ARTICLE IN PRESS
JID: ECOSTA [m3Gsc;May 27, 2021;10:15]

Econometrics and Statistics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Econometrics and Statistics

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

Edgeworth expansions for multivariate random sums

Farrukh Javed[a], Nicola Loperfido[b,*], Stepan Mazur[c]

[a] Örebro University School of Business Fakultetsgatan 1, SE-70182 Örebro, Sweden
[b] Università degli Studi di Urbino "Carlo Bo" Via Saffi 42, 61029 Urbino, Italy
[c] Örebro University School of Business Fakultetsgatan 1, SE-70182 Örebro, Sweden

ARTICLE INFO

Article history:
Received 7 October 2020
Revised 3 February 2021
Accepted 21 April 2021
Available online xxx

MSC:
Primary 60K35
Secondary 60K35

Keywords:
Edgeworth expansion
Fourth cumulant
Random sum
Skew-normal

ABSTRACT

The sum of a random number of independent and identically distributed random vectors has a distribution which is not analytically tractable, in the general case. The problem has been addressed by means of asymptotic approximations embedding the number of summands in a stochastically increasing sequence. Another approach relies on fitting flexible and tractable parametric, multivariate distributions, as for example finite mixtures. Both approaches are investigated within the framework of Edgeworth expansions. A general formula for the fourth-order cumulants of the random sum of independent and identically distributed random vectors is derived and it is shown that the above mentioned asymptotic approach does not necessarily lead to valid asymptotic normal approximations. The problem is addressed by means of Edgeworth expansions. Both theoretical and empirical results suggest that mixtures of two multivariate normal distributions with proportional covariance matrices satisfactorily fit data generated from random sums where the counting random variable and the random summands are Poisson and multivariate skew-normal, respectively.

1. Introduction

Let $\{\mathbf{x}_i, i \in \mathbb{N}\}$ be a sequence of independent and identically distributed (i.i.d.) d-dimensional random vectors. Also, let N be a random variable independent of the sequence whose support is the set of nonnegative integers. Finally, let $\mathbf{s} = \mathbf{x}_1 + \ldots + \mathbf{x}_N$ be the sum of the first N components of the sequence, with the convention that $\mathbf{s}$ is the d -dimensional null vector when $N = 0$. We refer to $\mathbf{s}$ as to the random sum of $\{\mathbf{x}_i, i \in \mathbb{N}\}$ with aggregating variable N. The distributions of $\mathbf{s}$, N and $\mathbf{x}_i$ are the compound, primary and secondary distributions, respectively (Lin, 2006). Compound distributions have been mainly studied in actuarial science, where the random vector $\mathbf{s}$, $\mathbf{x}_i$ and N are the aggregated claim, the i-th claim size and the claim counts, respectively (Ambagaspitiya, 1999). Random sums occur in many other research fields, including physics, biology, reliability, queuing and finance (Gnedenko and Korolev, 1996), (Kalashnikov, 1997), (Klebanov et al., 2006). Analytic expressions for compound distributions are available only for certain types of primary and secondary distributions (Bowers et al., 1997, Chapter 13) and an approximation is required for a general case (Lin, 2006).

Asymptotic methods are often used to approximate the compound distributions, when the aggregating variable is expected to be high enough. Following an argument similar to the one used in central limit theorems, the aggregating variable is embedded in a sequence of stochastically increasing random variables. More precisely, let $\{N_q, q \in \mathbb{N}\}$ be a stochastically increasing sequence of random variables independent of $\{\mathbf{x}_i, i \in \mathbb{N}\}$, whose support is the set of nonnegative integers. Also,

* Corresponding author.
E-mail addresses: Farrukh.Javed@oru.se (F. Javed), nicola.loperfido@uniurb.it (N. Loperfido), Stepan.Mazur@oru.se (S. Mazur).

https://doi.org/10.1016/j.ecosta.2021.04.005
2452-3062/

Please cite this article as: F. Javed, N. Loperfido and S. Mazur, Edgeworth expansions for multivariate random sums, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2021.04.005

let $\boldsymbol{\mu}_q$ and $\boldsymbol{\Sigma}_q$ be the mean vector and the positive definite covariance matrix of the q-th element $\mathbf{s}_q$ of the sequence $\{\mathbf{s}_q = \mathbf{x}_1 + \ldots + \mathbf{x}_{N_q}, q \in \mathbb{N}\}$. Finally, let $\{\mathbf{z}_q = \boldsymbol{\Sigma}_q^{-1/2}(\mathbf{s}_q - \boldsymbol{\mu}_q), q \in \mathbb{N}\}$ be the associated sequence of standardized random vectors, where $\boldsymbol{\Sigma}_q^{-1/2}$ is the positive definite square root of the inverse of $\boldsymbol{\Sigma}_q$. Conditions for the convergence of $\{\mathbf{z}_q, q \in \mathbb{N}\}$ to a standard normal distribution have been given in the univariate case in the pioneering work of (Robbins, 1948). In Klebanov et al. (2006), the authors focused on geometric summation which leads to geometric stable distributions. In Daley et al. (2007) the tail behaviour has been investigated when the secondary distribution is multivariate subexponential.

Alternatively, the compound distributions can be approximated by properly chosen parametric distributions, as it is often done in actuarial sciences. There have been many proposals for fitting univariate compund distributions with parametric ones (see Burnecki et al. (2011)). Multivariate compound distributions have been approximated with skew-normal distributions and their generalizations (Bolance et al., 2008), Eling (2012). Unfortunately, the information matrix of the distribution in Bolance et al. (2008) is singular under normality, thus preventing the use of standard likelihood-based methods for testing the hypothesis of normality (Franceschini and Loperfido, 2014), while the skew-normal distribution suffers from the inferential problems discussed in Pewsey (2001). Multivariate compound distributions have also been approximated with finite mixture distributions (Bernardi et al., 2012), which may require the estimation of many parameters. For example, a mixture of k d-dimensional normal distributions is parametrized by $k(d^2 + 3d + 2)/2$ real values, unless some simplifying assumptions are made.

In an interesting thread of research, the compound distributions can also be approximated by means of Edgeworth expansions, which possess several attractive properties (Gray et al., 1975), (Bhattacharya and Ghosh, 1978), (Barndorff-Nielsen and Cox, 1979). The simplest, nontrivial Edgeworth expansion of a compound distribution depends on its first three cumulants, which have a simple analytical form. The third cumulant of a d-dimensional vector $\mathbf{y}$ with mean $\boldsymbol{\mu}$ and finite third-order moments is the $d^2 \times d$ matrix $\mathbf{K}_3(\mathbf{y}) = \mathbb{E}\left[(\mathbf{y} - \boldsymbol{\mu}) \otimes (\mathbf{y} - \boldsymbol{\mu}) \otimes (\mathbf{y} - \boldsymbol{\mu})^{\mathrm{T}}\right]$, where $\otimes$ stands for the Kronecker product. Let ν_j and $\boldsymbol{\xi}_j$ $(j = 1, 2, 3)$ be the $j-$th cumulants of N and $\mathbf{x}_i$, so that the first, second and third cumulants of $\mathbf{s}$ are $\mathbb{E}(\mathbf{s}) = \nu_1 \boldsymbol{\xi}_1$, $Var(\mathbf{s}) = \nu_1 \boldsymbol{\xi}_2 + \nu_2 \boldsymbol{\xi}_1 \boldsymbol{\xi}_1^{\mathrm{T}}$ and

$$\mathbf{K}_3(\mathbf{s}) = \nu_1 \boldsymbol{\xi}_3 + \nu_2 \left(\boldsymbol{\xi}_2 \otimes \boldsymbol{\xi}_1 + \mathrm{vec}(\boldsymbol{\xi}_2) \boldsymbol{\xi}_1^{\mathrm{T}} + \boldsymbol{\xi}_1 \otimes \boldsymbol{\xi}_2 \right) + \nu_3 \boldsymbol{\xi}_1 \otimes \boldsymbol{\xi}_1^{\mathrm{T}} \otimes \boldsymbol{\xi}_1,$$

(Loperfido et al., 2018), where $\mathrm{vec}(\boldsymbol{\xi}_2)$ is the vector obtained by stacking the columns of the matrix $\boldsymbol{\xi}_2$ on top of each other. Note that the Edgeworth expansions are closely connected to both asymptotic and parametric approximations. Van Hulle (2005) used Edgeworth expansions to show that the approximation of a multivariate normal distribution with another distribution with the same mean vector and covariance matrix, as measured by the Kullback-Leibler divergence, tends to improve when the third cumulants of the latter distribution become negligible. Loperfido (2019) used theoretical as well as empirical arguments based on Edgeworth expansions to show that data generated by a distribution might be satisfactorily fitted by another distribution, as long as the two distributions have the same first three cumulants.

Intuition suggests that the Edgeworth approximation of a given distribution improves when more terms are included in the approximation itself. In particular, a third-order Edgeworth approximation might be improved by including the fourth-order cumulants of the approximated distribution. Inclusion of fourth cumulants is particularly useful when the performance of the chosen statistical method heavily depends on the fourth cumulants of the sampled distribution, as exemplified by the following cases. Mardia (1974) showed that the performance of a likelihood test based on the erroneous assumption of normality crucially depends on a measure of multivariate kurtosis which is a simple function of fourth cumulants. Fourth cumulants of the sampled distribution impair the performance of widely used likelihood tests on covariance matrices, when normality is assumed (Yanagihara et al., 2005). Rezvandehy and Deutsch (2018) addressed preferential sampling in spatial statistics by means of fourth cumulants. Yanagihara (2007) stressed the importance of fourth cumulants in the multivariate linear model. (Arevalillo and Navarro, 2012) investigated the effect of fourth cumulants on the Fisher discriminant function, and found it to be nonnegligible. (Mutschler, 2018) provides analytical expressions for higher order cumulants for non-Gaussian or nonlinear (pruned) solutions to DSGE models to provide means to gain more information for calibration and estimation.

In this paper we derive a general formula for the fourth-order cumulants of the random sum of independent and identically distributed random vectors and show that the asymptotic approach based on stochastically increasing sequences does not necessarily lead to valid asymptotic normal approximations. We address the problem by means of Edgeworth expansions. Both theoretical and empirical arguments suggest that the distribution of a random sum where the summands and their number are skew-normal and Poisson is satisfactorily approximated by a mixture of two multivariate normal distributions with proportional covariance matrices.

Sections 2 and 3 contain the theoretical results regarding the fourth cumulant of a multivariate random sum and Mardia's measure of multivariate kurtosis. These results are used in Section 4 to highlight some difficulties in the convergence of multivariate random sums to a multivariate normal distribution, which are addressed in Section 5 by means of Edgeworth expansions. Section 6 and the Appendix contain some concluding remarks and all theorems' proofs, respectively.

2. Fourth cumulant

Let $\mathbf{x} = (x_1, \ldots, x_d)^{\mathrm{T}}$ be a d -dimensional random vector with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^{\mathrm{T}}$ and covariance matrix $\boldsymbol{\Sigma} = \{\sigma_{ij}\}$, $i, j = 1, \ldots, d$. We assume that the fourth moments of $\mathbf{x}$ are finite, i.e. $\mathbb{E}(|x_i x_j x_h x_k|) < +\infty$ for $i, j, h, k = 1, \ldots, d$. The fourth cumulant $\mathcal{K}_4=\{\kappa_{ijhk}\}$ of $\mathbf{x}$ is the d-dimensional, symmetric tensor of order 4 whose elements are the fourth-order derivatives of the cumulant generating function of $\mathbf{x}$: $\kappa_{ijhk}=\log\mathbb{E}[\exp(\iota\mathbf{t}^{\mathrm{T}}\mathbf{x})]/\partial t_i\partial t_j\partial t_h\partial t_k$, where $\iota^2=-1$ and $\mathbf{t}=(t_1,\dots,t_d)^{\mathrm{T}}$ is an arbitrary vector of constants. An equivalent representation of κ_{ijhk} is

$$\mathbb{E}\Big[(x_i-\mu_i)\big(x_j-\mu_j\big)(x_h-\mu_h)(x_k-\mu_k)\Big]-\sigma_{ij}\sigma_{hk}-\sigma_{ih}\sigma_{jk}-\sigma_{ik}\sigma_{jh}. \tag{2.1}$$

The elements κ_{ijhk} might be arranged into the $d^2\times d^2$ block matrix $\mathbf{K}_4(\mathbf{x})=\{\tilde{\kappa}_{pq}\}$, where $\tilde{\kappa}_{pq}=\log\mathbb{E}\Big[\exp\big(\iota\mathbf{t}^{\mathrm{T}}\mathbf{x}\big)\Big]/\partial t_p\partial t_q\partial\mathbf{t}\partial\mathbf{t}^{\mathrm{T}}$ for $p,q=1,\dots,d$. The matrix $\mathbf{K}_4(\mathbf{x})$ is the unfolded version of $\mathcal{K}_4$ (Qi and Wang, 2007) and can be represented as

$$\mathbf{K}_4(\mathbf{x})=\mathbb{E}\Big[\mathbf{y}\otimes\mathbf{y}^{\mathrm{T}}\otimes\mathbf{y}\otimes\mathbf{y}^{\mathrm{T}}\Big]-\Big(\mathbf{I}_{d^2}+\mathbf{K}_{d,d}\Big)(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})-\mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma})^{\mathrm{T}}, \tag{2.2}$$

where $\mathbf{y}=\mathbf{x}-\boldsymbol{\mu}$, $\mathrm{vec}(\boldsymbol{\Sigma})$ is the vectorization of $\boldsymbol{\Sigma}$ and $\mathbf{K}_{d,d}$ is the $d^2\times d^2$ commutation matrix (Magnus and Neudecker, 1979). We refer to the matrix $\mathbf{K}_4(\mathbf{x})$ as to the fourth cumulant of $\mathbf{x}$. Some spectral properties of $\mathbf{K}_4(\mathbf{x})$ are examined by (Loperfido, 2011). Basic properties of the fourth cumulant that can be found in Kollo and von Rosen (2005) and Loperfido (2014), among others.

In the next theorem, we present the fourth cumulant of the random sum of random vectors $\mathbf{s}$. It is shown that the fourth cumulant of $\mathbf{s}$ can be represented via the first fourth cumulants of the counting random variable N and the random summands $\mathbf{x}_i$.

