Some remarks on Koziol's kurtosis

Nicola, Loperfido

Dipartimento di Economia, Società e Politica,
Università degli Studi di Urbino"Carlo Bo",
Via Saffi 42, 61029 Urbino (PU), ITALY

Abstract

James Koziol, in his 1987 and 1989 papers, proposed to use the sums of either the squared fourth-order cumulants or moments as test statistics for multivariate normality. His proposals are by far less popular than Mardia's measure of multivariate kurtosis, that is the fourth moment of the Mahalanobis distance of a random vector from its mean. We investigate some properties of Koziol's measures of multivariate kurtosis which motivate their use in statistical practice. Firstly, we show some of their connections with Mahalanobis angles. Secondly, we use inequalities to highlight their connections with other measures of multivariate skewness and kurtosis. Thirdly, we obtain their analytical formulae for some well-known multivariate statistical models. Simple examples illustrate the interpretation of Koziol's measures of multivariate kurtosis and detect a wrong statement abouth them which appeared in the statistical literature. We suggest that Mardia's and Koziol's measures of kurtosis should be used together to detect interesting data structures.

Keywords: Cantelli's inequality, Moment inequality, Multivariate kurtosis, Multivariate skewness.
2010 MSC: Primary 62H99, Secondary 60E15

1. Introduction

The kurtosis of a random variable X with mean μ, variance σ^2 and finite fourth moment is its fourth standardized moment

$$\beta_2(X) = \mathrm{E}\left\{\left(\frac{X-\mu}{\sigma}\right)^4\right\}.$$

Since the kurtosis of a normal random variable equals three, it is sometimes more convenient to work with the excess kurtosis, that is the fourth standardized cumulant $\gamma_2(X) = \beta_2(X) - 3$. Distributions whose excess kurtosis is negative, null or positive are commonly referred to as to platykurtic, mesokurtic and leptokurtic distributions. The random variables X and $aX + b$ have the same kurtosis, where a and b are constants with a being nonnull. The kurtosis attains its minimum, that is one, if and only if the corresponding distribution assigns probability 0.5 to two real values. Kurtosis also gives some insight into the tails of a distribution, as shown by the inequality

$$\Pr\left(\left|\frac{X-\mu}{\sigma}\right| \geq \lambda\right) \leq \frac{\beta_2(X) - 1}{\lambda^4 - 2\lambda^2 + \beta_2(X)},$$

where $\lambda > 1$ (see [33]).

Generalization of kurtosis to the multivariate case is not straightforward: a p-dimensional random vector might have up to $p(p+1)(p+2)(p+3)/24$ distinct fourth-order moments and cumulants. Let $\mathbf{x} = \left(X_1, ..., X_p\right)^\top$ be a $p-$dimensional random vector with mean $\boldsymbol{\mu}$, nonsingular variance $\boldsymbol{\Sigma}$ and finite fourth order moments: $\mathrm{E}\left(X_i^4\right) < +\infty$ for $i \in \{1, \ldots, p\}$. Also, let $\mathbf{z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$ be the standardization of $\mathbf{x}$, where $\boldsymbol{\Sigma}^{-1/2}$ is the symmetric, positive definite

*Email address: `nicola.loperfido@uniurb.it`

square root of the concentration matrix $\mathbf{\Sigma}^{-1}$. A fourth-order, standardized moment of $\mathbf{x}$ is the expected product of any four components of $\mathbf{z}$: $\mu_{ijhk} = \mathrm{E}\left(Z_i Z_j Z_h Z_k\right)$, $i, j, h, k \in \{1, \ldots, p\}$. The expectation μ_{ijhk} is commonly referred to as to the $ijhk$-th standardized moment of $\mathbf{x}$. The $ijhk$-th standardized cumulant of $\mathbf{x}$ is the derivative evaluated at $\mathbf{t} = \mathbf{0}$

$$\kappa_{ijhk} = (\iota)^4 \left. \frac{\partial^4 \log E\left[\exp\left(\iota \mathbf{t}^\top \mathbf{z}\right)\right]}{\partial t_i \partial t_j \partial t_h \partial t_k} \right|_{\mathbf{t}=\mathbf{0}},$$

with $\mathbf{0}$ being the $p \times 1$ zero vector, $\iota = \sqrt{-1}$ and $\mathbf{t}^\top = \left(t_1, ..., t_p\right)$. Assume that $i \leq j \leq h \leq k$ for $i, j, h, k \in \{1, \ldots, p\}$ without loss of generality. Then the $ijhk$-th standardized cumulant of $\mathbf{x}$ equals its $ijhk$-th standardized moment diminished by (i) 3, (ii) 1 and (iii) 0 when (i) $i = j = h = k$, (ii) $i = j < h = k$ and (iii) the remaining cases occur, respectively.

Since the number of possibly distinct fourth-order moments and cumulants grows very quickly with the vector's dimension, it is convenient to summarize them by appropriate scalar functions. Mardia [24] proposed to measure the kurtosis of a random vector with the fourth moment of its Mahalanobis distance from the mean:

$$\beta_{2,M}(\mathbf{x}) = \mathrm{E}\left[\left\{(\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right\}^2\right] = \sum_{i,j} \mu_{iijj},$$

which equals $p(p+2)$ for a p-dimensional, normal random vector. Mardia [25] proposed to measure excess kurtosis with the difference

$$\gamma_{2,M}(\mathbf{x}) = \beta_{2,M}(\mathbf{x}) - p(p+2) = \sum_{i,j} \kappa_{iijj}.$$

Mardia's overall index of multivariate (excess) kurtosis is the sum of the excess kurtoses when each of κ_{iijj}, for $i,j \in \{1, \ldots, p\}$, is seen as a multivariate excess kurtosis. Mardia's (excess) kurtosis of a random variable coincides with the (excess) kurtosis of the random variable itself. Moreover, the above mentioned properties of univariate kurtosis naturally generalize to Mardia's kurtosis. Firstly, Mardia's kurtosis is invariant with respect to one-to-one, affine transformations. Secondly, Mardia's kurtosis of a random vector is never smaller than the squared number of the vector's components, with equality holding if and only if the distribution is symmetric and its support is an ellipsoid. Thirdly, the following inequality holds for any real ε greater than p, according to [21]:

$$\Pr\left\{(\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \geq \varepsilon\right\} \leq \frac{\beta_{2,M}(\mathbf{x}) - p^2}{\varepsilon^2 - 2p\varepsilon + \beta_{2,M}(\mathbf{x})}.$$

Mardia's kurtosis is the best known and most used measure of multivariate kurtosis [8, 13, 39]. The popularity of Mardia's kurtosis is partly due to its analytical tractability in many well-known statistical models [25, 39]. Its popularity is also due to its close connection with measures of multivariate skewness, which has been extensively investigated by means of inequalities [14, 29, 32].

Despite its merits and popularity, Mardia's kurtosis suffers from several limitations. Koziol [16] noticed that not all fourth-order standardized moments are taken into account in Mardia's kurtosis. Koziol [17] illustrated the problem with a bivariate, folded normal distribution with some nonnull fourth-order cumulants but with null Mardia's excess kurtosis. The problem might be conveniently addressed by considering the sum of all squared fourth-order cumulants, as proposed in Koziol [15]:

$$\gamma_{2,K}(\mathbf{x}) = \sum_{i,j,h,k} \kappa_{ijhk}^2.$$

Koziol [15] motivated this kurtosis measure within the framework of Neyman's smooth goodness of fit tests ([31]). It often appears in projection pursuit ([30]), multivariate normality testing ([27]) and in the asymptotic distribution of the likelihood ratio statistic for testing hypotheses on covariance matrices ([38]). Koziol [16] proposed to summarize fourth standardized moments, rather than cumulants, with the sum of their squares:

$$\beta_{2,K}(\mathbf{x}) = \sum_{i,j,h,k} \mu_{ijhk}^2.$$

We shall refer to $\beta_{2,K}(\mathbf{x})$ and $\gamma_{2,K}(\mathbf{x})$ as to Koziol's kurtosis and Koziol's excess kurtosis. Also, we shall refer to both $\beta_{2,K}(\mathbf{x})$ and $\gamma_{2,K}(\mathbf{x})$ as to Koziol's kurtoses. The phrasing "Koziol's excess kurtosis" is not completely satisfactory. It makes the unaware reader think that fourth standardized moments are higher than expected under normality, while the opposite might be true, as it happens with the multivariate uniform distribution. It would be better to refer to $\beta_{2,K}$ as to Koziol's moment kurtosis and to $\gamma_{2,K}$ as to Koziol's cumulant kurtosis. The choice of referring to $\gamma_{2,K}(\mathbf{x})$ as to Koziol's excess kurtosis is consistent with the widely used notation and terminology used for univariate excess kurtosis γ_2 and Mardia's excess kurtosis $\gamma_{2,M}$. Moreover, it is instrumental for highlighting the mutual connections between these kurtosis measures.

The sample counterparts of Koziol's kurtoses are defined as follows. Let $\mathbf{m}$ and $\mathbf{S}$ be the sample mean vector and the nonsingular sample covariance matrix of the $n \times p$ data matrix $\mathbf{X}$. Also, let $\mathbf{x}_i^\top$ and $\mathbf{x}_j^\top$ be the i-th and j-th rows of $\mathbf{X}$, for $i, j \in \{1, \ldots, p\}$. The sample counterpart of $\beta_{2,K}(\mathbf{x})$ is

$$b_{2,K}(\mathbf{X}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{(\mathbf{x}_i - \mathbf{m})^\top \mathbf{S}^{-1}\left(\mathbf{x}_j - \mathbf{m}\right)\right\}^4 .$$

Similarly, the sample counterpart of $\beta_{2,M}(\mathbf{x})$ is

$$b_{2,M}(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n}\left\{(\mathbf{x}_i - \mathbf{m})^\top \mathbf{S}^{-1}(\mathbf{x}_i - \mathbf{m})\right\}^2 .$$

The sample counterpart of $\gamma_{2,K}(\mathbf{x})$ can be obtained from $b_{2,K}(\mathbf{X})$ and $b_{2,M}(\mathbf{X})$:

$$g_{2,K}(\mathbf{X}) = b_{2,K}(\mathbf{X}) - 6b_{2,M}(\mathbf{X}) + 6p + p^2.$$

Best and Rayner [5] used simulations to compare powers of the bivariate normality tests based on $b_{2,M}(\mathbf{X})$ and $g_{2,K}(\mathbf{X})$. Henze [8] investigated the asymptotic behaviour of $b_{2,K}(\mathbf{X})$ under fairly general conditions and in some relevant special cases, including normality, ellipticity and mixing. Klar [12] derived the asymptotic distribution of $b_{2,K}(\mathbf{X})$ and $g_{2,K}(\mathbf{X})$ both under multivariate normality and under distributions with nonnull fourth cumulants, using orthogonal multivariate polynomials. Henze [9] surveyed previous literature on $b_{2,K}(\mathbf{X})$ and $g_{2,K}(\mathbf{X})$ within the framework of invariant tests for multivariate normality. The statistic $g_{2,K}(\mathbf{X})$ is often used in conjunction with Mardia's skewness ([24]) when testing multivariate normality [15, 27, 36].

