Journal of Multivariate Analysis 123 (2014) 386–394

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Note(s)

A note on the fourth cumulant of a finite mixture distribution

Nicola Loperfido

Dipartimento di Economia, Società e Politica, Università degli Studi di Urbino "Carlo Bo", via Saffi 42, 61029 Urbino, Italy

ARTICLE INFO

Article history:
Received 18 May 2012
Available online 27 September 2013

AMS subject classifications:
62H15
15A69

Keywords:
Cluster analysis
Covariance testing
Finite mixture model
Generalized skew-normal
Kurtosis
Tensor rank

ABSTRACT

The paper shows that the fourth cumulant of a finite mixture distribution might be decomposed into the mean of the components' fourth cumulants and the fourth cumulant of the components' means, when the mixture's components have the same second and third cumulants. Statistical applications include robustness properties of likelihood-based testing procedures and kurtosis-based projection methods. Practical relevance of theoretical results in the paper are illustrated with two well-known data sets.

1. Introduction

Let $x = (X_1, \ldots, X_d)^T$ be a d-dimensional random vector with mean $\mu = (\mu_1, \ldots, \mu_d)^T$, covariance matrix $\Sigma = \left\{\sigma_{ij}\right\}$ and finite fourth-order moments: $E\left(\left|X_i X_j X_h X_k\right|\right) < +\infty$, for $i, j, h, k = 1, \ldots, d$. The fourth cumulant $\mathcal{K}_4 = \left\{\kappa_{ijhk}\right\}$ of x is the d-dimensional, symmetric tensor of order 4 whose elements are the fourth-order derivatives of the cumulant generating function of x: $\kappa_{ijhk} = \log E\left[\exp\left(\iota t^T x\right)\right] / \partial t_i \partial t_j \partial t_h \partial t_k$, where $\iota = \sqrt{-1}$ and $t^T = (t_1, \ldots, t_d)$. An equivalent representation of κ_{ijhk} is

$$E\left[(X_i - \mu_i)\left(X_j - \mu_j\right)(X_h - \mu_h)(X_k - \mu_k)\right] - \sigma_{ij}\sigma_{hk} - \sigma_{ih}\sigma_{jk} - \sigma_{ik}\sigma_{jh}.$$

The elements κ_{ijhk} might be arranged into the $d^2 \times d^2$ block matrix $\kappa_4(x) = \left\{M_{pq}\right\}$, where $M_{pq} = \log E\left[\exp\left(\iota t^T x\right)\right] / \partial t_p \partial t_q \partial t \partial t^T$ for $p, q = 1, \ldots, d$. The matrix $\kappa_4(x)$ is the unfolded version of $\mathcal{K}_4$ (see, for example [41]) and can be represented as

$$E\left(y \otimes y^T \otimes y \otimes y^T\right) - \left(I_{d^2} + K_{d,d}\right)(\Sigma \otimes \Sigma) - vec(\Sigma)\, vec^T(\Sigma),$$

where $y = x - \mu$, $\otimes$ denotes the Kronecker product, $vec(\Sigma)$ is the vectorization of Σ and $K_{d,d}$ is the $d^2 \times d^2$ commutation matrix [25]. With a slight abuse of notation, we shall refer to the matrix $\kappa_4(x)$ as to the fourth cumulant of x. Loperfido [23] examines some spectral properties of $\kappa_4(x)$.

In the general case, the number of distinct elements in $\kappa_4(x)$ increases very quickly with the dimension of x. If x is d- dimensional, $\kappa_4(x)$ might contain up to $d(d+1)(d+2)(d+3)/24$ distinct elements (see, for example, [16]). This suggests that statistical applications of the fourth order cumulant might greatly benefit from its parsimonious modelling, especially when they deal with multivariate kurtosis, as measured by functions of the fourth standardized cumulant. As a first example, Mardia [27], Malkovich and Afifi [26], Henze [13] investigate different measures of multivariate kurtosis for

E-mail addresses: nicola.loperfido@uniurb.it, nicola.loperfido@econ.uniurb.it.

http://dx.doi.org/10.1016/j.jmva.2013.09.007

testing the hypothesis of multivariate normality. As a second example, kurtosis is used in Independent Component Analysis to recover the independent components themselves, when they are assumed to be leptokurtic (see, for example, [15]). As a third example, likelihood-based procedures for testing hypotheses on covariance matrices might be very sensitive to the kurtosis of the sampled distribution, when the latter is erroneously assumed to be multivariate normal [14,28,43,45].

Finite mixtures of multivariate distributions have often been used to achieve parsimonious modelling. Let $F_1, \ldots, F_g$ be d-dimensional cumulative distribution functions, and let $\pi_1, \ldots, \pi_g$ be nonnegative real numbers which add up to one. The weighted average $F = \pi_1 F_1 + \cdots \pi_g F_g$ is said to be a finite mixture distribution (or model), whose i-th component and i-th weight are F_i and π_i, respectively. Let μ_i and Ω_i be the mean and the variance of the i-th mixture's component F_i, respectively, for $i = 1, \ldots, g$. It is well-known that the mixture's mean μ, i.e. its first cumulant, is the mean of the components' means. It is also well-known that the mixture's covariance Σ, i.e. its second cumulant, is the mean of the components' covariances plus the covariance of the components' means. The above representations of the mean and the variance are appealing in that they are easily expressed both in words and in matrix notation. Higher-order cumulants of finite mixture distributions might be obtained via the law of total cumulance [5]. Unfortunately, it leads to results which are neither easily interpretable nor admit simple representations in the matrix form, thus limiting their use in statistical modelling.

In recent years, finite mixtures of elliptical distributions with proportional scatter matrices have been used to explore the statistical properties of kurtosis-based multivariate procedures. Projections which either maximize or minimize kurtosis have been used both in cluster analysis and in outlier detection [36–39]. In tensor terminology, they might be regarded as the best rank-one approximations to the fourth standardized cumulant. Tyler et al. [44] and Peña et al. [40] used a kurtosis matrix independently introduced by Cardoso [7] and Mori et al. [34] to uncover several features of multivariate data. The same matrix might be regarded as the sum of the d diagonal blocks of the fourth cumulant, which are $d \times d$ symmetric matrices [19]. Both approaches have good statistical properties when sampling from a finite mixture of elliptical distributions with proportional scatter matrices. At present time, however, no one investigated their robustness to violation of the underlying assumptions, not even in the special case of a mixture of two multivariate normal distributions with the same variance.

This paper addresses the above mentioned problems within the framework of finite mixtures whose components have identical second and third cumulants. They include several classes of well-known finite mixture models, most notably finite mixtures of normal distributions with equal covariance matrices. McLachlan and Peel [32] report many applications of such models and remark that often the component-covariances are restricted to being the same ([32], page 83). Their widespread use is partly due to the little inferential problems they pose, compared to other finite normal mixtures [33]. An additional advantage, from this paper's perspective, is their very parsimonious modelling of the fourth cumulant. However, the class of finite mixtures with equal second and third cumulants is much wider, since it also includes location mixtures. Skewness-based projection pursuit might be helpful in detecting clusters, when the sampled distribution is a location mixture of two multivariate, symmetric distributions [24]. Principal points of location mixtures of spherically symmetric distributions have nice theoretical properties [20,30,31]. Section 3 in this paper discusses location mixtures of multivariate skew-normal distributions.

