ELSEVIER

Journal of Statistical Planning and Inference 86 (2000) 129–141

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Testing multinormality based on low-dimensional projection ☆

Jiajuan Liang[a,b], Runze Li[c], Hongbin Fang[a], Kai-Tai Fang[a,b,*]

[a]*Department of Mathematics, Hong Kong Baptist University, Hong Kong*

[b]*Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing, People's Republic of China*

[c]*Department of Statistics, University of North Carolina at Chapel Hill, NC 27599-3260, USA*

Received 17 August 1998; received in revised form 15 July 1999; accepted 1 September 1999

Abstract

A method based on properties of left-spherical matrix distributions and affine invariant statistics is employed to construct projection tests for multivariate normality. The projection tests are indirectly dependent on the dimension of raw data. As a result, the projection tests can be performed for arbitrary dimension d and sample size n even if $n < d$ in high-dimensional case as soon as the projection dimension is suitably chosen. By Monte Carlo simulation, we show that the projection tests significantly improve the power of existing tests for multinormality in the case of high dimension with a small sample size. Analysis on a practical example shows that the projection tests are useful complements to existing tests for multinormality. © 2000 Elsevier Science B.V. All rights reserved.

MSC: 62E25; 62G10; 62H15

Keywords: Left-spherical matrix distributions; Mardia's skewness and kurtosis; Projection tests; Testing multinormality

1. Introduction

The importance of testing multivariate normality (multinormality, or MVN for short) has been witnessed by many methodologies in statistical data analysis. Mardia (1980) covers a large number of references before 1980 in this area. A considerable number of new statistics for testing MVN have been proposed since 1980, see, e.g., Hawkins (1981), Koziol (1982), Csörgo (1986), Ahn (1992), Mudholkar et al. (1992), Bowman and Foster (1993), Cox and Wermuth (1994), Mudholkar et al. (1995), Zhu et al. (1995), and Henze and Wagner (1997). Horswell and Looney (1992), and Romeu and

☆ This work was supported by National Institute on Drug Abuse Grant DA01070, USA, and the Hong Kong Baptist University Grant FRG/96-97/II-15.

* Corresponding author.

PII: S0378-3758(99)00168-8

Ozturk (1993), respectively, made an extensive Monte Carlo study on the performance of some statistics for testing MVN. Based on their simulation results, Romeu and Ozturk concluded that the Mardia's skewness and kurtosis statistics (Mardia, 1970) possess good power performance against a wide range of alternative distributions.

Because of the well-known properties of the multivariate normal distribution, statistics for testing MVN are usually required to possess *affine invariance*. An advantage in using affine invariant statistics for testing MVN is that estimation of unknown parameters $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in the normal distribution $\mathrm{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ can be avoided. Affine invariant statistics for testing MVN are usually related to the sample covariance matrix. When the dimension d is large and the sample size n is small and close to d, the sample covariance matrix tends to singularity. As a result, the power performance of many existing affine invariant statistics in testing MVN is impaired in the case of high dimension with a small sample size. When $d > n$, the affine invariant statistics such as those mentioned in the above literature cannot be directly used to test the joint d-variate MVN.

The purpose of this paper is to develop a procedure for constructing projection tests for MVN based on affine invariant statistics so that the affine invariant statistics can still be effectively used in the case of high dimension with a small sample size or even if $d \geqslant n$. The idea for constructing the projection tests is based on the properties of a *left-spherical matrix distribution* (LSMD for short) (Dawid, 1977), and the fact that affine invariant statistics for testing MVN have unchanged null distributions in the family of LSMD. Läuter (1996) and Läuter et al. (1996) proposed similar idea to construct linear score tests for the mean vector $\boldsymbol{\mu}$ in the normal distribution $\mathrm{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Section 2 below presents the details for constructing the projection tests. Section 3 gives a simulation study and analysis on a practical example. Some discussions and further applications are given in the last section.

2. The testing procedure

2.1. Theoretical background

The procedure for constructing the projection tests for MVN is based on some properties of LSMD. Some further properties of LSMD can be found in Fang and Zhang (1990). An $n \times d$ random matrix $\boldsymbol{X}$ is said to have an LSMD if $\boldsymbol{\Gamma X} \stackrel{\mathrm{d}}{=} \boldsymbol{X}$ for any orthogonal matrix $\boldsymbol{\Gamma}$ which is independent of $\boldsymbol{X}$, where the sign "$\stackrel{\mathrm{d}}{=}$" means that the two sides of the equality have the same distribution. Using the notations in Fang and Zhang (1990), we denote by $\boldsymbol{X} \sim \mathrm{LS}^{+}_{n\times d}(\phi)$ if $\boldsymbol{X}$ has an LSMD and $P(\det(\boldsymbol{X}'\boldsymbol{X}) = 0) = 0$. In this case the characteristic function of $\boldsymbol{X}$ has the form $\phi(\boldsymbol{T}'\boldsymbol{T})$ for some real function $\phi(\cdot)$. A random matrix $\boldsymbol{X} \sim \mathrm{LS}^{+}_{n\times d}(\phi)$ possesses many interesting properties. For example, if $\boldsymbol{Z} \sim \mathrm{LS}^{+}_{(n-1)\times q}(\phi)$, then there exists a $q \times q$ random matrix $\boldsymbol{V}$ such that

$$\boldsymbol{Z} \stackrel{\mathrm{d}}{=} \boldsymbol{U}\boldsymbol{V}, \tag{2.1}$$

where $\boldsymbol{V}$ is independent of $\boldsymbol{U} \sim \mathscr{U}_{n-1,q}$, the uniform distribution on the Stiefel manifold $\mathcal{O}(n-1,q) = \{\boldsymbol{H}_{(n-1)\times q}\colon \boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_q\}$ (Theorem 3.1.2 of Fang and Zhang, 1990, p. 94). The following two corollaries can be easily obtained by the definition of LSMD, property (2.1) and Theorem 2 of Läuter (1996).

