SUT Journal of Mathematics
Vol. 50, No. 2 (2014), 483–511

Measures of multivariate skewness and kurtosis in high-dimensional framework

Kazuyuki Koizumi, Takuma Sumikawa and Tatjana Pavlenko

(Received December 1, 2014; Revised February 24, 2015)

Abstract. Skewness and kurtosis characteristics of a multivariate p-dimensional distribution introduced by Mardia (1970) have been used in various testing procedures and demonstrated attractive asymptotic properties in large sample settings. However these characteristics are not designed for high-dimensional problems where the dimensionality, p can largely exceeds the sample size, N. Such type of high-dimensional data are commonly encountered in modern statistical applications. This the suggests that new measures of skewness and kurtosis that can accommodate high-dimensional settings must be derived and carefully studied. In this paper, we show that, by exploiting the dependence structure, new expressions for skewness and kurtosis are introduced as an extension of the corresponding Mardia's measures, which uses the potential advantages that the block-diagonal covariance structure has to offer in high dimensions. Asymptotic properties of newly derived measures are investigated and the cumulant based characterizations are presented along with of applications to a mixture of multivariate normal distributions and multivariate Laplace distribution, for which the explicit expressions of skewness and kurtosis are obtained. Test statistics based on the new measures of skewness and kurtosis are proposed for testing a distribution shape, and their limit distributions are established in the asymptotic framework where $N \to \infty$ and p is fixed but large, including $p > N$. For the dependence structure learning, the gLasso based technique is explored followed by AIC step which we propose for optimization of the gLasso candidate model. Performance accuracy of the test procedures based on our estimators of skewness and kurtosis are evaluated using Monte Carlo simulations and the validity of the suggested approach is shown for a number of cases when $p > N$.

AMS 2010 *Mathematics Subject Classification.* 62H10, 62D05.

Key words and phrases. Multivariate skewness; Multivariate kurtosis; Normality test; Covariance structure approximation; Block diagonal matrix; AIC.

§1. Introduction

Modern experimental technology provides a possibility for collection and exchange of massive sets of data which naturally poses a challenge the quantitative analyses of such type of data. Examples include statistical analyses of complex and high-dimensional systems where simultaneous measuring a large number of feature variables is associated with small sample size. This is a common situation in e.g. gene expression data which usually comprises measurements on thousands of features, but the sample size is often in several hundreds; image data are usually obtained by measuring dozens of thousands of variables at the same time whereas the size of available samples usually remains in hundreds.

An important component of statistical analyses of this type of high-dimensional problems are characteristics of the shape of the population distribution underlying the data, such as e.g. measures of spherical or elliptical symmetry. For example, the dimensionality reduction technique, sliced inverse regression suggested in Cook and Li [2] and its extension (see, Liang [14]) are shown to be powerful alternatives to standard feature selection procedures in high-dimensional setting, however under the assumption of sphericity or elliptic symmetry of the underlying distribution. Another example to give is on learning structural sparsity in high-dimensional data by gLasso technique; this is a very efficient tool, however the interpretations of zeros in the inverse covariance matrix and convergence properties of the obtained estimators are given under multivariate normality (which is a special case of elliptically symmetric distribution).

Further needs for development of new symmetry measures can also be motivated by the following circumstances. Over the last decades, several new distribution families have been introduced for modeling skewed data, e.g. the asymmetric multivariate Laplace distribution is studied with applications in Kots et al. [12], multivariate skew t-distribution is considered in Kots and Nadarajar [13] and various types of skew elliptical distributions are presented in the monograph Genton [5]. Typically these families are characterized by the scale and symmetry parameters and estimation of these parameters creates problems when sample size is small relative to the dimensionality. Hence, it would be highly desirable to develop new characteristics along with the corresponding testing procedures that have stable performance accuracy in high dimensions.

Most common population measures characterizing the distribution shape are given by *multivariate skewness* and *multivariate kurtosis* defined by Mardia

[15] as

$$\beta_1 = E[\{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\}^3], \tag{1.1}$$

$$\beta_2 = E[\{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}^2], \tag{1.2}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent and identically distributed random p-dimensional vectors with expectation $\boldsymbol{\mu}$ and non-singular covariance matrix Σ.

Recall that the multivariate skewness is a measure of the asymmetry of a distribution about its mean and its value far from zero indicates stronger asymmetry of the underlying distribution than that with close to zero skewness value. The multivariate kurtosis measures the peakedness of a distribution scaled by its covariance (Both these measures are always non-negative, unlike the univariate case). We also note that $\beta_1 = 0$ and $\beta_2 = p(p+2)$ hold under multivariate normality.

Several suggestions and generalizations for modifying of (1.1) and (1.2) are considered since Mardia's pioneering work. For recent results, we refer to e.g. Miyagawa et al. [17] who proposed the sample measure of multivariate kurtosis of the form containing Mardia [15] and Srivastava [20]. Further, Koizumi et al. [10] suggest two extensions of Jarque-Bera test, which is an omnibus type test for assessing multivariate normality. They are constructed by combining estimators of Mardia's, multivariate skewness and kurtosis, or by Srivastava's multivariate skewness and kurtosis introduced in Srivastava [20]. For reviews of testing the multivariate normality using skewness and kurtosis, see e.g., Henze [6] or Mecklin and Mundfrom [16]. Clearly, these procedures, while demonstrating attractive asymptotic properties and good performance accuracy in large sample case, are not applicable in high dimensions since the sample based covariance matrix is singular when $p > n$ and hence can not be inverted. The main goal of this paper is to design such measures of skewness and kurtosis which can tackle the challenge of high-dimensionality in combination with small sample size. Our crucial idea in designing these measures is to point out advantages offered by the dependence structure, and show how these advantages can be exploited.

We introduce new measures of skewness and kurtosis under the block-diagonal covariance structure with a constraint that the block dimension does not exceed the sample size and no other constraints on the covariance structure within the blocks. We then use asymptotic theory assuming p fixed and $n \to \infty$, including the case when $p > n$ and establish the asymptotic distributions of our suggested sample based skewness and kurtosis under the additional normality assumption. Furthermore, while the original estimators of β_1 and β_2 as well as corresponding test procedures were developed keeping p fixed and letting $n \to \infty$, our new estimates are so constructed that they are also valid under the standard high-dimensional asymptotic, i.e. when both p and $n \to \infty$. As the use of our estimators of β_1 and β_2, particularly high-dimensional case, to tackle the problem of testing normality is very natural application, a simulation study with a variety of parameter settings is performed. The accuracy of the test statistics is evaluated for size control, inclusive of the cases when the dimensionality p far exceeds sample size. Since our theoretical results, e.g. moments of the suggested estimators, are largely asymptotic, we provide numerical evaluation of asymptotic accuracy. In particular, with the use of simulations, we derive two improved versions of sample based skewness. Finally, the block-diagonal covariance structure: this assumption, while being very beneficial for theoretical consideration in high-dimensions, seem to be too strong for practical applications. Of course, in e.g. genome data the block-diagonal covariance structure is a natural model representing interactions of tightly linked genes; pathway-level analysis can provide biologically meaningful hypothetical block structure of functionally related genes. However, it is important to derive a structure estimation technique which work without information from outside the datasets. This motivates our proposal on the structure learning technique which consists of two stages; gLasso procedure is first used for obtaining a set of candidate sparse structures and then Akaike's Information Criterion (AIC) is applied to optimize the block-diagonal structure approximation.

The reminder of the paper is organized as follows. In Section 2, we define population measures of skewness and kurtosis along with corresponding cumulant-based expressions, establish explicit inequality relating skewness and kurtosis, and provide some examples of applications of newly derived characteristics. In Section 3, we derive main asymptotic results under multivariate normality and show how the suggested estimators of skewness and kurtosis can be improved by the exact bias correction and normalizing transformations. In Section 4, we suggest an algorithm for estimating the covariance structure and show its optimal properties. Section 5 summarizes Monte Carlo simulation experiments and the validity of suggested test procedures under different parameter settings.

§2. New measures of multivariate skewness and kurtosis and their characterizations

In this section, we derive our new measures of multivariate skewness and kurtosis. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be independent, identically distributed, p-dimensional random vectors, with $\boldsymbol{x}, \boldsymbol{y} \sim \mathcal{F}$, where $\mathcal{F}$ denotes the distribution function with $E(\boldsymbol{x}) = E(\boldsymbol{y}) = \boldsymbol{\mu}$ and $Cov(\boldsymbol{x}) = Cov(\boldsymbol{y}) = \Sigma$. Assume further that $\boldsymbol{x}$ and $\boldsymbol{y}$ can be partitioned into non-empty, disjoint independent subsets as

$$\boldsymbol{x} = (\boldsymbol{x}^{(1)\prime}, \boldsymbol{x}^{(2)\prime}, \ldots, \boldsymbol{x}^{(k)\prime})' \text{ and } \boldsymbol{y} = (\boldsymbol{y}^{(1)\prime}, \boldsymbol{y}^{(2)\prime}, \ldots, \boldsymbol{y}^{(k)\prime})',$$

where $\boldsymbol{x}^{(l)} = (x_1^{(l)}, x_2^{(l)}, \ldots, x_{p_l}^{(l)})'$, $\boldsymbol{y}^{(l)} = (y_1^{(l)}, y_2^{(l)}, \ldots, y_{p_l}^{(l)})'$ are p_l-vectors and $\sum_{l=1}^{k} p_l = p$. Then $\boldsymbol{\mu} = (\boldsymbol{\mu}^{(1)'}, \boldsymbol{\mu}^{(2)'}, \ldots, \boldsymbol{\mu}^{(k)'})'$ and $\Sigma = \mathrm{diag}(\Sigma_1, \Sigma_2, \ldots, \Sigma_k)$, where Σ_l is $p_l \times p_l$ matrix. We define our measures of skewness and kurtosis as