Theorem 2.1. *Let N be a nonnegative counting random variable with finite fourth moment. Also, let $\mathbf{x}_1,\dots,\mathbf{x}_N$ be i.i.d. random vectors with finite fourth moment and independent of N. Then the fourth cumulant of $\mathbf{s}$ is*

$$\begin{aligned}\mathbf{K}_4(\mathbf{s})&=\nu_1\boldsymbol{\xi}_4+\nu_2\Big[\Big(\mathbf{I}_{d^2}+\mathbf{K}_{d,d}\Big)\big(\boldsymbol{\xi}_2\otimes\boldsymbol{\xi}_2\big)+\mathrm{vec}(\boldsymbol{\xi}_2)\mathrm{vec}(\boldsymbol{\xi}_2)^{\mathrm{T}}\Big]+\nu_2\Big(\boldsymbol{\xi}_3\otimes\boldsymbol{\xi}_1^{\mathrm{T}}+\boldsymbol{\xi}_3^{\mathrm{T}}\otimes\boldsymbol{\xi}_1+\boldsymbol{\xi}_1^{\mathrm{T}}\otimes\boldsymbol{\xi}_3+\boldsymbol{\xi}_1\otimes\boldsymbol{\xi}_3^{\mathrm{T}}\Big)\\&+\nu_3\Big(\boldsymbol{\xi}_2\otimes\boldsymbol{\xi}_1\boldsymbol{\xi}_1^{\mathrm{T}}+\mathrm{vec}(\boldsymbol{\xi}_2)\boldsymbol{\xi}_1^{\mathrm{T}}\otimes\boldsymbol{\xi}_1^{\mathrm{T}}+\boldsymbol{\xi}_1^{\mathrm{T}}\otimes\boldsymbol{\xi}_2\otimes\boldsymbol{\xi}_1+\boldsymbol{\xi}_1\otimes\boldsymbol{\xi}_2\otimes\boldsymbol{\xi}_1^{\mathrm{T}}+\boldsymbol{\xi}_1\mathrm{vec}(\boldsymbol{\xi}_2)^{\mathrm{T}}\otimes\boldsymbol{\xi}_1+\boldsymbol{\xi}_1\boldsymbol{\xi}_1^{\mathrm{T}}\otimes\boldsymbol{\xi}_2\Big)\\&+\nu_4\Big(\boldsymbol{\xi}_1\boldsymbol{\xi}_1^{\mathrm{T}}\otimes\boldsymbol{\xi}_1\boldsymbol{\xi}_1^{\mathrm{T}}\Big),\end{aligned}$$

where ν_j and $\boldsymbol{\xi}_j$ denote the j–th ($j=1,2,3,4$) cumulants of N and $\mathbf{x}_i$, respectively.

The next corollary is the direct consequence of Theorem 2.1 and considers the fourth cumulant of the random sum of random variables. This result coincides with the formula obtained in Cummins and Wiltbank (1983).

Corollary 2.1. *Let N be a nonnegative counting random variable with finite fourth moment. Let $x_1,\dots,x_N$ be i.i.d. random variables with a finite fourth moment, independent of N. Then the fourth cumulant of s is*

$$\mathbf{K}_4(s)=\nu_1\xi_4+3\nu_2\xi_2^2+4\nu_2\xi_1\xi_3+6\nu_3\xi_1^2\xi_2+\nu_4\xi_1^4,$$

where ν_j and ξ_j denote the j–th ($j=1,2,3,4$) cumulants of N and x_i, respectively.

When the summands are symmetric random vectors and the number of summands has a symmetric distribution the fourth cumulant of $\mathbf{s}$ has a simpler form. This result is again the direct consequence of Theorem 2.1.

Corollary 2.2. *Let N be a nonnegative counting variable with a symmetric distribution. Also, let $\mathbf{x}_1,\dots,\mathbf{x}_N$ be i.i.d. symmetric random vectors with a finite fourth moment, independent of N. Then the fourth cumulant of $\mathbf{s}$ is*

$$\mathbf{K}_4(\mathbf{s})=\nu_1\boldsymbol{\xi}_4+\nu_2\Big[\Big(\mathbf{I}_{d^2}+\mathbf{K}_{d,d}\Big)\big(\boldsymbol{\xi}_2\otimes\boldsymbol{\xi}_2\big)+\mathrm{vec}(\boldsymbol{\xi}_2)\mathrm{vec}(\boldsymbol{\xi}_2)^{\mathrm{T}}\Big]+\nu_4\Big(\boldsymbol{\xi}_1\boldsymbol{\xi}_1^{\mathrm{T}}\otimes\boldsymbol{\xi}_1\boldsymbol{\xi}_1^{\mathrm{T}}\Big).$$

A nonrandom sum of random vectors is just a random sum where the number of summands N is a degenerate random variable: $\mathbb{P}(N=n)=1$. It follows that $\nu_1=n$, $\nu_2=\nu_3=\nu_4=0$ and the fourth cumulant of the random sum is just the product of the first cumulant of the number of summands and the fourth cumulant of a summand: $\mathbf{K}_4(\mathbf{s})=\nu_1\boldsymbol{\xi}_4$, consistently with ordinary properties of cumulants.

3. Mardia's kurtosis

The kurtosis and the excess kurtosis of a random variable x with mean μ, variance σ^2 and finite fourth moment $\mathrm{E}(x^4)$ are often measured by its fourth standardized moment and its fourth standardized cumulant:

$$\beta_2=\mathrm{E}\left[\left(\frac{x-\mu}{\sigma}\right)^4\right],\ \gamma_2=\beta_2-3.$$

As remarked by Kollo (2008), the default measures of multivariate kurtosis and multivariate excess kurtosis have been proposed in Mardia (1970) and Mardia (1974). Mardia's kurtosis and Mardia's excess kurtosis of a d-dimensional random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$, non-singular covariance matrix $\boldsymbol{\Sigma}$ and finite fourth-order moments are given by

$$\beta_{2,d}^M(\mathbf{x})=\mathbb{E}\left\{\left[(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]^2\right\}\quad\text{and}\quad\gamma_{2,d}^M(\mathbf{x})=\beta_{2,d}^M(\mathbf{x})-d(d+2).$$

Kollo and Srivastava (2004) first represented Mardia's measures by means of moment matrices. More precisely, they showed that Mardia's kurtosis and Mardia's excess kurtosis of a random vector are the traces of its fourth standardized moment and its fourth standardized cumulant. Following the moment matrices approach in Kollo and Srivastava (2004), we represent Mardia's kurtosis and Mardia's kurtosis of a random vector as quadratic forms involving the concentration matrix as well as the fourth central moment and the fourth cumulant. This statement is made more precise in the next theorem.

Lemma 3.1. *Let $\mathbf{\Sigma}$, $\overline{\mathbf{M}}_4(\mathbf{x})$ and $\mathbf{K}_4(\mathbf{x})$ be the variance, the fourth central moment and the fourth cumulant of the d-dimensional random vector $\mathbf{x}$. Then Mardia's kurtosis and Mardia's excess kurtosis are*

$$\beta^M_{2,d}(\mathbf{x}) = \mathrm{vec}(\mathbf{\Sigma}^{-1})^{\mathrm{T}}\overline{\mathbf{M}}_4(\mathbf{x})\mathrm{vec}(\mathbf{\Sigma}^{-1}) \quad \text{and} \quad \gamma^M_{2,d}(\mathbf{x}) = \mathrm{vec}(\mathbf{\Sigma}^{-1})^{\mathrm{T}}\mathbf{K}_4(\mathbf{x})\mathrm{vec}(\mathbf{\Sigma}^{-1}).$$

4. Asymptotic behaviour

In this section we derive a closed-form expression for the Mardia's kurtosis of the random sums of normal random vectors. It is a simple function of the first four cumulants of the aggregating variable, the Mahalanobis distance of the mean from the origin and the vector's dimension. We use this result to show that the same standardized random sum might not converge to a normal distribution, even when it is embedded in a sequence of stochastically increasing aggregating variables. For the sake of simplicity, the theorem is stated for random vectors having the identity matrix as the covariance matrix, but it can be straightforwardly generalized to any positive definite covariance matrix.

Theorem 4.1. *Let ν_i be the i-th cumulant of the nonnegative random variable N, for $i = 1, 2, 3, 4$. Also, let $\mathbf{s} = \mathbf{x}_1 + \ldots + \mathbf{x}_N$ be a random sum of independent, d-dimensional normal random vectors $\mathbf{x}_i \sim \mathcal{N}_d(\boldsymbol{\mu}, \mathbf{I}_d)$, where $\boldsymbol{\mu} \in \mathbb{R}^d$. Then, by letting $q = \nu_2\left(\nu_1^2 + \nu_1\nu_2\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^{-1}$ the Mardia's kurtosis of $\mathbf{s}$ is*

$$\beta_{2,d}(\mathbf{s}) = \left(\nu_2 + \nu_1^2\right)\left[2\left(\frac{d-1}{\nu_1^2} + \frac{q^2}{\nu_2^2}\right) + \left(\frac{d}{\nu_1} - q\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2\right] + (\nu_3 + \nu_1\nu_2)\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)\left(\frac{2d-2}{\nu_1^2} + \frac{5q^2}{\nu_2^2}\right) + \left(\nu_4 + 3\nu_2^2\right)\frac{q^2}{\nu_2^2}\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2.$$

Theorem 4.1 provides some insights into the asymptotic behaviour of random sums. Let $\{N_q, q \in \mathbb{N}\}$ be a stochastically increasing sequence of random variables whose support is the set of nonnegative integers and are independent of $\{\mathbf{x}_i \sim \mathcal{N}_d(\boldsymbol{\xi}, \mathbf{\Psi})\}$, where $i \in \mathbb{N}$, $\boldsymbol{\xi} \in \mathbb{R}^d$, $\mathbf{\Psi} \in \mathbb{R}^d \times \mathbb{R}^d$, $\mathbf{\Psi} = \mathbf{\Psi}^{\mathrm{T}} > 0$. Also, let $\boldsymbol{\mu}_q$ and $\mathbf{\Sigma}_q$ be the mean vector and the positive definite covariance matrix of the q-th element $\mathbf{s}_q$ in the sequence $\{\mathbf{s}_q = \mathbf{x}_1 + \ldots + \mathbf{x}_{N_q}, q \in \mathbb{N}\}$. Finally let $\{\mathbf{z}_q = \mathbf{\Sigma}_q^{-1/2}(\mathbf{s}_q - \boldsymbol{\mu}_q), q \in \mathbb{N}\}$ be the associated sequence of standardized random vectors, where $\mathbf{\Sigma}_q^{-1/2}$ is the positive definite square root of the inverse of $\mathbf{\Sigma}_q$. A necessary, but not sufficient condition for the convergence of $\{\mathbf{z}_q, q \in \mathbb{N}\}$ to a standard normal distribution is the convergence of the sequence of Mardia's kurtoses $\{\beta_{2,d}(\mathbf{x}_1 + \ldots + \mathbf{x}_{N_q})\}$ to $d(d+2)$, that is the Mardia's kurtosis of a d -dimensional normal vector with positive definite covariance matrix. This condition is fullfilled if $\boldsymbol{\xi}$ is a null vector. The following corollary shows that $\{\beta_{2,d}(\mathbf{x}_1 + \cdots + \mathbf{x}_{N_q})\}$ might not converge to $d(d+2)$ if $\boldsymbol{\xi}$ is not a null vector, thus highlighting some differences between the convergence of random and nonrandom sums.

Corollary 4.1. *Let $\{\mathbf{x}_i \sim \mathcal{N}_d(\boldsymbol{\mu} \neq \mathbf{0}_d, \mathbf{I}_d)\}$ and $\{a_i\}$ be a random sequence of mutually independent random vectors and a strictly increasing sequence of positive integers, for $i = 1, 2, \ldots$. Also, let $\beta_{2,d}(\mathbf{w})$ be the Mardia's kurtosis of a d-dimensional random vector $\mathbf{w}$. Finally, let $\{N_i = Y + a_i\}$ and $\{M_i = a_iY\}$, where $P(Y = 1) = P(Y = 2) = 0.5$. Then*

$$\lim_{i\longrightarrow+\infty}\beta_{2,d}\left(\mathbf{x}_1 + \ldots + \mathbf{x}_{N_i}\right) = d(d+2) \text{ and } \lim_{i\longrightarrow+\infty}\beta_{2,d}\left(\mathbf{x}_1 + \ldots + \mathbf{x}_{M_i}\right) = +\infty.$$

5. Edgeworth expansions

When the normal approximation to a given distribution is not satisfactory we can improve it by means of Edgeworth expansions. We can approximate either the characteristic function or the probability density function of a random vector by means of its moments or cumulants (Kollo and von Rosen, 2005). In particular, from Corollary 2.1.5.1 and Theorem 2.1.7 of Kollo and von Rosen (2005), the cumulant generating function and characteristic function of $\mathbf{s}$ can be expressed as

$$\psi_{\mathbf{s}}(\mathbf{t}) = \sum_{k=1}^{n}\frac{\iota^k}{k!}(\mathbf{t}^{\mathrm{T}})^{\otimes k}\mathrm{vec}(\mathbf{K}_k(\mathbf{s}))^{\mathrm{T}} + r_n,$$

$$\varphi_{\mathbf{s}}(\mathbf{t}) = 1 + \sum_{k=1}^{n}\frac{\iota^k}{k!}(\mathbf{t}^{\mathrm{T}})^{\otimes k}\mathrm{vec}(\mathbf{M}_k(\mathbf{s}))^{\mathrm{T}} + r_n,$$

respectively, where $\mathbf{t}^{\otimes k}$ stands for $\mathbf{t} \otimes \mathbf{t} \otimes \cdots \otimes \mathbf{t}$, and r_n is the remainder term. The density function of $\mathbf{s}$ can be expressed as

$$f_{\mathbf{s}}(\mathbf{s}) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{-\iota\mathbf{t}^{\mathrm{T}}\mathbf{s}}\varphi_{\mathbf{s}}(\mathbf{t})\mathbf{dt}.$$