Corollary 2.1. *Let* $\boldsymbol{Y} \sim \mathrm{LS}^+_{(n-1)\times d}(\phi)$ *and the random matrix* $\boldsymbol{D}_{d\times q} = f(\boldsymbol{Y}'\boldsymbol{Y})$ $(1 \leqslant q < d)$ *be a function of* $\boldsymbol{Y}'\boldsymbol{Y}$ *and uniquely determined by* $\boldsymbol{Y}'\boldsymbol{Y}$. *Assume that* $P\{\mathrm{rank}(\boldsymbol{D}) = q\} = 1$. *Let*

$$\boldsymbol{Z} = \boldsymbol{Y}\boldsymbol{D}. \tag{2.2}$$

Then $\boldsymbol{Z} \sim \mathrm{LS}^+_{(n-1)\times q}(\phi_1)$ *for some* $\phi_1(\cdot)$.

Corollary 2.2. *Let* $\boldsymbol{X}_{n\times d} = (\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)'$, *where* $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ *are i.i.d. and* $\boldsymbol{x}_1 \sim \mathrm{N}_d(\boldsymbol{\mu},\boldsymbol{\Sigma})$ $(\boldsymbol{\Sigma} > \boldsymbol{0}$, *positive definite*). *Assume that* $\boldsymbol{B}_{d\times d}$ *is an arbitrary constant matrix and* $\boldsymbol{b} \in \mathbb{R}^d$ *is a constant vector. Denote by* $\boldsymbol{W}_{n\times d} = (\boldsymbol{w}_1,\ldots,\boldsymbol{w}_n)'$ *with*

$$\boldsymbol{w}_i = \boldsymbol{B}\boldsymbol{x}_i + \boldsymbol{b}, \quad i = 1,\ldots,n \tag{2.3}$$

and $\boldsymbol{Y}_{(n-1)\times d} = (\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{n-1})' = \boldsymbol{A}\boldsymbol{W}$, *where* $\boldsymbol{A}_{(n-1)\times n}$ *is a constant matrix satisfying*

$$\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{I}_{n-1} \quad \text{and} \quad \boldsymbol{A}\boldsymbol{1}_n = \boldsymbol{0}, \tag{2.4}$$

where $\boldsymbol{1}_n$ *is a column vector of ones. Let the random matrices* $\boldsymbol{Z}$ *and* $\boldsymbol{D}$ *be defined as in* (2.2). *Then the following assertions are true.*

(1) $\boldsymbol{Z} \sim \mathrm{LS}^+_{(n-1)\times q}(\phi_2)$ *for some* $\phi_2(\cdot)$;

(2) *If* $T(\boldsymbol{Z})$ *is a statistic based on* $\boldsymbol{Z}$ *and satisfies*

$$T(\boldsymbol{Z}) \stackrel{\mathrm{d}}{=} T(\boldsymbol{Z}\boldsymbol{C}) \tag{2.5}$$

for any constant matrix $\boldsymbol{C}_{q\times q}$, *then* $T(\boldsymbol{Z})$ *has the same distribution in the family of LSMD. In particular*

$$T(\boldsymbol{Z}) \stackrel{\mathrm{d}}{=} T(\boldsymbol{Z}_0), \quad \boldsymbol{Z}_0 \sim \mathrm{N}_{(n-1)\times q}(\boldsymbol{0}, \boldsymbol{I}_{n-1} \otimes \boldsymbol{I}_q). \tag{2.6}$$

The constant matrix $\boldsymbol{A}$ in Corollary 2.2 centerizes the raw data. It is obvious that the choice of $\boldsymbol{A}$ is not unique. The Helmert's transformation as used by some authors (Mardia, 1980) is a convenient choice. It is defined as follows: let $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}$: $\boldsymbol{Y} = (\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{n-1})'$ with $\boldsymbol{y}_i$ given by

$$\boldsymbol{y}_i = (\boldsymbol{x}_1 + \cdots + \boldsymbol{x}_i - i\boldsymbol{x}_{i+1})/\sqrt{i(i+1)}, \quad i = 1,\ldots,n-1, \tag{2.7}$$

we have $\boldsymbol{A}_{(n-1)\times n} = (a_{ij})$ with a_{ij} given by

$$a_{ij} = \begin{cases} \frac{1}{\sqrt{i(i+1)}}, & j = 1,\ldots,i, \\ \frac{-i}{\sqrt{i(i+1)}}, & j = i+1, \\ 0, & \text{otherwise} \end{cases} \tag{2.8}$$

for $i = 1,\ldots,n-1$. Then $\boldsymbol{A}$ satisfies (2.4).

Now, we derive the statistics for testing MVN based on Corollaries 2.1 and 2.2. Denote by $T_{n,d}(\boldsymbol{X}) = T_{n,d}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n)$, which is an affine invariant statistic. Let

$$\boldsymbol{Y} = \boldsymbol{AX} \quad \text{and} \quad \boldsymbol{Z} = (z_1,\dots,z_{n-1})' = \boldsymbol{YD}, \tag{2.9}$$

where the matrices $\boldsymbol{D}$ and $\boldsymbol{A}$ satisfy the same conditions as those in Corollary 2.2. We call

$$T_{n-1,q}(\boldsymbol{Z}) = T_{n-1,q}(z_1,\dots,z_{n-1}), \tag{2.10}$$

the q-dimensional ($1 \leqslant q \leqslant d$) *projection statistic* of $T_{n,d}(\boldsymbol{X})$. There are a number of existing affine invariant statistics for testing MVN in the literature. We will implement Mardia's (1970) skewness and kurtosis coefficients $b_{1d}(\boldsymbol{X})$ and $b_{2d}(\boldsymbol{X})$, which are defined by

$$b_{1d} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij}^3 \quad \text{and} \quad b_{2d} = \frac{1}{n}\sum_{i=1}^{n} r_i^4, \tag{2.11}$$

respectively, where $n > d$ and

$$\begin{aligned} r_{ij} &= (\boldsymbol{x}_i - \bar{\boldsymbol{x}})'\boldsymbol{S}_n^{-1}(\boldsymbol{x}_j - \bar{\boldsymbol{x}}), \qquad r_i^2 = (\boldsymbol{x}_i - \bar{\boldsymbol{x}})'\boldsymbol{S}_n^{-1}(\boldsymbol{x}_i - \bar{\boldsymbol{x}}), \\ \bar{\boldsymbol{x}} &= \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i, \qquad \boldsymbol{S}_n = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})'. \end{aligned} \tag{2.12}$$