The literature on Koziol's kurtoses is much more limited than the literature on Mardia's kurtosis. The only paper deriving closed form expressions for Koziol's kurtosis of some nonnormal distributions appears to be Henze [8]. Finally, to the best of our knowledge, only Franceschini and Loperfido [6] investigated the connections between Koziol's kurtosis and other measures of multivariate kurtosis. This paper addresses these problems by investigating some properties of Koziol's measures of multivariate kurtosis which will hopefully motivate their use in statistical practice. In particular, the paper shows some of their advantages over Mardia's kurtosis, uses inequalities to highlight their connections with other measures of shape and obtains their analytical formulae for some well-known multivariate statistical models. Simple examples illustrate the interpretation of Koziol's measures of multivariate kurtosis and detect a wrong statement about them which appeared in the statistical literature.

The paper is structured as follows. Section 2 and Section 3 deal with the interpretation and the representation of Koziol's kurtosis, including a counterexample to a wrong statement which appeared in the literature. Section 4 connects Koziol's kurtoses to other measures of multivariate kurtosis and to Mardia's skewness. Section 5 gives analytical formulae for Koziol's kurtoses of some well-known multivariate distributions. Section 6 and 7 contain a simulation study and concluding remarks. Theorems' proofs are in the Appendix.

2. Interpretation

This section deals with the interpretation of Koziol's kurtosis, with special emphasis on its comparison with Mardia's kurtosis. Theory is illustrated with simple examples based on bivariate distributions with finite support.

In order to get a better insight into Mardia's and Koziol's kurtoses, we consider the bivariate random vector $\mathbf{u} = (U_1, U_2)^\top$ whose distribution places equal probability mass on the bivariate real vectors

$$\begin{pmatrix} \cos\omega \\ \sin\omega \end{pmatrix}, \begin{pmatrix} \cos\omega \\ -\sin\omega \end{pmatrix}, \begin{pmatrix} -\cos\omega \\ \sin\omega \end{pmatrix}, \begin{pmatrix} -\cos\omega \\ -\sin\omega \end{pmatrix},$$
$$\begin{pmatrix} \sin\omega \\ \cos\omega \end{pmatrix}, \begin{pmatrix} \sin\omega \\ -\cos\omega \end{pmatrix}, \begin{pmatrix} -\sin\omega \\ \cos\omega \end{pmatrix}, \begin{pmatrix} -\sin\omega \\ -\cos\omega \end{pmatrix},$$

where ω is a real value. The marginal distribution of the i-th component of $\mathbf{u}$ is

$$\Pr(U_i = \pm\cos\omega) = \Pr(U_i = \pm\sin\omega) = 0.25,$$

where $\Pr(X = \pm x) = \varkappa \in \mathbb{R}$ is shorthand for $\Pr(X = x) = \Pr(X = -x) = \varkappa$. The mean, variance, skewness (that is the third standardized moment) and kurtosis of U_i are

$$\mathrm{E}(U_i) = 0, \ \mathrm{V}(U_i) = 0.5, \ \gamma_1(U_i) = 0 \text{ and } \beta_2(U_i) = 2 - \sin^2(2\omega).$$

The fourth-order standardized moments of $\mathbf{u}$ are

$$\mu_{1111} = \mu_{2222} = 2 - \sin^2(2\omega), \mu_{1122} = \mu_{1212} = \mu_{1221} = \mu_{2211} = \mu_{2121} = \mu_{2112} = \sin^2(2\omega),$$
$$\mu_{1222} = \mu_{2122} = \mu_{2212} = \mu_{2221} = \mu_{2111} = \mu_{1211} = \mu_{1131} = \mu_{1112} = 0.$$

Mardia's kurtosis of $\mathbf{u}$ equals four regardless of the value taken by the parameter ω:

$$\beta_{2,M}(\mathbf{u}) = 2\left\{2 - \sin^2(2\omega)\right\} + 2\sin^2(2\omega) = 4.$$

Despite that, ω controls the clustering of the vectors in the support of $\mathbf{u}$: for $\omega = \pi/8$ the vectors are evenly spaced on the unit circle, for $\omega = \pi/2$ four pairs of vectors overlap on opposite sides of both axes, for $\omega = \pi/16$ there are four clusters of two vectors symmetrically placed around the origin. Does this mean that fourth-order moments of $\mathbf{u}$ are unable to discriminate between different patterns of its outcomes? In other words, does multivariate kurtosis give an insight into the cluster structure of $\mathbf{u}$?

We address the problem by considering the cosines of the angles between two vectors in the support of $\mathbf{u}$. Let $\mathbf{v}$ be a bivariate random vector independent of $\mathbf{u}$ and identically distributed. The cosine of the angle between $\mathbf{u}$ and $\mathbf{v}$ is

$$\frac{\mathbf{u}^\top\mathbf{v}}{\sqrt{(\mathbf{u}^\top\mathbf{u})(\mathbf{v}^\top\mathbf{v})}} = \mathbf{u}^\top\mathbf{v}.$$

The distribution of $\mathbf{u}^\top\mathbf{v}$ is

$$\Pr\left(\mathbf{u}^\top\mathbf{v} = \pm 1\right) = \Pr\left\{\mathbf{u}^\top\mathbf{v} = \pm\cos(2\omega)\right\} = \Pr\left\{\mathbf{u}^\top\mathbf{v} = \pm\sin(2\omega)\right\} = \frac{\Pr(\mathbf{u}^\top\mathbf{v} = 0)}{2} = \frac{1}{8}.$$

The mean, variance, skewness and the kurtosis of $\mathbf{u}^\top\mathbf{v}$ are

$$\mathrm{E}\left(\mathbf{u}^\top\mathbf{v}\right) = 0, \ \mathrm{V}\left(\mathbf{u}^\top\mathbf{v}\right) = 0.5, \ \gamma_1\left(\mathbf{u}^\top\mathbf{v}\right) = 0, \beta_2\left(\mathbf{u}^\top\mathbf{v}\right) = \frac{\cos(8\omega) + 7}{4}.$$

The first and the second derivatives of the kurtosis of $\mathbf{u}^\top\mathbf{v}$ with respect to ω are

$$\frac{\partial\beta_2(\mathbf{u}^\top\mathbf{v})}{\partial\omega} = -2\sin(8\omega), \ \frac{\partial^2\beta_2(\mathbf{u}^\top\mathbf{v})}{\partial^2\omega} = -16\cos(8\omega).$$

The kurtosis of $\mathbf{u}^\top\mathbf{v}$ achieves its minimum 1.5 when $\omega = \pi/8$, that is when the vectors in the support are evenly spaced on the unit circle. On the other hand, the kurtosis of $\mathbf{u}^\top\mathbf{v}$ achieves its maximum 2 when $\omega = \pi/2$, that is when four pairs of vectors in the support overlap on opposite sides of the axes. Koziol's kurtosis is exactly four times the kurtosis of $\mathbf{u}^\top \mathbf{v}$, as it happens for any symmetric and bivariate distribution whose support is the unit circle:

$$\beta_{2,K}(\mathbf{x}) = 2\left\{2 - \sin^2(2\omega)\right\}^2 + 6\sin^4(2\omega) = \cos(8\omega) + 7.$$

We conclude that Koziol's kurtosis gives some insight into the clustering structure of $\mathbf{u}$, while Mardia's kurtosis does not.

We generalize the above argument to any dimension using angles between random vectors. The Mahalanobis angle between two random outcomes $\mathbf{x}$ and $\mathbf{y}$ from a p-dimensional distribution with mean $\boldsymbol{\mu}$ and nonsingular covariance $\boldsymbol{\Sigma}$ is

$$\frac{(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})}{\sqrt{(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})}},$$

that is the cosine of the standardizations of $\mathbf{x}$ and $\mathbf{y}$ ([26]). High values of Mahalanobis angles indicate that the random vectors belong to the same orthant, while low values indicate that the random vectors belong to opposite orthants with respect to the origin. As a direct consequence, nonnegligible probabilities of Mahalanobis angles with high absolute values suggest the presence of linearly separable clusters. Mardia [26] introduced Mahalanobis angles to help the interpretation of its measure of multivariate skewness and kurtosis, while Juan and Prieto [11] used them for outlier detection.

Unfortunately, both moments and the distribution of Mahalanobis distances do not seem to be analytically tractable. We address the problem by replacing the denominator of a Mahalanobis angle with its expectation, that is the dimension of the random vector, and refer to this ratio as to the quasi-Mahalanobis angle. More formally, the quasi-Mahalanobis angle between two random outcomes $\mathbf{x}$ and $\mathbf{y}$ from a p-dimensional distribution with mean $\boldsymbol{\mu}$ and nonsingular covariance $\boldsymbol{\Sigma}$ is

$$\frac{(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})}{p}.$$

When the variance of $(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ gets smaller, that is when Mardia's kurtosis gets closer to the squared dimension of the random vector, the approximation of a Mahalanobis angle with a quasi-Mahalanobis angle improves. In particular, Mahalanobis and quasi-Mahalanobis angles coincide whenever the support of the distribution is the p-dimensional unit sphere, as it happens with the von Mises-Fisher distribution.

The mean, variance, skewness and kurtosis of quasi-Mahalanobis angles have a straightforward interpretation, since

$$\mathrm{E}\left\{(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\} = 0,\ \mathrm{E}\left[\left\{(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\}^2\right] = p,$$

$$\mathrm{E}\left[\left\{(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\}^3\right] = \beta_{1,M}(\mathbf{x}),\ \mathrm{E}\left[\left\{(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\}^4\right] = \beta_{2,K}(\mathbf{x}).$$

The first two identities are derived in the proof of Theorem 1. The third and fourth moments of $(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})$ are just Mardia's skewness ([24]) and Koziol's kurtosis, respectively. Koziol's kurtosis is then p^2 times the kurtosis of a quasi-Mahalanobis angle of the corresponding distribution. Consistently with the interpretation of univariate kurtosis, high values of Koziol's kurtosis indicate that extreme quasi-Mahalanobis angles may occur with nonnegligible probability.

The following inequality highlights another connection between quasi-Mahalanobis angles and Koziol's kurtosis. It basically states that extreme values of the former are less likely to occur when the latter is low.

Theorem 1. *Let* $\mathbf{x}$ *be a* $p-$*dimensional random vector with mean* $\boldsymbol{\mu}$*, nonsingular covariance* $\boldsymbol{\Sigma}$ *and finite fourth-order moments. Also, let* $\mathbf{y}$ *be a random vector independent of* $\mathbf{x}$ *and identically distributed. Then the following inequality holds true for any real value* $\alpha > 1/\sqrt{p}$*:*

$$\Pr\left\{\left|\frac{(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})}{p}\right| \geq \alpha\right\} \leq \frac{\beta_{2,K}(\mathbf{x}) - p^2}{\alpha^4 p^4 - 2\alpha^2 p^3 + \beta_{2,K}(\mathbf{x})}.$$

As mentioned before, Mahalanobis angles and quasi-Mahalanobis angles coincide when the support of the underlying distribution is the p-dimensional unit sphere. When the same distribution is symmetric its Mardia's kurtosis equals p^2 and therefore provides very little information on the distribution itself. Koziol's kurtoses do not suffer from the same limitation, due to their connections with quasi-Mahalanobis angles. We conclude that Mardia's and Koziol's kurtoses complement each other and that the latter should be used when dealing with symmetric distributions on the p-dimensional unit sphere, as for example the von Mises-Fisher distribution. Ley and Verdebout [18] thoroughly review directional distributions and directional statistics. Koziol's kurtoses and Mahalanobis angles also give some insights into the interpretation of kurtosis as a measure of peakedness. Problems with this interpretation have been discussed in the univariate case [4, 22] but not in the multivariate one. The examples and the theorems in this section show that it is quite easy to compare the Mahalanobis angles of two symmetric distributions on the p-dimensional unit sphere by means of Koziol's kurtoses. On the other hand, there is not a natural and intuitive way to define the peakedness of a distribution whose support is the p-dimensional unit sphere.

