Tests for multivariate normality based on canonical correlations

Måns Thulin

Abstract We propose new affine invariant tests for multivariate normality, based on independence characterizations of the sample moments of the normal distribution. The test statistics are obtained using canonical correlations between sets of sample moments in a way that resembles the construction of Mardia's skewness measure and generalizes the Lin–Mudholkar test for univariate normality. The tests are compared to some popular tests based on Mardia's skewness and kurtosis measures in an extensive simulation power study and are found to offer higher power against many of the alternatives.

Keywords Goodness-of-fit · Kurtosis · Multivariate normality · Skewness · Test for normality

1 Introduction

Many classical multivariate statistical methods are based on the assumption that the data comes from a multivariate normal distribution. Consequently, the use of such methods should be followed by an investigation of the assumption of normality. A number of tests for multivariate normality can be found in the literature, but the field has not been investigated to the same extent as have tests for univariate normality.

Let $\gamma = \mathrm{E}(X - \mu)^3/\sigma^3$ denote the skewness of a univariate random variable X and $\kappa = \mathrm{E}(X - \mu)^4/\sigma^4 - 3$ denote its (excess) kurtosis. Both these quantities are 0 for

M. Thulin (✉)
Department of Mathematics, Uppsala University, P.O. Box 480, 751 06 Uppsala, Sweden
e-mail: thulin@math.uu.se

the normal distribution but nonzero for many other distributions, and some common tests for univariate normality are therefore based on γ and κ.

Different analog multivariate measures of skewness and kurtosis have been proposed, perhaps most notably by Mardia (1970). Said measures have been used for various tests for multivariate normality in the last few decades. Some of these tests, in particular the tests that use Mardia's skewness and kurtosis measures as test statistics, have proved to have high power in many simulation studies (e.g. Mecklin and Mundfrom 2004, 2005) and new tests for normality based on multivariate skewness and kurtosis continue to be published today (Doornik and Hansen 2008; Kankainen et al. 2007).

In many inferential situations, some types of departures from normality are a more serious concern than are others. For instance, MANOVA is known to be sensitive to deviations from normality in the form of asymmetry, but to be relatively robust against deviations in the form of heavy tails. Using skewness and kurtosis allows us to construct tests that are directed toward some particular class of alternatives: skewness is used to detect asymmetric alternatives whereas kurtosis is used to detect alternatives with either short or long tails. This typically results in tests that, in comparison to omnibus tests that are directed to all alternatives, have higher power against the class of alternatives that they are directed to.

While more directed toward certain alternatives, such tests may however still be prone to reject alternatives from other classes. The sample skewness and sample kurtosis are correlated, which for instance can cause a skewness-based test to reject normality for a symmetric distribution with heavy tails. Henze (2002) and others have argued that this is a reason to avoid directed tests for normality. Directed tests will however in general have comparatively low power against alternatives that they are not directed to, lowering the risk of rejecting normality because of an unimportant deviation from normality. It is arguably better to have a test that has high power against interesting alternatives and lower power against uninteresting alternatives, rather than a test that has medium high power against all alternatives.

In this paper six new directed tests for normality, all related to multivariate skewness or kurtosis, are proposed. Their common basis is independence characterizations of sample moments of the multivariate normal distribution.

In Sect. 2 we reexamine Mardia's measure of multivariate skewness, which leads to two new classes of tests for multivariate normality. In Sect. 3 we state explicit expressions for covariances between multivariate sample moments in terms of moments of $X = (X_1, \ldots, X_p)'$. This will allow us to estimate the moments involved and to test whether these sample moments are correlated.

In Sect. 4 we study the first class of new tests for normality, all of which are related to multivariate skewness. These can be viewed as multivariate generalizations of the univariate Z_2' test (Thulin 2010), which in turn is a modified version of the Lin and Mudholkar (1980) test. In Sect. 5 we study the second class of tests, related to multivariate kurtosis. These, in turn, are generalizations of the Thulin (2010) Z_3' modification of a test proposed by Mudholkar et al. (2002). The tests are applied to the *Iris* data in Sect. 6. The results of a simulation study comparing the new tests with tests based on Mardia's skewness and kurtosis measures is presented in Sect. 7, which is followed by a discussion in Sect. 8. The text concludes with an appendix

containing proofs and tables. Additional tables and figures are included in two online supplements.

2 Mardia's multivariate skewness and kurtosis measures revisited

2.1 Multivariate skewness

A well-known characterization of the multivariate normal distributions is that the i.i.d. p-variate variables $X_1, \ldots, X_n$ are normal if and only if the sample mean vector $\bar{X} = (\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_p)'$ and the sample covariance matrix S are independent. Our aim is to test this independence in order to assess the normality of a population. As testing independence is difficult, we will resort to testing correlations instead.