For convenience of applying existing statistical tables, the following standardized forms are often used in practice (Mardia, 1970; Romeu and Ozturk, 1993):

$$\begin{aligned} &\text{standardized form of } b_{1d}\text{:} \quad sb_{1d} = nb_{1d}/6 \xrightarrow{\mathscr{D}} \chi^2(f), \\ &\text{standardized form of } b_{2d}\text{:} \quad sb_{2d} = \frac{b_{2d} - d(d+2)(n-1)/(n+1)}{\sqrt{8d(d+2)/n}} \xrightarrow{\mathscr{D}} \mathrm{N}(0,1), \end{aligned} \tag{2.13}$$

where $\chi^2(f)$ is the chi-squared distribution with f degrees of freedom, $f = d(d+1)(d+2)/6$ and the sign "$\xrightarrow{\mathscr{D}}$" means "converge in distribution" ($n \to \infty$). Large values of $|sb_{1d}|$ or $|sb_{2d}|$ indicate evidence of departure from MVN.

The q-dimensional projection statistics ($q \geqslant 2$) of sb_{1d} and sb_{2d} based on $\boldsymbol{Z}_{(n-1)\times q}$ resulted from (2.9), and their asymptotic null distributions (i.e., $\boldsymbol{Z}$ had an LSMD), are given by $sb_{1q}(\boldsymbol{Z})$ and $sb_{2q}(\boldsymbol{Z})$, respectively:

$$sb_{1q}(\boldsymbol{Z}) = (n-1)b_{1q}(\boldsymbol{Z})/6 \xrightarrow{\mathscr{D}} \chi^2(f_q) \tag{2.14}$$

and

$$sb_{2q}(\boldsymbol{Z}) = \frac{b_{2q}(\boldsymbol{Z}) - q(q+2)(n-2)/n}{\sqrt{8q(q+2)/(n-1)}} \xrightarrow{\mathscr{D}} \mathrm{N}(0,1), \tag{2.15}$$

where $f_q = q(q+1)(q+2)/6$, $b_{1q}(\boldsymbol{Z})$ and $b_{2q}(\boldsymbol{Z})$ are computed by the formulas (2.11) and (2.12) with $\boldsymbol{X}_{n\times p}$ replaced by $\boldsymbol{Z}_{(n-1)\times q}$. If $\boldsymbol{Z}$ has an LSMD, then

$$sb_{1q}(\boldsymbol{Z}) \stackrel{\mathrm{d}}{=} sb_{1q}(\boldsymbol{Z}_0) \quad \text{and} \quad sb_{2q}(\boldsymbol{Z}) \stackrel{\mathrm{d}}{=} sb_{2q}(\boldsymbol{Z}_0), \tag{2.16}$$

where $\boldsymbol{Z}_0$ is the same as that in (2.6). Since both $|sb_{1q}(\boldsymbol{Z}_0)|$ and $|sb_{2q}(\boldsymbol{Z}_0)|$ have a tendency to take small values, in other words, we will reject the hypothesis that $\boldsymbol{Z}$ has an LSMD for large values of $|sb_{1q}(\boldsymbol{Z})|$ or $|sb_{2q}(\boldsymbol{Z})|$. In the special case of $q=1$ in (2.9), by a similar discussion, the classical skewness $\mathrm{Sk}_{n-1}(z)$ and kurtosis $\mathrm{Ku}_{n-1}(z)$ given by

$$\mathrm{Sk}_{n-1}(z) = \frac{1}{n-1}\sum_{i=1}^{n-1}(z_i-\bar{z})^3 \Big/ \left[\frac{1}{n-1}\sum_{i=1}^{n-1}(z_i-\bar{z})^2\right]^{3/2}, \tag{2.17}$$

and

$$\mathrm{Ku}_{n-1}(z) = \frac{1}{n-1}\sum_{i=1}^{n-1}(z_i-\bar{z})^4 \Big/ \left[\frac{1}{n-1}\sum_{i=1}^{n-1}(z_i-\bar{z})^2\right]^{2}, \tag{2.18}$$

respectively, can be applied to test the hypothesis that $z_{(n-1)\times 1}$ has a spherical distribution $S_{n-1}^{+}(\phi)$ (Fang et al., 1990). If $z \sim S_{n-1}^{+}(\phi)$, then (Lemma 3.2 of Fang et al., 1990)

$$\mathrm{Sk}_{n-1}(z) \stackrel{\mathrm{d}}{=} \mathrm{Sk}_{n-1}(z_0), \qquad \mathrm{Ku}_{n-1}(z) \stackrel{\mathrm{d}}{=} \mathrm{Ku}_{n-1}(z_0), \tag{2.19}$$

where $z_0 \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{I}_{n-1})$. Large values of $|\mathrm{Sk}_{n-1}(z)|$ or $|\mathrm{Ku}_{n-1}(z)-3|$ imply that z is not spherically distributed.

It is noted that, a basic problem in the projection tests based on $\boldsymbol{Z}$ is: how to choose the matrix $\boldsymbol{D}$ and the projection dimension q in (2.9) and (2.10)? We will attack this problem by Monte Carlo simulation and analysis on a practical example in Section 3.