$$\beta_{h,1} \equiv \sum_{l=1}^{k} E\left[\{(\boldsymbol{x}^{(l)} - \boldsymbol{\mu}^{(l)})'\Sigma_l^{-1}(\boldsymbol{y}^{(l)} - \boldsymbol{\mu}^{(l)})\}^3\right] = \sum_{l=1}^{k} \beta_{h,1}^{(l)},$$

$$\beta_{h,2} \equiv \sum_{l=1}^{k} E\left[\{(\boldsymbol{x}^{(l)} - \boldsymbol{\mu}^{(l)})'\Sigma_l^{-1}(\boldsymbol{x}^{(l)} - \boldsymbol{\mu}^{(l)})\}^2\right] = \sum_{l=1}^{k} \beta_{h,2}^{(l)},$$

respectively, which are natural extensions of original Mardia [15]. We note that $\beta_{h,1} = 0$ and $\beta_{h,2} = \sum_{l=1}^{k} p_l(p_l + 2)$ hold under assumption that $\mathcal{F}$ is the distribution function of $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$. Observe that both $\beta_{h,1}$ and $\beta_{h,2}$ are invariant measures with respect to nonsingular transformation

$$\boldsymbol{x} = A^*\boldsymbol{u} + \boldsymbol{b}, \tag{2.1}$$

where $A^* = \mathrm{diag}(A_1, A_2, \ldots, A_k)$, A_l is a non-singular $p_l \times p_l$ matrix ($l = 1, 2, \ldots, k$) and $\boldsymbol{b} \in \mathbb{R}^p$. Using invariance property, we, without loss of generality, assume that $\boldsymbol{\mu} = \boldsymbol{0}$ and $\Sigma = I_p$, and derive the relationship of skewness and kurtosis. For any $a_0, a_1, a_2 \in \mathbb{R}$,

$$E\left[\left(a_0 + a_1 \sum_{i=1}^{p_l} x_i^{(l)} + a_2 \sum_{i=1}^{p_l} x_i^{(l)^2}\right)^2\right] \geq 0 \tag{2.2}$$

hold. Due to independence of $\boldsymbol{x}^{(l)}$'s and by the assumptions on $\boldsymbol{\mu}$ and Σ, we obtain

$$E\left[\left(\sum_{i=1}^{p_l} x_i^{(l)}\right)^2\right] = p_l, \ E\left[\left(\sum_{i=1}^{p_l} x_i^{(l)^2}\right)^2\right] = \beta_{h,2}^{(l)}.$$

Further, by applying technique from Kollo and Srivastava [11] for each l block after some vector algebra, we obtain

$$E\left[\left(\sum_{i=1}^{p_l} x_i^{(l)}\right)\left(\sum_{i=1}^{p_l} x_i^{(l)^2}\right)\right] = p_l \beta_{h,1}^{(l)}.$$

Now, by putting $a_0 = p_l^2$, $a_1 = p_l \beta_{h,1}^{(l)}$, $a_2 = p_l$ into (2.2), we see that $\beta_{h,2}^{(l)} \geq p_l^2 + \beta_{h,1}^{(l)}$, which in turn provides the following inequality:

$$\beta_{h,2} \geq \sum_{l=1}^{k} p_l^2 + \beta_{h,1}.$$

We further obtain expression for population measures $\beta_{h,1}$ and $\beta_{h,2}$ in terms of cumulants. We first define $K_{h,1}$ and $K_{h,2}$ as

$$K_{h,1} = \sum_{l=1}^{k} \sum_{r,r'=1}^{p_l} \sum_{s,s'=1}^{p_l} \sum_{t,t'=1}^{p_l} (\kappa_{11(l)}^{(rr')})^{-1} (\kappa_{11(l)}^{(ss')})^{-1} (\kappa_{11(l)}^{(tt')})^{-1} \kappa_{111(l)}^{(rst)} \kappa_{111(l)}^{(r's't')}, \tag{2.3}$$

$$K_{h,2} = \sum_{l=1}^{k} \sum_{r,r'=1}^{p_l} \sum_{s,s'=1}^{p_l} (\kappa_{11(l)}^{(rr')})^{-1} (\kappa_{11(l)}^{(ss')})^{-1} \kappa_{1111(l)}^{(rr'ss')}, \tag{2.4}$$

where $\kappa_{1111(l)}^{(rstu)}$ denotes the cumulant of order $(1,1,1,1)$ for the random variables $(x_r^{(l)}, x_s^{(l)}, x_t^{(l)}, x_u^{(l)})$ and $r, s, t, u = 1, 2, \ldots, p_l$. Now, by calculating moments and by using the relationship between cumulants and moments (see, Kendall and Stuart [8], p.84), we obtain

$$K_{h,1} = \beta_{h,1}, \tag{2.5}$$

$$K_{h,2} = \beta_{h,2} - \sum_{l=1}^{k} p_l(p_l + 2). \tag{2.6}$$

We mention two examples where our suggested measures might be useful.

Example 1: *Mixture of multivariate normal distributions with common covariance matrices.*

The random vector $\boldsymbol{x}$ is said to have a mixture multivariate normal distribution if $\boldsymbol{x}$ has the probability density function (p.d.f.)

$$\epsilon\phi(\boldsymbol{x}; \boldsymbol{\mu}_1, \Sigma) + \epsilon'\phi(\boldsymbol{x}; \boldsymbol{\mu}_2, \Sigma), \tag{2.7}$$

for some p-dimensional vector $\boldsymbol{\mu}_j$, $j = 1, 2$ and some non-singular matrix Σ with $0 < \epsilon < 1$, $\epsilon' = 1 - \epsilon$, and ϕ is the p.d.f. of normal distribution.

Day [3] has proposed to use Mahalanobis distance, $\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ as a measure of non-normality for this population and has derived test of normality based on an estimator of Δ. We assume that $\boldsymbol{\mu}_j = (\boldsymbol{\mu}_j^{(1)'}, \boldsymbol{\mu}_j^{(2)'}, \ldots, \boldsymbol{\mu}_j^{(k)'})'$, $j = 1, 2$ and $\Sigma = \text{diag}(\Sigma_1, \Sigma_2 \ldots, \Sigma_k)$, and show for this mixture that $K_{h,1}$ and $K_{h,2}$ can be expressed as functions of Δ_l which is the Mahalanobis distance for l's block, defined as

$$\Delta_l^2 = (\boldsymbol{\mu}_1^{(l)} - \boldsymbol{\mu}_2^{(l)})'\Sigma_l^{-1}(\boldsymbol{\mu}_1^{(l)} - \boldsymbol{\mu}_2^{(l)}).$$

By making suitable non-singular transformation for each block and by independence of $\boldsymbol{x}^{(l)}$'s, we find that the p.d.f. in (2.7) reduces to

$$\begin{aligned}&\prod_{l=1}^{k}\Big[\epsilon\phi(\boldsymbol{x}^{(l)};\boldsymbol{\mu}_1^{(l)},\Sigma_l)+\epsilon'\phi(\boldsymbol{x}^{(l)};\boldsymbol{\mu}_2^{(l)},\Sigma_l)\Big]\\ &=\prod_{l=1}^{k}\{\epsilon\phi(x_1^{(l)}-\Delta_l)+\epsilon'\phi(x_1^{(l)})\}\prod_{i=2}^{p_l}\phi(x_i^{(l)}).\end{aligned}$$

Observe that $K_{h,1}$ and $K_{h,2}$ remain unchanged since they are invariant under linear transformations.

We show for this population that $K_{h,1}$ and $K_{h,2}$ are some function of Δ_l so that a possible test of multivariate normality can be considered. By independence of $\boldsymbol{x}^{(l)}$'s, the moment generating function of $\boldsymbol{x}$, $m_{\boldsymbol{x}}(\boldsymbol{t})$, can be expressed as

$$\begin{aligned}m_{\boldsymbol{x}}(\boldsymbol{t})&=E[e^{\boldsymbol{t}'\boldsymbol{x}}]=E[e^{\boldsymbol{t}^{(1)'}\boldsymbol{x}^{(1)}+\boldsymbol{t}^{(2)'}\boldsymbol{x}^{(2)}+\cdots+\boldsymbol{t}^{(k)'}\boldsymbol{x}^{(k)}}]\\ &=E[e^{\boldsymbol{t}^{(1)'}\boldsymbol{x}^{(1)}}]E[e^{\boldsymbol{t}^{(2)'}\boldsymbol{x}^{(2)}}]\cdots E[e^{\boldsymbol{t}^{(k)'}\boldsymbol{x}^{(k)}}]\\ &=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{\boldsymbol{t}^{(1)'}\boldsymbol{x}^{(1)}}\{\epsilon\phi(x_1^{(1)}-\Delta_1)+\epsilon'\phi(x_1^{(1)})\}\prod_{i=2}^{p_1}\phi(x_i^{(1)})d\boldsymbol{x}^{(1)}\cdots\\ &\qquad\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{\boldsymbol{t}^{(k)'}\boldsymbol{x}^{(k)}}\{\epsilon\phi(x_1^{(k)}-\Delta_k)+\epsilon'\phi(x_1^{(k)})\}\prod_{i=2}^{p_k}\phi(x_i^{(k)})d\boldsymbol{x}^{(k)}\end{aligned}$$

for $\boldsymbol{t}=(\boldsymbol{t}^{(1)'},\boldsymbol{t}^{(2)'},\ldots,\boldsymbol{t}^{(k)'})'\in\mathbb{R}^p$ and $\boldsymbol{t}^{(l)}=(t_1^{(l)},t_2^{(l)},\ldots,t_{p_l}^{(l)})'$. Since the expressions of each product term are the same it is sufficient to evaluate it for one block. We get

$$\begin{aligned}&\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{\boldsymbol{t}^{(l)'}\boldsymbol{x}^{(k)}}\{\epsilon\phi(x_1^{(l)}-\Delta_l)+\epsilon'\phi(x_1^{(l)})\}\prod_{i=2}^{p_l}\phi(x_i^{(l)})d\boldsymbol{x}^{(l)}\\ &=\prod_{i=2}^{p_l}\exp\Big[\frac{t_i^{(l)^2}}{2}\Big]\Big(\epsilon'\exp\Big[\frac{t_1^{(l)^2}}{2}\Big]+\epsilon\int_{-\infty}^{\infty}e^{t_1^{(l)}x_1^{(l)}}\phi(x_1^{(l)}-\Delta_l)dx_1^{(l)}\Big)\\ &=\prod_{i=1}^{p_l}\exp\Big[\frac{t_i^{(l)^2}}{2}\Big](\epsilon'+\epsilon\exp\big[t_1^{(l)}\Delta_l\big]).\end{aligned}$$