Assume that $X_1, \ldots, X_n$ are i.i.d. p-variate random variables with nonsingular covariance matrix Σ. Let $\bar{X} = (\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_p)'$ be the sample mean vector and let

$$S = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1p} \\ S_{12} & S_{22} & \cdots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{1p} & S_{2p} & \cdots & S_{pp} \end{bmatrix}$$

be the sample covariance matrix with $S_{ij} = (n-1)^{-1} \sum_{k=1}^{n} (X_{k,i} - \bar{X}_i)(X_{k,j} - \bar{X}_j)$. Define

$$u = \text{vech}(S) = (S_{11}, S_{12}, \ldots, S_{1p}, S_{22}, S_{23}, \ldots, S_{2p}, S_{33}, \ldots, S_{p-1,p}, S_{pp})'$$

so that u is a vector containing the $q = p(p+1)/2$ distinct elements of S. Now, consider the covariance matrix of the vector $(\bar{X}', u')'$, in the following denoted $(\bar{X}, u)$:

$$\text{Cov}((\bar{X}, u)) = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} \tag{1}$$

where $\Lambda_{11} = \text{Cov}(\bar{X})$, $\Lambda_{22} = \text{Cov}(u)$, $\Lambda_{21} = \Lambda_{12}'$ and Λ_{12} contains covariances of the type $\text{Cov}(\bar{X}_i, S_{jk})$, $i, j, k = 1, \ldots, p$. If $\bar{X}$ and u are uncorrelated, then $\Lambda_{12} = 0$.

Mardia (1970, 1974) noted that for univariate random variables, asymptotically $\text{cor}(\bar{X}, S^2) \approx \frac{1}{\sqrt{2}}\gamma$ if κ is assumed to be negligible. Based on this, he used $\text{Cov}((\bar{X}, u))$ to construct a multivariate skewness measure. Studying the canonical correlations (see e.g. Mardia et al. 1979) between $\bar{X}$ and u he proposed the measure

$$\beta_{1,p} = 2 \sum_{i=1}^{p} \lambda_i^2$$

where $\lambda_1, \ldots, \lambda_p$ are the canonical correlations. This expression reduces to $2\text{cor}(\bar{X}, S^2)^2 \approx \gamma^2$ for univariate random variables.

From the theory of canonical correlations we have that $\lambda_1^2, \ldots, \lambda_p^2$ are the eigenvalues of $\mathbf{\Lambda_{11}}^{-1}\mathbf{\Lambda_{12}}\mathbf{\Lambda_{22}}^{-1}\mathbf{\Lambda_{21}}$ and thus

$$\beta_{1,p} = 2tr(\mathbf{\Lambda_{11}}^{-1}\mathbf{\Lambda_{12}}\mathbf{\Lambda_{22}}^{-1}\mathbf{\Lambda_{21}}).$$

Taking these moments to order n^{-1} Mardia showed that

$$\beta_{1,p} \approx \mathrm{E}\left((X - \mu)'\Sigma^{-1}(Y - \mu)\right)^3$$

where X and Y are independent and identical random vectors. The sample counterpart of the above expression,

$$b_{1,p} = \frac{1}{n^2}\sum_{i,j=1}^{n}\left((X_i - \bar{X})'S^{-1}(X_j - \bar{X})\right)^3, \tag{2}$$

is commonly used as a measure for multivariate skewness and as a test statistic for a test for multivariate normality.

In Section 2.8 of McCullagh (1987) Mardia's approximation of $\beta_{1,p}$ is shown to be a natural generalization of γ^2. It is however not necessarily a good approximation of the canonical correlations between $\bar{X}$ and u. An important assumption underlying Mardia's skewness measure is that the fourth central moments of the distribution are negligible. Seeing as, for univariate variables, $\gamma^2 - 2 \leq \kappa$; see Dubkov and Malakhov (1976); this seems like a rather strong condition. For univariate random variables, Thulin (2010) noted that

$$\rho_2 = \mathrm{cor}(\bar{X}, S^2) = \frac{\gamma}{\sqrt{\kappa + 3 - \frac{n-3}{n-1}}}$$

and used $\hat{\rho}_2 = Z_2'$, the sample moment version of this quantity, as a test statistic for a test for normality, proposing a test that is a modified version of the test of Lin and Mudholkar (1980). In Thulins' simulation power study Z_2' was more powerful than $\hat{\gamma}$ against most of the alternatives under study. Consequently, for $p = 1$ it is better to use the explicit expression for $\mathrm{cor}(\bar{X}, S^2)$ rather than the approximation $\mathrm{cor}(\bar{X}, S^2) \approx \frac{1}{\sqrt{2}}\gamma$. It is therefore of interest to use Mardia's approach without any approximations, in the hope that this will render a more powerful test for normality. In Sect. 3 we give explicit expressions for $\mathrm{Cov}(\bar{X}_i, S_{jk})$ and $\mathrm{Cov}(S_{ij}, S_{kl})$, allowing us to study $\mathbf{\Lambda_{11}}^{-1}\mathbf{\Lambda_{12}}\mathbf{\Lambda_{22}}^{-1}\mathbf{\Lambda_{21}}$ without approximations and to construct new test statistics.

