: *Coding theory and design theory, Part II, IMA Vol. Math. Appl.*, vol. 21, Springer, New York, pp. 319–335.
- SRIVASTAVA, M S (1984). A measure of skewness and kurtosis and a graphical method for assessing multivariate normality. *Statistics & Probability Letters* **2**, 5, 263–267.
- SUMIKAWA, T, KOIZUMI, K and SEO, T (2013). Measures of multivariate skewness and kurtosis in highdimensional framework. *Hiroshima Statistical Research Group Technical Reports, Technical Report* **13**, 3, 22.
- TERDIK, GY (2002). Higher order statistics and multivariate vector Hermite polynomials for nonlinear analysis of multidimensional time series. *Teor Ver Matem Stat(Teor Imovirnost ta Matem Statyst)* **66**, 147–168.

Publisher's Note. Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

SREENIVASA RAO JAMMALAMADAKA
UNIVERSITY OF CALIFORNIA, SANTA
BARBARA, CA, USA
E-mail: sreenivas@ucsb.edu

EMANUELE TAUFER
UNIVERSITY OF TRENTO, TRENTO, ITALY
E-mail: emanuele.taufer@unitn.it

GYORGY H. TERDIK
UNIVERSITY OF DEBRECEN, DEBRECEN,
HUNGARY
E-mail: terdik.gyorgy@inf.unideb.hu

Paper received: 29 February 2020.