As remarked by several authors [3, 10] the interpretation of kurtosis is not straightforward. This is particularly true for Koziol's and Mardia's kurtoses, whose algebraic nature gives only a limited geometric insight. As an example, consider the fourth-order standardized moments which are components of Koziol's, but not Mardia's, measure of multivariate kurtosis: $E\left(Z_i Z_j Z_h Z_k\right)$ and $E\left(Z_i^2 Z_j Z_h\right)$, where $\mathbf{z} = \left(Z_1, ..., Z_p\right)^\top$ is a $p-$dimensional random vector, the indices i, j, h, k differ from each other and range from 1 to p. At present time, there is not a simple distributional property associated to either low or high values of these components. The interpretation of kurtosis is even more difficult in the presence of skewness, given the close connections between the two concepts. The problem has been addressed with quantile-based measures of kurtosis ([10]) and with symmetrizing linear projections [20, 22] in the univariate and the multivariate case, respectively. Unfortunately, the first approach is not easily generalizable to the multivariate case, while the second approach implies some loss of information.

3. Representation

Koziol's kurtoses admits several meaningful and useful representations. This section reviews them and shows a new one. Koziol's excess kurtosis $\gamma_{2,K}(\mathbf{x})$, that is the sum of all squared fourth-order cumulants, does not have a straightforward interpretation in terms of quasi-Mahalanobis angles, but it may be recovered from Mardia's kurtosis $\beta_{2,M}(\mathbf{x})$, Koziol's kurtosis $\beta_{2,K}(\mathbf{x})$ and the vector's dimension p via the identity ([16])

$$\beta_{2,K}(\mathbf{x}) = \gamma_{2,K}(\mathbf{x}) + 6\beta_{2,M}(\mathbf{x}) - 6p - 3p^2.$$

For example, Koziol's excess kurtosis of the vector $\mathbf{u}$ defined in the previous section is

$$\gamma_{2,K}(\mathbf{x}) = \cos(8\omega) + 7 - 6 \cdot 4 + 6 \cdot 2 + 3 \cdot 2^2 = \cos(8\omega) + 7,$$

which equals Koziol's kurtosis, as it happens for all bivariate distributions whose support is an ellipse.

Both Mardia's and Koziol's kurtoses are simple functions of the fourth standardized moment matrix

$$\mathbf{M}_{4,\mathbf{z}} = \mathrm{E}\left(\mathbf{z} \otimes \mathbf{z}^\top \otimes \mathbf{z} \otimes \mathbf{z}^\top\right),$$

where $\otimes$ denotes the Kronecker (tensor) product ([13]). For example, the fourth standardized moment matrix of $\mathbf{u}$ is

$$\begin{pmatrix} 2 - \sin^2(2\omega) & 0 & 0 & \sin^2(2\omega) \\ 0 & \sin^2(2\omega) & \sin^2(2\omega) & 0 \\ 0 & \sin^2(2\omega) & \sin^2(2\omega) & 0 \\ \sin^2(2\omega) & 0 & 0 & 2 - \sin^2(2\omega) \end{pmatrix}.$$

Mardia's kurtosis coincides with the trace of the fourth standardized moment ([14]) while Koziol's kurtosis coincides with the trace of the fourth squared standardized moment ([13]):

$$\mathrm{tr}\left(\mathbf{M}_{4,\mathbf{z}}\right) = \beta_{2,M}(\mathbf{x}), \ \mathrm{tr}\left(\mathbf{M}_{4,\mathbf{z}}^2\right) = \beta_{2,K}(\mathbf{x}).$$

Similarly, Mardia's excess kurtosis coincides with the trace of the fourth standardized cumulant while Koziol's excess kurtosis coincides with the trace of the fourth squared standardized cumulant:

$$\operatorname{tr}\left(\mathbf{K}_{4,\mathbf{z}}\right) = \gamma_{2,M}\left(\mathbf{x}\right), \operatorname{tr}\left(\mathbf{K}_{4,\mathbf{z}}^{2}\right) = \gamma_{2,K}\left(\mathbf{x}\right).$$

The fourth standardized cumulant is the $p^2 \times p^2$ block matrix $\mathbf{K}_{4,\mathbf{z}} = \left\{\mathbf{K}_{pq}\right\}$ where

$$\mathbf{K}_{pq} = \left.\frac{\partial^4 \log E\left[\exp\left(\iota \mathbf{t}^\top \mathbf{z}\right)\right]}{\partial t_p \partial t_q \partial \mathbf{t} \partial \mathbf{t}^\top}\right|_{\mathbf{t}=\mathbf{0}}.$$

Equivalently, the element in the $i-$th row and in the $j-$th column of the $(h,k)-$th block is the $ijhk-$cumulant κ_{ijhk}. For example, the fourth standardized cumulant matrix of $\mathbf{u}$ is

$$-\begin{pmatrix} \sin^2(2\omega)+1 & 0 & 0 & \cos^2(2\omega) \\ 0 & \cos^2(2\omega) & \cos^2(2\omega) & 0 \\ 0 & \cos^2(2\omega) & \cos^2(2\omega) & 0 \\ \cos^2(2\omega) & 0 & 0 & \sin^2(2\omega)+1 \end{pmatrix}.$$

Computation of Koziol's kurtosis and excess kurtosis with the above representations require the standardization of the random vector. The fourth central moment and the fourth cumulant of $\mathbf{x}$ are the $p^2 \times p^2$ matrices

$$\overline{\mathbf{M}}_{4,\mathbf{x}} = \mathrm{E}\left\{(\mathbf{x}-\boldsymbol{\mu}) \otimes (\mathbf{x}-\boldsymbol{\mu})^\top \otimes (\mathbf{x}-\boldsymbol{\mu}) \otimes (\mathbf{x}-\boldsymbol{\mu})^\top\right\}, \mathbf{K}_{4,\mathbf{x}} = \overline{\mathbf{M}}_{4,\mathbf{x}} - \operatorname{vec}(\boldsymbol{\Sigma}) \operatorname{vec}^\top(\boldsymbol{\Sigma}) - \left(\mathbf{I}_{p^2} + \mathbf{C}_{p,p}\right)(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}),$$

where $\mathbf{I}_{p^2}$ is the identity matrix of size p^2, $\mathbf{C}_{p,p}$ is the $p^2 \times p^2$ Commutation matrix and $vec(\boldsymbol{\Sigma})$ is the vectorization of $\boldsymbol{\Sigma}$. The matrices $\overline{\mathbf{M}}_{4,\mathbf{x}}$, $\mathbf{K}_{4,\mathbf{x}}$ and $\boldsymbol{\Sigma}^{-1}$ might be used to compute Koziol's kurtoses of a random vector without standardizing it, as shown in the following theorem.

Theorem 2. *Let $\boldsymbol{\Sigma}$, $\overline{\mathbf{M}}_{4,\mathbf{x}}$ and $\mathbf{K}_{4,\mathbf{x}}$ be the nonsingular covariance matrix, the fourth central moment and the fourth cumulant of the p-dimensional random vector $\mathbf{x}$. Then Koziol's kurtosis and Koziol's excess kurtosis are*

$$\beta_{2,K}(\mathbf{x}) = \operatorname{vec}^\top\left(\boldsymbol{\Sigma}^{-1}\right)\overline{\mathbf{M}}_{4,\mathbf{x}}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\overline{\mathbf{M}}_{4,\mathbf{x}}\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right),$$

$$\gamma_{2,d}(\mathbf{x}) = \operatorname{vec}^\top\left(\boldsymbol{\Sigma}^{-1}\right)\mathbf{K}_{4,\mathbf{x}}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\mathbf{K}_{4,\mathbf{x}}\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right).$$

Scalar functions of either the fourth standardized moment or the fourth standardized cumulant, such as Mardia's and Koziol's kurtoses, may not retain enough information about the kurtosis structure of the underlying distribution. Kollo [13] addressed the problem by summarizing these $p^2 \times p^2$ matrices with the $p \times p$ matrix

$$\mathbf{B} = \sum_{i,j} E\left(Z_i Z_j \mathbf{z}\mathbf{z}^\top\right),$$

that is the sum of the matrices $\mathbf{M}_{ij} = E\left(Z_i Z_j \mathbf{z}\mathbf{z}^\top\right)$ which constitute the blocks of the fourth standardized moment $\mathbf{M}_{4,\mathbf{z}} = \left\{\mathbf{M}_{ij}\right\}$. For example, Kollo's kurtosis matrix of $\mathbf{u}$ is

$$2\begin{pmatrix} 1 & \sin^2(2\omega) \\ \sin^2(2\omega) & 1 \end{pmatrix}.$$

Its squared Euclidean norm is $\cos(8\omega) - 4\cos(4\omega) + 11 \neq \cos(8\omega) + 7 = \beta_{2,K}(\mathbf{u})$. Hence the distribution of $\mathbf{u}$ provides a counter example to the incorrect statement $\beta_{2,K}(\mathbf{x}) = \|\mathbf{B}\|^2$ which appears in Kollo [13].

4. Inequalities

This section focuses on some inequalities between Koziol's kurtoses and the measure of multivariate kurtosis in Malkovich and Afifi [23] and the measure of multivariate skewness in Mardia [24]. It also investigates lower bounds for Koziol's kurtoses.

The following inequality holds true for a random variable X with mean μ, variance σ^2 and finite fourth moment:

$$\beta_2(X) = \mathrm{E}\left\{\left(\frac{X-\mu}{\sigma}\right)^4\right\} \geq 1 + \mathrm{E}^2\left\{\left(\frac{X-\mu}{\sigma}\right)^3\right\} = 1 + \beta_1(X).$$

Its extensions to the multivariate case and to specific classes of multivariate distributions are thoroughly reviewed in Ogasawara [32]. Special emphasis has been given to inequalities between Mardia's (excess) kurtosis and measures of multivariate skewness [14, 29, 32]. To the best of our knowledge, inequalities involving Koziol's kurtosis and measures of multivariate skewness have never been derived.

Mardia [24] generalized $\beta_1(X)$, that is the square of the third standardized moment, to the multivariate case as follows:

$$\beta_{1,M}(\mathbf{x}) = \mathrm{E}\left[\left\{(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\}^3\right],$$

where $\mathbf{x}$ and $\mathbf{y}$ are independent and identically distributed $p-$dimensional random vectors with mean $\boldsymbol{\mu}$, nonsingular variance $\boldsymbol{\Sigma}$ and finite third-order moments. Several authors [14, 29, 32] independently showed that Mardia's kurtosis is never smaller than the sum of Mardia's skewness and the vector's dimension:

$$\beta_{2,M}(\mathbf{x}) \geq \beta_{1,M}(\mathbf{x}) + p.$$

The following inequality shows that Mardia's skewness of a random vector is never greater than the product of its Koziol's kurtosis and its dimension.