We shall show that, when mixture's components have identical second and third cumulants, the fourth cumulant of the mixture equals the mean of the components' fourth cumulants, plus the fourth cumulant of the components' means. Statistical applications deal with robustness of multivariate statistical procedures. First we shall assess the robustness of MANOVA statistics when the data are drawn from a normal mixture with two homoscedastic components. Then we shall use mixtures with two skew-normal components to assess the robustness of the kurtosis-based procedures proposed by Peña and Prieto [36–39], Tyler et al. [44] and Peña et al. [40]. Other theorems in the paper, regarding fourth multivariate cumulants and moments, are instrumental in proving the above results as well as being interesting in their own right.

The rest of the paper is organized as follows. Section 2 contains the main results. Section 3 discusses some statistical applications. Section 4 illustrates their practical relevance with two well-known data sets. All proofs are deferred to the Appendix.

2. Main results

The following theorem represents the fourth moment $\mu_4(x-c)$ of the difference $x-c$, where x is a d-dimensional random vector and c is a real vector of the same dimension, as a function of the first four moments of x: $\mu = \mu_1 = E(x)$, $\mu_2 = E\left(xx^T\right)$, $\mu_3 = E\left(x \otimes x^T \otimes x\right)$ and $\mu_4 = E\left(x \otimes x^T \otimes x \otimes x^T\right)$.

Theorem 1. *Let μ_1, μ_2, μ_3, μ_4 be the first, second, third and fourth moment of the d-dimensional random vector x. Then the fourth moment of $x-c$, where c is a d-dimensional real vector, is*

$$\begin{aligned}&\mu_4 - \mu_3^T \otimes c - \mu_3 \otimes c^T - c^T \otimes \mu_3 - c \otimes \mu_3^T + \mu_2 \otimes cc^T + vec\left(\mu_2\right) \otimes c^T \otimes c^T \\ &\quad + K_{d,d}\left(cc^T \otimes \mu_2\right) + K_{d,d}\left(\mu_2 \otimes cc^T\right) + c \otimes c \otimes vec^T\left(\mu_2\right) + cc^T \otimes \mu_2 \\ &\quad - \mu_1 c^T \otimes cc^T - c\mu_1^T \otimes cc^T - cc^T \otimes \mu_1 c^T - cc^T \otimes c\mu_1^T + cc^T \otimes cc^T.\end{aligned}$$

As a direct consequence, the fourth central moment of a random vector might be represented via the first noncentral moments of the vector itself. More precisely, let μ_1, μ_2, μ_3, μ_4 be the first, second, third and fourth moment of the d-

dimensional random vector x. Then the fourth central moment of x is

$$\mu_4 - \mu_3^T \otimes \mu_1 - \mu_3 \otimes \mu_1^T - \mu_1^T \otimes \mu_3 - \mu_1 \otimes \mu_3^T + \mu_2 \otimes \mu_1 \mu_1^T + vec\,(\mu_2) \otimes \mu_1^T \otimes \mu_1^T + K_{d,d}\left(\mu_1 \mu_1^T \otimes \mu_2\right) + K_{d,d}\left(\mu_2 \otimes \mu_1 \mu_1^T\right) + \mu_1 \otimes \mu_1 \otimes vec^T\,(\mu_2) + \mu_1 \mu_1^T \otimes \mu_2 - 3\mu_1 \mu_1^T \otimes \mu_1 \mu_1^T.$$

Cumulants of the linear transformation $y = Ax$ admit simple representations in terms of matrix operations. For example, the first cumulant $E(y) = AE(x)$ is evaluated via matrix multiplication only. The second cumulant $V(y) = AV(x)A^T$ is evaluated using both the matrix multiplication and transposition. The fourth cumulant $\kappa_4\,(y)$ is evaluated using the matrix multiplication, transposition and the tensor product, as shown in Theorem 2.

Theorem 2. *Let $\kappa_4\,(x)$ be the fourth cumulant of a d-dimensional random vector x and let A be an $h \times d$ real matrix. Then the fourth cumulant of Ax is $\kappa_4\,(Ax) = (A \otimes A)\,\kappa_4\,(x)\left(A^T \otimes A^T\right)$.*

It is well-known that the second cumulant of a finite mixture might be decomposed into the mean of the components' covariances and the covariance of the components' means. The following theorem shows that a similar result holds for the fourth cumulant of a mixture, when the mixture's components have the same second and third cumulants.

Theorem 3. *Let the random vector x be the mixture of the distribution functions $F_1, \ldots, F_g$, with weights $\pi_1, \ldots, \pi_g$. Also, let μ_i, Ω, Γ, and $\kappa_{4,i}$ be the first, second, third and fourth cumulants of F_i. Finally, let Ξ and m be a random matrix and a random vector satisfying $\Pr\left(\Xi = \kappa_{4,i}\right) = \Pr\,(m = \mu_i) = \pi_i$, for $i = 1, \ldots, g$. Then the fourth cumulant of x is the sum of the expected value of Ξ and the fourth cumulant of m: $\kappa_4\,(x) = E\,(\Xi) + \kappa_4\,(m)$.*

As mentioned in the Introduction, statistical applications of the fourth cumulant would greatly benefit from its parsimonious modelling. The problem might be addressed by imposing tensor rank restrictions, as it is often done in multilinear algebra (see, for example, the review paper by Kolda and Bader [18]). The fourth cumulant $\kappa_4\,(x)$ of a d-dimensional random vector has (symmetric) tensor rank r if r is the smallest integer for which the following decomposition holds:

$$\kappa_4\,(x) = \sum_{i=1}^{r} c_i v_i \otimes v_i^T \otimes v_i \otimes v_i^T,$$

where $v_1, \ldots, v_r$ are nonnull d-dimensional real vectors and $c_1, \ldots, c_r$ and are nonzero real values: $v_i \in \mathbb{R}_0^d$, $c_i \in \mathbb{R}_0$, $i = 1, \ldots, r$ [8]. As remarked by Kilmer and Martin [17]: "There is no known closed-form solution to determine the rank r of a tensor a priori. Rank determination of a tensor is a widely-studied problem". The following theorem shows that the problem has a very simple solution for the fourth cumulant of the mixture of two multivariate normal distributions with the same variance.

Theorem 4. *Let the distribution of the random vector x be the mixture, with weights π_1 and π_2, of $N_d\,(\mu_1, \Omega)$ and $N_d\,(\mu_2, \Omega)$. Then its fourth cumulant is*

$$\pi_1 \pi_2 \left(\pi_1^2 - 4\pi_1 \pi_2 + \pi_2^2\right)(\mu_1 - \mu_2) \otimes (\mu_1 - \mu_2)^T \otimes (\mu_1 - \mu_2) \otimes (\mu_1 - \mu_2)^T.$$

3. Statistical applications

This section examines two statistical applications of previous section's results. We shall first use some well-known kurtosis measures for assessing the effect of nonnormality on some likelihood-based testing procedures. Let $y_1, \ldots, y_n$ be a random sample from $N_d\,(\mu, \Sigma)$. Also, let A and $\bar{y}$ be the SSP matrix and the mean vector, respectively:

$$A = (n-1)\,S = \sum_{i=1}^{n} (y_i - \bar{y})\,(y_i - \bar{y})^T, \quad \bar{y} = \sum_{i=1}^{n} y_i.$$

A relevant inferential problem is testing the null hypothesis $H_{0,1} : \Sigma = \Sigma_0$, where Σ_0 is a known $d \times d$ positive definite symmetric matrix, versus the local alternative $H_{1,1} : \Sigma = \Sigma_0 + \left(1/\sqrt{n}\right) B$, where B is a $d \times d$ positive semi-definite symmetric matrix. The corresponding likelihood-ratio statistic is

$$L_1 = \left(\frac{e}{n-1}\right)^{d(n-1)/2} \left|\Sigma_0^{-1} A\right|^{(n-1)/2} \exp\left[-\frac{1}{2} tr\left(\Sigma_0^{-1} A\right)\right].$$