Utilizing the Proposition 2.1 of Kollo and Srivastava (2004) and Theorem 1 of Loperfido (2014), we get the following relation between first four cumulants of $\mathbf{s}$:

$$\begin{aligned}
\mathbf{M}_1(\mathbf{s}) &= \mathbf{K}_1(\mathbf{s}),\\
\mathbf{M}_2(\mathbf{s}) &= \mathbf{K}_2(\mathbf{s}) + \mathbf{K}_1(\mathbf{s})\mathbf{K}_1(\mathbf{s})^{\mathrm{T}},\\
\mathbf{M}_3(\mathbf{s}) &= \mathbf{K}_3(\mathbf{s}) + \left(\mathbf{K}_2(\mathbf{s}) + \mathbf{K}_1(\mathbf{s})\mathbf{K}_1(\mathbf{s})^{\mathrm{T}}\right) \otimes \mathbf{K}_1(\mathbf{s}) + \mathbf{K}_1(\mathbf{s}) \otimes \left(\mathbf{K}_2(\mathbf{s}) + \mathbf{K}_1(\mathbf{s})\mathbf{K}_1(\mathbf{s})^{\mathrm{T}}\right)\\
&\quad + \mathrm{vec}\left(\mathbf{K}_2(\mathbf{s}) + \mathbf{K}_1(\mathbf{s})\mathbf{K}_1(\mathbf{s})^{\mathrm{T}}\right)\mathbf{K}_{1,\mathbf{s}}^{\mathrm{T}} - 2\mathbf{K}_1(\mathbf{s})\mathbf{K}_1(\mathbf{s})^{\mathrm{T}} \otimes \mathbf{K}_1(\mathbf{s}),\\
\mathbf{M}_4(\mathbf{s}) &= \mathbf{K}_4(\mathbf{s}) + (\mathbf{I}_{d^2} + \mathbf{K}_{p,p})\left[\left(\mathbf{K}_2(\mathbf{s}) + \mathbf{K}_1(\mathbf{s})\mathbf{K}_1(\mathbf{s})^{\mathrm{T}}\right) \otimes \left(\mathbf{K}_2(\mathbf{s}) + \mathbf{K}_{1,\mathbf{s}}\mathbf{K}_{1,\mathbf{s}}^{\mathrm{T}}\right)\right] + \mathrm{vec}\left(\mathbf{K}_2(\mathbf{s}) + \mathbf{K}_1(\mathbf{s})\mathbf{K}_1(\mathbf{s})^{\mathrm{T}}\right)\\
&\quad \times \mathrm{vec}\left(\mathbf{K}_2(\mathbf{s}) + \mathbf{K}_1(\mathbf{s})\mathbf{K}_1(\mathbf{s})^{\mathrm{T}}\right)^{\mathrm{T}} + \mathbf{M}_3(\mathbf{s})^{\mathrm{T}} \otimes \mathbf{K}_1(\mathbf{s}) + \mathbf{M}_3(\mathbf{s}) \otimes \mathbf{K}_1(\mathbf{s})^{\mathrm{T}} + \mathbf{K}_1(\mathbf{s})^{\mathrm{T}} \otimes \mathbf{M}_3(\mathbf{s})\\
&\quad + \mathbf{K}_1(\mathbf{s}) \otimes \mathbf{M}_3(\mathbf{s})^{\mathrm{T}} - \left(\mathbf{K}_2(\mathbf{s}) + \mathbf{K}_1(\mathbf{s})\mathbf{K}_1(\mathbf{s})^{\mathrm{T}}\right) \otimes \mathbf{K}_1(\mathbf{s})\mathbf{K}_1(\mathbf{s})^{\mathrm{T}} - \mathrm{vec}\left(\mathbf{K}_2(\mathbf{s}) + \mathbf{K}_1(\mathbf{s})\mathbf{K}_1(\mathbf{s})^{\mathrm{T}}\right) \otimes \mathbf{K}_1(\mathbf{s})^{\mathrm{T}} \otimes \mathbf{K}_1(\mathbf{s})^{\mathrm{T}}\\
&\quad - \mathbf{K}_{d,d}\left(\mathbf{K}_1(\mathbf{s})\mathbf{K}_1(\mathbf{s})^{\mathrm{T}} \otimes \left(\mathbf{K}_2(\mathbf{s}) + \mathbf{K}_1(\mathbf{s})\mathbf{K}_1(\mathbf{s})^{\mathrm{T}}\right)\right) - \mathbf{K}_{d,d}\left(\left(\mathbf{K}_2(\mathbf{s}) + \mathbf{K}_1(\mathbf{s})\mathbf{K}_1(\mathbf{s})^{\mathrm{T}}\right) \otimes \mathbf{K}_1(\mathbf{s})\mathbf{K}_1(\mathbf{s})^{\mathrm{T}}\right)\\
&\quad - \mathbf{K}_1(\mathbf{s}) \otimes \mathbf{K}_1(\mathbf{s}) \otimes \mathrm{vec}\left(\mathbf{K}_2(\mathbf{s}) + \mathbf{K}_1(\mathbf{s})\mathbf{K}_1(\mathbf{s})^{\mathrm{T}}\right)^{\mathrm{T}} - \mathbf{K}_1(\mathbf{s})\mathbf{K}_1(\mathbf{s})^{\mathrm{T}} \otimes \left(\mathbf{K}_2(\mathbf{s}) + \mathbf{K}_1(\mathbf{s})\mathbf{K}_1(\mathbf{s})^{\mathrm{T}}\right)\\
&\quad + 3\mathbf{K}_1(\mathbf{s})\mathbf{K}_1(\mathbf{s})^{\mathrm{T}} \otimes \mathbf{K}_1(\mathbf{s})\mathbf{K}_1(\mathbf{s})^{\mathrm{T}}.
\end{aligned}$$

Since the first three cumulants of $\mathbf{s}$ are known from Loperfido et al. (2018) and the fourth cumulant is delivered in Theorem 2.1, we get the following approximations of the characteristic and density functions

$$\varphi_{\mathbf{s}}(\mathbf{t}) \approx 1 + \sum_{k=1}^{4} \frac{\iota^k}{k!}(\mathbf{t}^{\mathrm{T}})^{\otimes k}\mathrm{vec}(\mathbf{M}_k(\mathbf{s}))^{\mathrm{T}},$$

$$f_{\mathbf{s}}(\mathbf{s}) \approx \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{-\iota\mathbf{t}^{\mathrm{T}}\mathbf{s}}\left(1 + \sum_{k=1}^{4} \frac{\iota^k}{k!}(\mathbf{t}^{\mathrm{T}})^{\otimes k}\mathrm{vec}(\mathbf{M}_k(\mathbf{s}))^{\mathrm{T}}\right)\mathbf{dt}.$$

The density function of $\mathbf{s}$ depends on d integrals, therefore, it is more natural to simplify it further. In particular, we consider the Edgeworth expansions which are based on simpler density function, say $f_{\mathbf{y}}(\mathbf{y})$, with finite fourth cumulant. Following Theorem 3.2.1 of Kollo and von Rosen (2005), we obtain the density function of $\mathbf{s}$ which is stated in the next theorem.

Theorem 5.1. *Under the assumption of Theorem 2.1, the density function $f_{\mathbf{s}}(\mathbf{y})$ can be represented via the density function $f_{\mathbf{y}}(\mathbf{y})$ as follows*

$$\begin{aligned}
f_{\mathbf{s}}(\mathbf{y}) &\approx f_{\mathbf{y}}(\mathbf{y}) - (\mathbf{K}_1(\mathbf{s}) - \mathbf{K}_1(\mathbf{y}))^{\mathrm{T}} f_{\mathbf{y}}^1(\mathbf{y}) + \frac{1}{2}\mathrm{vec}\left[\mathbf{K}_2(\mathbf{s}) - \mathbf{K}_2(\mathbf{y}) + (\mathbf{K}_1(\mathbf{s}) - \mathbf{K}_1(\mathbf{y}))(\mathbf{K}_1(\mathbf{s}) - \mathbf{K}_1(\mathbf{y}))^{\mathrm{T}}\right]^{\mathrm{T}}\mathrm{vec}\left(f_{\mathbf{y}}^2(\mathbf{y})\right)\\
&\quad - \frac{1}{6}\left[\mathrm{vec}(\mathbf{K}_3(\mathbf{s}) - \mathbf{K}_3(\mathbf{y}))^{\mathrm{T}} + 3\mathrm{vec}(\mathbf{K}_2(\mathbf{s}) - \mathbf{K}_2(\mathbf{y}))^{\mathrm{T}} \otimes (\mathbf{K}_1(\mathbf{s}) - \mathbf{K}_1(\mathbf{y}))^{\mathrm{T}} + (\mathbf{K}_1(\mathbf{s}) - \mathbf{K}_1(\mathbf{y}))^{\mathrm{T}\otimes 3}\right]\mathrm{vec}\left(f_{\mathbf{y}}^3(\mathbf{y})\right)\\
&\quad + \frac{1}{24}\Big\{\mathrm{vec}(\mathbf{K}_4(\mathbf{s}) - \mathbf{K}_4(\mathbf{y}))^{\mathrm{T}} + (\mathbf{K}_1(\mathbf{s}) - \mathbf{K}_1(\mathbf{y}))^{\mathrm{T}\otimes 4} + 6\mathrm{vec}((\mathbf{K}_2(\mathbf{s}) - \mathbf{K}_2(\mathbf{y})) \otimes (\mathbf{K}_1(\mathbf{s}) - \mathbf{K}_1(\mathbf{y}))\\
&\quad \times (\mathbf{K}_1(\mathbf{s}) - \mathbf{K}_1(\mathbf{y}))^{\mathrm{T}}\Big)^{\mathrm{T}} + 4\mathrm{vec}\left((\mathbf{K}_3(\mathbf{s}) - \mathbf{K}_3(\mathbf{y})) \otimes (\mathbf{K}_1(\mathbf{s}) - \mathbf{K}_1(\mathbf{y}))^{\mathrm{T}}\right)^{\mathrm{T}}\Big\}\mathrm{vec}\left(f_{\mathbf{y}}^4(\mathbf{y})\right),
\end{aligned}$$

where $f_{\mathbf{y}}^k(\mathbf{y}) = \frac{d^k f_{\mathbf{y}}(\mathbf{y})}{d\mathbf{y}^k}$, $k = 1, 2, 3, 4$.

The above mentioned Edgeworth expansion is based on the density function of $\mathbf{y}$. Usually $\mathbf{y}$ is taken to be normally distributed, i.e. $\mathbf{y} \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma})$. The main reason for that is based on the fact that many random statistics can be well approximated by normal distribution in the asymptotic setting. Therefore, the density function of $\mathbf{s}$ can be written via the density function of normal distribution by the Edgeworth expansion delivered in Theorem 5.1.

As remarked in Gupta and Kollo (2003), Edgeworth expansions might be used to approximate a distribution with a nonnormal one with the same first cumulants. Loperfido (2019) showed that for any random sum with Poisson agregating variable and skew-normal summands there is a mixture of two normal distributions with proportional covariance matrices having its same first three cumulants. These theoretical results suggest that the latter distribution might satisfactorily fit data generated from the former one. In order to empirically assess this conjecture we simulated 100 units from the random sum where the number of summands is Poisson with mean 2 and the summands are i.i.d. skew-normal random vectors with the null vector as the location parameter, the identity matrix as the scatter parameter and the unit vector as the shape parameter: $\mathbf{x}_i \sim SN_d(\mathbf{0}_d, \mathbf{I}_d, \mathbf{1}_d)$, for $d = 2, 3, 4, 5, 6, 7, 8, 9, 10$. The pdf of $\mathbf{x}_i = (x_{i1}, ..., x_{id})^{\mathrm{T}}$ is

$$f(x_{i1}, ..., x_{id}) = 2\Phi\left(\sum_{j=1}^{d} x_{ij}\right)\prod_{j=1}^{d}\phi\left(x_{ij}\right),$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the pdf and the cdf of a standard normal distribution, respectively. The simulated data are available as supplementary material to this paper. We used the R package (Scrucca et al., 2016) to fit a mixture model using the BIC criterion. For all the nine datasets, the chosen model was the mixture with two normal components and proportional covariance matrices. This empirical result is consistent with the theoretical ones.

6. Concluding remarks

This paper derives the fourth cumulant of a multivariate random sum and highlights some difficulties in the asymptotic approximation of its distribution with a multivariate normal one. The paper addresses the problem by means of Edgeworth expansions. For example, the random sum of i.i.d. skew-normal random vectors with the Poisson distribution as the aggregating variable has the same first three cumulants of a mixture of two normal distributions with proportional covariance matrices, for appropriate choice of parameters. The paper shows that data generated from the former distribution might be satisfactorily fitted by the latter.

The theoretical results in this paper are probabilistic in nature, but they also have potential statistical applications. For example, the estimation of a covariance matrix is often of interest in both Statistics and Econometrics (Braione M. and Storti, 2017); (Morana, 2019) and its default estimate is the sample covariance matrix. The covariance of the vectorized sample covariance matrix is closely related to the fourth centered moment and the fourth cumulant of the sampled distribution (Loperfido, 2011). Unfortunately, these matrices might contain up to $p(p+1)(p+2)(p+3)/24$ distinct elements, thus hampering the estimation procedure. The results in this paper might ease the problem when applied to a parametric model for aggregate claims.

Fourth-order moments are well-known to be very sensitive to outliers, which is a blessing as well as a curse. It is a curse when unnoticed outliers hamper the usage of Edgeworth expansions based on fourth-order moments. It is a blessing when using kurtosis measures based on fourth-order moments to detect outliers (see, for example, Livesey (2007)). The theoretical results in this paper pave the way in this direction, using an approach which might be informally described as follows. First, assume a parametric model for both the claim number and the claim size, as done in Loperfido (2019). Second, use the results in this paper to obtain the analytical formula of a multivariate kurtosis measure, as for example Mardia's kurtosis. Third, obtain robust estimates of the parameters and plug them into the analytical formula of this measure. Fourth, compare this estimate with the sample counterpart of the kurtosis measure. Outliers will make the latter statistic much bigger than the former one. The statistical significance of their difference hints at the presence of outliers. The plug-in estimate of the fourth cumulant might lead the robust and accurate density expansions sought by Ronchetti (2020).

It would be interesting to know whether convergence to normality of a multivariate random sum holds for well-known choices of the summands and their number. A first choice could be the random sum of skew-normal random vectors with a Poisson random variable as the aggregating variable. A second choice could be the random sum of asymmetric generalized Laplace random vectors with a negative binomial random variable as the aggregating variable. The third cumulants of both choices have been derived by Loperfido et al. (2018). We are currently working on the derivation of their fourth cumulants.

In this paper, the fourth cumulant of a multivariate random sum is derived under several assumptions, as for example the aggregating variable and the summands being independent. More general assumptions are discussed in Cummins and Wiltbank (1983) and Loperfido et al. (2018). In particular, the aggregating mechanism might be modelled by a random vector, rather than by a random variable. At present time, the derivation of the fourth cumulant of a multivariate random sum under these more general assumptions appers to be quite a formidable task.

Acknowledgements

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

Farrukh Javed and Stepan Mazur acknowledge financial support from the internal research grants at Örebro University and from the project "Models for macro and financial economics after the financial crisis" (Dnr: P18-0201) funded by Jan Wallander and Tom Hedelius Foundation.