Theorem 3. *Let* $\mathbf{x}$ *be a p-dimensional random vector with finite fourth moments and positive definite covariance matrix. Then the squared Mardia's skewness of* $\mathbf{x}$ *is never greater than p times its Koziol's kurtosis:*

$$\beta_{1,M}^2(\mathbf{x}) \leq p \cdot \beta_{2,K}(\mathbf{x}).$$

Data projections achieving either maximal or minimal kurtosis appear in independent component analysis, projection pursuit, outlier detection, cluster analysis and normality testing (see [7]). Mardia's kurtosis of a random vector is never smaller than the maximum kurtosis achievable by a linear projection of the random vector itself ([6]). Similarly, the following inequality shows that Koziol's excess kurtosis is never smaller than Malkovich and Afifi's kurtosis [23], that is the maximum value of the squared fourth cumulant achievable by a linear projection of the random vector itself:

$$\gamma_{2,A}(\mathbf{x}) = \sup_{\mathbf{c}\in\mathbb{R}_0^p} \gamma_2\left(\mathbf{c}^\top \mathbf{x}\right),$$

where $\mathbb{R}_0^p$ is the set of p-dimensional and nonnull real vectors ([23]).

Theorem 4. *Let* $\mathbf{x}$ *be a* $p-$*dimensional random vector with finite fourth-order moments and positive definite covariance matrix. Then Malkovich and Afifi's kurtosis is not greater than Koziol's excess kurtosis:*

$$\gamma_{2,A}(\mathbf{x}) \leq \gamma_{2,K}(\mathbf{x}).$$

We illustrate the above inequality with the random vector $\mathbf{u}$. An argument similar to the one in Franceschini and Loperfido [7] implies that the fourth standardized cumulant of a linear projection of $\mathbf{u}$ might be represented as

$$-\begin{pmatrix} \cos^2\theta \\ \sin\theta\cos\theta \\ \sin\theta\cos\theta \\ \sin^2\theta \end{pmatrix}^\top \begin{pmatrix} \sin^2(2\omega)+1 & 0 & 0 & \cos^2(2\omega) \\ 0 & \cos^2(2\omega) & \cos^2(2\omega) & 0 \\ 0 & \cos^2(2\omega) & \cos^2(2\omega) & 0 \\ \cos^2(2\omega) & 0 & 0 & \sin^2(2\omega)+1 \end{pmatrix} \begin{pmatrix} \cos^2\theta \\ \sin\theta\cos\theta \\ \sin\theta\cos\theta \\ \sin^2\theta \end{pmatrix},$$

where $\theta \in [0, 2\pi]$ is a real vaue. It might be simplified into

$$\frac{\cos(4\theta - 4\omega) + \cos(4\theta + 4\omega) - 6}{4},$$

whose square is always smaller than the Koziol's kurtosis of $\mathbf{u}$, consistently with the inequality in Theorem 4:

$$\left\{ \frac{\cos(4\theta - 4\omega) + \cos(4\theta + 4\omega) - 6}{4} \right\}^2 < 7 + \cos(8\omega).$$

We conclude this section with a remark on a lower bound for Koziol's kurtosis. Results in the previous section imply that the kurtosis of a quasi-Mahalanobis angle is

$$\beta_2 \left\{ \frac{(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})}{p} \right\} = \frac{\mathrm{E}\left[\left\{ (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right\}^4 \right]}{\mathrm{E}^2 \left[\left\{ (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right\}^2 \right]} = \frac{\beta_{2,K}(\mathbf{x})}{p^2}.$$

Since univariate kurtosis is never smaller than one, Koziol's kurtosis of a random vector is never smaller than the squared dimension of the random vector itself. Equality holds true if and only if the distribution assigns equal probability to the two real vectors with identical norm which constitute its support.

5. Models

This section derives analytical expressions for Koziol's kurtosis and Koziol's excess kurtosis of two well-known statistical models, thus providing a better insight into their connections with Mardia's kurtosis, as well as into their relative merits for testing nonnormality.

Independent component analysis (ICA) is a multivariate statistical technique aimed at recovering independent, unobserved signals by appropriate data projections. Default ICA methods rely on multivariate kurtosis ([28]). The basic ICA model is $\mathbf{x} = \mathbf{b} + \mathbf{A}\mathbf{s}$, where $\mathbf{x}$ is a p-dimensional random vector, $\mathbf{b}$ is a p-dimensional real vector, $\mathbf{A}$ is a $p \times p$ invertible real matrix and the components of $\mathbf{s} = \left(S_1, ..., S_p\right)^\top$ are mutually independent, standardized random variables of which at most one is normal ([28]). The following theorem connects Koziol's kurtosis of $\mathbf{x}$, Koziol's excess kurtosis of $\mathbf{x}$ and Mardia's kurtosis of $\mathbf{x}$.

Theorem 5. *Let $\beta_2(S_1),..., \beta_2\left(S_p\right)$ be the fourth standardized moments of the mutually independent random variables $S_1,..., S_p$. Also, let $\mathbf{x}$ be an affine, one-to-one transformation of $\mathbf{s} = \left(S_1, ..., S_p\right)^\top$: $\mathbf{x} = \mathbf{b} + \mathbf{A}\mathbf{s}$, where $\mathbf{b}$ is a p-dimensional real vector and $\mathbf{A}$ is a $p \times p$ matrix of full rank. Then Koziol's kurtosis, Mardia's kurtosis and Koziol's excess kurtosis of $\mathbf{x}$ are*

$$\beta_{2,K}(\mathbf{x}) = 3p(p-1) + \sum_{i=1}^{p} \beta_2^2(S_i),\ \beta_{2,M}(\mathbf{x}) = p(p-1) + \sum_{i=1}^{p} \beta_2(S_i),\ \gamma_{2,K}(\mathbf{x}) = \sum_{i=1}^{p} \{\beta_2(S_i) - 3\}^2.$$

The above results allow for a better insight into the limitations of $\beta_{2,M}$, $\beta_{2,K}$ and $\gamma_{2,K}$ when checking normality. Let S_1 and S_2 be two independent random variables whose distribution is

$$\Pr(S_i = -1) = P(S_i = 1) = \frac{1}{2\alpha_i},\ \Pr(S_i = 0) = 1 - \frac{1}{\alpha_i},$$

where $\alpha_i \geq 1$ and $i \in \{1, 2\}$. The first four moments of S_i are

$$\mathrm{E}(S_i) = \mathrm{E}\left(S_i^3\right) = 0,\ \mathrm{E}\left(S_i^2\right) = \mathrm{E}\left(S_i^4\right) = \frac{1}{\alpha_i},$$

so that the kurtosis of S_i is $\beta_2(S_i) = \alpha_i$. Also, let

$$X_1 = b_1 + a_{11}S_1 + a_{12}S_2,\, X_2 = b_2 + a_{21}S_1 + a_{22}S_2,$$

where $b_1, b_2, a_{11}, a_{12}, a_{21}, a_{22} \in \mathbb{R}$ and $a_{11}a_{22} \neq a_{12}a_{21}$. Mardia's kurtosis, Koziol's kurtosis and Koziol's excess kurtosis of $\mathbf{x} = (X_1, X_2)^\top$ are

$$\beta_{2,M}(\mathbf{x}) = \alpha_1 + \alpha_2 + 2,\, \beta_{2,K}(\mathbf{x}) = \alpha_1^2 + \alpha_2^2 + 6 \text{ and } \gamma_{2,K}(\mathbf{x}) = (\alpha_1 - 3)^2 + (\alpha_2 - 3)^2.$$

Despite $\mathbf{x}$ being a nonnormal random vector, these measures of multivariate kurtosis equal the values of the corresponding measures for a bivariate normal random vector when $\alpha_1 + \alpha_2 = 6$, $\alpha_1^2 + \alpha_2^2 = 18$ and $\alpha_1 = \alpha_2 = 3$. However, $\beta_{2,M}(\mathbf{x})$ and $\beta_{2,K}(\mathbf{x})$ might equal Mardia's kurtosis and Koziol's kurtosis of a bivariate normal random vector if some fourth-order cumulants of $\mathbf{x}$ differ from zero, while the same does not happen with $\gamma_{2,K}(\mathbf{x})$.

The probability density function of an extended skew-normal distribution (ESN) is

$$\frac{\phi_d(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Omega})}{\Phi(\tau)} \Phi\left(\tau\sqrt{1 + \boldsymbol{\alpha}^\top \boldsymbol{\Omega} \boldsymbol{\alpha}} + \boldsymbol{\alpha}^\top \mathbf{x} - \boldsymbol{\alpha}^\top \boldsymbol{\xi}\right),$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution and $\phi_d(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Omega})$ is the probability density function of a p-dimensional normal distribution with mean $\boldsymbol{\xi}$ and nonsingular variance $\boldsymbol{\Omega}$. The parameters $\boldsymbol{\alpha}$ and τ are a real p-dimensional vector and a real scalar, respectively ([2]). We write $\mathbf{x} \sim \mathcal{ESN}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \tau)$ to denote a p-dimensional random vector with the above probability density function, which is normal if and only if $\boldsymbol{\alpha}$ is a null vector. The following theorem shows that Koziol's excess kurtosis of an extended skew-normal random vector is just the square of its Mardia's excess kurtosis.

Theorem 6. *Let $\gamma_{2,M}(\mathbf{x})$ and $\gamma_{2,K}(\mathbf{x})$ be Mardia's and Koziol's excess kurtoses of an extended skew-normal random vector $\mathbf{x} \sim \mathcal{ESN}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \tau)$. Then the following identities hold true:*

$$\gamma_{2,K}(\mathbf{x}) = \gamma_{2,M}^2(\mathbf{x}) = \zeta_4^2(\tau)\left\{\frac{\boldsymbol{\delta}^\top \Omega^{-1} \boldsymbol{\delta}}{1 + \zeta_2(\tau)\boldsymbol{\delta}^\top \Omega^{-1} \boldsymbol{\delta}}\right\}^4,$$

$$\zeta_2(\tau) = \frac{\partial^2 \log \Phi(\tau)}{\partial^2 \tau} \zeta_4(\tau) = \frac{\partial^4 \log \Phi(\tau)}{\partial^4 \tau},\, \boldsymbol{\delta} = \frac{\boldsymbol{\Omega}\boldsymbol{\alpha}}{\sqrt{1 + \boldsymbol{\alpha}^\top \boldsymbol{\Omega} \boldsymbol{\alpha}}}$$

and $\Phi(\cdot)$ denotes the cumulative distribution function of a standard normal random variable.

6. Simulations

The paper by Best and Rayner [5] contains the only simulation study assessing the performance of Koziol's excess kurtosis as a test statistic for multivariate normality. It is limited to the bivariate case and comparisons are made only with respect to Mardia's kurtosis test for normality. To the best of our knowledge, the kurtosis-based tests for multivariate normality proposed by Mardia [24], Malkovich and Afifi [23], Koziol [15] and Koziol [16], have never been compared by means of simulation studies. This section addresses the problem within the framework of skew-normal distributions and independent components models.

Best and Rayner [5] simulated 5000 samples of size not greater than 50 from several bivariate, nonnormal distributions and compared the powers of several tests for bivariate normality, including those based on Koziol's excess kurtosis and Mardia's kurtosis. They concluded that the former should be preferred over the latter. The simulation study in this section overcomes several limitations of the one in [5]. First, it addresses multivariate normality in dimensions higher than two. Second, it uses sample sizes greater than 50. Third, it also considers the performance of Malkocich and Afifi's kurtosis test of multivariate normality.