Another relevant inferential problem is testing the null hypothesis $H_{0,2} : \Sigma = \sigma^2 I_d$ versus the local alternative $H_{1,2} : \Sigma = \sigma^2\left[I_d + \left(1/\sqrt{n}\right) B\right]$, where σ^2 is a positive constant. The corresponding likelihood-ratio statistic is

$$L_2 = \left\{\frac{|S|}{[tr\,(S)\,/d]^d}\right\}^{(n-1)/2}.$$

When the sampled distribution is normal, the null and nonnull asymptotic distributions of $T_i = -2\log L_i$ $(i = 1, 2)$ are central and noncentral chi-squared, respectively. Yanagihara et al. [45] investigated the asymptotic distributions of T_1 and T_2 under the more general assumption that the sampled distribution had finite fourth-order moments. Up to an error of order $o(1)$, they represented the means and the variances of T_1 and T_2 as simple functions of the quantities

$$\kappa_4^{(1)} = \sum_{i,j}^{d} \kappa_{iijj}, \qquad \kappa_{4,4}^{(1)} = \sum_{i,j,h,k}^{d} \kappa_{ijhk}^2, \qquad \kappa_{4,4}^{(2)} = \sum_{i,j,h,k}^{d} \kappa_{iihk}\kappa_{jjhk}.$$

As shown in the following theorem, these quantities are closely related to each other and have a simple analytical form, when the sampled distribution is a mixture of two normals with identical covariance matrices.

Theorem 5. *Let the distribution of the random vector x be the mixture, with weights π_1 and π_2, of $N_d(\mu_1, \Omega)$ and $N_d(\mu_2, \Omega)$, where Ω is a symmetric and positive definite $d \times d$ matrix. Then*

$$\left[\kappa_4^{(1)}\right]^2 = \kappa_{4,4}^{(1)} = \kappa_{4,4}^{(2)} = \frac{(\pi_1\pi_2)^2 (1 - 6\pi_1\pi_2)^2}{\left(1 + \pi_1\pi_2\delta^T\Omega^{-1}\delta\right)^4}.$$

As mentioned in the Introduction, kurtosis-based procedures have shown to possess good statistical properties when sampling from finite mixtures of normal distributions with equal variances. However, normal mixture models are not appropriate when nonnormality of the sampled distribution depends on the mixing process as well as on the components' skewness. When this happens, data might be adequately modelled by finite mixtures of skew-normal distributions [21,12,22,4]. We shall investigate the robustness of the aforementioned kurtosis-based procedures under this more general setting using a location mixture of two skew-normal distributions.

The distribution of a random vector x is multivariate skew-normal with location parameter ξ, scale parameter Ω and shape parameter α, that is $x \sim SN_d(\xi, \Omega, \alpha)$, if its pdf is $2\phi_d(x - \xi; \Omega)\,\Phi\left[\alpha^T(x - \xi)\right]$, where $\Phi(\cdot)$ is the cdf of a standardized normal variable and $\phi_d(x - \xi; \Omega)$ is the pdf of a d-dimensional normal distribution with mean ξ and variance Ω [3]. The fourth cumulant of x is a rank-one tensor: $\kappa_4(x) = h\eta \otimes \eta^T \otimes \eta \otimes \eta^T$, where $\eta = \Omega\alpha/\sqrt{1 + \alpha^T\Omega\alpha}$ and $h = \left(4/\pi^2\right)(2\pi - 6)$ [2]. Adcock et al. [1] review the main properties of the multivariate skew-normal distribution.

We can now use Theorem 3 and an argument very similar to the proof of Theorem 4 to obtain the fourth cumulant of the mixture, with weights π_1 and π_2, of $SN_d(\xi_1, \Omega, \alpha)$ and $SN_d(\xi_2, \Omega, \alpha)$, that is $h\eta \otimes \eta^T \otimes \eta \otimes \eta^T + k\lambda \otimes \lambda^T \otimes \lambda \otimes \lambda^T$, where $k = \pi_1\pi_2\left(\pi_1^2 - 4\pi_1\pi_2 + \pi_2^2\right)$ and $\lambda = \xi_1 - \xi_2$. In the general case, the tensor rank of the fourth cumulant is two. However, when $\eta = \lambda$ and $h = k$ (which happens when π_1 approximately equals either 0.5762 or 0.4238) the fourth cumulant is a null matrix, meaning that the two sources of nonnormality have opposite effects on the fourth cumulant. As a direct consequence, kurtosis-based procedures are unable to ascertain neither the presence nor the nature of nonnormality.

It was Pearson [35] who first posed the problem of distinguishing between inherently skewed distributions and finite mixtures. Since then, it has been addressed both in the statistical literature (see, for example, [32]) and in the medical one (see, for example, [42]). The problem also arises in the multivariate case: third-order cumulants of skew-normal distributions [10] and homoscedastic normal mixtures with two components [24] share the same tensor rank-one structure. We have shown that similar results hold for multivariate kurtosis, as measured by the fourth cumulant.

4. Numerical examples

This section illustrates some statistical implications of the theoretical results in the paper with well-known data sets. By Theorem 4, the fourth-order cumulants κ_{ijhk} of a mixture of two normal components with the same variances have the form $\gamma\lambda_i\lambda_j\lambda_h\lambda_k$ $(i, j, h, k = 1, \ldots, d)$ for some real value γ and some vector $\lambda = (\lambda_1, \ldots, \lambda_d)^T$. As seen in the previous section, the same holds for a multivariate skew-normal distribution. By standard consistency arguments, a $d^2 \times d^2$ fourth sample cumulant K might be well approximated by a matrix of the form $cv \otimes v^T \otimes v \otimes v^T$ $(c \in \mathbb{R}, v \in \mathbb{R}^d)$, when the sampled distribution belongs to one of the above families, and the sample size is large enough.

First, we shall find the constant c and the vector v which minimize the euclidean norm of the difference $K - cv \otimes v^T \otimes v \otimes v^T$. In tensor terminology, this means looking for the best symmetric rank-one approximation to the tensor K [8], which can be found via the Symmetric Higher Order Power Method [9]. Finally, we shall assess the accuracy of the approximation with the ratio $q = \left\|K - cv \otimes v^T \otimes v \otimes v^T\right\| / \|K\|$, where $\|A\|$ denotes the euclidean norm of the real matrix A, that is the square root of the sum of the squared elements of A (see, for example, [10]): the lower the value of q, the better the approximation. Intuitively, we can say that the rank-one approximations explain the $(1 - q) \cdot 100$ percent of the structure of K.

We shall first examine the rank-one approximation to the fourth cumulant of the Blue Crabs data set. It consists of 5 morphological measurements (frontal lobe size, rear width, carapace length, carapace width, and body depth), in millimetres, of 50 male crabs and 50 female crabs belonging to the blue species of Leptograpsus variegatus collected at Fremantle, Western Australia [6]. The gender-related clustering structure is apparent both from subject-matter considerations and from graphical analysis, and it has been modelled by a mixture of two multivariate normal distributions with the same covariance matrix (see, for example, [32], pp. 90–92). For this data set, we have $c = -10.058$, $v =$

$(0.2603, 0.1934, 0.6043, 0.6732, 0.2764)^T$ and $q = 0.0038$. The rank-one approximation is very accurate (it explains more than the 99% of the fourth cumulant's structure), consistently with Theorem 4 and the modelling assumptions.