Now we obtain the cumulant generating function

$$\sum_{l=1}^{k}\sum_{i=1}^{p_l}\frac{t_i^{(l)^2}}{2}+\sum_{l=1}^{k}\log(\epsilon'+\epsilon\exp\big[t_1^{(l)}\Delta_l\big]).\tag{2.8}$$

Since cross cumulants are equal to zero, we can obtain cumulants by differentiating (2.8). After a deal of calculations it can be found that

$$K_{h,1} = \sum_{l=1}^{k} \frac{(\epsilon\epsilon'(\epsilon-\epsilon')\Delta_l^3)^2}{(\epsilon^2+(2+\Delta_l^2)\epsilon\epsilon'+\epsilon'^2)^3} = \sum_{l=1}^{k} \frac{(\epsilon\epsilon'(\epsilon-\epsilon')\Delta_l^3)^2}{(1+\Delta_l^2\epsilon\epsilon')^3},$$

$$K_{h,2} = \sum_{l=1}^{k} \frac{\epsilon\epsilon'(\epsilon^2-4\epsilon\epsilon'+\epsilon'^2)\Delta_l^4}{(\epsilon^2+(2+\Delta_l^2)\epsilon\epsilon'+\epsilon'^2)^2} = \sum_{l=1}^{k} \frac{\epsilon\epsilon'(1-6\epsilon\epsilon')\Delta_l^4}{(1+\Delta_l^2\epsilon\epsilon')^2}.$$

Hence $K_{h,1}$ and $K_{h,2}$ (essentially $\beta_{h,1}$ and $\beta_{h,2}$) are the functions of Δ_l. Clearly, when $\Delta_l = 0$ $(l = 1, 2, \ldots k)$, $K_{h,1}$ and $K_{h,2}$ are zero in accordance with the property of original Mardia's measures. It is important to note that the original Day [3] test was derived in large sample context whereas our approach accommodates both large sample and high-dimensional cases.

Example 2: *Multivariate Laplace distribution with block diagonal covariance structure.*

The random vector $\boldsymbol{x}$ is said to have a multivariate Laplace distribution (denoted by $\boldsymbol{x} \sim ML_p(\boldsymbol{\mu}, \Sigma)$), where $E(\boldsymbol{x}) = \boldsymbol{\mu}$ and $Cov(\boldsymbol{x}) = \Sigma$. We assume that $\boldsymbol{\mu} = (\boldsymbol{\mu}^{(1)'}, \boldsymbol{\mu}^{(2)'}, \ldots, \boldsymbol{\mu}^{(k)'})'$ and $\Sigma = \mathrm{diag}(\Sigma_1, \Sigma_2 \ldots, \Sigma_k)$. Let $d_l = \boldsymbol{\mu}^{(l)'}\Sigma_l^{-1}\boldsymbol{\mu}^{(l)}$, $l = 1, 2, \ldots, k$. By using the relationships (2.5) and (2.6) and applying the technique by Kollo and Srivastava [11], we obtain

$$\beta_{h,1}^{(l)} = d_l(d_l^2 - 6d_l + 3(p_l+2)), \tag{2.9}$$

$$\beta_{h,2}^{(l)} = (p_l + 2d_l)(p_l+2) - 3d_l^2 \tag{2.10}$$

for each block[1]. Hence we obtain

$$\beta_{h,1} = \sum_{l=1}^{k} d_l(d_l^2 - 6d_l + 3(p_l+2)), \tag{2.11}$$

$$\beta_{h,2} = \sum_{l=1}^{k} \{(p_l + 2d_l)(p_l+2) - 3d_l^2\}. \tag{2.12}$$

When $d_l = 0, l = 1, 2, \ldots k$, $\beta_{h,1} = 0$ and $\beta_{h,2} = \sum_{l=1}^{k} p_l(p_l+2)$.

We now define sample counter-parts of $\beta_{h,1}$ and $\beta_{h,2}$. Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ be sample observation vectors of size N from a multivariate population with the mean vector $\boldsymbol{\mu}$ and the covariance matrix Σ, where $\boldsymbol{x}_j = (\boldsymbol{x}_j^{(1)'}, \boldsymbol{x}_j^{(2)'}, \ldots, \boldsymbol{x}_j^{(k)'})$ $= (x_{1j}^{(1)}, \ldots, x_{p_1 j}^{(1)}, x_{1j}^{(2)}, \ldots, x_{p_k j}^{(k)})$ $(j = 1, 2, \ldots, N)$. Let also $\overline{\boldsymbol{x}} = (\overline{\boldsymbol{x}}^{(1)'}, \overline{\boldsymbol{x}}^{(2)'}, \ldots,$

[1] observe that the corrected result for $\beta_{h,2}^{(l)}$ is obtained when justifying (2.10). For the sake of space we do not present the details of calculations here.

$\overline{\boldsymbol{x}}^{(k)\prime}$) and $S = \mathrm{diag}(S_1, S_2, \ldots, S_k)$ denote the sample mean vector and the sample covariance matrix, respectively, based on sample size N. Then

$$\overline{\boldsymbol{x}}^{(l)} = \frac{1}{N}\sum_{j=1}^{N}\boldsymbol{x}_j^{(l)} = (\overline{x}_1^{(1)}, \ldots, \overline{x}_{p_1}^{(1)}, \overline{x}_1^{(2)}, \ldots, \overline{x}_{p_k}^{(k)}),$$

$$S_l = \frac{1}{N}\sum_{j=1}^{N}(\boldsymbol{x}_j^{(l)} - \overline{\boldsymbol{x}}^{(l)})(\boldsymbol{x}_j^{(l)} - \overline{\boldsymbol{x}}^{(l)})'.$$

Sample measures of multivariate skewness and kurtosis are defined as

$$b_{h,1} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{k}\{(\boldsymbol{x}_i^{(l)} - \overline{\boldsymbol{x}}^{(l)})' S_l^{-1}(\boldsymbol{x}_j^{(l)} - \overline{\boldsymbol{x}}^{(l)})\}^3, \tag{2.13}$$

$$b_{h,2} = \frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{k}\{(\boldsymbol{x}_i^{(l)} - \overline{\boldsymbol{x}}^{(l)})' S_l^{-1}(\boldsymbol{x}_i^{(l)} - \overline{\boldsymbol{x}}^{(l)})\}^2, \tag{2.14}$$

respectively.

§3. Asymptotic properties of $b_{h,1}$ and $b_{h,2}$ and improved test statistics

By using similar way of Mardia [15], we obtained the following lemma.

Lemma 1. *When p_l and k are fixed, the expectation of $b_{h,1}$ in* (2.13) *and the expectation and the variance of $b_{h,2}$ in* (2.14) *when the population is $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ and Σ has a block diagonal structure are given by*

$$E[b_{h,1}] = \frac{1}{N}\sum_{l=1}^{k} p_l(p_l+1)(p_l+2) + o(N^{-1}), \tag{3.1}$$

$$E[b_{h,2}] = \frac{N-1}{N+1}\sum_{l=1}^{k} p_l(p_l+2), \tag{3.2}$$

$$Var[b_{h,2}] = \frac{8}{N}\sum_{l=1}^{k} p_l(p_l+2) + o(N^{-1}). \tag{3.3}$$

Proof. We note that probability density function of $\boldsymbol{x}$ is

$$f(\boldsymbol{x}) = \prod_{l=1}^{k}\frac{1}{(2\pi)^{\frac{p_l}{2}}|\Sigma_l|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\boldsymbol{x}^{(l)} - \boldsymbol{\mu}^{(l)})'\Sigma_l^{-1}(\boldsymbol{x}^{(l)} - \boldsymbol{\mu}^{(l)})\right].$$

Hence, we find the independence of $\boldsymbol{x}^{(l)}$ and $\boldsymbol{x}^{(l')}$ $(l \neq l',\ l, l' = 1, 2, \ldots, k)$. Similar to Mardia [15], we rewrite (2.13) as

$$b_{h,1} = \sum_{l=1}^{k} \sum_{r,r'=1}^{p_l} \sum_{s,s'=1}^{p_l} \sum_{t,t'=1}^{p_l} S_{(l)}^{rr'} S_{(l)}^{ss'} S_{(l)}^{tt'} M_{111(l)}^{(rst)} M_{111(l)}^{(r's't')}, \tag{3.4}$$

where

$$S_l^{-1} = \{S_{(l)}^{ij}\} \text{ and } M_{111(l)}^{(rst)} = \frac{1}{N} \sum_{i=1}^{N} (x_{ri}^{(l)} - \overline{x}_r^{(l)})(x_{si}^{(l)} - \overline{x}_s^{(l)})(x_{ti}^{(l)} - \overline{x}_t^{(l)}).$$