Proof of Theorem 2.1.. The mean of $\mathbf{s}$ is equal to $\nu_1\boldsymbol{\xi}_1$. Then the fourth cumulant of $\mathbf{s}$ is given by

$$\mathbf{K}_4(\mathbf{s}) = \underbrace{\mathbb{E}\left[\left(\mathbf{s}-\nu_1\boldsymbol{\xi}_1\right)\otimes\left(\mathbf{s}-\nu_1\boldsymbol{\xi}_1\right)^{\mathrm{T}}\otimes\left(\mathbf{s}-\nu_1\boldsymbol{\xi}_1\right)\otimes\left(\mathbf{s}-\nu_1\boldsymbol{\xi}_1\right)^{\mathrm{T}}\right]}_{\boldsymbol{\Theta}_1} - \underbrace{\left(\mathbf{I}_{d^2}+\mathbf{K}_{d,d}\right)(Var[\mathbf{s}]\otimes Var[\mathbf{s}])}_{\boldsymbol{\Theta}_2} - \underbrace{vec(Var[\mathbf{s}])vec(Var[\mathbf{s}])^{\mathrm{T}}}_{\boldsymbol{\Theta}_3}. \tag{6.1}$$

The first component of the expression may be represented as

$$\boldsymbol{\Theta}_1 = \mathbb{E}\left[(\mathbf{w}+\mathbf{u})\otimes(\mathbf{w}+\mathbf{u})^{\mathrm{T}}\otimes(\mathbf{w}+\mathbf{u})\otimes(\mathbf{w}+\mathbf{u})^{\mathrm{T}}\right]$$

with $\mathbf{w} = (\mathbf{x}_1 - \boldsymbol{\xi}_1) + \ldots + (\mathbf{x}_N - \boldsymbol{\xi}_1)$ and $\mathbf{u} = \boldsymbol{\xi}_1(N - \nu_1)$. Using the standard properties of the Kronecker product (Harville, 1997, Chapter 16) we obtain that

$$\begin{aligned}
\boldsymbol{\Theta}_1 &= \mathbb{E}\Big[\big(\mathbf{w}\otimes\mathbf{w}^T + \mathbf{w}\otimes\mathbf{u}^T + \mathbf{u}\otimes\mathbf{w}^T + \mathbf{u}\otimes\mathbf{u}^T\big)\otimes\big(\mathbf{w}\otimes\mathbf{w}^T + \mathbf{w}\otimes\mathbf{u}^T + \mathbf{u}\otimes\mathbf{w}^T + \mathbf{u}\otimes\mathbf{u}^T\big)\Big]\\
&= \mathbb{E}\Big[\big(\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{w}\otimes\mathbf{w}^T\big) + \big(\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{w}\otimes\mathbf{u}^T\big) + \big(\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{u}\otimes\mathbf{w}^T\big) + \big(\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{u}\otimes\mathbf{u}^T\big)\\
&\quad + \big(\mathbf{w}\otimes\mathbf{u}^T\otimes\mathbf{w}\otimes\mathbf{w}^T\big) + \big(\mathbf{w}\otimes\mathbf{u}^T\otimes\mathbf{w}\otimes\mathbf{u}^T\big) + \big(\mathbf{w}\otimes\mathbf{u}^T\otimes\mathbf{u}\otimes\mathbf{w}^T\big) + \big(\mathbf{w}\otimes\mathbf{u}^T\otimes\mathbf{u}\otimes\mathbf{u}^T\big)\\
&\quad + \big(\mathbf{u}\otimes\mathbf{w}^T\otimes\mathbf{w}\otimes\mathbf{w}^T\big) + \big(\mathbf{u}\otimes\mathbf{w}^T\otimes\mathbf{w}\otimes\mathbf{u}^T\big) + \big(\mathbf{u}\otimes\mathbf{w}^T\otimes\mathbf{u}\otimes\mathbf{w}^T\big) + \big(\mathbf{u}\otimes\mathbf{w}^T\otimes\mathbf{u}\otimes\mathbf{u}^T\big)\\
&\quad + \big(\mathbf{u}\otimes\mathbf{u}^T\otimes\mathbf{w}\otimes\mathbf{w}^T\big) + \big(\mathbf{u}\otimes\mathbf{u}^T\otimes\mathbf{w}\otimes\mathbf{u}^T\big) + \big(\mathbf{u}\otimes\mathbf{u}^T\otimes\mathbf{u}\otimes\mathbf{w}^T\big) + \big(\mathbf{u}\otimes\mathbf{u}^T\otimes\mathbf{u}\otimes\mathbf{u}^T\big)\Big].
\end{aligned}$$

Now we apply expectation on each term in the bracket

$$\begin{aligned}
\boldsymbol{\Theta}_1 &= \mathbb{E}\big[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{w}\otimes\mathbf{w}^T\big] + \mathbb{E}\big[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{w}\otimes\mathbf{u}^T\big] + \mathbb{E}\big[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{u}\otimes\mathbf{w}^T\big] + \mathbb{E}\big[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{u}\otimes\mathbf{u}^T\big]\\
&\quad + \mathbb{E}\big[\mathbf{w}\otimes\mathbf{u}^T\otimes\mathbf{w}\otimes\mathbf{w}^T\big] + \mathbb{E}\big[\mathbf{w}\otimes\mathbf{u}^T\otimes\mathbf{w}\otimes\mathbf{u}^T\big] + \mathbb{E}\big[\mathbf{w}\otimes\mathbf{u}^T\otimes\mathbf{u}\otimes\mathbf{w}^T\big] + \mathbb{E}\big[\mathbf{w}\otimes\mathbf{u}^T\otimes\mathbf{u}\otimes\mathbf{u}^T\big]\\
&\quad + \mathbb{E}\big[\mathbf{u}\otimes\mathbf{w}^T\otimes\mathbf{w}\otimes\mathbf{w}^T\big] + \mathbb{E}\big[\mathbf{u}\otimes\mathbf{w}^T\otimes\mathbf{w}\otimes\mathbf{u}^T\big] + \mathbb{E}\big[\mathbf{u}\otimes\mathbf{w}^T\otimes\mathbf{u}\otimes\mathbf{w}^T\big] + \mathbb{E}\big[\mathbf{u}\otimes\mathbf{w}^T\otimes\mathbf{u}\otimes\mathbf{u}^T\big)\\
&\quad + \mathbb{E}\big[\mathbf{u}\otimes\mathbf{u}^T\otimes\mathbf{w}\otimes\mathbf{w}^T\big] + \mathbb{E}\big[\mathbf{u}\otimes\mathbf{u}^T\otimes\mathbf{w}\otimes\mathbf{u}^T\big] + \mathbb{E}\big[\mathbf{u}\otimes\mathbf{u}^T\otimes\mathbf{u}\otimes\mathbf{w}^T\big] + \mathbb{E}\big[\mathbf{u}\otimes\mathbf{u}^T\otimes\mathbf{u}\otimes\mathbf{u}^T\big].
\end{aligned}$$

First, we evaluate $\mathbb{E}\big[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{w}\otimes\mathbf{w}^T\big]$. Since $\{\mathbf{x}_i\}_{i=1}^N$ and N are independently distributed we get that

$$\mathbb{E}\big[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{w}\otimes\mathbf{w}^T|N=n\big] = \mathbb{E}\left[\sum_{i=1}^n \mathbf{y}_i\otimes\sum_{i=1}^n \mathbf{y}_i^T\otimes\sum_{i=1}^n \mathbf{y}_i\otimes\sum_{i=1}^n \mathbf{y}_i^T|N=n\right],$$

where $\mathbf{y}_i = \mathbf{x}_i - \boldsymbol{\xi}_1$, $i = 1, \ldots, n$. Because $\mathbf{y}_1, \ldots, \mathbf{y}_n$ are independent and identically distributed, it holds that

$$\mathbb{E}\left[\sum_{i=1}^n \mathbf{y}_i\otimes\sum_{i=1}^n \mathbf{y}_i^T\otimes\sum_{i=1}^n \mathbf{y}_i\otimes\sum_{i=1}^n \mathbf{y}_i^T|N=n\right] = \overline{\mathbf{M}}_4\left(\sum_{i=1}^n \mathbf{x}_i|N=n\right).$$

By noticing that

$$\overline{\mathbf{M}}_4\left(\sum_{i=1}^n \mathbf{x}_i|N=n\right) = n\boldsymbol{\xi}_4 + n^2\big[(\mathbf{I}_{d^2} + \mathbf{K}_{d,d})(\boldsymbol{\xi}_2\otimes\boldsymbol{\xi}_2) + \mathrm{vec}(\boldsymbol{\xi}_2)\mathrm{vec}(\boldsymbol{\xi}_2)^T\big],$$

and by taking expectations over N we obtain that

$$\mathbb{E}\big[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{w}\otimes\mathbf{w}^T\big] = \nu_1\boldsymbol{\xi}_4 + (\nu_2 + \nu_1^2)\big[(\mathbf{I}_{d^2} + \mathbf{K}_{d,d})(\boldsymbol{\xi}_2\otimes\boldsymbol{\xi}_2) + \mathrm{vec}(\boldsymbol{\xi}_2)\mathrm{vec}(\boldsymbol{\xi}_2)^T\big].$$

Next, we evaluate $\mathbb{E}\big[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{w}\otimes\mathbf{u}^T\big]$ starting from the conditional expectation. Applying the properties of expected values and Kronecker products we get

$$\mathbb{E}\big[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{w}\otimes\mathbf{u}^T|N=n\big] = (n-\nu_1)\mathbb{E}\left[\sum_{i=1}^n \mathbf{y}_i\otimes\sum_{i=1}^n \mathbf{y}_i^T\otimes\sum_{i=1}^n \mathbf{y}_i|N=n\right]\otimes\boldsymbol{\xi}_1^T.$$

The above expectation is the third cumulant of the sum $\mathbf{x}_1 + \ldots + \mathbf{x}_n$. The random vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are independent and identically distributed, so that the third cumulant of their sum equals to the third cumulant of the i-th summand $\mathbf{x}_i$, multiplied by the number of summands n:

$$\mathbb{E}\big[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{w}\otimes\mathbf{u}^T|N=n\big] = n(n-\nu_1)\boldsymbol{\xi}_3\otimes\boldsymbol{\xi}_1^T.$$

By taking expectations over N we obtain

$$\mathbb{E}\big[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{w}\otimes\mathbf{u}^T\big] = \mathbb{E}\big[\mathbb{E}\big[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{w}\otimes\mathbf{u}^T|N=n\big]\big] = \mathbb{E}[N(N-\nu_1)]\boldsymbol{\xi}_3\otimes\boldsymbol{\xi}_1^T.$$

The expectation in the right-hand side of the above equation is just the variance (that is the second cumulant) of N. Hence, we obtain

$$\mathbb{E}\big[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{w}\otimes\mathbf{u}^T\big] = \nu_2\boldsymbol{\xi}_3\otimes\boldsymbol{\xi}_1^T.$$

Now we evaluate $\mathbb{E}\big[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{u}\otimes\mathbf{w}^T\big]$, starting from the identity

$$\mathbb{E}\big[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{u}\otimes\mathbf{w}^T|N=n\big] = \mathbb{E}\left[\sum_{i=1}^n \mathbf{y}_i\otimes\sum_{i=1}^n \mathbf{y}_i^T\otimes\boldsymbol{\xi}_1(n-\nu_1)\otimes\sum_{i=1}^n \mathbf{y}_i^T|N=n\right],$$

which may be simplified as follows, by remembering that $\mathbf{a}\otimes\mathbf{b}^T=\mathbf{b}^T\otimes\mathbf{a}$, for any two vectors $\mathbf{a}$ and $\mathbf{b}$:

$$\mathbb{E}\left[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{u}\otimes\mathbf{w}^T|N=n\right]=\mathbb{E}\left[\sum_{i=1}^{n}\mathbf{y}_i^T\otimes\sum_{i=1}^{n}\mathbf{y}_i\otimes\sum_{i=1}^{n}\mathbf{y}_i^T\otimes\boldsymbol{\xi}_1(n-\nu_1)|N=n\right].$$

The above expectation is the third cumulant of the sum $\mathbf{x}_1+\ldots+\mathbf{x}_n$. An argument similar to the one used before leads to

$$\mathbb{E}\left[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{u}\otimes\mathbf{w}^T\right]=\mathbb{E}\left[\mathbb{E}\left[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{u}\otimes\mathbf{w}^T|N=n\right]\right]=\nu_2\boldsymbol{\xi}_3^T\otimes\boldsymbol{\xi}_1.$$

Next, we consider the identity,

$$\mathbb{E}\left[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{u}\otimes\mathbf{u}^T|N=n\right]=\mathbb{E}\left[\sum_{i=1}^{n}\mathbf{y}_i\otimes\sum_{i=1}^{n}\mathbf{y}_i^T|N=n\right]\otimes\boldsymbol{\xi}_1(n-\nu_1)\otimes\boldsymbol{\xi}_1^T(n-\nu_1).$$

The above expectation is the second cumulant of the sum $\mathbf{x}_1+\ldots+\mathbf{x}_n$, therefore, it leads to

$$\mathbb{E}\left[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{u}\otimes\mathbf{u}^T\right]=\mathbb{E}\left[\mathbb{E}\left[\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{u}\otimes\mathbf{u}^T|N=n\right]\right]=(\nu_3+\nu_1\nu_2)\boldsymbol{\xi}_2\otimes\boldsymbol{\xi}_1\otimes\boldsymbol{\xi}_1^T,$$

where the last identity follows from the relation between cumulants and moments (Stuart and Ord, 1994, Chapter 3.12).