We shall now examine the rank-one approximation to the fourth cumulant of a subset of the Australian Athletes data set. It consists of 3 body composition measurements (body mass index, body fat index and lean body mass index) of 23 elite female netball players, collected by the Australian Institute of Sport (AIS). Clearly, they are not a random sample from the population of adult Australian women: their measurements were collected just because they were elite athletes competing at an international level. Azzalini and Dalla Valle [3], among others, modelled the nonrandomness of the AIS data by skew-normal distributions. For this data set, we have $c = -54.6217 \cdot 10^{-4}$, $v = (0.0983, 0.9885, 0.1151)^T$ and $q = 0.0158$. Again, the rank-one approximation is very accurate (it explains more than the 98% of the fourth cumulant's structure), consistently with the properties of the multivariate skew-normal distribution.

These numerical examples lead to the following remarks. In the first place, theoretical results in this paper might be able to model some fourth cumulant's structures in a parsimonious way. In the second place, the same fourth cumulant structure might be due to either heterogeneity or nonrandomness.

Acknowledgments

The author wish to thank an anonymous referee for her/his insightful comments and suggestions that helped to improve the paper.

Appendix

Proof of Theorem 1. By definition, the fourth moment of $x - c$ is

$$\mu_4(x-c) = E\left[(x-c)\otimes(x-c)^T\otimes(x-c)\otimes(x-c)^T\right].$$

First consider the Kronecker products $(x-c)\otimes(x-c)^T\otimes(x-c)\otimes(x-c)^T$:

$$\left(x\otimes x^T - x\otimes c^T - c\otimes x^T + c\otimes c^T\right)\otimes\left(x\otimes x^T - x\otimes c^T - c\otimes x^T + c\otimes c^T\right).$$

Parentheses elimination leads to

$$\begin{aligned}
&x\otimes x^T\otimes x\otimes x^T - x\otimes x^T\otimes c\otimes x^T - x\otimes x^T\otimes x\otimes c^T + x\otimes x^T\otimes c\otimes c^T\\
&\quad - x\otimes c^T\otimes x\otimes x^T + x\otimes c^T\otimes x\otimes c^T + x\otimes c^T\otimes c\otimes x^T - x\otimes c^T\otimes c\otimes c^T\\
&\quad - c\otimes x^T\otimes x\otimes x^T + c\otimes x^T\otimes x\otimes c^T + c\otimes x^T\otimes c\otimes x^T - c\otimes x^T\otimes c\otimes c^T\\
&\quad + c\otimes c^T\otimes x\otimes x^T - c\otimes c^T\otimes x\otimes c^T - c\otimes c^T\otimes c\otimes x^T + c\otimes c^T\otimes c\otimes c^T.
\end{aligned}$$

Since $a\otimes b^T = b^T\otimes a$ for any two vectors a and b, the above matrix equals

$$\begin{aligned}
&x\otimes x^T\otimes x\otimes x^T - x^T\otimes x\otimes x^T\otimes c - x\otimes x^T\otimes x\otimes c^T + x\otimes x^T\otimes c\otimes c^T\\
&\quad - c^T\otimes x\otimes x^T\otimes x + x\otimes x\otimes c^T\otimes c^T + x\otimes c\otimes c^T\otimes x^T - x\otimes c^T\otimes c\otimes c^T\\
&\quad - c\otimes x^T\otimes x\otimes x^T + c\otimes x\otimes x^T\otimes c^T + c\otimes c\otimes x^T\otimes x^T - x^T\otimes c\otimes c\otimes c^T\\
&\quad + c\otimes c^T\otimes x\otimes x^T - c^T\otimes c\otimes c^T\otimes x - c\otimes c^T\otimes c\otimes x^T + c\otimes c^T\otimes c\otimes c^T.
\end{aligned}$$

We shall now consider a more convenient representation of the matrix $x\otimes c\otimes c^T\otimes x^T$. The identities, $x\otimes c\otimes c^T\otimes x^T = (x\otimes c)\otimes\left(c^T\otimes x^T\right) = vec\left(cx^T\right)(c\otimes x)^T$, follow from standard properties of the Kronecker product and the vectorization operator. The commutation matrix $K_{d,d}$ is a $d^2\times d^2$ symmetric and orthogonal matrix [25], so that $x\otimes c\otimes c^T\otimes x^T = K_{d,d}\left[K_{d,d}vec\left(cx^T\right)(c\otimes x)^T\right]$. The $pq\times pq$ commutation matrix $K_{p,q}$ satisfies the matrix equation $vec\left(M^T\right) = K_{p,q}vec(M)$ for any $p\times q$ matrix M [25]. Hence we have $x\otimes c\otimes c^T\otimes x^T = K_{d,d}vec\left(xc^T\right)(c\otimes x)^T = K_{d,d}\left[(c\otimes x)\otimes\left(c^T\otimes x^T\right)\right]$. We shall use again the fact that $a\otimes b^T = b^T\otimes a$ for any two vectors a and b and the associative property of the Kronecker product to obtain $x\otimes c\otimes c^T\otimes x^T = K_{d,d}\left(c\otimes c^T\otimes x\otimes x^T\right)$. In a similar way, we can prove that $c\otimes x\otimes x^T\otimes c^T = K_{d,d}\left(x\otimes x^T\otimes c\otimes c^T\right)$. The expectation of $(x-c)\otimes(x-c)^T\otimes(x-c)\otimes(x-c)^T$ is conveniently expressed in terms of the first four moments of x after recalling that $\mu_1 = E(x)$, $\mu_2 = E\left(xx^T\right) = E\left(x\otimes x^T\right) = E\left(x^T\otimes x\right)$, $\mu_3 = E\left(x\otimes x^T\otimes x\right)$, $vec(\mu_2) = E(x\otimes x)$, $\mu_4 = E\left(x\otimes x^T\otimes x\otimes x^T\right)$, $\mu_3^T = E\left(x^T\otimes x\otimes x^T\right)$:

$$\begin{aligned}
\mu_4(x-c) = {}& \mu_4 - \mu_3^T\otimes c - \mu_3\otimes c^T + \mu_2\otimes c\otimes c^T - c^T\otimes\mu_3\\
&+ vec(\mu_2)\otimes c^T\otimes c^T + K_{d,d}\left(cc^T\otimes\mu_2\right) - \mu_1\otimes c^T\otimes c\otimes c^T\\
&- c\otimes\mu_3^T + K_{d,d}\left(\mu_2\otimes cc^T\right) + c\otimes c\otimes vec^T(\mu_2) - \mu_1^T\otimes c\otimes c\otimes c^T\\
&+ cc^T\otimes\mu_2 - cc^T\otimes\mu_1c^T - cc^T\otimes c\mu_1^T + cc^T\otimes cc^T.
\end{aligned}$$

Theorem 1 follows after rearranging the summands in the above matrix.