Since $b_{h,1}$ is invariant under a linear transformation, we assume that $\boldsymbol{\mu} = \boldsymbol{0}$ and $\Sigma = I_p$. S_l converges to Σ_l in probability, respectively $(l = 1, 2, \ldots, k)$. Hence, from (3.4), we obtain

$$b_{h,1} \xrightarrow{p} \sum_{l=1}^{k} \sum_{r}^{p_l} \sum_{s}^{p_l} \sum_{t}^{p_l} (M_{111(l)}^{(rst)})^2 \tag{3.5}$$

$$= \sum_{l=1}^{k} (M_{3(l)}^{(1)})^2 + \cdots + 3 \sum_{l=1}^{k} (M_{21(l)}^{(12)})^2 + \cdots + 6 \sum_{l=1}^{k} (M_{111(l)}^{(123)})^2 + \cdots \tag{3.6}$$

in probability, where $M_{111(l)}^{(rrr)} = M_{3(l)}^{(r)}$ and $M_{111(l)}^{(rss)} = M_{12(l)}^{(rs)}$ $(r \neq s)$. By using the normality of $(M_{3(l)}^{(1)}, \ldots, M_{21(l)}^{(12)}, \ldots, M_{111(l)}^{(123)}, \ldots)$, we can get

$$\begin{aligned} E[b_{h,1}] &= \frac{1}{N} p_1(p_1+1)(p_1+2) + \frac{1}{N} p_2(p_2+1)(p_2+2) \\ &\quad + \cdots + \frac{1}{N} p_k(p_k+1)(p_k+2) + o(N^{-1}) \\ &= \frac{1}{N} \sum_{l=1}^{k} p_l(p_l+1)(p_l+2) + o(N^{-1}). \end{aligned}$$

And we let $\boldsymbol{x}_{r(l)}^* = (x_{r1}^{(l)}, x_{r2}^{(l)}, \ldots, x_{rN}^{(l)})'$, $(r = 1, 2, \ldots, p_l)$. We consider an orthogonal transformation $\boldsymbol{z}_{r(l)}^* = H_l \boldsymbol{x}_{r(l)}^* = (z_{r1}^{(l)}, z_{r2}^{(l)}, \ldots, z_{rN}^{(l)})'$, where H_l is an orthogonal matrix with the first row as

$$\left(\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{N}}, \ldots, \frac{1}{\sqrt{N}} \right),$$

and the second row as

$$\left(-\frac{1}{\sqrt{N(N-1)}}, -\frac{1}{\sqrt{N(N-1)}}, \ldots, -\frac{1}{\sqrt{N(N-1)}}, \sqrt{\frac{N-1}{N}} \right).$$

Then we find that

$$\begin{aligned}
E[b_{h,2}] &= \sum_{l=1}^{k} E\left[\frac{1}{N}\sum_{i=1}^{N}\{(\boldsymbol{x}_i^{(l)} - \overline{\boldsymbol{x}}^{(l)})' S_l^{-1}(\boldsymbol{x}_i^{(l)} - \overline{\boldsymbol{x}}^{(l)})\}^2\right] \\
&= \sum_{l=1}^{k} E\left[\frac{1}{N}\sum_{i=1}^{N}\{(H_l\boldsymbol{x}_i^{(l)} - H_l\overline{\boldsymbol{x}}^{(l)})'(H_l S_l H_l')^{-1}(H_l\boldsymbol{x}_i^{(l)} - H_l\overline{\boldsymbol{x}}^{(l)})\}^2\right] \\
&= \sum_{l=1}^{k}(N-1)^2 E[y_l^2],
\end{aligned}$$

where

$$y_l = \boldsymbol{z}_2^{(l)\prime}\left(\sum_{s=2}^{N}\boldsymbol{z}_s^{(l)}\boldsymbol{z}_s^{(l)\prime}\right)^{-1}\boldsymbol{z}_2^{(l)},\ \ \boldsymbol{z}_s^{(l)} = (z_{1s}^{(l)}, z_{2s}^{(l)}, \dots, z_{p_l s}^{(l)})',\ \ s = 2, 3, \dots, N.$$

Since $\boldsymbol{z}_s^{(l)}$ is distributed as $\mathcal{N}_{p_l}(\boldsymbol{0}, I_{p_l})$ (see, e.g. Kendall and Stuart [7] p.229), by an well known result of y_l, we can get

$$E[b_{h,2}] = \sum_{l=1}^{k}(N-1)^2\frac{p_l(p_l+2)}{(N+1)(N-1)} = \frac{N-1}{N+1}\sum_{l=1}^{k}p_l(p_l+2).$$

Finally we consider the asymptotic variance of $b_{h,2}$. By similar way of Mardia [15], we evaluate the value of $Var[b_{h,2}]$ up to $o(N^{-1})$. Let $S_l = I_{p_l} + S_l^*$ so that $o(N^{-1})$ and $E[S_l^*] = 0$. Then we expand S_l^{-1} as

$$S_l^{-1} = (I + S_l^*)^{-1} = I - S_l^* + S_l^{*2} - \cdots .$$

Hence we get

$$\begin{aligned}
b_{h,2} =& \frac{1}{N}\sum_{l=1}^{k}\sum_{i=1}^{N}\{(\boldsymbol{x}_i^{(l)} - \overline{\boldsymbol{x}}^{(l)})'(\boldsymbol{x}_i^{(l)} - \overline{\boldsymbol{x}}^{(l)})\}^2 \\
&- \frac{2}{N}\sum_{l=1}^{k}\sum_{i=1}^{N}(\boldsymbol{x}_i^{(l)} - \overline{\boldsymbol{x}}^{(l)})'(\boldsymbol{x}_i^{(l)} - \overline{\boldsymbol{x}}^{(l)})(\boldsymbol{x}_i^{(l)} - \overline{\boldsymbol{x}}^{(l)})' S_l^*(\boldsymbol{x}_i^{(l)} - \overline{\boldsymbol{x}}^{(l)}) + \cdots . \qquad (3.7)
\end{aligned}$$

Put

$$\begin{aligned}
M_{i_1,\dots,i_t,(l)}^{(j_1,\dots,j_t)} &= \frac{1}{N}\sum_{i=1}^{N}\prod_{r=1}^{t}(x_{j_r i}^{(l)} - \overline{x}_{j_r}^{(l)})^{i_r}, \\
M_{2(l)}^{(i)*} &= S_{ii(l)}^*,\ M_{11(l)}^{(ij)*} = S_{ij(l)}^*,\ S_l^* = \{S_{ij(l)}^*\},
\end{aligned}$$

we can rewrite (3.7) as

$$
\begin{aligned}
b_{h,2} =& \sum_{l=1}^{k}\sum_{i=1}^{p_l} M_{4(l)}^{(i)} + \sum_{l=1}^{k}\sum_{i\neq j}^{p_l} M_{22(l)}^{(ij)} - 2\sum_{l=1}^{k}\sum_{i=1}^{p_l} M_{2(l)}^{(i)*} M_{4(l)}^{(i)} \\
& - 2\sum_{l=1}^{k}\sum_{i\neq j}^{p_l} M_{2(l)}^{(j)*} M_{22(l)}^{(ij)} - 2\sum_{l=1}^{k}\sum_{i=1}^{p_l}\sum_{j\neq k}^{p_l} M_{11(l)}^{(jk)*} M_{211(l)}^{(ijk)} - \cdots .
\end{aligned}
$$

By using asymptotic formula in Mardia [15], we can obtain

$$
Var[b_{h,2}] = \frac{8}{N}\sum_{l=1}^{k} p_l(p_l+2) + o(N^{-1}).
$$

□

For asymptotic distributions of skewness and kurtosis, we derive the following theorem:

Theorem 1. *Let* $b_{h,1}$ *and* $b_{h,2}$ *in* (2.13) *and* (2.14) *are sample measures of multivariate skewness and multivariate kurtosis on the basis of random samples of size* N *drawn from* $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ *and* Σ *has a block diagonal structure. Then, for large* N*,*

$$
z_{h,1} = \frac{N}{6} b_{h,1} \tag{3.8}
$$

has a χ^2*-distribution with* $\sum_{l=1}^{k} p_l(p_l+1)(p_l+2)/6$ *degrees of freedom and*

$$
z_{h,2} = \frac{b_{h,2} - \dfrac{N-1}{N+1}\sum\limits_{l=1}^{k} p_l(p_l+2)}{\sqrt{\dfrac{8}{N}\sum\limits_{l=1}^{k} p_l(p_l+2)}} \tag{3.9}
$$

is distributed as $\mathcal{N}(0,1)$*.*

Proof. From (3.6), we can rewrite $b_{h,1}$ as

$$
b_{h,1} = \sum_{l=1}^{k}(M_{3(l)}^{(1)})^2 + \cdots + 3\sum_{l=1}^{k}(M_{21(l)}^{(12)})^2 + \cdots + 6\sum_{l=1}^{k}(M_{111(l)}^{(123)})^2 + \cdots \tag{3.10}
$$

$$
= \sum_{l=1}^{k}\left\{(M_{3(l)}^{(1)})^2 + \cdots + 3(M_{21(l)}^{(12)})^2 + \cdots + 6(M_{111(l)}^{(123)})^2 + \cdots\right\}. \tag{3.11}
$$

We consider the following statistic for each block

$$\frac{N\{(M^{(1)}_{3(l)})^2+\cdots+3(M^{(12)}_{21(l)})^2+\cdots+6(M^{(123)}_{111(l)})^2+\cdots\}}{6}. \tag{3.12}$$

By using the limiting distributions of quadratic form, (3.12) has a χ^2-distribution with $p_l(p_l+1)(p_l+2)/6$ degrees of freedom, respectively. We note that these statistics are mutually independent, we can obtain

$$z_{h,1}=\frac{Nb_{h,1}}{6}\sim\chi^2_f$$

for large N, where $f=\sum_{l=1}^{k}p_l(p_l+1)(p_l+2)/6$.