In order to evaluate $\mathbb{E}\left[\mathbf{w}\otimes\mathbf{u}^T\otimes\mathbf{w}\otimes\mathbf{w}^T\right]$, we recall that $\mathbf{w}^T\otimes\mathbf{w}=\mathbf{w}\mathbf{w}^T=\mathbf{w}\otimes\mathbf{w}^T$ and use arguments similar to the above ones to obtain

$$\mathbb{E}\left[\mathbf{w}\otimes\mathbf{u}^T\otimes\mathbf{w}\otimes\mathbf{w}^T\right]=\mathbb{E}\left[\mathbf{u}^T\otimes\mathbf{w}\otimes\mathbf{w}\otimes\mathbf{w}^T\right]=\mathbb{E}[N(N-\nu_1)]\boldsymbol{\xi}_1^T\otimes\boldsymbol{\xi}_3=\nu_2\boldsymbol{\xi}_1^T\otimes\boldsymbol{\xi}_3.$$

Next, we consider $\mathbb{E}\left[\mathbf{w}\otimes\mathbf{u}^T\otimes\mathbf{w}\otimes\mathbf{u}^T\right]$, which can be simplified as

$$\begin{aligned}\mathbb{E}\left[\mathbf{w}\otimes\mathbf{w}\otimes\mathbf{u}^T\otimes\mathbf{u}^T|N=n\right]&=\mathbb{E}\left[\sum_{i=1}^{n}\mathbf{y}_i\otimes\sum_{i=1}^{n}\mathbf{y}_i\otimes\boldsymbol{\xi}_1^T(n-\nu_1)\otimes\boldsymbol{\xi}_1^T(n-\nu_1)|N=n\right]\\&=\mathbb{E}\left[\sum_{i=1}^{n}\mathbf{y}_i\otimes\sum_{i=1}^{n}\mathbf{y}_i|N=n\right]\otimes\boldsymbol{\xi}_1^T(n-\nu_1)\otimes\boldsymbol{\xi}_1^T(n-\nu_1)\\&=n(n-\nu_1)^2\mathrm{vec}(\boldsymbol{\xi}_2)\otimes\boldsymbol{\xi}_1^T\otimes\boldsymbol{\xi}_1^T,\end{aligned}$$

where we have used the fact that for any vector $\mathbf{a}$, $\mathbf{a}\otimes\mathbf{a}$ is equal to $\mathrm{vec}(\mathbf{a}\mathbf{a}^T)$. Now taking expectation over N would lead to

$$\begin{aligned}\mathbb{E}\left[\mathbf{w}\otimes\mathbf{w}\otimes\mathbf{u}^T\otimes\mathbf{u}^T\right]&=\mathbb{E}\left[\mathbb{E}\left[\mathbf{w}\otimes\mathbf{w}\otimes\mathbf{u}^T\otimes\mathbf{u}^T|N=n\right]\right]\\&=(\nu_3+\nu_1\nu_2)\mathrm{vec}(\boldsymbol{\xi}_2)\otimes\boldsymbol{\xi}_1^T\otimes\boldsymbol{\xi}_1^T.\end{aligned}$$

Next we consider $\mathbb{E}\left[\mathbf{w}\otimes\mathbf{u}^T\otimes\mathbf{u}\otimes\mathbf{w}^T\right]$. Using again the fact that $\mathbf{a}\otimes\mathbf{b}^T=\mathbf{b}^T\otimes\mathbf{a}$, for any two vectors $\mathbf{a}$ and $\mathbf{b}$, we get that

$$\begin{aligned}\mathbb{E}\left[\mathbf{w}\otimes\mathbf{u}^T\otimes\mathbf{u}\otimes\mathbf{w}^T|N=n\right]&=\mathbb{E}\left[\mathbf{u}^T\otimes\mathbf{w}\otimes\mathbf{w}^T\otimes\mathbf{u}|N=n\right]\\&=\mathbb{E}\left[\boldsymbol{\xi}_1^T(n-\nu_1)\otimes\sum_{i=1}^{n}\mathbf{y}_i\otimes\sum_{i=1}^{n}\mathbf{y}_i^T\otimes\boldsymbol{\xi}_1(n-\nu_1)|N=n\right]\\&=n(n-\nu_1)^2\boldsymbol{\xi}_1^T\otimes\boldsymbol{\xi}_2\otimes\boldsymbol{\xi}_1.\end{aligned}$$

Taking expected value over N leads us to

$$\begin{aligned}\mathbb{E}\left[\mathbf{w}\otimes\mathbf{u}^T\otimes\mathbf{u}\otimes\mathbf{w}^T\right]&=\mathbb{E}\left[\mathbb{E}\left[\mathbf{w}\otimes\mathbf{u}^T\otimes\mathbf{u}\otimes\mathbf{w}^T|N=n\right]\right]\\&=(\nu_3+\nu_1\nu_2)\boldsymbol{\xi}_1^T\otimes\boldsymbol{\xi}_2\otimes\boldsymbol{\xi}_1.\end{aligned}$$

We move on to consider $\mathbb{E}\left[\mathbf{w}\otimes\mathbf{u}^T\otimes\mathbf{u}\otimes\mathbf{u}^T\right]$. It holds that

$$\mathbb{E}\left[\mathbf{w}\otimes\mathbf{u}^T\otimes\mathbf{u}\otimes\mathbf{u}^T|N=n\right]=(n-\nu_1)^3\mathbb{E}\left[\sum_{i=1}^{n}\mathbf{y}_i|N=n\right]\otimes\boldsymbol{\xi}_1^T\otimes\boldsymbol{\xi}_1\otimes\boldsymbol{\xi}_1^T=\mathbf{0}.$$

As a direct consequence, we get that $\mathbb{E}[\mathbf{u}\otimes\mathbf{w}^T\otimes\mathbf{u}\otimes\mathbf{u}^T]=\mathbb{E}[\mathbf{u}\otimes\mathbf{u}^T\otimes\mathbf{w}\otimes\mathbf{u}^T]=\mathbb{E}[\mathbf{u}\otimes\mathbf{u}^T\otimes\mathbf{u}\otimes\mathbf{w}^T]=\mathbf{0}$.

Next, we consider $\mathbb{E}[\mathbf{u}\otimes\mathbf{w}^{\mathrm{T}}\otimes\mathbf{w}\otimes\mathbf{w}^{\mathrm{T}}]$. It holds that

$$\mathbb{E}\Big[\mathbf{u}\otimes\mathbf{w}^{\mathrm{T}}\otimes\mathbf{w}\otimes\mathbf{w}^{\mathrm{T}}\Big|N=n\Big]=\mathbb{E}\left[\boldsymbol{\xi}_1(n-\nu_1)\otimes\sum_{i=1}^{n}\mathbf{y}_i^{\mathrm{T}}\otimes\sum_{i=1}^{n}\mathbf{y}_i\otimes\sum_{i=1}^{n}\mathbf{y}_i^{\mathrm{T}}|N=n\right]=n(n-\nu_1)\boldsymbol{\xi}_1\otimes\boldsymbol{\xi}_3^{\mathrm{T}}.$$

Hence, we get

$$\mathbb{E}\Big[\mathbf{u}\otimes\mathbf{w}^{\mathrm{T}}\otimes\mathbf{w}\otimes\mathbf{w}^{\mathrm{T}}\Big]=\mathbb{E}[N(N-\nu_1)]\boldsymbol{\xi}_1\otimes\boldsymbol{\xi}_3^{\mathrm{T}}=\nu_2\boldsymbol{\xi}_1\otimes\boldsymbol{\xi}_3^{\mathrm{T}}.$$

Now we evaluate $\mathbb{E}\Big[\mathbf{u}\otimes\mathbf{w}^{\mathrm{T}}\otimes\mathbf{w}\otimes\mathbf{u}^{\mathrm{T}}\Big]$. We have that

$$\begin{aligned}\mathbb{E}\Big[\mathbf{u}\otimes\mathbf{w}^{\mathrm{T}}\otimes\mathbf{w}\otimes\mathbf{u}^{\mathrm{T}}|N=n\Big]&=\mathbb{E}\left[\boldsymbol{\xi}_1(n-\nu_1)\otimes\sum_{i=1}^{n}\mathbf{y}_i^{\mathrm{T}}\otimes\sum_{i=1}^{n}\mathbf{y}_i\otimes\boldsymbol{\xi}_1^{\mathrm{T}}(n-\nu_1)|N=n\right]\\&=(n-\nu_1)^2\boldsymbol{\xi}_1\otimes\mathbb{E}\left[\sum_{i=1}^{n}\mathbf{y}_i^{\mathrm{T}}\otimes\sum_{i=1}^{n}\mathbf{y}_i|N=n\right]\otimes\boldsymbol{\xi}_1^{\mathrm{T}}\\&=n(n-\nu_1)^2\boldsymbol{\xi}_1\otimes\boldsymbol{\xi}_2\otimes\boldsymbol{\xi}_1^{\mathrm{T}}.\end{aligned}$$

Consequently, we obtain

$$\begin{aligned}\mathbb{E}\Big[\mathbf{u}\otimes\mathbf{w}^{\mathrm{T}}\otimes\mathbf{w}\otimes\mathbf{u}^{\mathrm{T}}\Big]&=\mathbb{E}\Big[\mathbb{E}\Big[\mathbf{u}\otimes\mathbf{w}^{\mathrm{T}}\otimes\mathbf{w}\otimes\mathbf{u}^{\mathrm{T}}|N=n\Big]\Big]\\&=(\nu_3+\nu_1\nu_2)\boldsymbol{\xi}_1\otimes\boldsymbol{\xi}_2\otimes\boldsymbol{\xi}_1^{\mathrm{T}}.\end{aligned}$$

Next, we consider $\mathbb{E}\Big[\mathbf{u}\otimes\mathbf{w}^{\mathrm{T}}\otimes\mathbf{u}\otimes\mathbf{w}^{\mathrm{T}}\Big]$, starting from

$$\begin{aligned}\mathbb{E}\Big[\mathbf{u}\otimes\mathbf{w}^{\mathrm{T}}\otimes\mathbf{u}\otimes\mathbf{w}^{\mathrm{T}}|N=n\Big]&=\mathbb{E}\Big[\mathbf{u}\otimes\mathbf{w}^{\mathrm{T}}\otimes\mathbf{w}^{\mathrm{T}}\otimes\mathbf{u}|N=n\Big]\\&=\mathbb{E}\left[\boldsymbol{\xi}_1(n-\nu_1)\otimes\sum_{i=1}^{n}\mathbf{y}_i^{\mathrm{T}}\otimes\sum_{i=1}^{n}\mathbf{y}_i^{\mathrm{T}}\otimes\boldsymbol{\xi}_1(n-\nu_1)|N=n\right]\\&=n(n-\nu_1)^2\boldsymbol{\xi}_1\otimes\mathrm{vec}(\boldsymbol{\xi}_2)^{\mathrm{T}}\otimes\boldsymbol{\xi}_1.\end{aligned}$$

With expectation over N, we get

$$\begin{aligned}\mathbb{E}\Big[\mathbf{u}\otimes\mathbf{w}^{\mathrm{T}}\otimes\mathbf{u}\otimes\mathbf{w}^{\mathrm{T}}\Big]&=\mathbb{E}\Big[\mathbb{E}\Big[\mathbf{u}\otimes\mathbf{w}^{\mathrm{T}}\otimes\mathbf{u}\otimes\mathbf{w}^{\mathrm{T}}|N=n\Big]\Big]\\&=(\nu_3+\nu_1\nu_2)\boldsymbol{\xi}_1\otimes\mathrm{vec}(\boldsymbol{\xi}_2)^{\mathrm{T}}\otimes\boldsymbol{\xi}_1.\end{aligned}$$

We move on to evaluate $\mathbb{E}\Big[\mathbf{u}\otimes\mathbf{u}^{\mathrm{T}}\otimes\mathbf{w}\otimes\mathbf{w}^{\mathrm{T}}\Big]$. It holds that

$$\begin{aligned}\mathbb{E}\Big[\mathbf{u}\otimes\mathbf{u}^{\mathrm{T}}\otimes\mathbf{w}\otimes\mathbf{w}^{\mathrm{T}}|N=n\Big]&=\mathbb{E}\left[\boldsymbol{\xi}_1(n-\nu_1)\otimes\boldsymbol{\xi}_1^{\mathrm{T}}(n-\nu_1)\otimes\sum_{i=1}^{n}\mathbf{y}_i\otimes\sum_{i=1}^{n}\mathbf{y}_i^{\mathrm{T}}|N=n\right]\\&=(n-\nu_1)^2\boldsymbol{\xi}_1\otimes\boldsymbol{\xi}_1^{\mathrm{T}}\otimes\mathbb{E}\left[\sum_{i=1}^{n}\mathbf{y}_i\otimes\sum_{i=1}^{n}\mathbf{y}_i^{\mathrm{T}}|N=n\right]\\&=n(n-\nu_1)^2\boldsymbol{\xi}_1\otimes\boldsymbol{\xi}_1^{\mathrm{T}}\otimes\boldsymbol{\xi}_2.\end{aligned}$$

Hence, we obtain

$$\begin{aligned}\mathbb{E}\Big[\mathbf{u}\otimes\mathbf{u}^{\mathrm{T}}\otimes\mathbf{w}\otimes\mathbf{w}^{\mathrm{T}}\Big]&=\mathbb{E}\Big[\mathbb{E}\Big[\mathbf{u}\otimes\mathbf{u}^{\mathrm{T}}\otimes\mathbf{w}\otimes\mathbf{w}^{\mathrm{T}}|N=n\Big]\Big]\\&=(\nu_3+\nu_1\nu_2)\boldsymbol{\xi}_1\otimes\boldsymbol{\xi}_1^{\mathrm{T}}\otimes\boldsymbol{\xi}_2.\end{aligned}$$

Next, we consider $\mathbb{E}\Big[\mathbf{u}\otimes\mathbf{u}^{\mathrm{T}}\otimes\mathbf{u}\otimes\mathbf{u}^{\mathrm{T}}\Big]$, that is

$$\mathbb{E}\Big[\boldsymbol{\xi}_1(N-\nu_1)\otimes\boldsymbol{\xi}_1^{\mathrm{T}}(N-\nu_1)\otimes\boldsymbol{\xi}_1(N-\nu_1)\otimes\boldsymbol{\xi}_1^{\mathrm{T}}(N-\nu_1)\Big]=(\nu_4+3\nu_2^2)\boldsymbol{\xi}_1\otimes\boldsymbol{\xi}_1^{\mathrm{T}}\otimes\boldsymbol{\xi}_1\otimes\boldsymbol{\xi}_1^{\mathrm{T}}.$$