2.2. Testing multinormality by the projection statistics

Let $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ be i.i.d. observations from a population with an unknown distribution function $F(\boldsymbol{x})$ ($\boldsymbol{x} \in \mathbb{R}^d$, $d \geqslant 2$). We want to test, on the basis of $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$, the null hypothesis

$$\mathrm{H}_{0d}\colon F(\boldsymbol{x}) \in \mathcal{N}_d, \tag{2.20}$$

where $\mathcal{N}_d$ is the class of all nondegenerate d-dimensional normal distribution $\mathrm{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ($\boldsymbol{\Sigma} > \boldsymbol{0}$). The alternative hypothesis of H_{0d} in (2.20) can be stated as $\mathrm{H}_{1d}\colon F(\boldsymbol{x}) \notin \mathcal{N}_d$. The transformation

$$\boldsymbol{X}_{n\times d} \stackrel{(1)}{\Rightarrow} \boldsymbol{Y}_{(n-1)\times d} = \boldsymbol{AX} \stackrel{(2)}{\Rightarrow} \boldsymbol{Z}_{(n-1)\times q} = \boldsymbol{YD}, \tag{2.21}$$

defined by Corollary 2.2 makes it feasible to transfer a test for (2.20) to a test for the hypothesis

$$\mathrm{H}_{0q}\colon \boldsymbol{Z} \text{ has a left-spherical matrix distribution}, \tag{2.22}$$

against the alternative hypothesis H_{1q}: H_{0q} in (2.22) is not true. It is obvious that rejection for (2.22) results in rejection for (2.20). But acceptance for (2.22) does not necessarily imply acceptance for (2.20). We call a test for (2.22) a *necessary test* for

(2.20). In the special case of $q = 1$, by a similar discussion, a test for (2.20) can be transferred to a test for

$$\begin{aligned} \mathrm{H}_{01}:\ & \text{the random vector } z \text{ resulted from (2.21) for } q = 1 \\ & \text{has a spherical distribution,} \end{aligned} \tag{2.23}$$

against H_{11}: H_{01} in (2.23) is not true.

In constructing the projection tests by (2.21), the choices of the constant matrix $\boldsymbol{A}$ for $\boldsymbol{Y}=\boldsymbol{AX}$, the matrix $\boldsymbol{D}$, and the projection dimension q may have some influence on the performance of the projection tests. Complete theoretical solutions to these problems seem to be intractable. For convenience in computation, we suggest choosing $\boldsymbol{A}$ as the Helmert's transformation (2.7). The matrix $\boldsymbol{D}$ is determined by a similar way to that considered by Läuter (1996) and Läuter et al. (1996). The idea is that, choosing $\boldsymbol{D}$ as composed of the first q-columns of the matrix $\boldsymbol{D}_0$ which is the solution to the eigenvalue problem

$$\boldsymbol{Y}'\boldsymbol{Y}\boldsymbol{D}_0 = \mathrm{diag}(\boldsymbol{Y}'\boldsymbol{Y})\boldsymbol{D}_0\boldsymbol{\Lambda}, \tag{2.24}$$

where $\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1,\ldots,\lambda_d)$ with $\lambda_1 > \cdots > \lambda_d$ with probability 1 and $\mathrm{diag}(\boldsymbol{Y}'\boldsymbol{Y})$ denotes the diagonal matrix consisted of the diagonal elements of $\boldsymbol{Y}'\boldsymbol{Y}$. Some conditions need to be imposed on the matrix $\boldsymbol{D}_0$ to ensure the unique solution to problem (2.24). When all of the variables are measured in the same scale, based on simulation, we suggest choosing $\boldsymbol{D}$ in (2.21) as composed of the first q columns of the matrix $\boldsymbol{D}_1$ which is the solution to the eigenvalue problem

$$\boldsymbol{Y}'\boldsymbol{Y}\boldsymbol{D}_1 = \boldsymbol{D}_1\boldsymbol{G}, \tag{2.25}$$

where $\boldsymbol{G} = \mathrm{diag}(g_1,\ldots,g_d)$ with $g_1 > \cdots > g_d$ and $\boldsymbol{D}_1$ is assumed to have positive diagonal elements with probability 1 to ensure the unique solution to problem (2.25). It seems to be intractable to give the best choice for the value of q which is suitable for any choice of $\boldsymbol{D}$.

In the next section we perform a Monte Carlo study on the effect in choosing different q's by constricting $\boldsymbol{D}$ to the first q columns of $\boldsymbol{D}_1$ in (2.25) and the matrix $\boldsymbol{A}$ to Helmert's transformation (2.7).