On using results given by (3.2) and (3.3) and the central limit theorem, $z_{h,2}$ in (3.9) has a standard normal distribution. □

One of possible applications of Theorem 1 is to use $z_{h,1}$ and $z_{h,2}$ for testing multivariate normality in high-dimensional setting. We are interested in testing $H_0:\mathcal{F}$ is the distribution function of $\mathcal{N}_p(\boldsymbol{\mu},\Sigma)$ versus H_1: not H_0 when $\Sigma=\mathrm{diag}(\Sigma_1,\Sigma_2,\ldots,\Sigma_k)$ and $\boldsymbol{\mu}=(\boldsymbol{\mu}^{(1)\prime},\boldsymbol{\mu}^{(2)\prime},\ldots,\boldsymbol{\mu}^{(k)\prime})'$ using N observations $\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_N$ coming from the population with distribution $\mathcal{F}$. Our simulation experiments presented in Section 5 indicate that the test performance accuracy is poor; this is due to the order of asymptotic moments in (3.1) and (3.3). Clearly when $p=o(N)$, the effect of the residual terms in (3.1) and (3.3) will be pronounced. Therefore, to improve test accuracy we suggest modification of $z_{h,1}$ and $z_{h,2}$. First we need to derive exact moments of $b_{h,1}$ and $b_{h,2}$ which are given in the following lemma:

Lemma 2. *When p_l and k are fixed, the exact expectation of $b_{h,1}$ in* (2.13) *and the exact expectation and the exact variance of $b_{h,2}$ in* (2.14) *when the population is $\mathcal{N}_p(\boldsymbol{\mu},\Sigma)$ and Σ has a block diagonal structure are given by*

$$E[b_{h,1}]=\frac{1}{(N+1)(N+3)}\sum_{l=1}^{k}p_l(p_l+2)\{(N+1)(p_l+1)-6\}, \tag{3.13}$$

$$E[b_{h,2}]=\frac{N-1}{N+1}\sum_{l=1}^{k}p_l(p_l+2), \tag{3.14}$$

$$Var[b_{h,2}]=\sum_{l=1}^{k}\frac{8p_l(p_l+2)}{(N+1)^2(N+3)(N+5)}(N-p_l-1)(N-p_l+1). \tag{3.15}$$

Proof. To use Khatri and Pillai's [9] results, (2.13) and (2.14) are expressed as follows:

$$b_{h,1} = N \sum_{l=1}^{k} \sum_{i=1}^{N} \sum_{j=1}^{N} R_{ij(l)}^3,$$
$$b_{h,2} = N \sum_{l=1}^{k} \sum_{i=1}^{N} R_{ii(l)}^2,$$

where

$$R_{ij(l)} = (\boldsymbol{x}_i^{(l)} - \overline{\boldsymbol{x}}^{(l)})'(NS_l)^{-1}(\boldsymbol{x}_j^{(l)} - \overline{\boldsymbol{x}}^{(l)}). \tag{3.16}$$

Since $b_{h,1}$ and $b_{h,2}$ are invariant under linear transformation in (2.1), we assume without loss of generality $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I_p$. Put $\boldsymbol{x}_{r(l)}^* = (x_{r1}^{(l)}, x_{r2}^{(l)}, \ldots, x_{rN}^{(l)})'$ $(r = 1, 2, \ldots, p_l)$, then transform $\boldsymbol{x}_{r(l)}^*$ to $\boldsymbol{\xi}_{r(l)}^* = (\xi_{r1}^{(l)}, \xi_{r2}^{(l)}, \ldots, \xi_{rN}^{(l)})'$ so that

$$\xi_{r,i-1}^{(l)} = \sqrt{\frac{i-1}{i}} \Big(- x_{ri}^{(l)} - \frac{1}{i-1} \sum_{i'=1}^{i-1} x_{ri'}^{(l)} \Big), \; i = 2, 3, \ldots, N$$
$$\xi_{rN}^{(l)} = \sqrt{N} \overline{x}_r^{(l)}.$$

This is called Helmert orthogonal transformation. By this transformation, we get

$$\boldsymbol{x}_i^{(l)} - \overline{\boldsymbol{x}}^{(l)} = -a_i \boldsymbol{\xi}_{i-1(l)} + \sum_{i'=i}^{N-1} b_{i'} \boldsymbol{\xi}_{i'(l)}, \; i = 1, 2, \ldots N,$$
$$\sum_{i=1}^{N} (\boldsymbol{x}_i^{(l)} - \overline{\boldsymbol{x}}^{(l)})(\boldsymbol{x}_i^{(l)} - \overline{\boldsymbol{x}}^{(l)})' = \sum_{i=1}^{N-1} \boldsymbol{\xi}_{i(l)} \boldsymbol{\xi}_{i(l)}' = T_l T_l',$$

where

$$\boldsymbol{\xi}_{0(l)} = \mathbf{0}, \; \boldsymbol{\xi}_{N(l)} = -a_N \boldsymbol{\xi}_{N-1(l)}, \; \boldsymbol{\xi}_{i(l)} = (\xi_{1i}^{(l)}, \xi_{2i}^{(l)}, \ldots, \xi_{p_l i}^{(l)})',$$
$$a_i = \sqrt{\frac{i-1}{i}}, \; b_i = \frac{1}{\sqrt{i(i+1)}}.$$

Then exact moments of $\boldsymbol{z}_i^{(l)} = T_l^{-1} \boldsymbol{\xi}_{i(l)}$, $i = 1, 2, \ldots, N$ are given by Khatri and Pillai [9]. And $R_{ij(l)}$ in (3.16) is expressed as

(3.17)

$$R_{ij(l)} = \Big(- a_i \boldsymbol{z}_{i-1}^{(l)} + \sum_{i'=i}^{N-1} b_{i'} \boldsymbol{z}_{i'}^{(l)} \Big)' \Big(- a_j \boldsymbol{z}_{j-1}^{(l)} + \sum_{i'=j}^{N-1} b_{i'} \boldsymbol{z}_{i'}^{(l)} \Big), \; i, j = 1, 2, \ldots, N.$$

Hence from the moments of $\boldsymbol{z}_1^{(l)}, \boldsymbol{z}_2^{(l)}, \ldots, \boldsymbol{z}_{N-1}^{(l)}$ exact moments of $b_{h,1}$ and $b_{h,2}$ can be obtained. □

Now, by using Lemma 2 and Theorem 1, we propose the following improved statistics.

Theorem 2. *Let $b_{h,1}$ and $b_{h,2}$ in (2.13) and (2.14) are sample measures of multivariate skewness and multivariate kurtosis on the basis of random samples of size N drawn from $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ and Σ has a block diagonal structure. Then, for large N and $N - p_l - 1 > 0$,*

$$z^*_{h,1} = \frac{N}{6} \sum_{l=1}^{k} \frac{(p_l+1)(N+1)(N+3)}{N\{(N+1)(p_l+1)-6\}} b_{h,1} \tag{3.18}$$

has a χ^2-distribution with $\sum_{l=1}^{k} p_l(p_l+1)(p_l+2)/6$ degrees of freedom and

$$z^*_{h,2} = \frac{\{(N+1)b_{h,2} - \sum_{l=1}^{k} p_l(p_l+2)(N-1)\}\sqrt{(N+3)(N+5)}}{\sqrt{8\sum_{l=1}^{k} p_l(p_l+2)(N-3)(N-p_l-1)(N-p_l+1)}} \tag{3.19}$$

is distributed as $\mathcal{N}(0,1)$.

Observe that $b_{h,1}$ is an estimator for the population parameter $\beta_{h,1}$ which is zero not only in case of normality but also for the wider class of all elliptically symmetric distributions; see, e.g. Baringhaus and Henze [1]. Therefore, the test for multivariate normality based on $z^*_{h,1}$ must be considered only against alternative distributions having positive multivariate skewness. Following this, we propose another modification of $z^*_{h,1}$ based on Wilson-Hilferty transformation (Wilson and Hilferty [21]), an effective and simple transform of $z^*_{h,1}$ to standard normal distribution.

Theorem 3. *Let $b_{h,1}$ in (2.13) be a sample measure of multivariate skewness on the basis of random samples of size N drawn from $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ Then*

$$z_{wh} = \left\{ \left(\frac{z^*_{h,1}}{f} \right)^{\frac{1}{3}} - 1 + \frac{2}{9f} \right\} \Big/ \sqrt{\frac{2}{9f}}, \quad f = \frac{1}{6} \sum_{l=1}^{k} p_l(p_l+1)(p_l+2) \tag{3.20}$$

is distributed as $\mathcal{N}(0,1)$ when $f \to \infty$ after $N \to \infty$.

Proof. The statistic (3.18) converge in distribution to χ^2-distribution with $\sum_{l=1}^{k} p_l(p_l+1)(p_l+2)/6$ degrees of freedom under large N. By evaluating the leading term of characteristic function of (3.18) with large f and under large N, we obtain (3.20). $f \to \infty$ means essentially the number of block $k \to \infty$. □

§4. Covariance structure approximation

In this section, we propose a new method of estimation for block diagonal structure. Let $\boldsymbol{x}$ be a random p-vectors from $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ and $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ be sample observation vectors of size N from $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$. Assume $\Xi = \mathrm{diag}(\Xi_1, \Xi_2, \ldots, \Xi_k) = \Sigma^{-1}$ where Ξ_l is a $p_l \times p_l$ matrix, $p_l < N$, $l = 1, 2, \ldots, k$ and $\sum_{l=1}^{k} p_l = p$. Our purpose is to get a estimator of Ξ.

To ensure that the estimator of Ξ exists and be sparsity we make the following assumptions about the covariance matrix Σ.

Existence. There exist such a constant $\varepsilon > 0$ that

$$0 < \varepsilon \le \phi_{\min}(\Sigma) \le \phi_{\max}(\Sigma) < \frac{1}{\varepsilon},$$

where $\phi_{\min}(\Sigma)$ and $\phi_{\max}(\Sigma)$ are the smallest and the largest eigenvalues of Σ, respectively. This condition ensures that Ξ exists.

Sparsity. Let $A = \{(i, j) : \xi_{ij} \neq 0, i > j\}$ denote the set of non-zero off-diagonal entries of Ξ. For the number of A-elements, we assume that

$$\#A < \frac{p(p-1)}{2},$$

where $\#A$ means the number of set A. This assumption is to ensure sparsity of Ξ.