2.2 Multivariate kurtosis

Mardia (1970, 1974) proposed the multivariate kurtosis measure

$$\beta_{2,p} = \mathrm{E}\left((X - \mu)'\Sigma^{-1}(Y - \mu)\right)^2$$

with sample counterpart

$$b_{2,p} = \frac{1}{n} \sum_{i=1}^{n} \left((X_i - \bar{X})' S^{-1} (X_i - \bar{X}) \right)^2. \tag{3}$$

In the univariate setting

$$\rho_3 = \text{cor}\left(\bar{X}, \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} (X_i - \bar{X})^3 \right)$$

$$= \frac{\kappa}{\sqrt{\lambda + 9\frac{n}{n-1}(\kappa + \gamma^2) + \frac{6n^2}{(n-1)(n-2)}}}, \tag{4}$$

where $\lambda = \frac{\mu_6}{\sigma^6} - 15\kappa - 10\gamma^2 - 15$ is the sixth standardized cumulant (Thulin 2010). In a simulation power study (Thulin 2010) found the test for normality based on $\hat{\rho}_3 = Z_3'$, the sample counterpart of (4), to have a better overall performance than the popular $\hat{\kappa} = b_2 = b_{2,1}$ test. It is therefore of interest to find a multivariate generalization of Z_3', in hopes that it will yield a test with higher power than $b_{2,p}$.

Similarly to what was done above for the covariance, let

$$S_{ijk} = \frac{n}{(n-1)(n-2)} \sum_{r=1}^{n} (X_{r,i} - \bar{X}_i)(X_{r,j} - \bar{X}_j)(X_{r,k} - \bar{X}_k)$$

and

$$\boldsymbol{v} = (S_{111}, S_{112}, \ldots, S_{pp(p-1)}, S_{ppp})',$$

a vector of length $p + p(p-1) + p(p-1)(p-2)/6$. We will construct tests based on the fact that $\bar{X}$ and $\boldsymbol{v}$ are independent if $X_1, \ldots, X_n$ are normal. The covariance matrix of $(\bar{X}, \boldsymbol{v})$ can be written as

$$\text{Cov}((\bar{X}, \boldsymbol{v})) = \begin{bmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} \\ \boldsymbol{\Psi}_{21} & \boldsymbol{\Psi}_{22} \end{bmatrix} \tag{5}$$

where $\boldsymbol{\Psi}_{11} = \text{Cov}(\bar{X})$, $\boldsymbol{\Psi}_{22} = \text{Cov}(\boldsymbol{v})$, $\boldsymbol{\Psi}_{21} = \boldsymbol{\Psi}_{12}'$ and $\boldsymbol{\Psi}_{12}$ contains covariances of the type $\text{Cov}(\bar{X}_i, S_{jkl})$, $i, j, k, l = 1, \ldots, p$. If X and $\boldsymbol{v}$ are uncorrelated, $\boldsymbol{\Psi}_{12} = \boldsymbol{0}$.

3 Explicit expressions for the covariances

In the following theorems we state explicit expressions for the elements of $\text{Cov}((\bar{X}, \boldsymbol{u}))$ and $\text{Cov}((\bar{X}, \boldsymbol{v}))$ in terms of moments of $(X_1, \ldots, X_p)$. These covariances can be obtained by tedious but routine calculations of the moments involved, that are much simplified by the use of tensor notation, as described in McCullagh (1987). All five

covariances can be found scattered in the literature, expressed using cumulants: (6)–(8) are all given in Section 4.2.3 of McCullagh (1987), (9) is found in Problem 4.5 of McCullagh (1987) and (10) is expression (7) in Kaplan (1952).

Theorem 1 *Assume that $X_1, \ldots, X_n$ are i.i.d. p-variate random variables with $\mathrm{E}|X_i X_j X_k X_l| < \infty$ for $i, j, k, l = 1, 2, \ldots, p$. Let $\mu_{i_1, \ldots, i_s} = \mathrm{E}(X_{i_1} - \mu_{i_1})(X_{i_2} - \mu_{i_2}) \cdots (X_{i_s} - \mu_{i_s})$. Then, for $n \geq 2p + p(p-1)/2$ and $i, j, k, l = 1, 2, \ldots, p$*

(i) the elements of $\mathbf{\Lambda}_{11}$ are

$$\mathrm{Cov}(\bar{X}_i, \bar{X}_j) = \frac{1}{n}\mu_{ij}, \tag{6}$$

(ii) the elements of $\mathbf{\Lambda}_{12}$ and $\mathbf{\Lambda}_{21}$ are

$$\mathrm{Cov}(\bar{X}_i, S_{jk}) = \frac{1}{n}\mu_{ijk} \tag{7}$$

and
(iii) the elements of $\mathbf{\Lambda}_{22}$ are

$$\mathrm{Cov}(S_{ij}, S_{kl}) = \frac{1}{n}(\mu_{ijkl} - \mu_{ij}\mu_{kl}) + \frac{1}{n(n-1)}(\mu_{ik}\mu_{jl} + \mu_{il}\mu_{jk}). \tag{8}$$

Since $\mathbf{\Psi}_{11} = \mathbf{\Lambda}_{11}$, we only give the expressions for $\mathbf{\Psi}_{22}$ and $\mathbf{\Psi}_{12}$ in the following theorem.