3. Simulation and applications

3.1. Simulation on the effect of choosing different projection dimension q

The simulation in this subsection is focused on the case of high dimension with a small sample size: (1) $d = 10$ and $n = 26$; and (2) $d = 20$ and $n = 26$. The percentiles of the projection statistics $sb_{1q}(\boldsymbol{Z})$ and $sb_{2q}(\boldsymbol{Z})$ $(2 \leqslant q \leqslant d)$ are chosen as follows. By (2.16), we can choose the percentiles of $sb_{1q}(\boldsymbol{Z})$ and $sb_{2q}(\boldsymbol{Z})$ as those given by Romeu and Ozturk (1993, Table A1, A3–A4) when they are available. Otherwise, we generate samples $\boldsymbol{Z}_0 \sim \mathrm{N}_{(n-1)\times q}(\boldsymbol{0}, \boldsymbol{I}_{n-1} \otimes \boldsymbol{I}_q)$ for each fixed q to calculate $sb_{1q}(\boldsymbol{Z}_0)$ and $sb_{2q}(\boldsymbol{Z}_0)$. By simulation with 10,000 replications, we can obtain the percentiles of $sb_{1q}(\boldsymbol{Z})$ and $sb_{2q}(\boldsymbol{Z})$. Two groups of alternative distributions are chosen. The first group of alternative distributions consists of the subclasses of elliptical distributions whose density functions can be found in Fang et al. (1990, Chapter 3): (a1) the multivariate t-distribution $\boldsymbol{x} \sim Mt_d(m, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, we choose $m = 5$ in the simulation; (a2) the Kotz-type distribution, we choose $N = 2$, $s = 1$ and $r = 0.5$; (a3) the Pearson-type VII distribution, we choose $m = 1$ and $N = 15$; (a4) the Pearson-type II distribution, we choose $m = \frac{3}{2}$. Since the distributions of $sb_{1q}(\boldsymbol{Z})$ and $sb_{2q}(\boldsymbol{Z})$ are independent of the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ under the null hypothesis (2.20) when $\boldsymbol{Z}$ is obtained by (2.21), in the simulation we can choose $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_d$, without loss of generality, for both of the normal distribution and the elliptical distributions (a1)–(a4). The *TFWW algorithm* (Tashiro, 1977; Fang and Wang, 1994, pp. 166–170) is employed to generate samples from the above selected elliptical distributions. The second group of non-elliptical distributions is described as follows, where the sign $[x]$ denotes the integer part of x: (b1) the multivariate double exponential distribution is taken to be composed of i.i.d. univariate double exponential distribution, each has a density $f(x) = \exp(-|x|)/2$, $x \in R$; (b2) the multivariate chi-squared distribution is composed of i.i.d. univariate $\chi^2(1)$; (b3) multivariate exponential distribution is composed of i.i.d. univariate exponential variables, each of them has a density $f(x) = \exp(-x)$, $x > 0$; (b4) normal+χ^2: the multivariate distribution with i.i.d. marginals, $[d/3]$ marginals have a normal distribution and $d - [d/3]$ marginals have a chi-squared distribution $\chi^2(1)$; (b5) the generalized lambda distribution GLD-1 as used by Romeu and Ozturk (1993) with lambda parameters corresponding to the classical skewness $\sqrt{\beta_1} = 0.89$ and kurtosis $\beta_2 = 3.2$. To demonstrate the performance of the projection statistics on the alternative distributions with pure (multivariate) skewness $\beta_{1d} > 0$ and the kurtosis β_{2d} (Mardia, 1970) being the same as that of the normal distribution $\mathrm{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we obtain the lambda parameters (Dudewicz and Karian, 1996, Table 1) of GLD-1 as $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (-1.1734, 0.1828, 0.007038, 0.2846)$ for $d = 10$ (with $\beta_{1,10} = 7.7281$) and $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (-1.1755, 0.1825, 0.007038, 0.2846)$ for $d = 20$ (with $\beta_{1,20} = 15.6246$). Empirical samples from these non-elliptical distributions are generated according to their marginal distributions. The multivariate GLD-1 is generated by the same way as in Romeu and Ozturk (1993). For each generated set of samples (the $n \times d$ sample matrix $\boldsymbol{X}$), we firstly use the statistics $sb_{1d}(\boldsymbol{X})$ and $sb_{2d}(\boldsymbol{X})$ given by (2.11)–(2.13) to test multinormality of $\boldsymbol{X}$ before using (2.2) and then perform (2.21) to carry out the projection tests by the statistics $sb_{1q}(\boldsymbol{Z})$ and $sb_{2q}(\boldsymbol{Z})$ given by (2.14)–(2.15) for each $q = 2, \ldots, d$. By simulation with 2000 replications, we obtain the power values (at $\alpha = 0.05$) of the four statistics

$$sb_{1d}(\boldsymbol{X}), \quad sb_{2d}(\boldsymbol{X}), \quad sb_{1q}(\boldsymbol{Z}), \quad sb_{2q}(\boldsymbol{Z}), \tag{3.1}$$

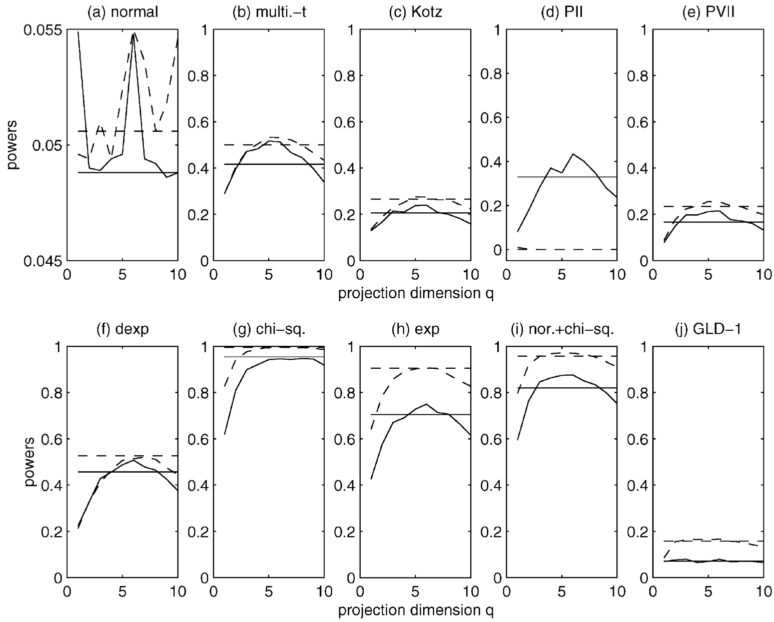
in testing multinormality of the two groups of alternative distributions. For the special case $q = 1$, the one-dimensional projection tests are done by the classical skewness and kurtosis statistics $\mathrm{Sk}_{n-1}(z)$ and $\mathrm{Ku}_{n-1}(z)$ given by (2.17)–(2.18). The percentiles of $\mathrm{Sk}_{n-1}(z)$ and $\mathrm{Ku}_{n-1}(z)$ are chosen from Pearson and Hartley (1972, Table 34). For convenience in viewing the changing trend of the power under different choices of q's, we demonstrate all of the power values of the four statistics (3.1) in Fig. 1.