Then, Pavlenko et al. [18] proposed a gLasso estimator of Ξ as the minimizer of the penalized negative log-likelihood

$$\widehat{\Xi}_\lambda = \arg\min_{\Xi > 0} \{\mathrm{tr}(\Xi \widehat{\Sigma}) - \log|\Xi| + \lambda ||\Xi^-||_1\},$$

where $\widehat{\Sigma}$ is the maximum likelihood estimator of Σ, $\Xi^- = \Xi - \mathrm{diag}(\Xi)$, $||\Xi^-||_1 = \sum_{i<j} |\xi_{ij}|$ is ℓ_1-norm of Ξ^-, λ is a non-negative tuning parameter, and λ is the order $\sqrt{\log p/N}$ (see, Rothman et al. [19]). This estimator is similar to the original gLasso introduced in Friedman et al. [4] (they used $||\Xi||_1$ instead of $||\Xi^-||_1$).

Further, following the modification to fast convergence be considered by Pavlenko et al. [18]. Let $\mathcal{K}$ denote the inverse of correlation matrix and Γ denote the diagonal matrix of the standard deviations. Then, a gLasso estimator of $\mathcal{K}$ be defined as

$$\widetilde{\mathcal{K}}_\lambda = \arg\min_{\mathcal{K} > 0} \{\mathrm{tr}(\mathcal{K} \widehat{\mathcal{K}}^{-1}) - \log|\mathcal{K}| + \lambda ||\mathcal{K}^-||_1\}, \tag{4.1}$$

where $\widehat{\mathcal{K}}^{-1}$ is the estimated correlation matrix. Since $\mathcal{K} = (\kappa_{i,j}) = \Gamma \Xi \Gamma$, the estimator of Ξ be given by

$$\widetilde{\Xi}_\lambda = \widehat{\Gamma}^{-1} \widetilde{\mathcal{K}}_\lambda \widehat{\Gamma}^{-1}, \tag{4.2}$$

where $\widehat{\Gamma}$ is a sample estimator of Γ. We call this procedure gLasso-method.

However, these estimators cannot necessarily estimate Ξ to the block diagonal structure. Then, we propose an AIC-method of making Ξ the block diagonal matrix by using Akaike's information criterion (AIC). AIC is defined as

$$\text{AIC} = -2\log L(\widehat{\Xi}|\boldsymbol{X}) + 2d, \tag{4.3}$$

where $\log L(\cdot)$ means log-likelihood function, $\widehat{\Xi}$ is maximum likelihood estimator of Ξ and d is the number of free parameters. The model which makes AIC the minimum is considered to be the optimal model. Our method of estimation for block diagonal structure is following:

(A.1) We calculate $\widetilde{\Xi}_\lambda$by gLasso estimator in (4.1) and (4.2).

(A.2) Candidate models are determined from obtained $\widetilde{\Xi}_\lambda$.

(A.3) AICs for all candidate models are calculated by (4.3).

(A.4) We select the optimal model by values of AICs.

Hence, a block diagonal estimation of Ξ be attained.

An example of the proposed AIC-method is given. Parameters are the following:
$p = 6, N = 10, \lambda = 0.29$ and population is $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ where $\boldsymbol{\mu} = \boldsymbol{0}, \Sigma = \text{diag}(\Sigma_1, \Sigma_2, \Sigma_3)$,
$\Sigma_l = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ $(l = 1, 2, 3)$ and $\rho = 0.85$. Then

$$\widetilde{\Xi}_\lambda = \begin{pmatrix} 0.73 & -0.35 & 0 & 0 & 0 & 0.03 \\ -0.35 & 0.68 & 0 & 0 & 0 & 0.04 \\ 0 & 0 & 0.83 & -0.26 & 0 & 0 \\ 0 & 0 & -0.26 & 0.77 & 0.11 & 0 \\ 0 & 0 & 0 & 0.11 & 0.72 & -0.40 \\ 0.03 & 0.04 & 0 & 0 & -0.40 & 0.60 \end{pmatrix}.$$

is calculated by glasso package in R. Next, we consider how to decide candidate models. When we decide candidate models, we need the following rule:

(R.1) (The number of 0 in each block matrix) ≤ 2.

(R.2) If the number of 0 is not contained in block matrix which has not overlapped under (R.1), the size of this matrix do not make small.

(R.3) If block matrix which satisfy (R.1) has overlapped, we fix one block matrix and make others small.

Under these rules, we find four candidate models in this case. For example,

$$\begin{pmatrix} 0.73 & -0.35 & 0 & 0 & 0 & 0 \\ -0.35 & 0.68 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.83 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.77 & 0.11 & 0 \\ 0 & 0 & 0 & 0.11 & 0.72 & -0.40 \\ 0 & 0 & 0 & 0 & -0.40 & 0.60 \end{pmatrix}$$

is model$(2,1,3)$(=model(p_1,p_2,p_3)) and there are model$(2,2,2)$, model$(2,3,1)$ and model$(2,1,2,1)$. We calculate AIC for each candidate model, respectively. In this case, AIC in (4.3) becomes

$$\mathrm{AIC} = N\sum_{l=1}^{3}(p_l \log 2\pi - \log|S_l^{-1}| + p_l) + 2d,$$

where $S = \mathrm{diag}(S_1, S_2, S_3)$ is the maximum likelihood estimator of $\Sigma = \mathrm{diag}(\Sigma_1, \Sigma_2, \Sigma_3)$, d is the number of free parameters of a model. S^{-1} and AIC of the model$(2,1,3)$ be calculated as

$$S^{-1}(2,1,3) = \begin{pmatrix} 5.64 & -5.02 & 0 & 0 & 0 & 0 \\ -5.02 & 5.09 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.08 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2.01 & 3.02 & -2.10 \\ 0 & 0 & 0 & 3.02 & 9.79 & -7.70 \\ 0 & 0 & 0 & -2.10 & -7.70 & 6.56 \end{pmatrix},$$

$$\mathrm{AIC}(2,1,3) = 161.4.$$

In similar way, AICs of model$(2,2,2)$, model$(2,3,1)$ and model$(2,1,2,1)$ are calculated as

$$\mathrm{AIC}(2,2,2) = 151.2,\ \mathrm{AIC}(2,3,1) = 170.5,\ \mathrm{AIC}(2,1,2,1) = 184.5.$$

Since AIC$(2,2,2)$ is the smallest value in this example, model$(2,2,2)$ is the optimal model. In this case, true model is selected.

§5. Simulation studies

5.1. Performance accuracy of new multivariate skewness and kurtosis

In this subsection, we investigate accuracies of test procedures based on our newly defined estimators $b_{h,1}$ and $b_{h,2}$. Monte Carlo simulation are used to evaluate the size of our test statistics by calculating attained significance level (ASL) as

$$\mathrm{ASL} = \frac{\#\{H_0 \text{ is rejected}\}}{\text{Total number of replications}}.$$

For each block, we assume that $\boldsymbol{\mu}^{(l)} = \mathbf{0}$ and $\Sigma_l = I_{p_l}$ $(l = 1, 2, \ldots, k)$ without loss of generality by invariance property.

An assortment of block sizes p_l's is considered for each p in combination with N, which qualitatively represents both large sample and high-dimensional cases:

$$p = 200, 300, 400, 500, 1000, 2000, \;\; p_l = 5, 10, 20, \;\; N = 50, 100, 200, 400, 800.$$

In our numerical experiment, we carry out $10,000$ and $1,000$ replications for the case of $N < 400$ and $N \geq 400$, respectively. But for the cases of $p = 1000, 2000$, we carry out $1,000$ replications for all parameters.

For each set of p and p_l, draw a sample of N independent observations from corresponding distribution under the null hypothesis. Replicate this r times, and for each $z_{h,1}$, $z^*_{h,1}$, z_{wh}, $z_{h,2}$ and $z^*_{h,2}$ calculate

$$\mathrm{ASL} = \frac{\#\{H_0 \text{ is rejected}\}}{r}.$$

The results listed in Tables 1-7 demonstrate that ASL is closely approaching the true test size when N is large and, what is most important, provide good accuracy for $p > N$ and even for $p \gg N$, stably over various of block sizes. This is except for $z_{h,1}$ and $z_{h,2}$, whose poor performance can be explained by the effects of the bias term of $o(N^{-1})$ in the expectation and variance, see (3.1) and (3.3), respectively. Hence our numerical experiments support the results of Theorems 1–3, thereby justifying validity of newly defined statistics for testing multivariate normality in high-dimensions.

It is important to not that our estimators skewness and kurtosis can be applied for directional tests, i.e. for testing symmetry or peakedness of a distribution. We note that both $z^*_{h,1}$ and z_{wh} improve corresponding original estimators, $z_{h,1}$ for all the sets of simulation parameters. We also note that $z^*_{h,2}$ is an improvement of $z_{h,2}$ when $N \leq 400$, and for $N = 800$, the accuracy of both $z_{h,2}$ and $z^*_{h,2}$ is almost the same. Hence, we can recommend $z^*_{h,1}$ and z_{wh} when the test of symmetry of a distribution is of interests. When $N \leq 400$, we recommend $z^*_{h,2}$ for the kurtosis test.

5.2. Correct selection rate of AIC-method

In this subsection, we investigate correct selection rate (CSR) of AIC-method and gLasso-method by simulation studies, respectively. CSR of AIC-method calculated by using algorithm (A.1)-(A.4) in Section 4 is the probability of selecting the true model. CSR of gLasso-method calculated by (4.1) and (4.2) is the probability of selecting the true model. We decide candidate models under the condition (R.1)-(R.3) in Section 4. As a numerical experiment, we carry out 100 replications. Simulation parameters are the following: $p = 10$, $N = 10, 20$, $\lambda = \sqrt{\log p/N}$. We consider two cases for the covariance structure of population.