Combining all the terms involved in $\boldsymbol{\Theta}_1$, we finally get

$$\begin{aligned}\boldsymbol{\Theta}_1=&\ \nu_1\boldsymbol{\xi}_4+(\nu_2+\nu_1^2)\Big[(\mathbf{I}_{d^2}+\mathbf{K}_{d,d})(\boldsymbol{\xi}_2\otimes\boldsymbol{\xi}_2)+\mathrm{vec}(\boldsymbol{\xi}_2)\mathrm{vec}(\boldsymbol{\xi}_2)^{\mathrm{T}}\Big]+\nu_2\boldsymbol{\xi}_3\otimes\boldsymbol{\xi}_1^{\mathrm{T}}+\nu_2\boldsymbol{\xi}_3^{\mathrm{T}}\otimes\boldsymbol{\xi}_1+(\nu_3+\nu_1\nu_2)\boldsymbol{\xi}_2\otimes\boldsymbol{\xi}_1\otimes\boldsymbol{\xi}_1^{\mathrm{T}}\\&+\nu_2\boldsymbol{\xi}_1^{\mathrm{T}}\otimes\boldsymbol{\xi}_3+(\nu_3+\nu_1\nu_2)\mathrm{vec}(\boldsymbol{\xi}_2)\otimes\boldsymbol{\xi}_1^{\mathrm{T}}\otimes\boldsymbol{\xi}_1^{\mathrm{T}}+(\nu_3+\nu_1\nu_2)\boldsymbol{\xi}_1^{\mathrm{T}}\otimes\boldsymbol{\xi}_2\otimes\boldsymbol{\xi}_1+\mathbf{0}+\nu_2\boldsymbol{\xi}_1\otimes\boldsymbol{\xi}_3^{\mathrm{T}}+(\nu_3+\nu_1\nu_2)\boldsymbol{\xi}_1\otimes\boldsymbol{\xi}_2\\&\otimes\boldsymbol{\xi}_1^{\mathrm{T}}+(\nu_3+\nu_1\nu_2)\boldsymbol{\xi}_1\otimes\mathrm{vec}(\boldsymbol{\xi}_2)^{\mathrm{T}}\otimes\boldsymbol{\xi}_1+\mathbf{0}+(\nu_3+\nu_1\nu_2)\boldsymbol{\xi}_1\otimes\boldsymbol{\xi}_1^{\mathrm{T}}\otimes\boldsymbol{\xi}_2+\mathbf{0}+\mathbf{0}+(\nu_4+3\nu_2^2)\boldsymbol{\xi}_1\otimes\boldsymbol{\xi}_1^{\mathrm{T}}\otimes\boldsymbol{\xi}_1\otimes\boldsymbol{\xi}_1^{\mathrm{T}}\end{aligned}$$

$$= \nu_1\xi_4 + (\nu_2+\nu_1^2)\left[(\mathbf{I}_{d^2}+\mathbf{K}_{d,d})(\xi_2\otimes\xi_2)+\mathrm{vec}(\xi_2)\mathrm{vec}(\xi_2)^{\mathrm{T}}\right]+\nu_2\left[\xi_3\otimes\xi_1^{\mathrm{T}}+\xi_3^{\mathrm{T}}\otimes\xi_1+\xi_1^{\mathrm{T}}\otimes\xi_3+\xi_1\otimes\xi_3^{\mathrm{T}}\right]$$
$$+ (\nu_3+\nu_1\nu_2)\left[\xi_2\otimes\left(\xi_1\xi_1^{\mathrm{T}}\right)+\left(\mathrm{vec}(\xi_2)\xi_1^{\mathrm{T}}\right)\otimes\xi_1^{\mathrm{T}}+\xi_1^{\mathrm{T}}\otimes\xi_2\otimes\xi_1+\xi_1\otimes\xi_2\otimes\xi_1^{\mathrm{T}}+\left(\xi_1\mathrm{vec}(\xi_2)^{\mathrm{T}}\right)\otimes\xi_1\right.$$
$$\left.+\left(\xi_1\xi_1^{\mathrm{T}}\right)\otimes\xi_2\right]+(\nu_4+3\nu_2^2)\left(\xi_1\xi_1^{\mathrm{T}}\right)\otimes\left(\xi_1\xi_1^{\mathrm{T}}\right).$$

Utilizing the results for variance of $\mathbf{s}$ already derived in Loperfido et al. (2018), we get that

$$\Theta_2 = (\mathbf{I}_{d^2}+\mathbf{K}_{d,d})\left[\left(\nu_1\xi_2+\nu_2\xi_1\xi_1^{\mathrm{T}}\right)\otimes\left(\nu_1\xi_2+\nu_2\xi_1\xi_1^{\mathrm{T}}\right)\right],$$
$$\Theta_3 = \mathrm{vec}\left(\nu_1\xi_2+\nu_2\xi_1\xi_1^{\mathrm{T}}\right)\mathrm{vec}\left(\nu_1\xi_2+\nu_2\xi_1\xi_1^{\mathrm{T}}\right)^{\mathrm{T}}.$$

Putting all the components of equation (6.1), we finally get the expression as

$$\mathbf{K}_4(\mathbf{s}) = \nu_1\xi_4 + (\nu_2+\nu_1^2)\left[(\mathbf{I}_{d^2}+\mathbf{K}_{d,d})(\xi_2\otimes\xi_2)+\mathrm{vec}(\xi_2)\mathrm{vec}(\xi_2)^{\mathrm{T}}\right]+\nu_2\left[\xi_3\otimes\xi_1^{\mathrm{T}}+\xi_3^{\mathrm{T}}\otimes\xi_1+\xi_1^{\mathrm{T}}\otimes\xi_3+\xi_1\otimes\xi_3^{\mathrm{T}}\right]$$
$$+ (\nu_3+\nu_1\nu_2)\left[\xi_2\otimes\left(\xi_1\xi_1^{\mathrm{T}}\right)+\left(\mathrm{vec}(\xi_2)\xi_1^{\mathrm{T}}\right)\otimes\xi_1^{\mathrm{T}}+\xi_1^{\mathrm{T}}\otimes\xi_2\otimes\xi_1+\xi_1\otimes\xi_2\otimes\xi_1^{\mathrm{T}}+\left(\xi_1\mathrm{vec}(\xi_2)^{\mathrm{T}}\right)\otimes\xi_1\right.$$
$$\left.+\left(\xi_1\xi_1^{\mathrm{T}}\right)\otimes\xi_2\right]+(\nu_4+3\nu_2^2)\left(\xi_1\xi_1^{\mathrm{T}}\right)\otimes\left(\xi_1\xi_1^{\mathrm{T}}\right)-(\mathbf{I}_{d^2}+\mathbf{K}_{d,d})\left[\left(\nu_1\xi_2+\nu_2\xi_1\xi_1^{\mathrm{T}}\right)\otimes\left(\nu_1\xi_2+\nu_2\xi_1\xi_1^{\mathrm{T}}\right)\right]$$
$$-\mathrm{vec}\left(\nu_1\xi_2+\nu_2\xi_1\xi_1^{\mathrm{T}}\right)\mathrm{vec}\left(\nu_1\xi_2+\nu_2\xi_1\xi_1^{\mathrm{T}}\right)^{\mathrm{T}}.$$

Standard matrix algebra yields the identities

$$\left(\nu_1\xi_2+\nu_2\xi_1\xi_1^{\mathrm{T}}\right)\otimes\left(\nu_1\xi_2+\nu_2\xi_1\xi_1^{\mathrm{T}}\right)=\nu_1^2\xi_2\otimes\xi_2+\nu_1\nu_2\xi_2\otimes\xi_1\xi_1^{\mathrm{T}}+\nu_1\nu_2\xi_1\xi_1^{\mathrm{T}}\otimes\xi_2+\nu_2^2\xi_1\xi_1^{\mathrm{T}}\otimes\xi_1\xi_1^{\mathrm{T}}$$

and

$$\mathrm{vec}\left(\nu_1\xi_2+\nu_2\xi_1\xi_1^{\mathrm{T}}\right)vec\left(\nu_1\xi_2+\nu_2\xi_1\xi_1^{\mathrm{T}}\right)^{\mathrm{T}}$$
$$=\nu_1^2\mathrm{vec}(\xi_2)\mathrm{vec}(\xi_2)^{\mathrm{T}}+\nu_1\nu_2\mathrm{vec}(\xi_2)\otimes\xi_1^{\mathrm{T}}\otimes\xi_1^{\mathrm{T}}+\nu_1\nu_2\xi_1^{\mathrm{T}}\otimes\xi_1^{\mathrm{T}}\otimes\mathrm{vec}(\xi_2)+\nu_2^2\xi_1\xi_1^{\mathrm{T}}\otimes\xi_1\xi_1^{\mathrm{T}}.$$

Straightforward application of basic properties of the commutation matrix (Kollo and von Rosen, 2005, p. 80) yields the identities

$$\mathbf{K}_{d,d}\left(\xi_2\otimes\xi_1\xi_1^{\mathrm{T}}\right)=\xi_1\otimes\xi_2\otimes\xi_1^{\mathrm{T}},\ \mathbf{K}_{d,d}\left(\xi_1\xi_1^{\mathrm{T}}\otimes\xi_2\right)=\xi_1^{\mathrm{T}}\otimes\xi_2\otimes\xi_1,$$

$$\mathrm{vec}(\xi_2)\xi_1^{\mathrm{T}}\otimes\xi_1^{\mathrm{T}}=\mathrm{vec}(\xi_2)\otimes\xi_1^{\mathrm{T}}\otimes\xi_1^{\mathrm{T}}=\mathrm{vec}(\xi_2)\otimes\mathrm{vec}\left(\xi_1\xi_1^{\mathrm{T}}\right)^{\mathrm{T}}=\mathbf{K}_{d,d}\left(\mathrm{vec}(\xi_2)\xi_1^{\mathrm{T}}\otimes\xi_1^{\mathrm{T}}\right),$$

$$\left(\xi_1^{\mathrm{T}}\otimes\xi_1^{\mathrm{T}}\right)\mathrm{vec}(\xi_2)=\xi_1^{\mathrm{T}}\otimes\xi_1^{\mathrm{T}}\otimes\mathrm{vec}(\xi_2)=\mathrm{vec}\left(\xi_1\xi_1^{\mathrm{T}}\right)^{\mathrm{T}}\otimes\mathrm{vec}(\xi_2)=\mathbf{K}_{d,d}\left[\left(\xi_1^{\mathrm{T}}\otimes\xi_1^{\mathrm{T}}\right)\mathrm{vec}(\xi_2)\right].$$

Putting al above together we obtain the expression for the fourth cumulant stated in the theorem. □

Proof of Lemma 3.1.. Let $\mathbf{z}=\Sigma^{-1/2}(\mathbf{x}-\boldsymbol{\mu})$ be the standardized version of $\mathbf{x}$, where $\boldsymbol{\mu}$ is the mean of $\mathbf{x}$ and $\Sigma^{-1/2}$ is the positive definite square root of the concentration matrix Σ^{-1}:

$$\Sigma^{-1/2}=\Sigma^{-\mathrm{T}/2},\ \ \Sigma^{-1/2}>0,\ \ \Sigma^{-1/2}\Sigma^{-1/2}=\Sigma^{-1}.$$

The Mardia's kurtosis of $\mathbf{x}$ coincides with the trace of the fourth moment of $\mathbf{z}$ (Kollo and Srivastava, 2004):

$$\beta_{2,d}^{M}(\mathbf{x})=tr[\mathbf{M}_4(\mathbf{z})].$$

The trace of the fourth moment $\mathbf{M}_4(\mathbf{w})$ of any d-dimensional random vector $\mathbf{w}$ admits the representation $\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}}\mathbf{M}_4(\mathbf{w})\mathrm{vec}(\mathbf{I}_d)$, where $\mathbf{I}_d$ is the d-dimensional identity matrix (Loperfido, 2011). Also, the fourth moment of $\mathbf{Qw}$ is

$$(\mathbf{Q}\otimes\mathbf{Q})\mathbf{M}_4(\mathbf{w})\left(\mathbf{Q}^{\mathrm{T}}\otimes\mathbf{Q}^{\mathrm{T}}\right),$$

where $\mathbf{Q}$ is a $k\times d$ real matrix (Franceschini and Loperfido, 2012). 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 $\mathbf{x}$:

$$\mathbf{M}_4(\mathbf{z})=\left(\Sigma^{-1/2}\otimes\Sigma^{-1/2}\right)\overline{\mathbf{M}}_4(\mathbf{x})\left(\Sigma^{-1/2}\otimes\Sigma^{-1/2}\right).$$

We recall a fundamental property of the tensor product and the vectorization operator (see, for example, Rao and Rao (1998, page 201)):

$$\mathrm{vec}(\mathbf{ABC}) = \left(\mathbf{C}^{\mathrm{T}} \otimes \mathbf{A}\right)\mathrm{vec}(\mathbf{B}),$$

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

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

Hence we have $\beta_{2,d}(\mathbf{x}) = \mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{\mathrm{T}}\overline{\mathbf{M}}_4(\mathbf{x})\mathrm{vec}(\boldsymbol{\Sigma}^{-1})$ and this completes the first part of the proof.

We now prove the second part of the theorem. Let $\boldsymbol{\gamma}_i$ and $\boldsymbol{\sigma}_i$ be the i-th columns of $\boldsymbol{\Sigma}^{-1}$and $\boldsymbol{\Sigma}$, so that

$$\mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{\mathrm{T}}\mathrm{vec}(\boldsymbol{\Sigma}) = \boldsymbol{\gamma}_1^{\mathrm{T}}\boldsymbol{\sigma}_1 + \cdots + \boldsymbol{\gamma}_d^{\mathrm{T}}\boldsymbol{\sigma}_d = d.$$

The last equality is a direct consequence of $\boldsymbol{\Sigma}^{-1}$ being the inverse of $\boldsymbol{\Sigma}$: $\boldsymbol{\gamma}_1^{\mathrm{T}}\boldsymbol{\sigma}_1 = 1$ for $i = 1, \ldots, d$. The same property of $\boldsymbol{\Sigma}^{-1}$, together with the above mentioned property of the vectorization operator and the tensor product imply

$$\mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{\mathrm{T}}(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma}^{-1}) = \mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{\mathrm{T}}\mathrm{vec}\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\right) = \mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{\mathrm{T}}\mathrm{vec}(\boldsymbol{\Sigma}) = d.$$

We now recall two properties of the commutation matrix $\mathbf{K}_{p,q} \in \mathbb{R}^{pq} \times \mathbb{R}^{pq}$ (Magnus and Neudecker, 1979): $\mathrm{vec}(\mathbf{A}^{\mathrm{T}}) = \mathbf{K}_{p,q}\mathrm{vec}(\mathbf{A})$ for any $p \times q$ matrix $\mathbf{A}$ and $\mathbf{K}_{p,q}^{\mathrm{T}} = \mathbf{K}_{q,p}$. These properties and the symmetry of $\boldsymbol{\Sigma}^{-1}$ imply $\mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{\mathrm{T}}\mathbf{K}_{d,d} = \mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{\mathrm{T}}$.