Proof of Theorem 2. The fourth cumulant of a d-dimensional random vector x with variance Σ is $\kappa_4(x) = \overline{\mu}_4(x) - \left(I_{d^2} + K_{d,d}\right)(\Sigma \otimes \Sigma) - vec(\Sigma)\, vec^T(\Sigma)$, where $\overline{\mu}_4(x)$ is the fourth central moment of x. As a direct consequence, the fourth cumulant $\kappa_4(Ax)$ of Ax is

$$\overline{\mu}_4(Ax) - \left(I_{d^2} + K_{h,h}\right)\left[\left(A\Sigma A^T\right) \otimes \left(A\Sigma A^T\right)\right] - vec\left(A\Sigma A^T\right) vec^T\left(A\Sigma A^T\right),$$

since $var(Ax) = A\Sigma A^T$. Franceschini and Loperfido [11] showed that $\overline{\mu}_4(Ax) = (A \otimes A)\,\overline{\mu}_4(x)\left(A^T \otimes A^T\right)$. The identity $vec(M_1M_2M_3) = \left(M_3^T \otimes M_1\right) vec(M_2)$ holds for matrices $M_1 \in \mathbb{R}^p \times \mathbb{R}^q$, $M_2 \in \mathbb{R}^q \times \mathbb{R}^r$, $M_3 \in \mathbb{R}^r \times \mathbb{R}^s$ [25] and leads to $vec\left(A\Sigma A^T\right) = (A \otimes A)\, vec(\Sigma)$. The identity $\left(A\Sigma A^T\right) \otimes \left(A\Sigma A^T\right) = (A \otimes A)(\Sigma \otimes \Sigma)\left(A^T \otimes A^T\right)$ follows from repeated application of a fundamental property of the Kronecker product [29]: if matrices M_1, M_2, M_3 and M_4 are of appropriate size, then $(M_1 \otimes M_2)(M_3 \otimes M_4) = M_1M_3 \otimes M_2M_4$. The commutation matrix $K_{d,d}$ is at the same time symmetric and orthogonal [25], so that $K_{d,d}K_{d,d} = I_{d^2}$ and

$$K_{h,h}(A \otimes A) = \left[K_{h,h}(A \otimes A) K_{d,d}\right] K_{d,d} = (A \otimes A) K_{d,d}.$$

The right-hand side of the above equation follows from another fundamental property of the Kronecker product [25]: $M_1 \otimes M_2 = K_{p,r}(M_2 \otimes M_1) K_{s,q}$, when $M_1 \in \mathbb{R}^p \times \mathbb{R}^q$ and $M_2 \in \mathbb{R}^r \times \mathbb{R}^s$.

The above identities lead to the following representation of the fourth cumulant of Ax:

$$\begin{aligned}\kappa_4(Ax) &= (A \otimes A)\,\overline{\mu}_4(x)\left(A^T \otimes A^T\right) - (A \otimes A)(\Sigma \otimes \Sigma)\left(A^T \otimes A^T\right) \\ &\quad - (A \otimes A) K_{d,d}(\Sigma \otimes \Sigma)\left(A^T \otimes A^T\right) - (A \otimes A)\, vec(\Sigma)\, vec^T(\Sigma)\left(A^T \otimes A^T\right) \\ &= (A \otimes A)\left[\overline{\mu}_4(x) - \left(I_{d^2} + K_{d,d}\right)(\Sigma \otimes \Sigma) - vec(\Sigma)\, vec^T(\Sigma)\right]\left(A^T \otimes A^T\right).\end{aligned}$$

Hence $\kappa_4(Ax) = (A \otimes A)\,\kappa_4(x)\left(A^T \otimes A^T\right)$ and this completes the proof.

Proof of Theorem 3. By assumption, the components' variances are equal, implying that all components have the same dimension, which we shall denote by d. In the proof, we shall make repeated use of the identity

$$\kappa_4(y) = \overline{\mu}_4(y) - vec[var(y)]\, vec^T[var(y)] - \left(I_{d^2} + K_{d,d}\right)[var(y) \otimes var(y)],$$

where $\kappa_4(y)$, $\overline{\mu}_4(y)$ and $var(y)$ are the fourth cumulant, the fourth central moment and the variance of the random vector y.

Let $\mu_{i,j}$ ($\overline{\mu}_{i,j}$) the j-th (central) moment of the i-th mixture component, with cdf F_i and weight π_i, for $j = 1, 2, 3, 4$ and $i = 1, \ldots, g$. Ordinary properties of moments and cumulants imply that $\mu_{i,1} = \mu_i$, $\overline{\mu}_{i,1} = 0_d$, $\overline{\mu}_{i,2} = \Omega$, $\overline{\mu}_{i,3} = \Gamma$ and $\kappa_{i,4} = \overline{\mu}_{i,4} - vec(\Omega)\, vec^T(\Omega) - \left(I_{d^2} + K_{d,d}\right)(\Omega \otimes \Omega)$. Hence the expected value of Ξ is

$$E(\Xi) = \sum_{i=1}^{g} \pi_i \kappa_{i,4} = \left(\sum_{i=1}^{g} \pi_i \overline{\mu}_{i,4}\right) - vec(\Omega)\, vec^T(\Omega) - \left(I_{d^2} + K_{d,d}\right)(\Omega \otimes \Omega).$$

The mean, the variance, the fourth central moment and the fourth cumulant of m are

$$E(m) = \sum_{i=1}^{g} \pi_i \mu_i = \mu, \qquad \Psi = \sum_{i=1}^{g} \pi_i \delta_i \delta_i^T, \qquad \overline{\mu}_4(m) = \sum_{i=1}^{g} \pi_i \delta_i \delta_i^T \otimes \delta_i \delta_i^T,$$

$$\kappa_4(m) = \sum_{i=1}^{g} \pi_i \delta_i \delta_i^T \otimes \delta_i \delta_i^T - vec(\Psi)\, vec^T(\Psi) - \left(I_{d^2} + K_{d,d}\right)(\Psi \otimes \Psi),$$

where $\delta_i = \mu_i - \mu$. The mean, the variance the fourth central moment and the fourth cumulant of x are

$$E(x) = E(m) = \mu, \qquad \Sigma = \Omega + \Psi, \qquad \overline{\mu}_4(m) = \sum_{i=1}^{g} \pi_i \mu_{i,4}(x - \mu),$$

$$\kappa_4(x) = \overline{\mu}_4(x) - vec(\Sigma)\, vec^T(\Sigma) - \left(I_{d^2} + K_{d,d}\right)(\Sigma \otimes \Sigma).$$

The fourth moment about μ of a random vector with cdf F_i is $\mu_{i,4}(x - \mu) = \mu_{i,4}\left[\left(x - \mu_{i,1}\right) + \delta_i\right]$. A straightforward application of Theorem 1 leads to the following expression for $\mu_{i,4}(x - \mu)$:

$$\begin{aligned}&\overline{\mu}_{i,4} - \overline{\mu}_{i,3}^T \otimes \delta_i - \overline{\mu}_{i,3} \otimes \delta_i^T - \delta_i^T \otimes \overline{\mu}_{i,3} - \delta_i \otimes \overline{\mu}_{i,3}^T + \overline{\mu}_{i,2} \otimes \delta_i \delta_i^T \\ &\quad + vec\left(\overline{\mu}_{i,2}\right)(\delta_i \otimes \delta_i)^T + K_{d,d}\left(\delta_i \delta_i^T \otimes \overline{\mu}_{i,2}\right) + K_{d,d}\left(\overline{\mu}_{i,2} \otimes \delta_i \delta_i^T\right) \\ &\quad + \delta_i \otimes \delta_i \otimes vec^T\left(\overline{\mu}_{i,2}\right) + \delta_i \delta_i^T \otimes \overline{\mu}_{i,2} - \overline{\mu}_{i,1} \delta_i^T \otimes \delta_i \delta_i^T \\ &\quad - \delta_i \overline{\mu}_{i,1}^T \otimes \delta_i \delta_i^T - \delta_i \delta_i^T \otimes \overline{\mu}_{i,1} \delta_i^T - \delta_i \delta_i^T \otimes \delta_i \overline{\mu}_{i,1}^T + \delta_i \delta_i^T \otimes \delta_i \delta_i^T.\end{aligned}$$