It can be seen from plot (a) of Fig. 1(A) and (B) that the two projection statistics $sb_{1q}(\boldsymbol{Z})$ and $sb_{2q}(\boldsymbol{Z})$ keep type I error rate within a suitable range for every q, while the power values for the nine selected non-normal alternative distributions depend on the choice of q. From Fig. 1(B) for dimension $d=20$, it seems that the best value of q is close to $q=[d/3]$ for the case of high dimension with a small sample size. The power of these cases is remarkably improved by the projection tests by $sb_{1q}(\boldsymbol{Z})$ and $sb_{2q}(\boldsymbol{Z})$. Moreover, the power of $\mathrm{Sk}_{n-1}(z)$ and $\mathrm{Ku}_{n-1}(z)$ (i.e., $q=1$) for most of the cases is almost equivalent to the power of $sb_{1d}(\boldsymbol{X})$ and $sb_{2d}(\boldsymbol{X})$, respectively. However, it can be seen from Fig. 1(A) for dimension $d=10$ that the power improvement by the projection tests by $sb_{1q}(\boldsymbol{Z})$ and $sb_{2q}(\boldsymbol{Z})$ is not so remarkable as those cases for $d=20$. In the cases shown by Fig. 1(A), it seems that the best choice of q is close to $q=[d/2]$. It is noted that, for the distribution PII in plot (d) in both Fig. 1(A) and (B), which has no skewness, the power values of both $sb_{1d}(\boldsymbol{X})$ and $sb_{1q}(\boldsymbol{Z})$ (for different choices of q) are controlled by the significant level $\alpha=0.05$. Similarly, for the distribution GLD-1 in plot (j) in both Fig. 1(A) and (B), which is purely skewed, the power values of both $sb_{2d}(\boldsymbol{X})$ and $sb_{2q}(\boldsymbol{Z})$ (for different choices of q) are also basically controlled by the significant level $\alpha=0.05$. This is consistent with the fact that skewness statistics could not detect kurtosis problems and viceversa. From plot (j) in both Fig. 1(A) and (B), it seems that Mardia's skewness coefficient is not very sensitive to multivariate skewness. When choosing the matrix $\boldsymbol{D}$ in (2.21) as composed of the first q columns of the matrix $\boldsymbol{D}_0$ in (2.24), we obtain similar empirical results to those by choosing the matrix $\boldsymbol{D}$ in (2.21) according to (2.25).

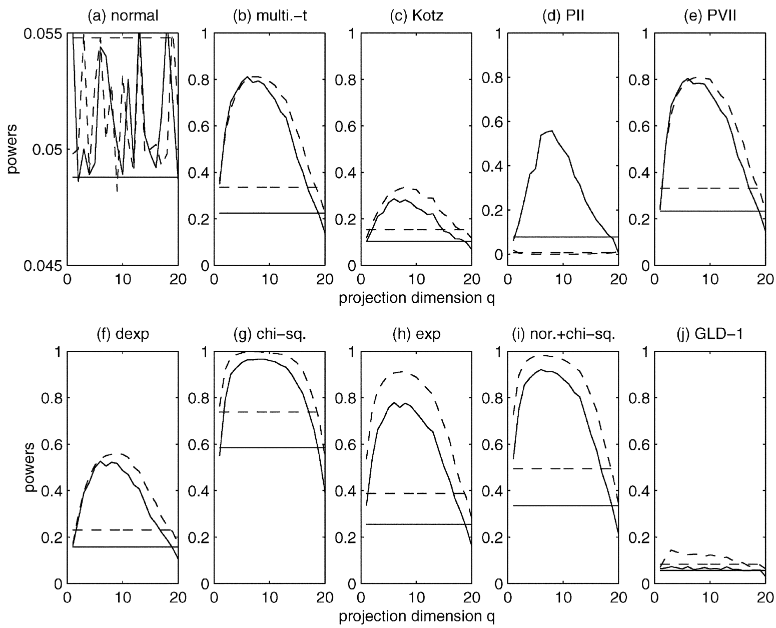
3.2. Applications

In implementing the projection tests in Section 2.2, we can also employ the idea in principal components analysis to give a suitable choice of the projection dimension q. One of the treatments in choosing the principal components is based on the *explanation of variability* (Muirhead, 1982). This procedure is as follows: let $\boldsymbol{G}$ be determined by (2.25) and $g_1 > \cdots > g_d$ be its diagonal elements. $\sum_{i=1}^{d} g_i$ is called the *total variability* (or the *total variation* in some literature (Jolliffe, 1986) and $e_q=\sum_{i=1}^{q} g_i/\sum_{i=1}^{d} g_i$ the *explanation of variability* of the first q $(1\leqslant q\leqslant d)$ principal components. If q is the first value which makes e_q to achieve above some given level, e.g., 0.95, then the first q principal components can be used as the projection directions for dimension reduction. In the following we implement this idea to analyze a practical data set by the four statistics (3.1).

Fig. 1. Power comparisons for different projection dimension q: dashed line stands for $sb_{1d}(\boldsymbol{X})$, real line for $sb_{2d}(\boldsymbol{X})$, dashed curve for $sb_{1q}(\boldsymbol{Z})$ and real curve for $sb_{2q}(\boldsymbol{Z})$. In both of (A) and (B), (a) stands for the normal distribution; (b) for the multivariate t-distribution; (c) for the Kotz's-type distribution; (d) for the Pearson-type II distribution; (e) for the Pearson-type VII distribution; (f) for the double exponential distribution; (g) for the chi-squared distribution; (h) for the exponential distribution; (i) for the distribution with normal and chi-squared marginal distributions; and (j) for the distribution GLD-1.



(A) for $d = 10$, $n = 26$, $\alpha = 5\%$



(B) for $d = 20$, $n = 26$, $\alpha = 5\%$

Table 1
p-values of the statistics Sk_n and Ku_n in testing univariate normality of the nine variables in Example 3.1

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
Sk_n	0.0103	0.0016	0.0616	0.0815	0.0006	0.2548	0.0278	0.0149	0.2257
Ku_n	0.0300	0.0023	0.2841	0.4656	0.0023	0.2799	0.0843	0.0080	0.2465

Example 3.1. The data are from 19 depressive patients acquired at the beginning and at the end of a six week therapy. The nine variables represent the changes of the absolute theta power of electroencephalogram (EEG) during the therapy in nine selected channels (denoted by X_1 to X_9, respectively). The full data can be found in Table 1 of Läuter et al. (1996).