The fourth cumulant $\mathbf{K}_4(\mathbf{x})$ of $\mathbf{x}$ is a function of its variance $\boldsymbol{\Sigma}$ and its fourth central moment $\overline{\mathbf{M}}_4(\mathbf{x})$:

$$\mathbf{K}_4(\mathbf{x}) = \overline{\mathbf{M}}_4(\mathbf{x}) - \mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma})^{\mathrm{T}} - \left(\mathbf{I}_{d^2} + \mathbf{K}_{d,d}\right)(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}),$$

where $\mathbf{I}_{d^2}$ is the d^2-dimensional identity matrix (see, for example, Kollo (2008)). The identities

$$\beta_{2,d}(\mathbf{x}) = \mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{\mathrm{T}}\overline{\mathbf{M}}_4(\mathbf{x})\mathrm{vec}(\boldsymbol{\Sigma}^{-1}), \quad \mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{\mathrm{T}}\mathrm{vec}(\boldsymbol{\Sigma}) = d,$$

$$\mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{\mathrm{T}}\mathbf{K}_{d,d} = \mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{\mathrm{T}} \quad \text{and} \quad \mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{\mathrm{T}}(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma}^{-1}) = d$$

which appear in previous part of the proof lead to

$$\mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{\mathrm{T}}\mathbf{K}_4(\mathbf{x})\mathrm{vec}(\boldsymbol{\Sigma}^{-1}) = \beta_{2,d}(\mathbf{x}) - d(d+2).$$

The right-hand side of the identity is just Mardia's excess kurtosis of $\mathbf{x}$ and this completes the proof. □

Proof of Theorem 4.1.. We first recall some fundamental properties of the Kronecker product and the vectorization operator (see, for example, Rao and Rao (1998, p. 194-201)): (P1) the Kronecker product is associative: $(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) = \mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C}$; (P2) if matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ are of appropriate size, then $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$; (P3) $\mathbf{A} = c\mathbf{A} = c \otimes \mathbf{A} = \mathbf{A} \otimes c$, where $c \in \mathbb{R}$ and $\mathbf{A} \in \mathbb{R}^p \times \mathbb{R}^q$; (P4) $\mathrm{vec}(\mathbf{ABC}) = \left(\mathbf{C}^{\mathrm{T}} \otimes \mathbf{A}\right)\mathrm{vec}(\mathbf{B})$, where $\mathbf{A} \in \mathbb{R}^p \times \mathbb{R}^q$, $\mathbf{B} \in \mathbb{R}^q \times \mathbb{R}^r$ and $\mathbf{C} \in \mathbb{R}^r \times \mathbb{R}^s$; (P5) $\mathrm{vec}\left(\mathbf{aa}^{\mathrm{T}}\right) = \mathrm{vec}\left(\mathbf{a} \otimes \mathbf{a}^{\mathrm{T}}\right) = \mathrm{vec}\left(\mathbf{a}^{\mathrm{T}} \otimes \boldsymbol{a}\right) = \mathrm{vec}(\mathbf{a} \otimes \mathbf{a}) = \mathrm{vec}\left(\mathbf{a}^{\mathrm{T}} \otimes \boldsymbol{a}^{\mathrm{T}}\right) = \mathbf{a} \otimes \mathbf{a}$, where $\mathbf{a} \in \mathbb{R}^p$.

The variance of $\mathbf{s}$ is $\boldsymbol{\Sigma} = \nu_1\mathbf{I}_d + \nu_2\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}$ (see Loperfido et al. (2018)). By the Sherman-Morrison formula, the inverse of $\boldsymbol{\Sigma}$ is

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\nu_1}\mathbf{I}_d - \frac{\nu_2\mathbf{I}_d\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}\mathbf{I}_d}{\nu_1^2 + \nu_1\nu_2\boldsymbol{\mu}^{\mathrm{T}}\mathbf{I}_d\boldsymbol{\mu}} = \frac{1}{\nu_1}\mathbf{I}_d - q\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}.$$

The following identities will be used several times in the proof:

$$\frac{1}{\nu_1} - q\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu} = \frac{1}{\nu_1} - \frac{\nu_2\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}}{\nu_1^2 + \nu_1\nu_2\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}} = \frac{\nu_1^2 + \nu_1\nu_2\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu} - \nu_1\nu_2\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}}{\nu_1^3 + \nu_1^2\nu_2\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}} = \frac{\nu_1 q}{\nu_2}.$$

By Theorem 2.1, the fourth central moment of $\mathbf{s}$ is

$$\overline{\mathbf{M}}_4(\mathbf{s}) = \left(\nu_2 + \nu_1^2\right)\left(\mathbf{K}_{d,d} + \mathbf{I}_{d^2} + \mathrm{vec}(\mathbf{I}_d)\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}}\right) + (\nu_3 + \nu_1\nu_2)[\,\mathbf{I}_d \otimes \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}} + \mathrm{vec}(\mathbf{I}_d)\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}$$
$$+\boldsymbol{\mu}^{\mathrm{T}} \otimes \mathbf{I}_d \otimes \boldsymbol{\mu} + \boldsymbol{\mu} \otimes \mathbf{I}_d \otimes \boldsymbol{\mu}^{\mathrm{T}} + \boldsymbol{\mu}\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}} \otimes \boldsymbol{\mu} + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}} \otimes \mathbf{I}_d\,] + \left(\nu_4 + 3\nu_2^2\right)\left(\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}\right).$$

By Lemma 3.1 the Mardia's kurtosis of $\mathbf{s}$ is

$$\beta_{2,d}(\mathbf{s}) = \mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{\mathrm{T}}\overline{\mathbf{M}}_4(\mathbf{s})\mathrm{vec}(\boldsymbol{\Sigma}^{-1}) = \left(\nu_2 + \nu_1^2\right)Q_1 + (\nu_3 + \nu_1\nu_2)(Q_2 + 2Q_3 + 2Q_4) + \left(\nu_4 + 3\nu_2^2\right)Q_5,$$

where $Q_1 = \mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{\mathrm{T}}\left(\mathbf{K}_{d,d} + \mathbf{I}_{d^2} + \mathrm{vec}(\mathbf{I}_d)\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}}\right)\mathrm{vec}(\boldsymbol{\Sigma}^{-1})$, $Q_2 = \mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{\mathrm{T}}\left(\mathbf{I}_d \otimes \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}\right)\mathrm{vec}(\boldsymbol{\Sigma}^{-1})$, $Q_3 = \mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{\mathrm{T}}\left(\mathrm{vec}(\mathbf{I}_d)\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\mathrm{vec}(\boldsymbol{\Sigma}^{-1})$, $Q_4 = \mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{\mathrm{T}}\left(\boldsymbol{\mu}^{\mathrm{T}} \otimes \mathbf{I}_d \otimes \boldsymbol{\mu}\right)\mathrm{vec}(\boldsymbol{\Sigma}^{-1})$ and $Q_5 = \mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{\mathrm{T}}\left(\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}\right)\mathrm{vec}(\boldsymbol{\Sigma}^{-1})$ might be simplified by means of properties *(P1)-(P5)*.

We first evaluate the quadratic form

$$Q_1 = \left(\frac{1}{\nu_1}\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}} - q\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\left(\mathbf{K}_{d,d} + \mathbf{I}_{d^2} + \mathrm{vec}(\mathbf{I}_d)\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}}\right)\left(\frac{1}{\nu_1}\mathrm{vec}(\mathbf{I}_d) - q\boldsymbol{\mu} \otimes \boldsymbol{\mu}\right).$$

The $d^2 \times d^2$ commutation matrix $\mathbf{K}_{d,d}$ satisfies the identity $\mathbf{K}_{d,d}\mathrm{vec}(\mathbf{A}) = \mathrm{vec}(\mathbf{A}^{\mathrm{T}})$ for any $d \times d$ matrix $\mathbf{A}$ (see Magnus and Neudecker (1979)), so that

$$\begin{aligned}
&\left(\frac{1}{\nu_1}\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}} - q\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\mathbf{K}_{d,d}\left(\frac{1}{\nu_1}\mathrm{vec}(\mathbf{I}_d) - q\boldsymbol{\mu} \otimes \boldsymbol{\mu}\right) \\
&= \left(\frac{1}{\nu_1}\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}} - q\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\left(\frac{1}{\nu_1}\mathrm{vec}(\mathbf{I}_d) - q\boldsymbol{\mu} \otimes \boldsymbol{\mu}\right) \\
&= \frac{1}{\nu_1^2}\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}}\mathrm{vec}(\mathbf{I}_d) - \frac{2q}{\nu_1}\left(\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\mathrm{vec}(\mathbf{I}_d) + q^2\left(\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)(\boldsymbol{\mu} \otimes \boldsymbol{\mu}) \\
&= \frac{d}{\nu_1^2} - \frac{2q}{\nu_1}\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right) + q^2\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2.
\end{aligned}$$

A little algebra and the identity $\nu_1^{-1} - q\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu} = q\nu_1\nu_2^{-1}$ yield

$$\left(\frac{1}{\nu_1}\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}} - q\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\mathbf{K}_{d,d}\left(\frac{1}{\nu_1}\mathrm{vec}(\mathbf{I}_d) - q\boldsymbol{\mu} \otimes \boldsymbol{\mu}\right) = \frac{d-1}{\nu_1^2} + \left(\frac{1}{\nu_1} - q\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2 = \frac{d-1}{\nu_1^2} + \frac{q^2}{\nu_2^2}.$$

In a similar way it can be shown that

$$\left(\frac{1}{\nu_1}\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}} - q\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\mathrm{vec}(\mathbf{I}_d)\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}}\left(\frac{1}{\nu_1}\mathrm{vec}(\mathbf{I}_d) - q\boldsymbol{\mu} \otimes \boldsymbol{\mu}\right) = \left(\frac{d}{\nu_1} - q\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2,$$

thus leading to the simplified expression

$$Q_1 = 2\left(\frac{d-1}{\nu_1^2} + \frac{q^2}{\nu_2^2}\right) + \left(\frac{d}{\nu_1} - q\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2.$$

We now simplify the expression of the quadratic form

$$Q_2 = \left(\frac{1}{\nu_1}\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}} - q\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\left(\mathbf{I}_d \otimes \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}\right)\left(\frac{1}{\nu_1}\mathrm{vec}(\mathbf{I}_d) - q\boldsymbol{\mu} \otimes \boldsymbol{\mu}\right).$$

Repeated application of properties *(P1)-(P5)* yield

$$\begin{aligned}
Q_2 &= \frac{1}{\nu_1^2}\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}}\left(\mathbf{I}_d \otimes \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}\right)\mathrm{vec}(\mathbf{I}_d) + q^2\left(\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\left(\mathbf{I}_d \otimes \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}\right)(\boldsymbol{\mu} \otimes \boldsymbol{\mu}) - \frac{2q}{\nu_1}\left(\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\left(\mathbf{I}_d \otimes \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}\right)\mathrm{vec}(\mathbf{I}_d) \\
&= \mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}}\frac{\boldsymbol{\mu} \otimes \boldsymbol{\mu}}{\nu_1^2} + q^2\left(\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\left(\boldsymbol{\mu} \otimes \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right) - \frac{2q}{\nu_1}\left(\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)(\boldsymbol{\mu} \otimes \boldsymbol{\mu}) \\
&= \frac{\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}}{\nu_1^2} + q^2\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^3 - \frac{2q}{\nu_1}\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2.
\end{aligned}$$

Further application of the identity $\nu_1^{-1} - q\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu} = q\nu_1\nu_2^{-1}$ yields

$$Q_2 = \left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)\left(\frac{1}{\nu_1} - q\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2 = \frac{\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}}{\left(\nu_1 + \nu_2\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2} = \frac{q^2}{\nu_2^2}\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}.$$

The following identities allow for a simpler expression of Q_3:

$$\left(\mathrm{vec}(\mathbf{I}_d)\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\mathrm{vec}(\mathbf{I}_d) = \mathrm{vec}\left(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{I}_d\boldsymbol{\mu}\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}}\right) = \left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)\mathrm{vec}\left(\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}}\right) = \left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)\mathrm{vec}(\mathbf{I}_d).$$

We can then represent Q_3 as

$$\begin{aligned}
&\left(\frac{1}{\nu_1}\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}} - q\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\left(\mathrm{vec}(\mathbf{I}_d)\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\left(\frac{1}{\nu_1}\mathrm{vec}(\mathbf{I}_d) - q\boldsymbol{\mu} \otimes \boldsymbol{\mu}\right) \\
&= \frac{1}{\nu_1^2}\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}}\left(\mathrm{vec}(\mathbf{I}_d)\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\mathbf{I}_d^{\mathrm{V}} - \frac{2q}{\nu_1}\left(\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\left(\mathrm{vec}(\mathbf{I}_d)\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\mathrm{vec}(\mathbf{I}_d) + q^2\left(\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\left(\mathrm{vec}(\mathbf{I}_d)\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)(\boldsymbol{\mu} \otimes \boldsymbol{\mu}) \\
&= \frac{1}{\nu_1^2}\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}}\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)\mathrm{vec}(\mathbf{I}_d) - \frac{2q}{\nu_1}\left(\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)\mathrm{vec}(\mathbf{I}_d) + q^2\left(\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\left(\mathrm{vec}(\mathbf{I}_d)\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu} \otimes \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right) \\
&= \frac{d}{\nu_1^2}\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right) - \frac{2q}{\nu_1}\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2 + q^2\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^3 \\
&= \left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)\left[\frac{d}{\nu_1^2} - \frac{2q}{\nu_1}\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right) + q^2\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{(d-1)\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)}{\nu_1^2} + \left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)\left[\frac{1}{\nu_1^2} - \frac{2q}{\nu_1}\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right) + q^2\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2\right] \\
&= \left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)\left(\frac{d-1}{\nu_1^2} + \frac{q^2}{\nu_2^2}\right).
\end{aligned}
$$

A simpler expression of $Q_4 = (\nu_1^{-1}\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}} - q\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}})(\boldsymbol{\mu}^{\mathrm{T}} \otimes \mathbf{I}_d \otimes \boldsymbol{\mu})(\nu_1^{-1}\mathrm{vec}(\mathbf{I}_d) - q\boldsymbol{\mu} \otimes \boldsymbol{\mu})$ might be derived using the identities

$$
\left(\boldsymbol{\mu}^{\mathrm{T}} \otimes \mathbf{I}_d \otimes \boldsymbol{\mu}\right)\mathrm{vec}(\mathbf{I}_d) = \mathrm{vec}((\mathbf{I}_d \otimes \boldsymbol{\mu})\mathbf{I}_d\boldsymbol{\mu}) = \mathrm{vec}((\mathbf{I}_d \otimes \boldsymbol{\mu})\boldsymbol{\mu}) = \mathrm{vec}((\mathbf{I}_d \otimes \boldsymbol{\mu})(\boldsymbol{\mu} \otimes 1)) = \boldsymbol{\mu} \otimes \boldsymbol{\mu},
$$

$$
\left(\boldsymbol{\mu}^{\mathrm{T}} \otimes \mathbf{I}_d \otimes \boldsymbol{\mu}\right)(\boldsymbol{\mu} \otimes \boldsymbol{\mu}) = \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu} \otimes (\mathbf{I}_d \otimes \boldsymbol{\mu})\boldsymbol{\mu} = \left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)(\mathbf{I}_d \otimes \boldsymbol{\mu})(\boldsymbol{\mu} \otimes 1) = \left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)(\boldsymbol{\mu} \otimes \boldsymbol{\mu}).
$$

We can then represent Q_4 as

$$
\begin{aligned}
&\frac{1}{\nu_1^2}\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}}\left(\boldsymbol{\mu}^{\mathrm{T}} \otimes \mathbf{I}_d \otimes \boldsymbol{\mu}\right)\mathrm{vec}(\mathbf{I}_d) - \frac{2q}{\nu_1}\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}}\left(\boldsymbol{\mu}^{\mathrm{T}} \otimes \mathbf{I}_d \otimes \boldsymbol{\mu}\right)(\boldsymbol{\mu} \otimes \boldsymbol{\mu}) + q^2\left(\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\left(\boldsymbol{\mu}^{\mathrm{T}} \otimes \mathbf{I}_d \otimes \boldsymbol{\mu}\right)(\boldsymbol{\mu} \otimes \boldsymbol{\mu}) \\
&\quad = \frac{1}{\nu_1^2}\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right) - \frac{2q}{\nu_1}\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2 + q^2\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^3 = \left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)\left(\frac{1}{\nu_1} - q\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2 = \frac{\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}}{\left(\nu_1 + \nu_2\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2} = \frac{q^2}{\nu_2^2}\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right).
\end{aligned}
$$