Theorem 2 *Assume that $X_1, \ldots, X_n$ are i.i.d. p-variate random variables with $\mathrm{E}|X_\alpha X_\beta X_\gamma X_\delta X_\epsilon X_\zeta| < \infty$ for $\alpha, \ldots, \zeta = 1, 2, \ldots, p$. Let $\mu_{i_1, \ldots, i_s} = \mathrm{E}(X_{i_1} - \mu_{i_1})(X_{i_2} - \mu_{i_2}) \cdots (X_{i_s} - \mu_{i_s})$. Then, for $n \geq 2p + p(p-1) + p(p-1)(p-2)/6$ and $i, j, k, r, s, t = 1, 2, \ldots, p$*

(i) the elements of $\mathbf{\Psi}_{12}$ and $\mathbf{\Psi}_{21}$ are

$$\mathrm{Cov}(\bar{X}_i, S_{rst}) = \frac{1}{n}(\mu_{irst} - \mu_{ir}\mu_{st} - \mu_{is}\mu_{rt} - \mu_{it}\mu_{rs}) \tag{9}$$

and
(ii) the elements of $\mathbf{\Psi}_{22}$ are

$$\mathrm{Cov}(S_{ijk}, S_{rst}) = \frac{1}{n}\lambda_{ijkrst} + \frac{1}{n-1}\left(\sum^9 \mu_{ir}(\mu_{jkst} - \sum^3 \mu_{jk}\mu_{st}) + \sum^9 \mu_{ijr}\mu_{kst}\right)$$

$$+ \frac{n}{(n-1)(n-2)}\sum^6 \mu_{ir}\mu_{js}\mu_{kt} \tag{10}$$

where λ_{ijkrst} is given below and $\sum^k$ denotes summation over k distinct permutations of i, j, k, r, s, t. In particular, in $\sum^9 \mu_{ir}(\ldots)$ the summation is taken over

all permutations of i, j, k, r, s, t where i and either of j, k switch places and/or r and either of s, t switch places. In $\sum^9 \mu_{ijr}\mu_{kst}$ the summation is taken over all permutations except $\mu_{ijk}\mu_{rst}$. Finally, in $\sum^3 \mu_{jk}\mu_{st}$ and

$$\lambda_{ijkrst} = \mu_{ijkrst} - \sum^{15} \mu_{ij}\left(\mu_{krst} - \sum^3 \mu_{kr}\mu_{st}\right) - \sum^{10} \mu_{ijk}\mu_{rst} - \sum^{15} \mu_{ij}\mu_{kr}\mu_{st}$$

the sums are taken over all distinct permutations.

4 Tests based on $\bar{X}$ and u

4.1 Modifying Mardia's statistic

The factor 2 in Mardia's expression

$$\beta_{1,p} = 2tr(\mathbf{\Lambda_{11}}^{-1}\mathbf{\Lambda_{12}}\mathbf{\Lambda_{22}}^{-1}\mathbf{\Lambda_{21}})$$

is only of interest if we assume negligible fourth moments (in the sense of Mardia (1970)). We will therefore omit it in the following and instead study the quantity

$$tr(\mathbf{\Lambda_{11}}^{-1}\mathbf{\Lambda_{12}}\mathbf{\Lambda_{22}}^{-1}\mathbf{\Lambda_{21}}).$$

Let $\mathbf{L_{11}}, \mathbf{L_{22}}, \mathbf{L_{12}}$ and $\mathbf{L_{21}}$ be the sample counterparts of $\mathbf{\Lambda_{11}}, \mathbf{\Lambda_{22}}, \mathbf{\Lambda_{12}}$ and $\mathbf{\Lambda_{21}}$, where $\mu_{i_1,\dots,i_s} = \mathrm{E}(X_{i_1} - \mu_{i_1})(X_{i_2} - \mu_{i_2})\dots(X_{i_s} - \mu_{i_s})$ are estimated by the sample moments

$$m_{i_1,\dots,i_s} = n^{-1}\sum_{k=1}^{n}(x_{k,i_1} - \bar{x}_{i_1})(x_{k,i_2} - \bar{x}_{i_2})\dots(x_{k,i_s} - \bar{x}_{i_s}), \tag{11}$$

i.e. where the moments in Theorem 1 are replaced by their sample counterparts. The test statistic for the new test is

$$Z_{2,p}^{(HL)} = tr(\mathbf{L_{11}}^{-1}\mathbf{L_{12}}\mathbf{L_{22}}^{-1}\mathbf{L_{21}}). \tag{12}$$

The null hypothesis of normality is rejected if $Z_{2,p}^{(HL)}$ is sufficiently large.

$Z_{2,1}^{(HL)}$ coincides with $Z_2'^2$ from Thulin (2010) and is thus equivalent to the $|Z_2'|$ test presented there. $Z_{2,2}^{(HL)}$ is a polynomial of degree 10 in 13 moments and the full formula takes up more than two pages. It is however readily computed using a computer, as is $Z_{2,p}^{(HL)}$ for higher p.

It should be noted that differences in index notation complicate the situation somewhat here. Mardia's skewness is denoted $b_{1,p}$, with 1 as its index, whereas the univariate correlation statistic Z_2' has 2 as its index. When generalizing Z_2' to the multivariate setting we will keep the index 2, hoping that it won't be confused with Mardia's kurtosis measure $b_{2,p}$.