Some simplifications might be achieved by recalling that $\mu_{i,1} = \mu_i$, $\overline{\mu}_{i,1} = 0_d$, $\overline{\mu}_{i,2} = \Omega$ and $\overline{\mu}_{i,3} = \Gamma$. Then $\mu_{i,4}(x-\mu)$ becomes

$$\overline{\mu}_{i,4} - \Gamma^T \otimes \delta_i - \Gamma \otimes \delta_i^T - \delta_i^T \otimes \Gamma - \delta_i \otimes \Gamma^T + \Omega \otimes \delta_i \delta_i^T + vec(\Omega) \otimes \delta_i^T \otimes \delta_i^T$$
$$+ K_{d,d}\left(\delta_i\delta_i^T \otimes \Omega\right) + K_{d,d}\left(\Omega \otimes \delta_i\delta_i^T\right) + \delta_i \otimes \delta_i \otimes vec^T(\Omega) + \delta_i\delta_i^T \otimes \Omega + \delta_i\delta_i^T \otimes \delta_i\delta_i^T.$$

The average of $\mu_{1,4}(x-\mu), \ldots, \mu_{g,4}(x-\mu)$, with weights $\pi_1, \ldots, \pi_g$ is

$$\sum_{i=1}^{g} \pi_i \mu_{i,4}(x-\mu) = \sum_{i=1}^{g} \pi_i \overline{\mu}_{i,4} - \Gamma^T \otimes \sum_{i=1}^{g} \pi_i \delta_i - \Gamma \otimes \sum_{i=1}^{g} \pi_i \delta_i^T - \sum_{i=1}^{g} \pi_i \delta_i^T \otimes \Gamma$$
$$- \sum_{i=1}^{g} \pi_i \delta_i \otimes \Gamma^T + \Omega \otimes \sum_{i=1}^{g} \pi_i \delta_i \delta_i^T + vec(\Omega) \otimes \sum_{i=1}^{g} \pi_i \delta_i^T \otimes \delta_i^T + K_{d,d}\left(\sum_{i=1}^{g} \pi_i \delta_i \delta_i^T \otimes \Omega\right)$$
$$+ K_{d,d}\left(\Omega \otimes \sum_{i=1}^{g} \pi_i \delta_i \delta_i^T\right) + \sum_{i=1}^{g} \pi_i \delta_i \otimes \delta_i \otimes vec^T(\Omega) + \sum_{i=1}^{g} \pi_i \delta_i \delta_i^T \otimes \Omega + \sum_{i=1}^{g} \pi_i \delta_i \delta_i^T \otimes \delta_i \delta_i^T.$$

It might be expressed in terms of $\overline{\mu}_4(x)$, Ψ and $\overline{\mu}_4(m)$ as follows:

$$\overline{\mu}_4(x) = \sum_{i=1}^{g} \pi_i \overline{\mu}_{i,4} + \overline{\mu}_4(m) + \left(I_{d^2} + K_{d,d}\right)(\Omega \otimes \Psi)$$
$$+ \left(I_{d^2} + K_{d,d}\right)(\Psi \otimes \Omega) + vec(\Omega)\, vec^T(\Psi) + vec(\Psi)\, vec^T(\Omega),$$

by using the fact that $\pi_1\delta_1 + \cdots + \pi_i\delta_i = 0_d$. The identity $\Sigma = \Omega + \Psi$ implies that $\overline{\mu}_4(x) - \kappa_4(x)$ equals

$$[vec(\Omega) + vec(\Psi)][vec(\Omega) + vec(\Psi)]^T + \left(I_{d^2} + K_{d,d}\right)(\Omega \otimes \Omega + \Omega \otimes \Psi + \Psi \otimes \Omega + \Psi \otimes \Psi)$$
$$= vec(\Omega)\, vec^T(\Omega) + \left(I_{d^2} + K_{d,d}\right)(\Omega \otimes \Omega) + vec(\Psi)\, vec^T(\Psi) + \left(I_{d^2} + K_{d,d}\right)(\Psi \otimes \Psi)$$
$$+ \left(I_{d^2} + K_{d,d}\right)(\Omega \otimes \Psi) + \left(I_{d^2} + K_{d,d}\right)(\Psi \otimes \Omega) + vec(\Omega)\, vec^T(\Psi) + vec(\Psi)\, vec^T(\Omega).$$

In terms of $E(\Xi)$, $\kappa_4(m)$ and $\overline{\mu}_4(m)$ the difference $\overline{\mu}_4(x) - \kappa_4(x)$ is

$$\sum_{i=1}^{g} \pi_i \overline{\mu}_{i,4} - E(\Xi) + \overline{\mu}_4(m) - \kappa_4(m) + \left(I_{d^2} + K_{d,d}\right)(\Omega \otimes \Psi)$$
$$+ \left(I_{d^2} + K_{d,d}\right)(\Psi \otimes \Omega) + vec(\Omega)\, vec^T(\Psi) + vec(\Psi)\, vec^T(\Omega).$$

By substituting the value of $\overline{\mu}_4(x)$ in the above equation we obtain

$$\sum_{i=1}^{g} \pi_i \overline{\mu}_{i,4} + \left(I_{d^2} + K_{d,d}\right)(\Omega \otimes \Psi) + vec(\Psi)\, vec^T(\Omega)$$
$$+ \left(I_{d^2} + K_{d,d}\right)(\Psi \otimes \Omega) + vec(\Omega)\, vec^T(\Psi) + \overline{\mu}_4(m) - \kappa_4(x)$$
$$= \sum_{i=1}^{g} \pi_i \overline{\mu}_{i,4} - E(\Xi) - \kappa_4(m) + \overline{\mu}_4(m) + \left(I_{d^2} + K_{d,d}\right)(\Omega \otimes \Psi)$$
$$+ \left(I_{d^2} + K_{d,d}\right)(\Psi \otimes \Omega) + vec(\Omega)\, vec^T(\Psi) + vec(\Psi)\, vec^T(\Omega).$$

Simple algebra leads to the identity $\kappa_4(x) = E(\Xi) + \kappa_4(m)$, thus completing the proof.

Proof of Theorem 4. We shall first recall some fundamental properties of the Kronecker product (see, for example, [29], page 460): (P1) the Kronecker product is associative: $(A \otimes B) \otimes C = A \otimes (B \otimes C) = A \otimes B \otimes C$; (P2) if a and b are two vectors, then ab^T, $a \otimes b^T$ and $b^T \otimes a$ denote the same matrix. Let $\mu = \pi_1\mu_1 + \pi_2\mu_2$ and $\Sigma = \Omega + \pi_1\pi_2\delta\delta^T$ denote the mean and the variance of x. Also, let $\delta = \mu_1 - \mu_2$ and $q = \pi_1\pi_2(1 - 6\pi_1\pi_2)$. The second and third cumulants of $N_d(\mu_1, \Omega)$ equal those of $N_d(\mu_2, \Omega)$, hence satisfying the assumptions of Theorem 3. Also, the fourth cumulants of $N_d(\mu_1, \Omega)$ and $N_d(\mu_2, \Omega)$ are null matrices. It follows that the fourth cumulant of x equals the fourth cumulant of the random vector m, whose distribution places probability mass π_1 on μ_1 and probability mass $\pi_2 = 1 - \pi_1$ on μ_2. The fourth central moment of m is

$$\overline{\mu}_4(m) = E\left[(m-\mu) \otimes (m-\mu)^T \otimes (m-\mu) \otimes (m-\mu)^T\right]$$
$$= (\mu_1-\mu) \otimes (\mu_1-\mu)^T \otimes (\mu_1-\mu) \otimes (\mu_1-\mu)^T \pi_1$$
$$+ (\mu_2-\mu) \otimes (\mu_2-\mu)^T \otimes (\mu_2-\mu) \otimes (\mu_2-\mu)^T \pi_2$$
$$= \pi_2^4\pi_1 (\mu_1-\mu_2) \otimes (\mu_1-\mu_2)^T \otimes (\mu_1-\mu_2) \otimes (\mu_1-\mu_2)^T$$
$$+ \pi_1^4\pi_2 (\mu_1-\mu_2) \otimes (\mu_1-\mu_2)^T \otimes (\mu_1-\mu_2) \otimes (\mu_1-\mu_2)^T.$$