Now, we want to test multinormality of the nine variables $X_1,\dots,X_9$. At first, we use the classical skewness Sk_n and kurtosis statistics Ku_n to test the marginal normality of the nine variables so that we may find some possible sources of departure from joint normality before implementing the projection tests. Table 1 gives the simulation results on the p-values of the two statistics Sk_n and Ku_n, where the p-values are obtained by generating 10,000 normal samples $z \sim \mathrm{N}_n(\mathbf{0}, \boldsymbol{I}_n)$ $(n=19)$.

From Table 1, we can draw an initial conclusion: some of the nine variables $(X_1,\dots,X_9)$ show evidence of non-normality. Sources of departure from joint normality may come from the variables $X_1,\dots,X_5$, X_7 and X_8. Among them, X_2 and X_5 seem to show severe departure from normality. For illustration, we consider the three sets of variables:

$$\begin{aligned}
&(1)\ \text{the nine-dimensional random vector } (X_1,\dots,X_9);\\
&(2)\ \text{the eight-dimensional random vector } (X_1,\dots,X_5,X_7,X_8,X_9);\\
&(3)\ \text{the seven-dimensional random vector } (X_1,\dots,X_5,X_7,X_8).
\end{aligned} \tag{3.2}$$

Second, we perform (2.21) by: (a) $\boldsymbol{X} \overset{(1)}{\Rightarrow} \boldsymbol{Y} = \boldsymbol{AX}$ is realized by the Helmert's transformation (2.7); (b) $\boldsymbol{Y} \overset{(2)}{\Rightarrow} \boldsymbol{Z} = \boldsymbol{YD}$ is realized by choosing $\boldsymbol{D}$ as composed of the first q columns $(q=1,\dots,9)$ of the matrix $\boldsymbol{D}_1$ in (2.25). In the case $q=1$, the statistics $\mathrm{Sk}_{n-1}(z)$ and $\mathrm{Ku}_{n-1}(z)$ given by (2.17)–(2.18) are used to test spherical symmetry of z (hypothesis (2.23)). For each q, the observed values of the statistics

$$sb_{1q}(\boldsymbol{Z}), \quad sb_{2q}(\boldsymbol{Z}), \quad \mathrm{Sk}_{n-1}(z), \quad \mathrm{Ku}_{n-1}(z) \tag{3.3}$$

are calculated from the given data. Then we generate 10,000 normal samples $\boldsymbol{Z}_0 \sim \mathrm{N}_{(n-1)\times q}(\mathbf{0}, \boldsymbol{I}_{n-1} \otimes \boldsymbol{I}_q)$ $(n=19,\ q=1,\dots,9)$ for simulating the p-values of the four statistics (3.3) in testing hypothesis (2.20) and (2.23), respectively. The p-values of $sb_{1d}(\boldsymbol{X})$ and $sb_{2d}(\boldsymbol{X})$ in testing multinormality for the three sets of variables in (3.2) are obtained by generating 10,000 normal samples $\boldsymbol{X}_{n\times d} \sim \mathrm{N}_{n\times d}(\mathbf{0}, \boldsymbol{I}_n \otimes \boldsymbol{I}_d)$ for $d=7$, 8, 9 and $n=19$. Table 2 presents the simulation results and the explanation of variability e_q, where e_q is obtained by solving the eigenvalue problem (2.25).

Table 2
Explanation of variability and the corresponding p-values of the statistics $sb_{1d}(\boldsymbol{X})$, $sb_{2d}(\boldsymbol{X})$, $sb_{1q}(\boldsymbol{Z})$, $sb_{1q}(\boldsymbol{Z})$ and $\mathrm{Sk}_{n-1}(z)$, $\mathrm{Ku}_{n-1}(z)$

q	1	2	3	4	5	6	7	8	9
$(X_1,\dots,X_9)$ with p-value of $sb_{1d} = 0.0053$ & p-value of $sb_{2d} = 0.0000$									
e_q	0.7083	0.8665	0.9266	0.9623	0.9769	0.9873	0.9954	0.9981	1.0000
p-value of sb_{1q}	0.0386	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
p-value of sb_{2q}	0.3120	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0008
$(X_1,\dots,X_5,X_7,X_8,X_9)$ with p-value of $sb_{1d} = 0.0024$ & p-value of $sb_{2d} = 0.0064$									
e_q	0.7251	0.8819	0.9377	0.9726	0.9849	0.9946	0.9975	1.0000	—
p-value of sb_{1q}	0.0309	0.0001	0.0000	0.0000	0.0000	0.0000	0.0004	0.0052	—
p-value of sb_{2q}	0.2616	0.0002	0.0000	0.0000	0.0000	0.0000	0.0021	0.0042	—
$(X_1,\dots,X_5,X_7,X_8)$ with p-value of $sb_{1d} = 0.0003$ & p-value of $sb_{2d} = 0.0005$									
e_q	0.7518	0.9168	0.9566	0.9761	0.9890	0.9970	1.0000	—	—
p-value of sb_{1q}	0.0254	0.0011	0.0003	0.0000	0.0000	0.0001	0.0016	—	—
p-value of sb_{2q}	0.2123	0.0009	0.0001	0.0000	0.0002	0.0002	0.0018	—	—