4.2 Other test statistics from the theory of canonical correlations

Let Y and Z be normal random vectors with

$$\text{Cov}((Y, Z)) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

partitioned like (1). Let $\hat{\Sigma}_{11}, \hat{\Sigma}_{22}$ and $\hat{\Sigma}_{12} = \hat{\Sigma}'_{21}$ be the sample covariance matrices and $\hat{v}_1^2, \ldots, \hat{v}_p^2$ be the eigenvalues of $\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}$. In Section 10.3 of Kshirsagar (1972) the test statistic of the likelihood ratio test of $H_0 : \Sigma_{12} = 0$ versus $H_1 : \Sigma_{12} \neq 0$ is shown to be

$$-n \log \prod_{i=1}^{p}(1 - \hat{v}_i^2). \tag{13}$$

Now, let $\hat{\lambda}_1^2 \geq \hat{\lambda}_2^2 \geq \ldots \geq \hat{\lambda}_p^2$ be the eigenvalues of $L_{11}^{-1}L_{12}L_{22}^{-1}L_{21}$. Assuming that the necessary moments exist, $\bar{X}$ and u are asymptotically normal. Although L_{22} and L_{12} are not the usual sample covariance matrices, in the light of (13), this suggests the use of the following statistic for a test for normality:

$$Z_{2,p}^{(W)} = \prod_{i=1}^{p}(1 - \hat{\lambda}_i^2). \tag{14}$$

The null hypothesis of normality is rejected if $Z_{2,p}^{(W)}$ is sufficiently small.

   Another quantity that has been considered for a test of $H_0 : \Sigma_{12} = 0$, for instance by Bartlett (1939), is

$$Z_{2,p}^{(PB)} = \sum_{i=1}^{p} \frac{\hat{\lambda}_i^2}{1 - \hat{\lambda}_i^2}. \tag{15}$$

$Z_{2,p}^{(PB)}$ is similar to $Z_{2,p}^{(HL)}$, but weighs the correlation coefficients so that larger coefficients become more influential. The null hypothesis should be rejected for large values of $Z_{2,p}^{(PB)}$.

   Finally, we can consider the statistic

$$Z_{2,p}^{(max)} = \max(\hat{\lambda}_1^2, \ldots, \hat{\lambda}_p^2) = \hat{\lambda}_1^2, \tag{16}$$

large values of which imply non-normality. $Z_{2,p}^{(max)}$ seems perhaps like the most natural choice for a test statistic, as $\lambda_1 = 0$ implies that all canonical correlations are 0.

   The statistics $Z_{2,p}^{(HL)}$, $Z_{2,p}^{(W)}$, $Z_{2,p}^{(PB)}$ and $Z_{2,p}^{(max)}$ are all related to well-known statistics from multivariate analysis of variance; they are analogs of the Hotelling–Lawley trace, Wilk's Λ, the Pillai–Bartlett trace and Roy's greatest root, respectively. For $p = 1$ these statistics are all equivalent to the $|Z_2'|$ test from Thulin (2010).

4.3 Theoretical results

Some fundamental properties of the new test statistics are presented in the following theorem. Its proof is given in the Appendix.

Theorem 3 *Assume that $X_1, \ldots, X_n$ are i.i.d. p-variate random variables fulfilling the conditions of Theorem 1. Then, for $n \geq 2p + p(p-1)/2$ and $i, j, k = 1, 2, \ldots, p$*

(i) *$Z_{2,p}^{(HL)}, Z_{2,p}^{(W)}, Z_{2,p}^{(PB)}$ and $Z_{2,p}^{(max)}$ are affine invariant, i.e. invariant under nonsingular linear transformations $AX + b$ where A is a nonsingular $p \times p$ matrix and b is a p-vector,*

(ii) *The population canonical correlation $\lambda_1 = \max_{a,b} |\rho(a\bar{X}, bu)| = 0$ if $\mu_{ijk} = 0$ for all i, j, k and > 0 if $\mu_{ijk} \neq 0$ for at least one combination of i, j, k, and*

(iii) *$Z_{2,p}^{(HL)}, Z_{2,p}^{(W)}, Z_{2,p}^{(PB)}$ and $Z_{2,p}^{(max)}$ converge almost surely to the corresponding functions of the population canonical correlations $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$.*

Since the statistics are affine invariant, their distributions are the same for all p-variate normal distributions for a given sample size n. These null distributions are easily obtained using Monte Carlo simulation.

Since $\lambda_1 \geq \lambda_j$ for $j > 1$, $\lambda_1 = 0$ implies that all population canonical correlations are 0, as is the case for the normal distribution. The tests should therefore not be sensitive to distributions with that kind of symmetry. All four statistics are, by (ii) and (iii), however consistent against alternatives where $\mu_{ijk} \neq 0$ for at least one combination of i, j, k. In particular, they are sensitive to alternatives with skew marginal distributions.

5 Tests based on $\bar{X}$ and v

5.1 Test statistics

The ideas used in Sect. 4 for $\text{Cov}((\bar{X}, u))$ can also be used for $\text{Cov}((\bar{X}, v))$ in an analog manner, yielding multivariate generalizations of (4). This leads to two new tests for normality, as described below.