A simpler representation of $\overline{\mu}_4(m)$ might be obtained by recalling that $\pi_1 + \pi_2 = 1$ and $\delta = \mu_1 - \mu_2$: $\overline{\mu}_4(m) = \pi_1\pi_2\left(\pi_1^2 - \pi_1\pi_2 + \pi_2^2\right)\delta \otimes \delta^T \otimes \delta \otimes \delta^T$. The variance of m is $\pi_1\pi_2\delta\delta^T$, so that its fourth cumulant is

$$\begin{aligned}\kappa_4(m) &= \pi_1\pi_2\left(\pi_1^2 - \pi_1\pi_2 + \pi_2^2\right)\delta \otimes \delta^T \otimes \delta \otimes \delta^T - (\pi_1\pi_2)^2\left(I_{d^2} + K_{d,d}\right)\delta\delta^T \otimes \delta\delta^T \\ &\quad - (\pi_1\pi_2)^2\, vec\left(\delta\delta^T\right) vec^T\left(\delta\delta^T\right).\end{aligned}$$

Consider now the identities $\delta \otimes \delta^T \otimes \delta \otimes \delta^T = \delta \otimes \delta \otimes \delta^T \otimes \delta^T$, $vec\left(\delta\delta^T\right) = \delta \otimes \delta$, $\delta\delta^T \otimes \delta\delta^T = \delta \otimes \delta^T \otimes \delta \otimes \delta^T$ and $K_{d,d}\left(\delta\delta^T \otimes \delta\delta^T\right) = \delta \otimes \delta^T \otimes \delta \otimes \delta^T$, which are direct implications of standard properties of the Kronecker product, the vectorization operator and the commutation matrix. They lead to the following simplified expression for the fourth cumulant of m:

$$\begin{aligned}\kappa_4(m) &= \left[\pi_1\pi_2\left(\pi_1^2 - \pi_1\pi_2 + \pi_2^2\right) - 3(\pi_1\pi_2)^2\right]\left(\delta \otimes \delta^T \otimes \delta \otimes \delta^T\right) \\ &= \pi_1\pi_2\left(\pi_1^2 - 4\pi_1\pi_2 + \pi_2^2\right)\left(\delta \otimes \delta^T \otimes \delta \otimes \delta^T\right) = q\delta\delta^T \otimes \delta\delta^T.\end{aligned}$$

Proof of Theorem 5. We shall use again the notation in the previous proof. The matrix Ω is positive definite by assumption and the variance of x is $\Sigma = \Omega + \pi_1\pi_2\delta\delta^T$, by ordinary properties of mixture models. As a direct consequence, Σ is a full-rank matrix and the standardized random vector $z = \Sigma^{-1/2}(x - \mu)$ is well-defined. By Theorem 2 the fourth cumulant of z is $\kappa_4(z) = \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right)\kappa_4(x)\left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right)$. The quantity $\kappa_4^{(1)}$ is the sum of the diagonal elements of $\kappa_4(z)$, that is the trace of $\kappa_4(z)$ itself (see, for example, [11]), that is

$$\kappa_4^{(1)} = tr\left[\kappa_4(z)\right] = tr\left[\left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right)\kappa_4(x)\left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right)\right].$$

When matrices A and B are of appropriate size $tr(AB)$ equals $tr(BA)$. This fact, together with the above representation of $\kappa_4(x)$, leads to

$$\kappa_4^{(1)} = q \cdot tr\left[\left(\delta\delta^T \otimes \delta\delta^T\right)\left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right)\left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right)\right].$$

If matrices A, B, C and D are of appropriate size, then $(A \otimes B)(C \otimes D)$ equals $AC \otimes BD$. Also, recall that $\Sigma^{-1/2}\Sigma^{-1/2} = \Sigma^{-1}$ by definition and write $\kappa_4^{(1)} = q \cdot tr\left[\left(\delta\delta^T\Sigma^{-1}\right) \otimes \left(\delta\delta^T\Sigma^{-1}\right)\right]$. When both A and B are square matrices $tr(A \otimes B)$ equals $tr(A) \cdot tr(B)$. Hence $\kappa_4^{(1)} = q \cdot tr\left(\delta\delta^T\Sigma^{-1}\right)^2 = q \cdot \left(\delta^T\Sigma^{-1}\delta\right)^2$. The quantity $\kappa_{4,4}^{(1)}$ is the sum of all the squared elements of $\kappa_4(z)$, that is the trace of the product of $\kappa_4(z)$ and its transpose. Since $\kappa_4(z)$ is a symmetric matrix, $\kappa_{4,4}^{(1)}$ equals the trace of

$$\left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right)\kappa_4(x)\left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right)\left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right)\kappa_4(x)\left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right).$$

An argument similar to the one used to evaluate $\kappa_4^{(1)}$ leads us to the following equations:

$$\begin{aligned}\kappa_{4,4}^{(1)} &= tr\left[\kappa_4(x)\left(\Sigma^{-1} \otimes \Sigma^{-1}\right)\kappa_4(x)\left(\Sigma^{-1} \otimes \Sigma^{-1}\right)\right] \\ &= c \cdot tr\left[\left(\Sigma^{-1}\delta\delta^T \otimes \Sigma^{-1}\delta\delta^T\right)\left(\Sigma^{-1}\delta\delta^T \otimes \Sigma^{-1}\delta\delta^T\right)\right] \\ &= c \cdot tr\left[\left(\Sigma^{-1}\delta\delta^T\Sigma^{-1}\delta\delta^T\right) \otimes \left(\Sigma^{-1}\delta\delta^T\Sigma^{-1}\delta\delta^T\right)\right] \\ &= c \cdot tr\left[\left(\Sigma^{-1}\delta\delta^T\Sigma^{-1}\delta\delta^T\right)\right]^2 = c \cdot \left(\delta^T\Sigma^{-1}\delta\right)^4.\end{aligned}$$

The inverse of Σ might be easily evaluated by the Sherman–Morrison formula (see, for example, Mardia et al., page 459):

$$\Sigma^{-1} = \left(\Sigma = \Omega + \pi_1\pi_2\delta\delta^T\right)^{-1} = \Omega^{-1} - \frac{\Omega^{-1}\delta\delta^T\Omega^{-1}}{(\pi_1\pi_2)^{-1} + \delta^T\Omega^{-1}\delta}.$$

Hence, after a little algebra, we obtain

$$\delta^T\Sigma^{-1}\delta = \delta^T\left\{\Omega^{-1} - \frac{\Omega^{-1}\delta\delta^T\Omega^{-1}}{(\pi_1\pi_2)^{-1} + \delta^T\Omega^{-1}\delta}\right\}\delta = \left(1 + \pi_1\pi_2\delta^T\Omega^{-1}\delta\right)^{-1}.$$

The fourth standardized cumulant might be represented as $\kappa_4(z) = q \cdot \lambda \otimes \lambda^T \otimes \lambda \otimes \lambda^T$, where $\lambda = (\lambda_1, \ldots, \lambda_d)^T = \Sigma^{-1/2}\delta$, which is equivalent to $\kappa_{ijhk} = q\lambda_i\lambda_j\lambda_h\lambda_k$, for $i, j, h, k = 1, \ldots, d$. The identities

$$\kappa_{iihk}\kappa_{jjhk} = \left(q\lambda_i^2\lambda_h\lambda_k\right)\left(q\lambda_j^2\lambda_h\lambda_k\right) = \left(q\lambda_i\lambda_j\lambda_h\lambda_k\right)^2 = \kappa_{ijhk}^2$$

imply that $\kappa_{4,4}^{(1)}$ equals $\kappa_{4,4}^{(2)}$ and this completes the proof.