We simulated 10000 samples of $n = 25, 50, 75, 100$ units and $p = 2, 4, 6, 8, 10$ variables from the skew-normal distribution and from an ICA model. The parameters indexing the sampled densities have been chosen to account for different degrees of nonnormality. For each sample we computed the test statistics of Koziol's excess kurtosis, Koziol's kurtosis, Mardia's kurtosis and Malkovich and Afifi's kurtosis tests of multivariate normality. Finally, for each combination of units, variables and parameters we computed the percentage of sample rejected at the 0.05 level by the four tests.

First we simulated from the skew-normal distribution $SN_p\left(\mathbf{0}_p, \mathbf{I}_p, \alpha\mathbf{1}_p\right)$, where $\mathbf{0}_p$, $\mathbf{I}_p$ and $\alpha\mathbf{1}_p$ are the p-dimensional null vector, the p-dimensional identity matrix and the p-dimensional vector of ones. The parameter α has been set equal to either 5, 15 or 25 to account for increasing degrees of kurtosis. Next we simulated from the ICA model $\mathbf{x} = \mathbf{A}\mathbf{s}$, where $\mathbf{A}$ is the product of a Helmert matrix and a diagonal matrix, $\mathbf{s}$ is a p-dimensional random vector whose i-th component is a chi-square random variable with ωi degrees of freedom ($\omega = 20, 40, 60$), with higher degrees of freedom leading to higher kurtoses.

Tables 1 and 2 report the simulations' results, which are very similar to each other and might be summarized as follows. For each combination of units, variables and parameters the percentages of rejected samples do not change much from one test to another and no test is uniformly most powerful. The performances of the tests improve when the number of units increase and the number of variables decrease. The powers of the tests increase with the nonnormality of the sampled distribution, but only slightly so. Finally, the present simulations suggest that the four tests are unbiased.

7. Conclusion

This paper reviews and investigates the properties of Koziol's measures of multivariate kurtosis. Koziol's kurtoses are invariant with respect to one-to-one affine transformations. They are strictly related to Mahalanobis angles and might be meaningfully represented by means of the fourth standardized moment and cumulant matrices. Moreover, Koziol's kurtosis might be represented as a quadratic form involving the concentration matrix and the fourth central moment matrix. Koziol's kurtoses are related to other skewness and kurtosis measures by means of inequalities. Finally, they have simple analytical expressions in some well-known statistical models.

Despite their merits, Koziol's kurtoses suffer from several limitations. They are not defined for distributions whose fourth moments do not exist, as it might happen for the multivariate Pareto distribution ([34]). Moreover, Koziol's kurtoses might be unable to detect nonnormality, as it happens for the shape mixture of skew-normal distributions described in Loperfido [19]. Finally, Yanagihara [37] pointed out some limitations of the kurtosis measure in Koziol's kurtosis when estimating the residuals' kurtosis, within the framework of a multivariate, nonnormal regression model.

We recommend the joint use of Koziol's and Mardia's kurtoses in order to have a better insight into data structures. In the first place, Mardia's and Koziol's kurtoses are strictly related to large sample values of Mahalanobis distances and angles, which might be used to detect outliers ([11]). In the second place Koziol's kurtosis might be able to detect nonnormality when Mardia's kurtosis does not and vice versa, as illustrated in Section 5 with a simple ICA example. In the third place, they appear together when testing for multivariate normality ([27]). However, Koziol's and Mardia's tests for multivariate normality should be used cautiously. As remarked by Henze [9], any test for multivariate normality should be affine invariant as well as consistent. Koziol's and Mardia's measures fail to satisfy the latter property, even when used together with skewness measures as test statistics for multivariate normality.

The simulations studies in this paper are limited to two parametric families of multivariate distributions. Simulations studies with other distributions are needed to assess, understand and compare the performances of the main kurtosis-based tests for multivariate normality. A simulation study might consider a multivariate, symmetric and unimodal distribution, as for example the multivariate t distribution. Another simulation study might consider a multivariate, possibly skewed and multimodal distribution, as for example the multivariate beta distribution. Follow-up studies might consider hybrid approaches based on both distributions. The interpretation of simulations' results is easier when Koziol's and Mardia's kurtoses of the simulated distributions have symple analytical forms, as it is likely to happen with the multivariate t and beta distributions, since their higher moments have simple analytical forms, too. We hope that this paper will encourage further research along this line.

Acknowledgments

The author would like to thank the Editor, the Associate Editor and two anonymous Referees for their comments, which greatly helped in improving the quality of the present paper.

Table 1: Integer parts of the percentages of skew-normal samples rejected at the 0.05 level by Koziol's test of excess kurtosis ($\gamma_{2,K}$), Koziol's kurtosis test ($\beta_{2,K}$), Mardia's kurtosis test ($\beta_{2,M}$) and Malkovich and Afifi' kurtosis test ($\gamma_{2,A}$) for n = 25, 50, 75, 100 units, p = 2, 4, 6, 8, 10 variables and parameter's values α = 5, 15, 25.

Units	**Variables**	α = **5**				α = **15**				α = **25**			
		$\gamma_{2,K}$	$\beta_{2,K}$	$\beta_{2,M}$	$\gamma_{2,A}$	$\gamma_{2,K}$	$\beta_{2,K}$	$\beta_{2,M}$	$\gamma_{2,A}$	$\gamma_{2,K}$	$\beta_{2,K}$	$\beta_{2,M}$	$\gamma_{2,A}$
25	2	11	11	10	11	12	11	11	11	11	11	11	11
	4	8	8	7	8	8	8	7	8	8	8	7	8
	6	7	7	6	6	8	8	7	7	7	6	7	6
	8	6	5	5	6	6	6	6	6	6	6	6	6
	10	5	5	5	5	5	5	5	5	5	5	5	5
50	2	15	15	15	15	17	18	17	17	17	18	17	17
	4	11	11	11	11	11	11	10	11	11	11	10	11
	6	9	9	8	8	9	9	9	9	9	9	9	9
	8	8	7	8	7	8	8	7	6	8	8	7	6
	10	7	7	7	7	7	6	6	6	7	6	6	6
75	2	20	20	19	20	21	22	21	22	21	22	20	22
	4	13	13	12	13	14	14	12	14	14	14	13	13
	6	10	10	9	10	11	11	10	11	11	11	10	11
	8	8	8	7	7	9	9	9	8	8	9	8	8
	10	8	8	7	7	8	8	7	7	8	8	7	7
100	2	22	24	21	24	24	25	23	26	24	25	23	26
	4	16	15	14	15	17	17	15	17	17	17	15	17
	6	12	12	11	12	12	12	11	12	13	13	11	12
	8	9	10	9	9	9	9	9	8	9	10	9	9
	10	9	8	8	7	8	8	8	7	8	8	9	8

Table 2: Integer parts of the percentages of ICA samples rejected at the 0.05 level by Koziol's test of excess kurtosis ($\gamma_{2,K}$), Koziol's kurtosis test ($\beta_{2,K}$), Mardia's kurtosis test ($\beta_{2,M}$) and Malkovich and Afifi' kurtosis test ($\gamma_{2,A}$) for n = 25, 50, 75, 100 units, p = 2, 4, 6, 8, 10 variables and parameter's values ω = 20, 40, 60.

Units	**Variables**	ω= **20**				ω= **40**				ω= **60**			
		$\gamma_{2,K}$	$\beta_{2,K}$	$\beta_{2,M}$	$\gamma_{2,A}$	$\gamma_{2,K}$	$\beta_{2,K}$	$\beta_{2,M}$	$\gamma_{2,A}$	$\gamma_{2,K}$	$\beta_{2,K}$	$\beta_{2,M}$	$\gamma_{2,A}$
25	2	7	7	7	7	6	6	6	6	5	5	5	5
	4	7	7	6	7	6	6	5	5	5	6	5	5
	6	6	6	6	6	6	6	6	6	5	5	6	5
	8	6	6	5	5	5	5	5	5	5	5	5	5
	10	5	5	5	6	5	5	5	5	5	5	5	5
50	2	10	10	10	10	7	7	7	7	6	7	7	6
	4	10	9	9	9	7	7	7	7	6	6	6	6
	6	8	8	8	8	6	6	6	6	6	5	6	5
	8	7	7	7	6	5	5	6	5	5	5	5	4
	10	7	6	6	6	6	5	6	5	5	5	5	5
75	2	13	13	13	14	9	9	9	9	7	7	7	8
	4	11	10	10	10	8	8	8	8	7	7	6	7
	6	10	10	9	9	7	7	6	7	6	6	6	6
	8	8	8	8	6	6	6	6	6	5	5	5	5
	10	7	7	7	7	6	6	6	6	5	5	5	5
100	2	14	15	14	15	9	9	8	10	7	8	7	8
	4	12	12	12	12	8	8	7	8	7	6	7	6
	6	11	11	11	10	8	7	7	7	7	7	7	6
	8	9	9	8	7	6	6	6	5	5	6	5	5
	10	8	8	8	7	6	6	6	5	6	5	6	5

Appendix

Proof of Theorem 1. Let $\mathbf{z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu})$ and $\mathbf{w} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})$ be the standardization of $\mathbf{x}$ and $\mathbf{y}$, where $\boldsymbol{\Sigma}^{-1/2}$ is the positive definite symmetric square root of the concentration matrix $\boldsymbol{\Sigma}^{-1}$. Also, let $\mathrm{E}^2(X)$ denote the squared expectation of the random variable X: $\mathrm{E}^2(X) = \{\mathrm{E}(X)\}^2$. The following identities are direct implications of $\mathbf{z}$ and $\mathbf{w}$ being independent, standardized, p-dimensional random vectors:

$$\mathrm{E}\left(\mathbf{z}^\top\mathbf{w}\right) = \mathrm{E}\left(\sum_{i=1}^{p} Z_i W_i\right) = \sum_{i=1}^{p}\mathrm{E}(Z_iW_i) = \sum_{i=1}^{p}\mathrm{E}(Z_i)\,\mathrm{E}(W_i) = \sum_{i=1}^{p}\mathrm{E}^2(Z_i) = 0,$$

$$\mathrm{E}\left\{\left(\mathbf{z}^\top\mathbf{w}\right)^2\right\} = \sum_{i,j=1}^{p}\mathrm{E}\left(Z_iW_iZ_jW_j\right) = \sum_{i,j=1}^{p}\mathrm{E}\left(Z_iZ_j\right)\mathrm{E}\left(W_iW_j\right) = \sum_{i=1}^{p}\mathrm{E}\left(Z_i^2\right)\mathrm{E}\left(W_i^2\right) = p.$$