- (Case 1) $\boldsymbol{x} \sim \mathcal{N}_p(\mathbf{0}, \Sigma)$,
  $\Sigma = \mathrm{diag}(\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4, \Sigma_5)$, $\Sigma_l = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ $(l = 1, 2, 3, 4, 5)$, $\rho = 0.9$.
- (Case 2) $\boldsymbol{x} \sim \mathcal{N}_p(\mathbf{0}, \Sigma)$,
  $\Sigma = \mathrm{diag}(\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4)$, $\Sigma_s = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$ $(s = 1, 3)$, $\Sigma_t = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ $(t = 2, 4)$, $\rho = 0.9$.

Table 8 give CSR for the Case 1 by AIC-method and gLasso-method. Table 9 give CSR for the Case 2 by AIC-method and gLasso-method. From Tables 8 and 9, we note that our method improve gLasso-method by using AIC. Even when N is small, CSR of AIC-method is quite higher than the one of gLasso-method.

§6. Conclusion

In this paper, we considered tests for the multivariate normality when $p > N$. We proposed new definitions for multivariate skewness and kurtosis as natural extensions of Mardia's measures, and derived their asymptotic distributions under the multivariate normal population. Approximate accuracies of $z_{h,1}$, $z^*_{h,1}$, z_{wh}, $z_{h,2}$ and $z^*_{h,2}$ were evaluated by Monte Carlo simulation.

And we considered the problem to estimate for the covariance structure. There is gLasso-method in Pavlenko et al. [18] for this problem. We proposed an AIC-method which is an improvement of gLasso-method by using an information criterion AIC. Finally, correct selection rates of AIC-method were given by simulation.

Table 1 The ASL of $z_{h,1}$, $z^*_{h,1}$, z_{wh}, $z_{h,2}$ and $z^*_{h,2}$ for $\alpha = 0.1$

			Skewness			Kurtosis	
p	p_l	N	$z_{h,1}$	$z^*_{h,1}$	z_{wh}	$z_{h,2}$	$z^*_{h,2}$
200	5	50	0.002	0.128	0.133	0.058	0.104
		100	0.014	0.120	0.120	0.076	0.099
		200	0.040	0.106	0.109	0.087	0.100
		400	0.068	0.113	0.111	0.108	0.110
		800	0.072	0.090	0.094	0.097	0.099
	10	50	0.000	0.125	0.124	0.036	0.100
		100	0.004	0.120	0.129	0.065	0.099
		200	0.024	0.112	0.118	0.084	0.102
		400	0.051	0.101	0.130	0.091	0.101
		800	0.079	0.111	0.129	0.104	0.109
	20	50	0.000	0.099	0.095	0.010	0.106
		100	0.000	0.128	0.124	0.047	0.106
		200	0.007	0.123	0.124	0.073	0.101
		400	0.040	0.122	0.121	0.096	0.114
		800	0.066	0.119	0.117	0.098	0.106
300	5	50	0.000	0.131	0.124	0.060	0.106
		100	0.010	0.119	0.117	0.077	0.099
		200	0.033	0.111	0.109	0.086	0.100
		400	0.038	0.120	0.120	0.091	0.106
		800	0.033	0.110	0.120	0.088	0.100
	10	50	0.000	0.127	0.129	0.037	0.102
		100	0.001	0.128	0.124	0.070	0.102
		200	0.017	0.115	0.115	0.080	0.100
		400	0.043	0.111	0.108	0.090	0.100
		800	0.071	0.107	0.097	0.084	0.089
	20	50	0.000	0.100	0.094	0.010	0.107
		100	0.000	0.123	0.125	0.046	0.102
		200	0.003	0.118	0.127	0.069	0.099
		400	0.019	0.113	0.102	0.088	0.101
		800	0.054	0.128	0.106	0.095	0.100

Table 2 The ASL of $z_{h,1}$, $z^*_{h,1}$, z_{wh}, $z_{h,2}$ and $z^*_{h,2}$ for $\alpha = 0.1$

			Skewness			Kurtosis	
p	p_l	N	$z_{h,1}$	$z^*_{h,1}$	z_{wh}	$z_{h,2}$	$z^*_{h,2}$
400	5	50	0.000	0.126	0.127	0.054	0.103
		100	0.006	0.124	0.121	0.078	0.103
		200	0.025	0.114	0.118	0.086	0.098
		400	0.058	0.124	0.102	0.088	0.097
		800	0.058	0.083	0.114	0.100	0.101
	10	50	0.000	0.130	0.127	0.037	0.102
		100	0.001	0.125	0.124	0.073	0.110
		200	0.013	0.117	0.114	0.081	0.099
		400	0.035	0.111	0.112	0.084	0.093
		800	0.063	0.105	0.094	0.084	0.090
	20	50	0.000	0.105	0.097	0.009	0.106
		100	0.000	0.127	0.128	0.047	0.102
		200	0.001	0.123	0.130	0.072	0.103
		400	0.016	0.112	0.128	0.086	0.103
		800	0.037	0.116	0.114	0.091	0.100
500	5	50	0.000	0.127	0.134	0.055	0.102
		100	0.004	0.124	0.117	0.076	0.104
		200	0.023	0.114	0.108	0.088	0.101
		400	0.044	0.113	0.113	0.100	0.105
		800	0.063	0.110	0.103	0.102	0.104
	10	50	0.000	0.128	0.119	0.039	0.109
		100	0.000	0.128	0.122	0.063	0.098
		200	0.010	0.117	0.117	0.086	0.103
		400	0.015	0.110	0.109	0.093	0.102
		800	0.053	0.098	0.098	0.108	0.113
	20	50	0.000	0.098	0.099	0.009	0.103
		100	0.000	0.128	0.126	0.044	0.099
		200	0.001	0.115	0.124	0.073	0.103
		400	0.009	0.115	0.115	0.088	0.098
		800	0.051	0.126	0.126	0.098	0.107

Table 3 The ASL of $z_{h,1}$, $z^*_{h,1}$, z_{wh}, $z_{h,2}$ and $z^*_{h,2}$ for $\alpha = 0.05$

			Skewness			Kurtosis	
p	p_l	N	$z_{h,1}$	$z^*_{h,1}$	z_{wh}	$z_{h,2}$	$z^*_{h,2}$
200	5	50	0.000	0.072	0.072	0.021	0.055
		100	0.006	0.063	0.068	0.033	0.050
		200	0.018	0.056	0.056	0.043	0.052
		400	0.029	0.060	0.054	0.055	0.060
		800	0.038	0.047	0.044	0.049	0.051
	10	50	0.000	0.070	0.071	0.011	0.053
		100	0.001	0.063	0.072	0.026	0.049
		200	0.009	0.063	0.063	0.039	0.052
		400	0.029	0.058	0.080	0.041	0.045
		800	0.040	0.061	0.056	0.056	0.059
	20	50	0.000	0.050	0.048	0.002	0.055
		100	0.000	0.074	0.071	0.017	0.053
		200	0.002	0.068	0.073	0.031	0.052
		400	0.016	0.067	0.066	0.044	0.058
		800	0.031	0.066	0.058	0.048	0.056
300	5	50	0.000	0.076	0.072	0.023	0.057
		100	0.004	0.066	0.065	0.034	0.052
		200	0.015	0.060	0.060	0.043	0.052
		400	0.017	0.064	0.058	0.043	0.054
		800	0.014	0.057	0.058	0.042	0.051
	10	50	0.000	0.076	0.075	0.011	0.054
		100	0.000	0.073	0.068	0.031	0.054
		200	0.007	0.064	0.062	0.036	0.048
		400	0.022	0.057	0.055	0.040	0.048
		800	0.028	0.056	0.055	0.039	0.044
	20	50	0.000	0.053	0.049	0.001	0.054
		100	0.000	0.071	0.071	0.017	0.053
		200	0.001	0.064	0.071	0.029	0.049
		400	0.008	0.060	0.065	0.040	0.051
		800	0.027	0.068	0.049	0.050	0.055

Table 4 The ASL of $z_{h,1}$, $z^*_{h,1}$, z_{wh}, $z_{h,2}$ and $z^*_{h,2}$ for $\alpha = 0.05$

			Skewness			Kurtosis	
p	p_l	N	$z_{h,1}$	$z^*_{h,1}$	z_{wh}	$z_{h,2}$	$z^*_{h,2}$
400	5	50	0.000	0.070	0.073	0.021	0.052
		100	0.003	0.068	0.068	0.035	0.054
		200	0.011	0.060	0.063	0.040	0.049
		400	0.027	0.059	0.051	0.046	0.052
		800	0.028	0.036	0.063	0.051	0.052
	10	50	0.000	0.075	0.068	0.012	0.052
		100	0.000	0.069	0.068	0.030	0.057
		200	0.006	0.067	0.063	0.037	0.050
		400	0.016	0.056	0.057	0.033	0.041
		800	0.025	0.052	0.057	0.042	0.045
	20	50	0.000	0.052	0.049	0.001	0.054
		100	0.000	0.074	0.073	0.016	0.053
		200	0.001	0.070	0.073	0.032	0.051
		400	0.004	0.055	0.072	0.042	0.049
		800	0.019	0.055	0.050	0.049	0.056
500	5	50	0.000	0.075	0.075	0.019	0.053
		100	0.001	0.069	0.063	0.034	0.051
		200	0.009	0.062	0.060	0.042	0.051
		400	0.019	0.048	0.048	0.049	0.056
		800	0.032	0.049	0.049	0.048	0.054
	10	50	0.000	0.072	0.068	0.013	0.054
		100	0.000	0.068	0.068	0.027	0.052
		200	0.004	0.065	0.065	0.040	0.055
		400	0.006	0.061	0.061	0.043	0.052
		800	0.028	0.051	0.051	0.043	0.049
	20	50	0.000	0.049	0.048	0.002	0.053
		100	0.000	0.073	0.070	0.014	0.050
		200	0.000	0.067	0.069	0.030	0.053
		400	0.002	0.058	0.058	0.044	0.050
		800	0.018	0.068	0.068	0.050	0.057