In Table 2, the p-values 0.0000 of some projection tests imply that those p-values are less than 10^{-4} in 10^4 replications of simulation. The following remarks can be made from Table 2: (a) in testing the joint normality of $(X_1,\dots,X_9)$, the p-value of $sb_{1d}(\boldsymbol{X}) = 0.0053$ and the p-value of $sb_{2d}(\boldsymbol{X}) < 10^{-4}$, we conclude that the nine variables shows severe departure from joint normality at α=0.05. Most of the q-dimensional $(q = 1,\dots,9)$ projection tests except $\mathrm{Ku}_{n-1}(z)$ support this conclusion at $\alpha = 0.05$ and those q-dimensional $(q\geqslant 2)$ projection tests strongly support the conclusion drawn by $sb_{1d}(\boldsymbol{X})$ and $sb_{2d}(\boldsymbol{X})$ at $\alpha = 0.01$. If one would like to choose a projection dimension q according to the idea in principal components analysis under some given explanation of variability e_q, e.g., $e_q \geqslant 0.95$, the first q which e_q firstly exceeds 0.95 is $q = 4$. It can be seen that the four-dimensional projection tests by $sb_{14}(\boldsymbol{Z})$ and $sb_{24}(\boldsymbol{Z})$ give consistent conclusion with that drawn by $sb_{1d}(\boldsymbol{X})$ and $sb_{2d}(\boldsymbol{X})$ at $\alpha = 0.01$; (b) in testing multinormality for both of the two sets of variables given by (2) and (3) in (3.2), all of the q-dimensional $(q\geqslant 2)$ projection tests give consistent conclusions with those drawn by $sb_{1d}(\boldsymbol{X})$ and $sb_{2d}(\boldsymbol{X})$ at $\alpha = 0.01$. The analysis on the real data in Example 3.1 shows that the approach to testing multinormality by the projection tests is a useful complement to affine invariant statistics in testing high-dimensional normality.

4. Conclusions and further applications

Transformation (2.21) acts as a role of dimension reduction in testing high-dimensional normality. The projection tests are performed based on the transformed random matrix $\boldsymbol{Z}_{(n-1)\times q}$, which is indirectly dependent on the dimension d of the raw data $\boldsymbol{X}$. When $n < d$ in the raw data, many affine invariant statistics become unusable in testing multinormality while the projection tests are still applicable if one chooses the projection dimension $q < \min(d, n-1)$. Although marginal detection of univariate nonnormality can be performed to find any evidence of non-multinormality, the overall level of significance cannot be determined by the univariate tests. In addition, under the same sample size n, large dimension d seems to have no much influence on the efficiency of the projection tests once q is selected. This implies that the approach to testing multinormality by the projection tests is specially suitable in the case of high dimension with a small sample size. Besides applying the Mardia's skewness and kurtosis statistics to construct projection tests for multinormality, we can apply some other affine invariant statistics for the same purpose such as those studied by Hawkins (1981), Koziol (1982), and Henze and Wagner (1997).

References

Ahn, S.K., 1992. F-probability plot and its application to multivariate normality. Commun. Statist. –Theoret. Method 21, 997–1023.

Bowman, A.W., Foster, P.J., 1993. Adaptive smoothing and density-based tests of multivariate normality. J. Amer. Statist. Assoc. 88, 529–537.

Cox, D.R., Wermuth, N., 1994. Test of linearity, multivariate normality and adequacy of linear scores. Appl. Statist. 43, 347–355.

Csörgo, S., 1986. Testing for normality in arbitrary dimension. Ann. Statist. 14, 706–723.

Dawid, A.P., 1977. Spherical matrix distributions and a multivariate model. J. Roy. Statist. Soc. Ser. B 39, 254–261.

Dudewicz, E.J., Karian, Z.A., 1996. The extended generalized lambda distribution (EGLD) system for fitting distributions to data with moments, II: Tables. Amer. J. Math. Management Sci. 16, 271–332.

Fang, K.T., Kotz, S., Ng, K.W., 1990. Symmetric Multivariate and Related Distributions. Chapman & Hall, London.

Fang, K.T., Wang, Y., 1994. Number-Theoretic Methods in Statistics. Chapman & Hall, London.

Fang, K.T., Zhang, Y., 1990. Generalized Multivariate Analysis. Springer-Verlag and Science Press, Berlin/Beijing.

Hawkins, D.M., 1981. A new test for multivariate normality and homoscedasticity. Technometrics 23, 105–110.

Henze, N., Wagner, T., 1997. A new approach to the BHEP tests for multivariate normality. J. Multivariate Anal. 62, 1–23.

Horswell, R.L., Looney, S.W., 1992. A comparison of tests for multivariate normality that are based on measures of multivariate skewness and kurtosis. J. Statist. Comput. Simul. 42, 21–38.

Jolliffe, I.T., 1986. Principal Component Analysis. Springer, New York.

Koziol, J.A., 1982. A class of invariant procedures for assessing multivariate normality. Biometrika 69, 423–427.

Läuter, J., 1996. Exact t and F tests for analyzing studies with multiple endpoints. Biometrics 52, 964–970.

Läuter, J., Glimm, E., Kropf, S., 1996. New multivariate tests for data with an inherent structure. Biometrical J. 38 (1), 5–23.

Mardia, K.V., 1970. Measures of multivariate skewness and kurtosis with applications. Biometrika 57, 519–530.

Mardia, K.V., 1980. Tests of univariate and multivariate normality. In: Krishnaiah, P.R. (Ed.), Handbook of Statistics, Vol. 1. North-Holland Publishing Company, Amsterdam, pp. 279–320.

Mudholkar, G.S., McDermott, M., Srivastava, D.K., 1992. A test of p-variate normality. Biometrika 79, 850–854.

Mudholkar, G.S., Srivastava, D.K., Lin, C.Th., 1995. Some p-variate adaptations of Shapiro–Wilk test of normality. Commun. Statist. –Theoret. Methods 24, 953–985.

Muirhead, R.J., 1982. Aspects of Multivariate Statistical Theory. Wiley, New York.

Pearson, E.S., Hartley, H.O., 1972. Biometrika Tables for Statisticians, Vol. I. Cambridge University Press, London.

Romeu, J.L., Ozturk, A., 1993. A comparative study of goodness-of-fit tests for multivariate normality. J. Multivariate Anal. 46, 309–334.

Tashiro, D., 1977. On methods for generating uniform points on the surface of a sphere. Ann. Inst. Statist. Math. 29, 295–300.

Zhu, L.X., Wong, H.L., Fang, K.T., 1995. Testing multinormality based on sample entropy and projection pursuit. J. Statist. Plan. Inference 45, 373–385.