From Ogasawara [33] we know that the inequality

$$\Pr(|U| \geq \lambda) \leq \frac{\mathrm{E}\left(U^4\right)-1}{\lambda^4 - 2\lambda^2 + \mathrm{E}\left(U^4\right)}$$

holds true for a standard random variable U and a real value λ not smaller than one: $\mathrm{E}(U) = 0$, $\mathrm{E}\left(U^2\right) = 1$, $\lambda \geq 1$. By assumption $\alpha > 1/\sqrt{p}$ and $\mathbf{z}^\top\mathbf{w}/\sqrt{p}$ is a standard random variable. Hence we might let $U = \mathbf{z}^\top\mathbf{w}/\sqrt{p}$ and $\lambda = \alpha\sqrt{p}$ to obtain

$$\Pr\left(\left|\frac{\mathbf{z}^\top\mathbf{w}}{\sqrt{p}}\right| \geq \alpha\sqrt{p}\right) \leq \frac{\mathrm{E}\left\{\left(\mathbf{z}^\top\mathbf{w}/\sqrt{p}\right)^4\right\}-1}{\left(\alpha\sqrt{p}\right)^4 - 2\left(\alpha\sqrt{p}\right)^2 + \mathrm{E}\left\{\left(\mathbf{z}'\mathbf{w}/\sqrt{p}\right)^4\right\}} = \frac{\mathrm{E}\left\{(\mathbf{z}^\top\mathbf{w})^4\right\}-p^2}{\alpha^4p^4 - 2\alpha^2p^3 + \mathrm{E}\left\{(\mathbf{z}^\top\mathbf{w})^4\right\}}.$$

The definitions of $\mathbf{z}$ and $\mathbf{w}$, together with the results in Koziol [16] imply the identities $\mathbf{z}^\top\mathbf{w} = (\mathbf{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})$ and $\beta_{2,K}(\mathbf{x}) = \mathrm{E}\left\{(\mathbf{z}^\top\mathbf{w})^4\right\}$. Thus

$$\Pr\left\{\left|\frac{(\mathbf{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})}{p}\right| \geq \alpha\right\} \leq \frac{\beta_{2,K}(\mathbf{x}) - p^2}{\alpha^4p^4 - 2\alpha^2p^3 + \beta_{2,K}(\mathbf{x})}.$$

Proof of Theorem 2. Let $\mathbf{z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu})$ be the standardized version of $\mathbf{x}$, where $\boldsymbol{\mu}$ is the mean of $\mathbf{x}$ and $\boldsymbol{\Sigma}^{-1/2}$ is the symmetric, positive definite square root of the concentration matrix $\boldsymbol{\Sigma}^{-1}$. Koziol's kurtosis of $\mathbf{x}$ coincides with the trace of the squared fourth moment of $\mathbf{z}$ ([13]): $\beta_{2,K}(\mathbf{x}) = \mathrm{tr}\left(\mathbf{M}_{4,\mathbf{z}}^2\right)$. The trace of the fourth moment $\mathbf{M}_{4,\mathbf{w}}$ of any p-dimensional random vector $\mathbf{w}$ admits the representation $\mathrm{vec}^\top\left(\mathbf{I}_p\right)\mathbf{M}_{4,\mathbf{w}}\mathrm{vec}\left(\mathbf{I}_p\right)$, where $\mathbf{I}_p$ is the p-dimensional identity matrix. Also, the fourth moment of $\mathbf{Qw}$ is $(\mathbf{Q}\otimes\mathbf{Q})\,\mathbf{M}_{4,\mathbf{w}}\,(\mathbf{Q}^\top\otimes\mathbf{Q}^\top)$, where $\mathbf{Q}$ is a $k\times p$ real matrix ([6]). Hence the fourth moment $\mathbf{M}_{4,\mathbf{z}}$ of $\mathbf{z}$ is a simple function of the fourth moment of $\mathbf{y} = \mathbf{x}-\boldsymbol{\mu}$, i.e. the fourth central moment of $\mathbf{x}$:

$$\mathbf{M}_{4,\mathbf{z}} = \left(\boldsymbol{\Sigma}^{-1/2}\otimes\boldsymbol{\Sigma}^{-1/2}\right)\overline{\mathbf{M}}_{4,\mathbf{x}}\left(\boldsymbol{\Sigma}^{-1/2}\otimes\boldsymbol{\Sigma}^{-1/2}\right).$$

Hence Koziol's kurtosis of $\mathbf{x}$ might be represented as

$$\mathrm{vec}^\top\left(\mathbf{I}_p\right)\left(\boldsymbol{\Sigma}^{-1/2}\otimes\boldsymbol{\Sigma}^{-1/2}\right)\overline{\mathbf{M}}_{4,\mathbf{x}}\left(\boldsymbol{\Sigma}^{-1/2}\otimes\boldsymbol{\Sigma}^{-1/2}\right)\left(\boldsymbol{\Sigma}^{-1/2}\otimes\boldsymbol{\Sigma}^{-1/2}\right)\overline{\mathbf{M}}_{4,\mathbf{x}}\left(\boldsymbol{\Sigma}^{-1/2}\otimes\boldsymbol{\Sigma}^{-1/2}\right)\mathrm{vec}\left(\mathbf{I}_p\right).$$

We recall a fundamental property of the Kronecker product and the vectorization operator (see [35], 201): $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^\top\otimes\mathbf{A})\,\mathrm{vec}(\mathbf{B})$, where $\mathbf{A}\in\mathbb{R}^p\times\mathbb{R}^q$, $\mathbf{B}\in\mathbb{R}^q\times\mathbb{R}^r$ and $\mathbf{C}\in\mathbb{R}^r\times\mathbb{R}^s$. This property and the identity $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1}$ imply

$$\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right) = \left(\boldsymbol{\Sigma}^{-1/2}\otimes\boldsymbol{\Sigma}^{-1/2}\right)\mathrm{vec}\left(\mathbf{I}_p\right) \text{ and } \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right) = \mathrm{vec}^\top\left(\mathbf{I}_p\right)\left(\boldsymbol{\Sigma}^{-1/2}\otimes\boldsymbol{\Sigma}^{-1/2}\right).$$

The identity $(\mathbf{A}_1\otimes\mathbf{A}_2)(\mathbf{A}_3\otimes\mathbf{A}_4) = \mathbf{A}_1\mathbf{A}_3\otimes\mathbf{A}_2\mathbf{A}_4$, where $\mathbf{A}_1\in\mathbb{R}^p\times\mathbb{R}^q$, $\mathbf{A}_2\in\mathbb{R}^s\times\mathbb{R}^t$, $\mathbf{A}_3\in\mathbb{R}^q\times\mathbb{R}^v$, $\mathbf{A}_4\in\mathbb{R}^t\times\mathbb{R}^u$

(see [35], 194) implies

$$\left(\boldsymbol{\Sigma}^{-1/2}\otimes\boldsymbol{\Sigma}^{-1/2}\right)\left(\boldsymbol{\Sigma}^{-1/2}\otimes\boldsymbol{\Sigma}^{-1/2}\right)=\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right).$$

Hence we have

$$\beta_{2,K}\left(\mathbf{x}\right)=\mathrm{vec}^{\top}\left(\boldsymbol{\Sigma}^{-1}\right)\overline{\mathbf{M}}_{4,\mathbf{x}}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\overline{\mathbf{M}}_{4,\mathbf{x}}\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)$$

and this completes the first part of the proof. The second part of the proof is very similar. It suffices to recall that $\gamma_{2,K}\left(\mathbf{x}\right)=\mathrm{tr}\left(\mathbf{K}_{4,\mathbf{z}}^{2}\right)$ and that the fourth cumulant of $\mathbf{Qw}$ is $\left(\mathbf{Q}\otimes\mathbf{Q}\right)\mathbf{K}_{4,\mathbf{w}}\left(\mathbf{Q}^{\top}\otimes\mathbf{Q}^{\top}\right)$.

Proof of Theorem 3. Let $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ be the mean vector and the covariance matrix of $\mathbf{x}$. Also, let $\mathbf{y}$ be a random vector independent of $\mathbf{x}$ and identically distributed. Finally, let $\mathbf{z}=\boldsymbol{\Sigma}^{-1/2}\left(\mathbf{x}-\boldsymbol{\mu}\right)$ and $\mathbf{w}=\boldsymbol{\Sigma}^{-1/2}\left(\mathbf{y}-\boldsymbol{\mu}\right)$ be the standardization of $\mathbf{x}$ and $\mathbf{y}$, where $\boldsymbol{\Sigma}^{-1/2}$ is the positive definite symmetric square root of $\boldsymbol{\Sigma}$. We shall first apply the Cauchy-Schwarz inequality:

$$\mathrm{E}^{2}\left\{\left(\mathbf{z}^{\top}\mathbf{w}\right)^{3}\right\}=\mathrm{E}^{2}\left\{\left(\mathbf{z}^{\top}\mathbf{w}\right)^{2}\left(\mathbf{z}^{\top}\mathbf{w}\right)\right\}\leq\mathrm{E}\left\{\left(\mathbf{z}^{\top}\mathbf{w}\right)^{4}\right\}\mathrm{E}\left\{\left(\mathbf{z}^{\top}\mathbf{w}\right)^{2}\right\}.$$

According to [25] and [16] Mardia's skewness and Koziol's kurtosis of $\mathbf{x}$ might be represented as $\beta_{1,M}\left(\mathbf{x}\right)=\mathrm{E}\left\{\left(\mathbf{z}^{\top}\mathbf{w}\right)^{3}\right\}$ and $\beta_{2,K}\left(\mathbf{x}\right)=\mathrm{E}\left\{\left(\mathbf{z}^{\top}\mathbf{w}\right)^{4}\right\}$. Then the above inequality might be restated as

$$\beta_{1,M}^{2}\left(\mathbf{x}\right)\leq\beta_{2,K}\left(\mathbf{x}\right)\mathrm{E}\left\{\left(\mathbf{z}^{\top}\mathbf{w}\right)^{2}\right\}.$$

Let $\mathbf{u}=\left(U_{1},...,U_{p}\right)^{\top}$ be a random vector whose i-th component is $U_{i}=W_{i}Z_{i}$, so that

$$\mathrm{E}\left(U_{i}\right)=0,\ \mathrm{E}\left(U_{i}^{2}\right)=1,\ \mathrm{E}\left(U_{i}U_{j}\right)=0,\ \mathbf{z}^{\top}\mathbf{w}=U_{1}+...+U_{p}.$$

The second moment of $\mathbf{z}^{\top}\mathbf{w}$ is then

$$\mathrm{E}\left\{\left(\mathbf{z}^{\top}\mathbf{w}\right)^{2}\right\}=\sum_{i=1}^{p}\mathrm{E}\left(U_{i}^{2}\right)+2\sum_{i=1}^{p}\sum_{j\neq i}\mathrm{E}\left(U_{i}\right)\mathrm{E}\left(U_{j}\right)=p.$$

As a direct consequence, we can write $\beta_{1,M}^{2}\left(\mathbf{x}\right)\leq p\cdot\beta_{2,K}\left(\mathbf{x}\right)$ and complete the proof.