Let P_{11}, P_{22}, P_{12} and P_{21} be the sample counterparts of $\Psi_{11}, \Psi_{22}, \Psi_{12}$ and Ψ_{21}, where the $\mu_{i_1,\ldots,i_s}$ are estimated by the sample moments, as in (11) above. Let $\hat{\psi}_1^2 \geq \ldots \geq \hat{\psi}_p^2$ be the eigenvalues of $P_{11}^{-1}P_{12}P_{22}^{-1}P_{21}$.

The test statistics for the new tests are

$$Z_{3,p}^{(HL)} = tr(P_{11}^{-1}P_{12}P_{22}^{-1}P_{21}) = \sum_{i=1}^{p} \hat{\psi}_i^2, \tag{17}$$

$$Z_{3,p}^{(W)} = \prod_{i=1}^{p}(1 - \hat{\psi}_i^2), \tag{18}$$

We have also considered other statistics, but found these to have lower power than these two. Large values of $Z_{3,p}^{(HL)}$ and small values of $Z_{3,p}^{(W)}$ imply non-normality. Both statistics are equivalent to $|Z_3'|$ from Thulin (2010) for $p = 1$.

5.2 Theoretical results

The following theorem mimics Theorem 3 above. Its proof is given in the Appendix.

Theorem 4 *Assume that $X_1, \ldots, X_n$ are i.i.d. p-variate random variables fulfilling the conditions of Theorem 2. Then, for $n \geq 2p + p(p-1) + p(p-1)(p-2)/6$ and $i, j, k, r, s, t = 1, 2, \ldots, p$*

(i) *$Z_{3,p}^{(HL)}$ and $Z_{3,p}^{(W)}$ are affine invariant, i.e. invariant under nonsingular linear transformations $AX + b$ where A is a nonsingular $p \times p$ matrix and b is a p-vector,*

(ii) *The population canonical correlation $\psi_1 = \max_{a,b} |\rho(a\bar{X}, bv)| = 0$ if $\mu_{irst} - \mu_{ir}\mu_{st} - \mu_{is}\mu_{rt} - \mu_{it}\mu_{rs} = 0$ for all $i, r, s, t = 1, \ldots, p$ and > 0 otherwise, and*

(iii) *$Z_{3,p}^{(HL)}$ and $Z_{3,p}^{(W)}$ converge almost surely to the corresponding functions of the population canonical correlations $\psi_1 \geq \psi_2 \geq \ldots \geq \psi_p$.*

Using the affine invariance, the null distributions of the statistics can be obtained through Monte Carlo simulation.

By (ii) and (iii) both statistics are consistent against alternatives where $\mu_{irst} - \mu_{ir}\mu_{st} - \mu_{is}\mu_{rt} - \mu_{it}\mu_{rs} \neq 0$ for at least one combination of i, r, s, t.

6 Analysis of the Iris data set

In Table 1 we present the results for the new tests when applied to the famous *Iris* data set of Fisher (1936). The tests are applied to each of the three subsets of the *Iris* data: *Setosa*, *Versicolor* and *Virginica*. For each such subset $n = 50$ and $p = 4$. We also applied Mardia's skewness test $b_{1,p}$ (2), Mardia's kurtosis test $b_{2,p}$ (3) and the Mardia–Kent omnibus test T (Mardia and Kent 1991), in which the skewness and kurtosis measures are combined. To compute the critical values and p-values, we approximated the null distribution of each test statistic by 10,000 simulated samples

Table 1 Results for the Iris data set

	$b_{1,p}$	$b_{2,p}$	T	$Z_{2,p}^{(HL)}$	$Z_{2,p}^{(W)}$	$Z_{2,p}^{(PB)}$	$Z_{2,p}^{(max)}$	$Z_{3,p}^{(HL)}$	$Z_{3,p}^{(W)}$
5 % critical value	3.41	24.87	4,868.38	1.42	0.15	2.64	0.58	2.90	0.004
Setosa sample									
p value	0.13	0.03	0.09	0.32	0.31	0.31	0.25	0.54	0.46
Statistic	2.90	25.49	4,241.12	1.13	0.24	1.82	0.49	2.52	0.01
Versicolor sample									
p value	0.14	0.51	0.49	0.14	0.15	0.16	0.23	0.68	0.64
Statistic	2.84	21.97	2,593.51	1.28	0.20	2.15	0.50	2.43	0.02
Virginica sample									
p value	0.11	0.20	0.18	0.31	0.33	0.34	0.28	0.29	0.38
Statistic	2.97	23.34	3,561.22	1.13	0.25	1.77	0.48	2.68	0.01

from a normal distribution. The resulting critical values at the 5 % level are given in the table. Recall that for the $\mathbf{Z}_{2,p}^{(W)}$ and $\mathbf{Z}_{3,p}^{(W)}$ statistics, values of the statistic that are *smaller* than the critical value imply non-normality.

At the 5 % level, the only time at which normality is rejected is when $b_{2,p}$ is applied to the *Setosa* sample.

7 Simulation results

7.1 The simulation study

To evaluate the performance of the new $Z_{2,p}$ and $Z_{3,p}$ tests, a Monte Carlo study of their power was carried out. The tests were compared to $b_{1,p}, b_{2,p}$ (3) and the Mardia–Kent test T. The tests were compared for $n = 20$ and $n = 50$ for $p = 2$ and $p = 3$. For some alternatives, more combinations of n and p were used. Since the results for $p = 2$ and $p = 3$ were quite similar, we only present the results for $p = 3$ below. The results for $p = 2$ can be found in Supplement S1.