Table 5 The ASL of $z_{h,1}$, $z^*_{h,1}$, z_{wh}, $z_{h,2}$ and $z^*_{h,2}$ for $\alpha = 0.01$

			Skewness			Kurtosis	
p	p_l	N	$z_{h,1}$	$z^*_{h,1}$	z_{wh}	$z_{h,2}$	$z^*_{h,2}$
200	5	50	0.000	0.020	0.018	0.002	0.011
		100	0.001	0.015	0.018	0.006	0.012
		200	0.003	0.013	0.013	0.007	0.011
		400	0.004	0.008	0.014	0.010	0.011
		800	0.008	0.014	0.011	0.009	0.010
	10	50	0.000	0.018	0.021	0.001	0.012
		100	0.000	0.017	0.020	0.003	0.010
		200	0.001	0.014	0.014	0.006	0.012
		400	0.008	0.020	0.015	0.007	0.011
		800	0.006	0.009	0.015	0.014	0.017
	20	50	0.000	0.010	0.010	0.000	0.012
		100	0.000	0.019	0.018	0.001	0.012
		200	0.000	0.016	0.018	0.004	0.010
		400	0.002	0.016	0.015	0.010	0.010
		800	0.005	0.014	0.012	0.010	0.011
300	5	50	0.000	0.021	0.021	0.002	0.013
		100	0.001	0.017	0.017	0.004	0.011
		200	0.002	0.014	0.016	0.008	0.012
		400	0.003	0.016	0.010	0.008	0.011
		800	0.003	0.013	0.016	0.009	0.013
	10	50	0.000	0.019	0.020	0.001	0.011
		100	0.000	0.019	0.020	0.003	0.013
		200	0.001	0.016	0.015	0.006	0.010
		400	0.004	0.015	0.014	0.008	0.010
		800	0.003	0.007	0.017	0.011	0.012
	20	50	0.000	0.010	0.009	0.000	0.013
		100	0.000	0.021	0.018	0.002	0.012
		200	0.001	0.014	0.019	0.004	0.009
		400	0.001	0.013	0.018	0.006	0.010
		800	0.007	0.017	0.012	0.009	0.010

Table 6 The ASL of $z_{h,1}$, $z^*_{h,1}$, z_{wh}, $z_{h,2}$ and $z^*_{h,2}$ for $\alpha = 0.01$

			Skewness			Kurtosis	
p	p_l	N	$z_{h,1}$	$z^*_{h,1}$	z_{wh}	$z_{h,2}$	$z^*_{h,2}$
400	5	50	0.000	0.020	0.019	0.003	0.013
		100	0.000	0.018	0.019	0.005	0.011
		200	0.002	0.013	0.016	0.007	0.010
		400	0.005	0.012	0.010	0.008	0.009
		800	0.007	0.010	0.012	0.008	0.009
	10	50	0.000	0.020	0.019	0.001	0.013
		100	0.000	0.019	0.018	0.004	0.013
		200	0.001	0.019	0.014	0.007	0.010
		400	0.003	0.013	0.012	0.006	0.012
		800	0.002	0.011	0.010	0.004	0.005
	20	50	0.000	0.008	0.010	0.000	0.012
		100	0.000	0.020	0.020	0.001	0.011
		200	0.000	0.018	0.019	0.004	0.011
		400	0.001	0.016	0.020	0.008	0.012
		800	0.001	0.010	0.009	0.009	0.010
500	5	50	0.000	0.021	0.019	0.002	0.011
		100	0.000	0.019	0.016	0.005	0.010
		200	0.002	0.014	0.013	0.008	0.010
		400	0.004	0.007	0.007	0.008	0.009
		800	0.005	0.009	0.010	0.008	0.009
	10	50	0.000	0.019	0.018	0.001	0.013
		100	0.000	0.018	0.018	0.003	0.012
		200	0.000	0.018	0.018	0.007	0.011
		400	0.000	0.012	0.012	0.003	0.008
		800	0.006	0.010	0.010	0.004	0.005
	20	50	0.000	0.009	0.008	0.000	0.012
		100	0.000	0.019	0.019	0.001	0.010
		200	0.000	0.018	0.017	0.005	0.012
		400	0.000	0.011	0.011	0.006	0.013
		800	0.004	0.015	0.015	0.009	0.012

Table 7 The ASL of $z_{h,1}$, $z^*_{h,1}$, z_{wh}, $z_{h,2}$ and $z^*_{h,2}$ for $p_l = 20$

			Skewness			Kurtosis	
p	α	N	$z_{h,1}$	$z^*_{h,1}$	z_{wh}	$z_{h,2}$	$z^*_{h,2}$
1000	0.1	50	0.000	0.110	0.110	0.030	0.130
		100	0.000	0.134	0.134	0.049	0.110
		200	0.000	0.125	0.125	0.068	0.097
		400	0.000	0.122	0.121	0.060	0.099
	0.05	50	0.000	0.060	0.060	0.000	0.080
		100	0.000	0.077	0.075	0.020	0.059
		200	0.000	0.065	0.065	0.030	0.048
		400	0.000	0.064	0.063	0.035	0.051
	0.01	50	0.000	0.020	0.020	0.000	0.030
		100	0.000	0.017	0.017	0.000	0.019
		200	0.000	0.010	0.010	0.008	0.016
		400	0.000	0.012	0.011	0.009	0.011
2000	0.1	50	0.000	0.111	0.111	0.008	0.109
		100	0.000	0.149	0.149	0.041	0.106
		200	0.000	0.119	0.118	0.076	0.110
		400	0.000	0.115	0.115	0.093	0.106
	0.05	50	0.000	0.058	0.056	0.001	0.056
		100	0.000	0.082	0.082	0.017	0.049
		200	0.000	0.070	0.070	0.031	0.055
		400	0.000	0.080	0.080	0.040	0.058
	0.01	50	0.000	0.012	0.012	0.001	0.012
		100	0.000	0.018	0.018	0.001	0.014
		200	0.000	0.024	0.023	0.007	0.011
		400	0.000	0.020	0.020	0.008	0.011

Table 8 Comparison of CSR (case 1)

N	gLasso-method	AIC-method
10	0.19	0.87
20	0.56	0.92

Table 9 Comparison of CSR (case 2)

N	gLasso-method	AIC-method
10	0.31	0.64
20	0.68	0.94

Acknowledgements

The authors would like to thank the referee for his suitable comments and careful reading.

References

[1] Baringhaus, L. and Henze, N. (1992). Limit distributions for Mardia's measure of multivariate skewness. *The Annals of Statistics*, **20**, 1889–1902.

[2] Cook, R. and Li, B. (2002). Dimension reduction for conditional mean in regression. *The Annals of Statistics*, **30**, 455–474.

[3] Day, N. E. (1969). Estimating the components of a mixture of normal distributions. *Biometrika*. **56**, 463–474.

[4] Friedman, J., Hastie, T. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, **9**, 432–441.

[5] Genton, M. (2004). *Skew-elliptical Distributions and their applications.* A Journey beyond normality, Chapman & Hall/CRC, Boca Raton.

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

[7] Kendall, M. G. and Stuart, A. (1967). *The Advanced Theory of Statistics*, **2** (2nd ed.), Griffin, London.

[8] Kendall, M. G. and Stuart, A. (1969). *The Advanced Theory of Statistics*, **1** (3rd ed.), Griffin, London.

[9] Khatri, C. G. and Pillai, K. C. S. (1967). On the moments of traces of two matrices in multivariate analysis. *Annals of the Institute of Statistical Mathematics*, **19**, 143–156.

[10] Koizumi, K., Okamoto, N. and Seo, T. (2009). On Jarque-Bera tests for assessing multivariate normality. *Journal of Statistics: Advances in Theory and Applications*, **1**, 207–220.

[11] Kollo,T. and Srivastava, M. S. (2005). Estimation and testing of parameters in multivariate Laplace distribution. *Communications in Statistics -Theory and Methods*, **33**, 2363–2387.

[12] Kots, S., Kozubowski, T. and Podgorski, K. (2001). *The Laplace distribution, and generalizations: A revisit with applications to communications, economics, engineering, and finance*, Birkhauser, Boston.

[13] Kotz, S. and Nadarajah, S. (2004). *Multivariate t-distributions and their applications*, Cambridge University Press, Cambridge.

[14] Liang, J. (2012). T_3-plot for testing spherical symmetry for high-dimensional data with a small sample size. *to appear in Journal of Probability and Statistics*.

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

[16] Mecklin, C. J. and Mundfrom, D. J. (2004). An appraisal and bibliography of tests for multivariate normality. *International Statistical Review*, **72**, 123–138.

[17] Miyagawa, C., Koizumi, K. and Seo, T. (2011). A new multivariate kurtosis and its asymptotic distribution. *SUT Journal of Mathematics*, **47**, 55–71.

[18] Pavlenko, T., Björkström, A. and Tillander, A. (2012). Covariance structure approximation via gLasso in high-dimensional supervised classification. *Journal of Applied Statistics*, **39**, 1643–1666.

[19] Rothman, A., Bickel, P., Levina, E. and Zhu, J. (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, **2**, 494–515.

[20] Srivastava, M. S. (1984). A measure of skewness and kurtosis and a graphical method for assessing multivariate normality. *Statistics & Probability Letters*, **2**, 263–267.

[21] Wilson, E. B. and Hilferty, M. M. (1931). The distribution of chi-square. *Mathematics*, **17**, 684–688.

Kazuyuki Koizumi
Department of International College of Arts and Sciences,
Yokohama City University, Kanagawa, Japan
E-mail: `zumi@yokohama-cu.ac.jp`

Takuma Sumikawa
Department of Mathematical Information Science, Graduate School of Science,
Tokyo University of Science, Tokyo, Japan

Tatjana Pavlenko
Department of Mathematics,
KTH Royal Institute of Technology, Stockholm, Sweden