Proof of Theorem 4. By definition, Malkovich and Afifi's kurtosis of $\mathbf{x}$ is, see [23],

$$\gamma_{2,A}\left(\mathbf{x}\right)=\sup_{\mathbf{c}\in\mathbb{R}_{0}^{p}}\gamma_{2}^{2}\left(\mathbf{c}^{\top}\mathbf{x}\right),$$

where $\mathbb{R}_{0}^{p}$ is the set of p-dimensional and nonnull real vectors . It is invariant with respect to one-to-one affine transformations, so that

$$\gamma_{2,A}\left(\mathbf{x}\right)=\sup_{\mathbf{u}\in\mathbb{S}^{p-1}}\gamma_{2}^{2}\left(\mathbf{u}^{\top}\mathbf{z}\right),$$

where $\mathbb{S}^{p-1}$ is the set of p-dimensional real vectors with unit norm, $\mathbf{z}=\boldsymbol{\Sigma}^{-1/2}\left(\mathbf{x}-\boldsymbol{\mu}\right)$ is the standardization of $\mathbf{x}$, $\boldsymbol{\mu}$ is the mean of $\mathbf{x}$ and $\boldsymbol{\Sigma}^{-1/2}$ positive definite symmetric square root of the concentration matrix $\boldsymbol{\Sigma}^{-1}$. The fourth cumulant of $\mathbf{Ax}+\mathbf{b}$ is $\left(\mathbf{A}\otimes\mathbf{A}\right)\mathbf{K}_{4,\mathbf{x}}\left(\mathbf{A}^{\top}\otimes\mathbf{A}^{\top}\right)$, where $\mathbf{K}_{4,\mathbf{x}}$ is the fourth cumulant of $\mathbf{x}$, $\mathbf{A}$ is $k\times p$ real matrix and $\mathbf{b}$ is a k-dimensional real vector. The Malkovich and Afifi's kurtosis of $\mathbf{x}$ is then

$$\gamma_{2,A}\left(\mathbf{x}\right)=\sup_{\mathbf{u}\in\mathbb{S}^{p-1}}\left\{\left(\mathbf{u}\otimes\mathbf{u}\right)^{\top}\mathbf{K}_{4,\mathbf{z}}\left(\mathbf{u}\otimes\mathbf{u}\right)\right\}^{2}.$$

Koziol's excess kurtosis of $\mathbf{x}$ might be represented as the trace of its squared fourth standardized cumulant: $\gamma_{2,K}\left(\mathbf{x}\right)=$

$tr\left(\mathbf{K}_{4,\mathbf{z}}^{2}\right)$. Let λ be the dominant eigenvalue of $\mathbf{K}_{4,\mathbf{z}}$, that is its eigenvalue with maximum modulus:

$$\lambda^{2}=\sup_{\mathbf{v}\in\mathbb{S}^{p^{2}-1}}\left(\mathbf{v}^{\top}\mathbf{K}_{4,\mathbf{z}}\mathbf{v}\right)^{2},$$

where $\mathbb{S}^{p^{2}-1}$ is the set of p^{2}-dimensional real vectors with unit norm. Let $\mathbf{w}$ be a real, p-dimensional vector with unit norm, so that $\mathbf{w}\otimes\mathbf{w}$ is a p^{2}-dimensional real vector with unit norm: $\mathbf{w}\in\mathbb{S}^{p-1}$ and $\mathbf{w}\otimes\mathbf{w}\in\mathbb{S}^{p^{2}-1}$. Then the following inequality holds:

$$\left\{(\mathbf{w}\otimes\mathbf{w})^{\top}\mathbf{K}_{4,\mathbf{z}}(\mathbf{w}\otimes\mathbf{w})\right\}^{2}\leq\lambda^{2}.$$

As a direct consequence, the Malkovich and Afifi's kurtosis of $\mathbf{x}$ is never greater than λ^{2}: $\gamma_{2,A}(\mathbf{x})\leq\lambda^{2}$. The scalar λ^{2} is an eigenvector of the symmetric, positive semidefinite matrix $\mathbf{K}_{4,\mathbf{z}}^{2}$. Therefore λ^{2} never exceeds the trace of $\mathbf{K}_{4,\mathbf{z}}^{2}$: $\lambda^{2}\leq tr\left(\mathbf{K}_{4,\mathbf{z}}^{2}\right)$. By definition, Koziol's excess kurtosis of $\mathbf{x}$ is the sum of all squared fourth-order cumulants of $\mathbf{z}$, that is the trace of $\mathbf{K}_{4,\mathbf{z}}^{2}$: $\gamma_{2,K}(\mathbf{x})=tr\left(\mathbf{K}_{4,\mathbf{z}}^{2}\right)$. This identity and the above inequalities imply that the Malkovich and Afifi's kurtosis of $\mathbf{x}$ is never greater than its Koziol's excess kurtosis, $\gamma_{2,A}(\mathbf{x})\leq\gamma_{2,K}(\mathbf{x})$.

Proof of Theorem 5. Without loss of generality we can assume that $S_1,...,S_p$ are standardized random variables, so that $\mathrm{E}(\mathbf{x})=\mathbf{b}=\boldsymbol{\mu}$ and $\mathrm{cov}(\mathbf{x})=\mathbf{A}\mathbf{A}^{\top}=\boldsymbol{\Sigma}$. Also, let $\mathbf{y}=\mathbf{b}+\mathbf{A}\mathbf{z}$, where $\mathbf{z}=\left(Z_1,...,Z_p\right)^{\top}$ is a random vector independent of $\mathbf{s}$ and identically distributed. Finally, let $\widetilde{\mathbf{x}}=\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu})$ and $\widetilde{\mathbf{y}}=\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})$ be the standardization of $\mathbf{x}$ and $\mathbf{y}$, where $\boldsymbol{\Sigma}^{-1/2}$ is the positive definite symmetric square root of $\boldsymbol{\Sigma}$. It was shown in [16] that Koziol's kurtosis of $\mathbf{x}$ might be represented as

$$\beta_{2,K}(\mathbf{x})=\mathrm{E}\left\{\left(\widetilde{\mathbf{x}}^{\top}\widetilde{\mathbf{y}}\right)^{4}\right\}.$$

Simple matrix algebra, together with the above definitions, lead to the identities

$$\widetilde{\mathbf{x}}^{\top}\widetilde{\mathbf{y}}=(\mathbf{x}-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})=(\mathbf{A}\mathbf{s})^{\top}\left(\mathbf{A}\mathbf{A}^{\top}\right)^{-1}\mathbf{A}\mathbf{z}=\mathbf{s}^{\top}\mathbf{z}.$$

Let $\mathbf{w}=\left(W_1,...,W_p\right)^{\top}$ be a random vector whose i-th component is $W_i=S_iZ_i$, so that $\mathbf{s}^{\top}\mathbf{z}=W_1+...+W_p$. First we expand the fourth power of $\mathbf{s}^{\top}\mathbf{z}$ into the sum of products of four elements of $\mathbf{w}$:

$$\left(\mathbf{s}^{\top}\mathbf{z}\right)^{4}=\sum_{i=1}^{p}W_i^{4}+4\sum_{i=1}^{p}\sum_{j\neq i}W_i^{3}W_j+3\sum_{i=1}^{p}\sum_{j\neq i}W_i^{2}W_j^{2}.$$

Next, we shall take expectations and recall that the components of $\mathbf{w}$ are mutually independent:

$$\mathrm{E}\left\{\left(\mathbf{s}^{\top}\mathbf{z}\right)^{4}\right\}=\sum_{i=1}^{p}\mathrm{E}\left(W_i^{4}\right)+4\sum_{i=1}^{p}\sum_{j\neq i}\mathrm{E}\left(W_i^{3}\right)\mathrm{E}\left(W_j\right)+3\sum_{i=1}^{p}\sum_{j\neq i}\mathrm{E}\left(W_i^{2}\right)\mathrm{E}\left(W_j^{2}\right).\tag{a}$$

By assumption $\mathbf{s}$ and $\mathbf{z}$ are independent and identically distributed random vectors whose components are mutually independent and standardized random variables. Hence the mean, the variance and the kurtosis of W_i are

$$\mathrm{E}(W_i)=0,\ \mathrm{E}\left(W_i^{2}\right)=1,\ \mathrm{E}\left(W_i^{4}\right)=\beta_2^{2}(S_i).$$

The expectation in the right-hand side of (a) might then be simplified into

$$\mathrm{E}\left\{\left(\mathbf{s}^{\top}\mathbf{z}\right)^{4}\right\}=\sum_{i=1}^{p}\beta_2^{2}(S_i)+3p(p-1),$$

which equals Koziol's kurtosis, and this completes the first part of the proof.

By definition, Mardia's kurtosis of $\mathbf{x}$ is

$$\beta_{2,M}(\mathbf{x}) = \mathrm{E}\left[\left\{(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}^2\right].$$

Mardia's kurtosis is invariant with respect to affine, one-to-one transformations:

$$\beta_{2,M}(\mathbf{x}) = \mathrm{E}\left\{\left(\mathbf{s}^\top\mathbf{s}\right)^2\right\} = \sum_{i=1}^{p}\sum_{j=1}^{p}\mathrm{E}\left(S_i^2 S_j^2\right) = \sum_{i=1}^{p}\mathrm{E}\left(S_i^4\right) + \sum_{i=1}^{p}\sum_{j\neq i}^{p}\mathrm{E}\left(S_i^2 S_j^2\right).$$

By assumption, $S_1,\ldots,S_p$ are independent and standardized random variables whose fourth moments are $\beta_2(S_1),\ldots,$ $\beta_2\left(S_p\right)$, thus implying the identities

$$\beta_{2,M}(\mathbf{x}) = \sum_{i=1}^{p}\mathrm{E}\left(S_i^4\right) + \sum_{i=1}^{p}\sum_{j\neq i}^{p}\mathrm{E}\left(S_i^2\right)\mathrm{E}\left(S_j^2\right) = \sum_{i=1}^{p}\beta_2(S_i) + p(p-1).$$

We shall now prove the third part of the the theorem. The identity $\beta_{2,K}(\mathbf{x}) = \gamma_{2,K}(\mathbf{x}) + 6\beta_{2,M}(\mathbf{x}) - 6p - 3p^2$ ([16]) and the results in the previous part of the proof imply the following:

$$\gamma_{2,K}(\mathbf{x}) = \beta_{2,K}(\mathbf{x}) - 6\beta_{2,M}(\mathbf{x}) + 6p + 3p^2 = \sum_{i=1}^{p}\beta_2^2(S_i) + 3p(p-1) - 6\sum_{i=1}^{p}\beta_2(S_i) - 6p(p-1) + 6p + 3p^2$$
$$= \sum_{i=1}^{p}\beta_2^2(S_i) - 6\sum_{i=1}^{p}\beta_2(S_i) + 9p = \sum_{i=1}^{p}\left\{\beta_2(S_i) - 3\right\}^2.$$

Proof of Theorem 6. The mean, the variance and the fourth cumulant of an extended skew-normal random vector $\mathbf{x} \sim \mathcal{ESN}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \tau)$ are

$$\boldsymbol{\mu} = \boldsymbol{\xi} + \boldsymbol{\delta}\zeta_1(\tau) \text{ and } \boldsymbol{\Sigma} = \boldsymbol{\Omega} + \zeta_2(\tau)\boldsymbol{\delta}\boldsymbol{\delta}^\top,\ \mathbf{K}_{4,\mathbf{x}} = \zeta_4(\tau)\boldsymbol{\delta}\otimes\boldsymbol{\delta}^\top\otimes\boldsymbol{\delta}\otimes\boldsymbol{\delta}^\top,$$

where (see [1])

$$\zeta_i(\tau) = \frac{\partial^i \log\Phi(\tau)}{\partial^i\tau},\ \boldsymbol{\delta} = \frac{\boldsymbol{\Omega}\boldsymbol{\alpha}}{\sqrt{1+\boldsymbol{\alpha}^\top\boldsymbol{\Omega}\boldsymbol{\alpha}}}.$$