In many power studies for multivariate tests for normality alternatives with independent marginal distributions have been used. We believe that this can be misleading, as distributions with independent marginals are uncommon in practice and indeed of little interest in the multivariate setting, where the dependence structure of the marginals often is paramount. For this reason, we decided to focus mainly on alternatives with a more complex dependence structure in our study. One alternative with independent exponential marginals, which has been used in many previous power studies, is included for reference.

The alternatives used in the study are presented in Tables 2 and 3. Contour plots of the alternatives from Table 2 are given in Supplement S2. The asymmetric multivariate Laplace distribution mentioned in Table 3 is described in Kotz et al. (2000).

Table 2 Alternatives constructed using their marginal distributions

Distr. of Y_i	Construction of $(Y_1, \ldots, Y_p)$
Indep. $Exp(1)$	$Y_1, \ldots, Y_p$ i.i.d. $Exp(1)$
$LogN(0, a)$	$X_0 \sim LogN(0, a/2)$ indep. of $X_1, \ldots, X_p$ i.i.d. $LogN(0, a/2)$. $Y_i = X_0 \cdot X_p$
$Laplace(0, 1)$ (type I)	$X_0 \sim Exp(1)$ indep. of $X_1, \ldots, X_p$ i.i.d. $Exp(1)$. $Y_i = X_i - X_0$
$Laplace(0, 1)$ (type II)	$X_0 \sim N(0, 1)$ indep. of $X_{i,1}, X_{i,2}, X_{i,3}$, $i = 1, \ldots, p$ i.i.d. $N(0, 1)$. $Y_i = X_0 \cdot X_{i,1} + X_{i,2} \cdot X_{i,3}$
$Beta(a, b)$	$X_0 \sim \Gamma(b, 1)$ indep. of $X_1, \ldots, X_p$ i.i.d. $\Gamma(a, 1)$. $Y_i = X_i/(X_i + X_0)$
χ_2^2	$X_0 \sim \Gamma(0.5, 0.5)$ indep. of $X_1, \ldots, X_p$ i.i.d. $\Gamma(0.5, 0.5)$. $Y_i = X_0 + X_i$
χ_8^2	$X_0 \sim \Gamma(2, 0.5)$ indep. of $X_1, \ldots, X_p$ i.i.d. $\Gamma(2, 0.5)$. $Y_i = X_0 + X_i$

Table 3 Purely multivariate alternatives. Here Σ_r is a covariance matrix with unit variances and correlations r

Distribution	Description
$N(\mathbf{0}, \Sigma_{\mathbf{0}})$	Multivariate normal, symmetric
$N(\mathbf{0}, \Sigma_{\mathbf{0.9}})$	Multivariate normal, symmetric
$t(2)$	Multivariate t distribution, symmetric
$AL(\mathbf{0}, \Sigma_{\mathbf{0}})$	Symmetric multivariate Laplace
$AL(\mathbf{3}, \Sigma_{\mathbf{0}})$	Asymmetric multivariate Laplace
$AL(\mathbf{1}, \Sigma_{\mathbf{0.5}})$	Asymmetric multivariate Laplace
$\frac{9}{10}N(\mathbf{0}, \Sigma_{\mathbf{0}}) + \frac{1}{10}N(\mathbf{1}, \Sigma_{\mathbf{0}})$	Location polluted normal mixture
$\frac{9}{10}N(\mathbf{0}, \Sigma_{\mathbf{0}}) + \frac{1}{10}N(\mathbf{2}, \Sigma_{\mathbf{0}})$	Location polluted normal mixture
$\frac{9}{10}N(\mathbf{0}, \Sigma_{\mathbf{0}}) + \frac{1}{10}N(\mathbf{1}, \Sigma_{\mathbf{0.5}})$	Scale and rotation polluted normal mixture
$\frac{9}{10}N(\mathbf{0}, \Sigma_{\mathbf{0}}) + \frac{1}{10}N(\mathbf{2}, \Sigma_{\mathbf{0.5}})$	Scale and rotation polluted normal mixture
$\frac{3}{4}N(\mathbf{0}, \Sigma_{\mathbf{0}}) + \frac{1}{4}N(\mathbf{1}, \Sigma_{\mathbf{0}})$	Heavily location polluted normal mixture
$\frac{3}{4}N(\mathbf{0}, \Sigma_{\mathbf{0}}) + \frac{1}{4}N(\mathbf{2}, \Sigma_{\mathbf{0}})$	Heavily location polluted normal mixture
$\frac{3}{4}N(\mathbf{0}, \Sigma_{\mathbf{0}}) + \frac{1}{4}N(\mathbf{1}, \Sigma_{\mathbf{0.5}})$	Heavily scale and rotation polluted normal mixture
$\frac{3}{4}N(\mathbf{0}, \Sigma_{\mathbf{0}}) + \frac{1}{4}N(\mathbf{2}, \Sigma_{\mathbf{0.5}})$	Heavily scale and rotation polluted normal mixture

In order to see which alternatives that the different tests could be sensitive to, the population values of the statistics were determined for all alternatives. For most distributions the values were computed numerically, to one decimal place for Mardia's statistics and to two decimal places for the $Z_{2,p}$ and $Z_{3,p}$ tests. The population values are given in Table 3 in Supplement S1.