We evaluate Q_5 in a similar way:

$$
\begin{aligned}
Q_5 &= \left(\frac{1}{\nu_1}\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}} - q\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\left(\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}\right)\left(\frac{1}{\nu_1}\mathrm{vec}(\mathbf{I}_d) - q\boldsymbol{\mu} \otimes \boldsymbol{\mu}\right) \\
&= \frac{1}{\nu_1^2}\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}}\left(\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}\right)\mathrm{vec}(\mathbf{I}_d) - \frac{2q}{\nu_1}\mathrm{vec}(\mathbf{I}_d)^{\mathrm{T}}\left(\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}\right)(\boldsymbol{\mu} \otimes \boldsymbol{\mu}) + q^2\left(\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}^{\mathrm{T}}\right)\left(\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}} \otimes \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}\right)(\boldsymbol{\mu} \otimes \boldsymbol{\mu}) \\
&= \frac{\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2}{\nu_1^2} - \frac{2q}{\nu_1}\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^3 + q^2\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^4 = \left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2\left(\frac{1}{\nu_1} - q\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2 = \frac{q^2}{\nu_2^2}\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2.
\end{aligned}
$$

We conclude the proof by reprenting the Mardia's kurtosis of $\mathbf{s}$ by means of the simplified expressions for recalling the definitions of Q_1, Q_2, Q_3, $Q_4$and Q_5:

$$
\begin{aligned}
\beta_{2,d}(\mathbf{s}) &= \left(\nu_2 + \nu_1^2\right)\left[2\left(\frac{d-1}{\nu_1^2} + \frac{q^2}{\nu_2^2}\right) + \left(\frac{d}{\nu_1} - q\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2\right] + (\nu_3 + \nu_1\nu_2)\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)\left(\frac{2d-2}{\nu_1^2} + \frac{5q^2}{\nu_2^2}\right) \\
&\quad + \left(\nu_4 + 3\nu_2^2\right)\frac{q^2}{\nu_2^2}\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2.
\end{aligned}
$$

□

Proof of Corollary 4.1.. We first prove the theorem for $\{N_i = Y + a_i\}$. The first four cumulants of Y are $\mathbb{E}(Y) = 1.5$, $\mathbb{E}\left[(Y-1.5)^2\right] = 0.25$, $\mathbb{E}\left[(Y-1.5)^3\right] = 0$ and $\mathbb{E}\left[(Y-1.5)^4\right] - 3\mathbb{E}^2\left[(Y-1.5)^2\right] = -0.125$. As a direct consequence, the first four cumulants of N_i are $\nu_{i,1} = a_i + 1.5$, $\nu_{i,2} = 0.25$, $\nu_{i,3} = 0$, $\nu_{i,4} = -0.125$. Let $q_i = \nu_{i,2}\left(\nu_{i,1}^2 + \nu_{i,1}\nu_{i,2}\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^{-1}$, so that

$$
q_i = \frac{1}{4\nu_{i,1}^2 + \nu_{i,1}\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}} \quad \text{and} \quad \lim_{i \longrightarrow +\infty} q_i \cdot \nu_{i,1}^2 = \frac{1}{4}.
$$

The Mardia's kurtosis $\beta_{2,d}\left(\mathbf{x}_1 + \ldots + \mathbf{x}_{N_i}\right)$ of $\mathbf{x}_1 + \ldots + \mathbf{x}_{N_i}$ is then

$$
\left(0.25 + \nu_{i,1}^2\right)\left[2\left(\frac{d-1}{\nu_{i,1}^2} + 16q_i^2\right) + \left(\frac{d}{\nu_{i,1}} - q_i\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2\right] + \nu_{i,1}\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)\left(\frac{d-1}{2\nu_{i,1}^2} + 20q_i^2\right) + q_i^2\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^2.
$$

The definitions of q_i and $\nu_{i,1}$, together with standard calculus techniques, yield

$$
\lim_{i \longrightarrow +\infty} \beta_{2,d}\left(\mathbf{x}_1 + \ldots + \mathbf{x}_{N_i}\right) = d(d+2).
$$

We now prove the theorem for $\{M_i = a_i Y\}$. The first four cumulants of M_i are $\eta_{i,1} = 1.5a_i$, $\eta_{i,2} = 0.25a_i^2$, $\eta_{i,3} = 0$ and $\eta_{i,4} = -0.125a_i^4$. Let $r_i = \eta_{i,2}\left(\eta_{i,1}^2 + \eta_{i,1}\eta_{i,2}\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)^{-1}$, so that

$$
r_i = \frac{1}{a_i\left(2.25 + 0.375 \cdot \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\right)} \quad \text{and} \quad \lim_{i \longrightarrow +\infty} q_i \cdot \nu_{i,1} = \frac{5}{3}.
$$

By Theorem 4.1, the Mardia's kurtosis of $\mathbf{x}_1+\ldots+\mathbf{x}_{M_i}$ is

$$\beta_{2,d}\left(\mathbf{x}_1+\ldots+\mathbf{x}_{M_i}\right)=\left(\eta_{i,2}+\eta_{i,1}^2\right)\left[2\left(\frac{d-1}{\eta_{i,1}^2}+\frac{r_i^2}{\eta_{i,2}^2}\right)+\left(\frac{d}{\eta_{i,1}}-r_i\boldsymbol{\mu}^{\top}\boldsymbol{\mu}\right)^2\right]$$
$$+\left(\eta_{i,3}+\eta_{i,1}\eta_{i,2}\right)\left(\boldsymbol{\mu}^{\top}\boldsymbol{\mu}\right)\left(\frac{2d-2}{\eta_{i,1}^2}+\frac{5r_i^2}{\eta_{i,2}^2}\right)+\left(\eta_{i,4}+3\eta_{i,2}^2\right)\frac{r_i^2}{\eta_{i,2}^2}\left(\boldsymbol{\mu}^{\top}\boldsymbol{\mu}\right)^2.$$

By recalling the definitions of $\eta_{i,1}$, $\eta_{i,2}$, $\eta_{i,3}$ and $\eta_{i,4}$ the Mardia's kurtosis of $\mathbf{x}_1+\ldots+\mathbf{x}_{M_i}$ might be simplified into

$$2.5\left[2\left(\frac{d-1}{2.25}+\frac{r_i^2}{0.0625}\right)+\left(\frac{d}{1.5}-\frac{r_i}{a_i}\boldsymbol{\mu}^{\top}\boldsymbol{\mu}\right)^2\right]+$$

$$0.375a_{i,1}\left(\boldsymbol{\mu}^{\top}\boldsymbol{\mu}\right)\left(\frac{2d-2}{2.25}+\frac{5r_i^2}{0.0625}\right)+r_i^2\left(\boldsymbol{\mu}^{\top}\boldsymbol{\mu}\right)^2.$$

The assumption $\boldsymbol{\mu}\neq\mathbf{0}_d$, the definitions of r_i and $\eta_{i,1}$, together with standard calculus techniques, yield

$$\lim_{i\longrightarrow+\infty}\beta_{2,d}\left(\mathbf{x}_1+\ldots+\mathbf{x}_{M_i}\right)=+\infty.$$ □

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ecosta.2021.04.005.

References

Ambagaspitiya, R.S., 1999. On the distributions of two classes of correlated aggregate claims. Insurance: Mathematics and Economics 24 (3), 301–308. https://doi.org/10.1016/S0167-6687(99)00006-2.
Arevalillo, J.M., Navarro, H., 2012. A study of the e ect of kurtosis on discriminant analysis under elliptical populations. Journal of Multivariate Analysis 107, 53–63.
Barndorff-Nielsen, Cox, D.R., 1979. Edgeworth and saddle-point approximations with statistical applications. Journal of the Royal Statistical Society. Series B (Methodological) 41 (3), 279–312.
Bernardi, M., Maruotti, A., Petrella, L., 2012. Skew mixture models for loss distributions: A bayesian approach. Insurance: Mathematics and Economics 51, 617–623.
Bhattacharya, R.N., Ghosh, J.K., 1978. On the validity of the formal Edgeworth expansion. Annals of Statistics 6 (2), 434–451.
Bolance, C., Guillen, M., Pelican, E., Vernic, R., 2008. Skewed bivariate models and nonparametric estimation for the CTE risk measure. Insurance: Mathematics and Economics 43, 386–393.
Bowers, N., Gerber, H., Hickman, J., Jones, D., Nesbitt, C., 1997. Actuarial Mathematics, 2nd Society of Actuaries, Schaumburg, IL.
Braione M., B.L., Storti, G., 2017. A dynamic component model for forecasting high-dimensional realized covariance matrices. Econometrics and Statistics 1 (1), 40–61.
Burnecki, K., Janczura, J., Weron, R., 2011. Building loss models. In: Cizek, P., Härdle, W., Weron, R. (Eds.), Statistical tools for finance and insurance. Springer, Berlin, Heidelberg.
Cummins, D.J., Wiltbank, L.J., 1983. Estimating the total claims distribution using multivariate frequency and severity distributions. The Journal of Risk and Insurance 50, 377–403.
Daley, D.J., Omey, E., Vesilo, R., 2007. The tail behaviour of a random sum of subexponential random variables and vectors. Extremes 10, 21–39.
Eling, M., 2012. Fitting insurance claims to skewed distributions: are the skew-normal and skew-student good models? Insurance: Mathematics and Economics 51, 239–248.
Franceschini, C., Loperfido, N., 2012. On some inequalities between measures of multivariate kurtosis, with application to financial returns. In: Perna, C., Sibillo, M. (Eds.), Mathematical and statistical methods for actuarial sciences and finance. Springer, pp. 211–218.
Franceschini, C., Loperfido, N., 2014. Testing for normality when the sampled distribution is extended skew-normal. In: Corazza, M., Pizzi, C. (Eds.), Mathematical and statistical methods for actuarial sciences and finance. Springer, pp. 159–170.
Gnedenko, B., Korolev, V., 1996. Random Summation: Limit Theorems and Applications, 1st CRC Press.
Gray, H.L., Coberly, W.A., Lewis, T.O., 1975. On Edgeworth expansions with unknown cumulants. The Annals of Statistics 3 (3), 741–746.
Gupta, A.K., Kollo, T., 2003. Density expansions based on the multivariate skew normal distribution. Sankhya: The Indian Journal of Statistics 65 (4), 821–835.
Harville, D.A., 1997. Matrix Algebra from Statistician's Perspective. Springer, New York.
Kalashnikov, V., 1997. Geometric Sums: Bounds for Rare Events with Applications. Kluwer Academic, Dordrecht.
Klebanov, L.B., Kozubowski, T.J., Rachev, S.T., 2006. Ill-Posed Problems in Probability and Stability of Random Sums. Nova Science Publishers Inc, New York.
Kollo, T., 2008. Multivariate skewness and kurtosis measures with an application in ICA. Journal of Multivariate Analysis 99, 2328–2338.
Kollo, T., von Rosen, D., 2005. Advanced Multivariate Statistics with Matrices. Springer, Dordrecht.
Kollo, T., Srivastava, M.S., 2004. Estimation and testing of parameters in multivariate Laplace distribution. Communication in Statistics - Theory and Methods 33 (10), 2363–2387.
Lin, X.S., 2006. Encyclopedia of actuarial science. Wiley, pp. 314–317).
Livesey, J.H., 2007. Kurtosis provides a good omnibus test for outliers in small samples. Clinical Biochemistry 40, 1032–1036.
Loperfido, N., 2011. Spectral analysis of the fourth moment matrix. Linear Algebra and its Applications 435, 1837–1844.
Loperfido, N., 2014. A note on the fourth cumulant of a finite mixture distribution. Journal of Multivariate Analysis 123, 386–394.
Loperfido, N., 2019. Finite mixtures, projection pursuit and tensor rank: a triangulation. Advances in Data Analysis and Classification 31, 145–173.
Loperfido, N., Mazur, S., Podgórski, K., 2018. Third cumulant for multivariate aggregate claim models. Scandinavian Actuarial Journal 2018 (2), 109–128.
Magnus, J.R., Neudecker, H., 1979. The commutation matrix: some properties and applications. Annals of Statistics 7, 381–394.
Mardia, K.V., 1970. Measures of multivariate skewness and kurtosis with applications. Biometrika 57, 519–530.

Mardia, K.V., 1974. Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. Sankhya: The Indian Journal of Statistics, Series B 36 (2), 115–128.
Morana, C., 2019. Regularized semiparametric estimation of high dimensional dynamic conditional covariance matrices. Econometrics and Statistics 12, 42–65.
Mutschler, W., 2018. Higher-order statistics for DSGE models. Econometrics and Statistics 6, 44–56.
Pewsey, A., 2001. Problems of inference for Azzalini's skew-normal distribution. Journal of Applied Statistics 27, 859–870.
Qi, L., Sun, W., Wang, Y., 2007. Numerical multilinear algebra and its applications. Frontiers of Mathematics in China 2 (4), 501–526.
Rao, C.R., Rao, M.B., 1998. Matrix algebra and its applications to statistics and econometrics. World Scientific Publishing Co. Pte. Ltd., Singapore.
Rezvandehy, M., Deutsch, C.V., 2018. Declustering experimental variograms by global estimation with fourth order moments. Stochastic Environmental Research and Risk Assessment 32 (1), 261–277.
Robbins, H., 1948. The asymptotic distribution of the sum of a random number of random variables. Bulletin of the American Mathematical Society 54 (12), 1151–1161.
Ronchetti, E., 2020. Accurate and robust inference. Econometrics and Statistics 14, 74–88.
Scrucca, L., Fop, M., Murphy, T.B., Raftery, A.E., 2016. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. The R Journal 8 (1), 289–317.
Stuart, A., Ord, K., 1994. Kendall's Advanced Theory of Statistics, vol. 1: Distribution Theory, Sixth Hodder Arnold, London.
Van Hulle, M.M., 2005. Edgeworth approximation of multivariate differential entropy. Neural Computation 17 (9), 1903–1910.
Yanagihara, H., 2007. A family of estimators for multivariate kurtosis in a nonnormal linear regression model. Journal of Multivariate Analysis 98, 1–29.
Yanagihara, H., Tonda, T., Matsumoto, C., 2005. The effects of nonnormality on asymptotic distributions of some likelihood ratio criteria for testing covariance structures under normal assumption. Journal of Multivariate Analysis 96, 237–264.