The multivariate extended skew-normal class is closed under affine transformations, so that the distribution of the standardized random vector $\mathbf{z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu})$ is also extended skew-normal distributed: $\mathbf{z} \sim \mathcal{ESN}(\boldsymbol{\xi}_z, \boldsymbol{\Omega}_z, \boldsymbol{\alpha}_z, \tau)$, with

$$\boldsymbol{\delta}_z = \frac{\boldsymbol{\Omega}_z\boldsymbol{\alpha}_z}{\sqrt{1+\boldsymbol{\alpha}_z^\top\boldsymbol{\Omega}_z\boldsymbol{\alpha}_z}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\delta}.$$

The Kronecker product is associative: $(\mathbf{A}\otimes\mathbf{B})\otimes\mathbf{C} = \mathbf{A}\otimes(\mathbf{B}\otimes\mathbf{C}) = \mathbf{A}\otimes\mathbf{B}\otimes\mathbf{C}$ ([35], page 194). Also, if $\mathbf{a}$ and $\mathbf{b}$ are two vectors, then $\mathbf{a}\mathbf{b}^\top$, $\mathbf{a}\otimes\mathbf{b}^\top$ and $\mathbf{b}^\top\otimes\mathbf{a}$ denote the same matrix ([35], page 199). Hence the fourth cumulant of $\mathbf{z}$ is

$$\mathbf{K}_{4,\mathbf{z}} = \zeta_4(\tau)\boldsymbol{\delta}_z\otimes\boldsymbol{\delta}_z^\top\otimes\boldsymbol{\delta}_z\otimes\boldsymbol{\delta}_z^\top = \zeta_4(\tau)\left(\boldsymbol{\delta}_z\otimes\boldsymbol{\delta}_z^\top\right)\otimes\left(\boldsymbol{\delta}_z\otimes\boldsymbol{\delta}_z^\top\right) = \zeta_4(\tau)\left(\boldsymbol{\delta}_z^\top\boldsymbol{\delta}_z\right)\otimes\left(\boldsymbol{\delta}_z\boldsymbol{\delta}_z^\top\right).$$

Mardia's excess kurtosis of $\mathbf{x}$ is just the trace of its fourth standardized cumulant $\mathbf{K}_{4,\mathbf{z}}$. Also, the trace of the Kronecker product of two square matrices is the product of their traces ([35], page 195). We can then write

$$\gamma_{2,M}(\mathbf{x}) = \mathrm{tr}\left\{\zeta_4(\tau)\left(\boldsymbol{\delta}_z\boldsymbol{\delta}_z^\top\right)\otimes\left(\boldsymbol{\delta}_z\boldsymbol{\delta}_z^\top\right)\right\} = \zeta_4(\tau)\left(\boldsymbol{\delta}_z^\top\boldsymbol{\delta}_z\right)^2.$$

Mardia's excess kurtosis of $\mathbf{x}$ is (see [1])

$$\gamma_{2,M}(\mathbf{x}) = \zeta_4(\tau)\left\{\frac{\boldsymbol{\delta}^\top\Omega^{-1}\boldsymbol{\delta}}{1+\zeta_2(\tau)\boldsymbol{\delta}^\top\Omega^{-1}\boldsymbol{\delta}}\right\}^2,$$

The squared fourth standardized cumulant of $\mathbf{x}$ might be represented as follows, by recalling that for matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ of appropriate size, the identity $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = \mathbf{AC}\otimes\mathbf{BD}$ holds true ([35], page 194):

$$\mathbf{K}_{4,\mathbf{z}}^2 = \zeta_4^2(\tau)\left\{\left(\boldsymbol{\delta}_z\boldsymbol{\delta}_z^\top\right)\otimes\left(\boldsymbol{\delta}_z\boldsymbol{\delta}_z^\top\right)\right\}\left\{\left(\boldsymbol{\delta}_z\boldsymbol{\delta}_z^\top\right)\otimes\left(\boldsymbol{\delta}_z\boldsymbol{\delta}_z^\top\right)\right\} = \zeta_4^2(\tau)\left\{\left(\boldsymbol{\delta}_z\boldsymbol{\delta}_z^\top\boldsymbol{\delta}_z\boldsymbol{\delta}_z^\top\right)\otimes\left(\boldsymbol{\delta}_z\boldsymbol{\delta}_z^\top\boldsymbol{\delta}_z\boldsymbol{\delta}_z^\top\right)\right\} = \zeta_4^2(\tau)\left(\boldsymbol{\delta}_z^\top\boldsymbol{\delta}_z\right)^2\left\{\left(\boldsymbol{\delta}_z\boldsymbol{\delta}_z^\top\right)\otimes\left(\boldsymbol{\delta}_z\boldsymbol{\delta}_z^\top\right)\right\}.$$

Koziol's kurtosis of $\mathbf{x}$ is just the trace of its squared fourth standardized cumulant. The above identities imply the following relations:

$$\gamma_{2,K}(\mathbf{x}) = tr\left(\mathbf{K}_{4,\mathbf{z}}^2\right) = \zeta_4^2(\tau)\left(\boldsymbol{\delta}_z^\top\boldsymbol{\delta}_z\right)^4 = \gamma_{2,M}^2(\mathbf{x}) = \zeta_4^2(\tau)\left\{\frac{\boldsymbol{\delta}^\top\Omega^{-1}\boldsymbol{\delta}}{1+\zeta_2(\tau)\boldsymbol{\delta}^\top\Omega^{-1}\boldsymbol{\delta}}\right\}^4.$$

References

[1] C. Adcock, M. Eling, N. Loperfido, Skewed distributions in finance and actuarial science: a review, European Journal of Finance 21 (2015) 1253–1281.

[2] B. Arnold, R. Beaver, Hidden truncation models, Sankhya, series A 62 (2000) 22–35.

[3] B. Arnold, R. Groeneveld, Skewness and kurtosis orderings: An introduction, Lecture Notes-Monograph Series 22 (2000) 17–24.

[4] K. Balanda, H. Macgillivray, Kurtosis: A critical review, Amer. Statist. 42 (1988) 111–119.

[5] D. Best, J. Rayner, A test for bivariate normality, Statistics and Probability Letters 6 (1988) 407–412.

[6] C. Franceschini, N. Loperfido, On some inequalities between measures of multivariate kurtosis, with application to financial returns, in: C. Perna, M. Sibillo (Eds.), Mathematical and Statistical Methods for Actuarial Sciences and Finance, Springer, 2012, pp. 211–218.

[7] C. Franceschini, N. Loperfido, An algorithm for finding projections with extreme kurtosis, in: P. M. Perna C., R.-G. A. (Eds.), Studies in Theoretical and Applied Statistics: SIS2016-48th Meeting of the Italian Statistical Society, Salerno 8-10 June 2016, Springer, 2018, pp. 159–170.

[8] N. Henze, On mardia's kurtosis test for multivariate normality, Communications in statistics-Theory and Methods 23:4 (1994) 1031–1045.

[9] N. Henze, Invariant tests for multivariate normality: a critical review, Statistical Papers 43 (2002) 467–506.

[10] R. J. Jones M.C., P. A., Skewness-invariant measures of kurtosis, The American Statistician 65 (2011) 89–95.

[11] J. Juan, F. Prieto, Using angles to identify concentrated multivariate outliers, Technometrics 43 (2002) 311–322.

[12] B. Klar, A treatment of multivariate skewness, kurtosis and related statistics, Journal of Multivariate Analysis 83 (2002) 141–165.

[13] T. Kollo, Multivariate skewness and kurtosis measures with an application in ICA, Journal of Multivariate Analysis 99 (2008) 2328–2338.

[14] T. Kollo, M. Srivastava, Estimation and testing of parameters in multivariate laplace distribution, Communications in Statistics-Theory and Methods 33 (2005) 2363–2687.

[15] J. Koziol, An alternative formulation of Neyman's smooth goodness of fit tests under composite alternatives, Metrika 34 (1987) 17–24.

[16] J. Koziol, A note on measures of multivariate kurtosis, Biometrical Journal 31 (1989) 619–624.

[17] J. Koziol, Probability plots for assessing multivariate normality, Journal of the Royal Statistical Society Series D 42 (1993) 161–173.

[18] C. Ley, T. Verdebout, Modern Directional Statistics, Chapman and Hall/CRC, New York, 2017.

[19] N. Loperfido, Generalized skew-normal distributions, in: Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality, CRC/Chapman & Hall, 2004, pp. 65–80.

[20] N. Loperfido, Linear transformations to symmetry, Journal of Multivariate Analysis 129 (2014) 186–192.

[21] N. Loperfido, A probability inequality related to mardia's kurtosis, in: Mathematical and Statistical Methods for Actuarial Sciences and Finance, Springer, 2014, pp. 129–132.

[22] N. Loperfido, Kurtosis-based projection pursuit for outlier detection in financial time series, The European Journal of Finance to appear (2019).

[23] J. Malkovich, A. Afifi, On tests for multivariate normality, Journal of American Statistical Association 68 (1973) 176–179.

[24] K. Mardia, Measures of multivariate skewness and kurtosis with applications, Biometrika 57 (1970) 519–530.

[25] K. Mardia, Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies, Sankhyā B 36 (1974) 115–128.

[26] K. Mardia, Assessment of multinormality and the robustness of hotelling's T^2 test., Appl. Statist. 24 (1975) 163–171.

[27] K. V. Mardia, J. T. Kent, Rao score tests for goodness of fit and independence, Biometrika 78 (1991) 355–363.

[28] J. Miettinen, S. Taskinen, K. Nordhausen, H. Oja, Fourth moments and independent component analysis, Statistical Science 30 (2015) 372–390.

[29] T. Mòri, V. Rohatgi, G. Székely, On multivariate skewness and kurtosis, Theory of Probability and its Applications 38 (1993) 547–551.

[30] G. Nason, Three-dimensional projection pursuit, Applied Statistics 44 (1995) 1430.

[31] J. Neyman, Smooth test for goodness of fit, Skand Aktuarioskr 20 (1937) 149–199.

[32] H. Ogasawara, Extensions of Pearson's inequality between skewness and kurtosis to multivariate cases, Statistics and Probability Letters 30 (2017) 12–16.

[33] H. Ogasawara, Some improvements on Markov's theorem with extensions, The American Statistician (2018).
[34] H. Ogasawara, The multiple Cantelli inequalities, Statistical Methods and Applications 28 (2019) 495–506.
[35] C. Rao, M. Rao, Matrix algebra and its applications to statistics and econometrics, World Scientific Co. Pte. Ltd, Singapore, 1998.
[36] C. Urzúa, Omnibus tests for multivariate normality based on a class of maximum entropy distributions, Advances in Econometrics 12 (1997) 341–358.
[37] H. Yanagihara, A family of estimators for multivariate kurtosis in a nonnormal linear regression model, Journal of Multivariate Analysis 98 (2007) 1–29.
[38] H. Yanagihara, T. Tonda, C. Matsumoto, The effects of nonnormality on asymptotic distributions of some likelihood ratio criteria for testing covariance structures under normal assumption, Journal of Multivariate Analysis 96 (2005) 237–264.
[39] K. Zografos, On Mardia's and Song's measures of kurtosis in elliptical distributions, Journal of Multivariate Analysis 99 (2008) 858–879.