References

[1] C.A. Adcock, M. Eling, N. Loperfido, Skewed distributions in finance and actuarial science: a review, Eur. J. Fin. (2012) http://dx.doi.org/10.1080/1351847X.2012.720269.
[2] A. Azzalini, A. Capitanio, Statistical applications of the multivariate skew-normal distributions, J. R. Stat. Soc. Ser. B Stat. Methodol. 61 (1999) 579–602.
[3] A. Azzalini, A. Dalla Valle, The multivariate skew-normal distribution, Biometrika 83 (1996) 715–726.
[4] M. Bernardi, A. Maruotti, L. Petrella, Skew mixture models for loss distributions: a Bayesian approach, Insur. Math. Econ. 51 (2012) 617–623.
[5] D.R. Brillinger, The calculation of cumulants via conditioning, Ann. Inst. Statist. Math. 21 (1969) 215–218.
[6] N.A. Campbell, R.J. Mahon, A multivariate study of variation in two species of rock crab of genus Leptograpsus, Aus. J. Zoo. 22 (1974) 417–425.
[7] J.F. Cardoso, Source separation using higher order moments, in: Proc. ICASSP'89, 1989, pp. 2109–2112.
[8] P. Comon, G. Golub, L.H. Lim, B. Mourrain, Symmetric tensors and symmetric tensor rank, SIAM J. Matrix Anal. Appl. 30 (2008) 1254–1279.
[9] L. De Lathauwer, P. Comon, B. De Moor, J. Vandewalle, Higher-order power method. Application in independent component analysis, in: Proceedings of the International Symposium on Nonlinear Theory and Its Applications, Las Vegas (NV), 1995, pp. 91–96.
[10] G. De Luca, N. Loperfido, Modelling multivariate skewness in financial returns: a SGARCH approach, Eur. J. Fin. (2012) http://dx.doi.org/10.1080/1351847X.2011.640342.
[11] 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, 2011.
[12] S. Frühwirth-Schnatter, S. Pyne, Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions, Biostatistics 11 (2010) 317–336.
[13] N. Henze, The asymptotic behavior of a variant of multivariate kurtosis, Comm. Statist. Theory Methods 23 (1994) 1047–1061.
[14] L. Hu, P.M. Bentler, Y. Kano, Can test statistics in covariance structure analysis be trusted? Psychol. Bull. 112 (1992) 351–362.
[15] A. Hyvarinen, J. Karhunen, E. Oja, Independent Component Analysis, John Wiley & Sons, New York, 2001.
[16] M.C. Jones, R. Sibson, What is Projection Pursuit? (with discussion), J. R. Statist. Soc. A 150 (1987) 1–36.
[17] M.E. Kilmer, C.D. Martin, Factorization strategies for third-order tensors, Linear Algebra Appl. 435 (2011) 641–658.
[18] T. Kolda, B. Bader, Tensor decompositions and applications, SIAM Rev. 51 (2009) 455–500.
[19] T. Kollo, Multivariate skewness and kurtosis measures with an application in ICA, J. Multivariate Anal. 99 (2008) 2328–2338.
[20] H. Kurata, On principal points for location mixtures of multivariate spherically symmetric distributions, J. Statist. Plann. Inference 138 (2008) 3405–3418.
[21] T.I. Lin, J.C. Lee, S.Y. Yen, Finite mixture modelling using the skew normal distribution, Statist. Sinica 17 (2007) 909–927.
[22] T.C. Lin, T.I. Lin, Supervised learning of multivariate skew normal mixture models with missing information, Comput. Statist. 25 (2010) 183–201.
[23] N. Loperfido, Spectral analysis of the fourth moment matrix, Linear Algebra Appl. 435 (2011) 1837–1844.
[24] N. Loperfido, Skewness and the linear discriminant function, Statist. Probab. Lett. 88 (2013) 93–99.
[25] J.R. Magnus, H. Neudecker, The commutation matrix: some properties and applications, Ann. Statist. 7 (1979) 381–394.
[26] J.F. Malkovich, A.A. Afifi, On tests for multivariate normality, J. Amer. Statist. Assoc. 68 (1973) 176–179.
[27] K.V. Mardia, Measures of multivariate skewness and kurtosis with applications, Biometrika 57 (1970) 519–530.
[28] K.V. Mardia, Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies, Sankhya B 36 (1974) 115–128.
[29] K.V. Mardia, J.T. Kent, J.M. Bibby, Multivariate Analysis, Academic Press, London, 1979.
[30] S. Matsuura, H. Kurata, A principal subspace theorem for 2-principal points of general location mixtures of spherically symmetric distributions, Statist. Probab. Lett. 80 (2010) 1863–1869.
[31] S. Matsuura, H. Kurata, Principal points of a multivariate mixture distribution, J. Multivariate Anal. 102 (2011) 213–224.
[32] G. McLachlan, D. Peel, Finite Mixture Models, John Wiley and Sons, Inc, New York, 2000.
[33] V. Melnykov, R. Maitra, Finite mixture models and model-based clustering, Statist. Surv. 4 (2010) 80–116.
[34] T.F. Mòri, V.K. Rohatgi, G.J. Székely, On multivariate skewness and kurtosis, Theory Probab. Appl. 38 (1993) 547–551.
[35] K. Pearson, Contributions to the theory of mathematical evolution, II: skew variation, Philos. Trans. R. Soc. Lond. Ser. A 186 (1895) 343–414.
[36] D. Peña, F.J. Prieto, The kurtosis coefficient and the linear discriminant function, Statist. Probab. Lett. 49 (2000) 257–261.
[37] D. Peña, F.J. Prieto, Multivariate outlier detection and robust covariance estimation (with discussion), Technometrics 43 (2001) 286–310.
[38] D. Peña, F.J. Prieto, Cluster identification using projections, J. Amer. Statist. Assoc. 96 (2001) 1433–1445.
[39] D. Peña, F.J. Prieto, Combining random and specific directions for outlier detection and robust estimation of high-dimensional multivariate data, J. Comput. Graph. Statist. 16 (2007) 228–254.
[40] D. Peña, F.J. Prieto, J. Viladomat, Eigenvectors of a kurtosis matrix as interesting directions to reveal cluster structure, J. Multivariate Anal. 101 (2010) 1995–2007.
[41] L. Qi, W. Sun, Y. Wang, Numerical multilinear algebra and its applications, Front. Math. China 2 (2007) 501–526.
[42] N.J. Schork, A.B. Weder, M.A. Schork, On the asymmetry of biological frequency distributions, Gen. Epidem. 7 (1990) 427–446.
[43] J.R. Schott, Some tests for the equality of covariance matrices, J. Statist. Plann. Inference 94 (2001) 25–36.
[44] D.E. Tyler, F. Critchley, L. Dümbgen, H. Oja, Invariant co-ordinate selection (with discussion), J. R. Stat. Soc. B 71 (2009) 1–27.
[45] 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, J. Multivariate Anal. 96 (2005) 237–264.