Using R, the nine tests were applied to 1,000,000 samples from each alternative and each combination of n and p. The null distributions for all test statistics were estimated using 100,000 standard normal samples.

7.2 Results for symmetric alternatives

The results for alternatives with symmetric marginal distributions are presented in Table 4 in the "Appendix". Mardia's kurtosis test $b_{2,p}$ had the best overall performance against symmetric alternatives with long-tailed marginal distributions, with the Mardia-Kent T test as runner-up. The $Z_{3,p}$ tests had by far the best performance against symmetric alternatives with short-tailed marginal distributions, but performed poorly against heavy-tailed alternatives. It should therefore be regarded as being directed against short-tailed alternatives.

$b_{2,p}$ and the $Z_{3,p}$ tests were somewhat unexpectedly outperformed by the $b_{1,p}$ test and the $Z_{2,p}$ tests for the $Laplace(0, 1)$ (type I) distribution. This was likely caused by the fact that a distribution with that particular dependence structure (described in Table 2), while having symmetric marginal distributions, is not symmetric in a multivariate sense, as can be seen from the contour plot in Supplement S2 or in Table 3 in Supplement S1.

Finally, we investigated the size of the tests by computing their power against two normal distributions. All tests attained the desired size $\alpha = 0.05$.

Table 4 Power of tests for normality against symmetric alternatives, $\alpha = 0.05$, $p = 3$

Distribution	$b_{1,p}$	$b_{2,p}$	T	$Z_{2,p}^{(HL)}$	$Z_{2,p}^{(W)}$	$Z_{2,p}^{(PB)}$	$Z_{2,p}^{(max)}$	$Z_{3,p}^{(HL)}$	$Z_{3,p}^{(W)}$
Laplace(0, 1) (type I)									
$n = 20$	0.71	0.61	0.53	**0.75**	**0.76**	**0.75**	0.68	0.09	0.08
$n = 50$	**0.99**	0.94	0.86	**0.99**	**1.00**	**0.99**	**0.99**	0.24	0.23
Laplace(0, 1) (type II)									
$n = 20$	0.46	**0.58**	0.49	0.33	0.35	0.35	0.32	0.09	0.08
$n = 50$	0.72	**0.93**	0.83	0.51	0.53	0.54	0.52	0.17	0.16
Beta(1, 1)									
$n = 20$	0.02	0.01	0.02	0.06	0.06	0.06	0.06	0.28	**0.31**
$n = 50$	0.02	0.00	0.00	0.15	0.16	0.16	0.16	**0.96**	**0.97**
Beta(2, 2)									
$n = 20$	0.02	0.01	0.02	0.04	0.05	0.05	0.05	**0.10**	**0.10**
$n = 50$	0.02	0.00	0.01	0.10	0.10	0.09	0.09	**0.40**	**0.40**
t(2)									
$n = 20$	0.83	**0.89**	0.85	0.70	0.75	0.76	0.74	0.34	0.36
$n = 50$	0.98	**1.00**	**0.99**	0.95	0.95	0.95	0.95	0.77	0.77
$AL(\mathbf{0}, \mathbf{\Sigma_0})$									
$n = 20$	0.58	**0.71**	0.60	0.43	0.46	0.46	0.44	0.10	0.08
$n = 50$	0.80	**0.98**	0.90	0.61	0.62	0.63	0.62	0.16	0.14
$N(\mathbf{0}, \mathbf{\Sigma_0})$									
$n = 20$	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
$n = 50$	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
$N(\mathbf{0}, \mathbf{\Sigma_{0.9}})$									
$n = 20$	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
$n = 50$	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05

Bold figures mark tests with high power

7.3 Results for asymmetric alternatives

The results for alternatives with asymmetric marginal distributions are presented in Figs. 1 and 2 and Table 5 in the Appendix.

Mardias skewness test $b_{1,p}$ and the $Z_{2,p}$ tests are all directed to asymmetric alternatives, and outperformed the other tests. However, no directed test was uniformly more powerful than the other directed tests. For $p = 2$, the $Z_{2,p}^{(max)}$ had the best overall performance against asymmetric alternatives, while $b_{1,p}$ and the $Z_{2,p}^{(W)}$ and $Z_{2,p}^{(PB)}$ tests also displayed a good average performance. For $p = 3$ the performance of $Z_{2,p}^{(max)}$ was somewhat worse, whereas $b_{1,p}$, $Z_{2,p}^{(W)}$ and $Z_{2,p}^{(PB)}$ still showed good performance.

How varying n and p affects the power of the tests is investigated in Figs. 1 and 2. In Fig. 1a, we see that against a distribution with *Beta*(1, 2) marginal distributions, the $Z_{3,p}$ tests have the best performance for small n, whereas the $Z_{2,p}$ tests are superior for larger n. In Fig. 1, it is seen that against a distribution with *LogN*(0, 1) marginal
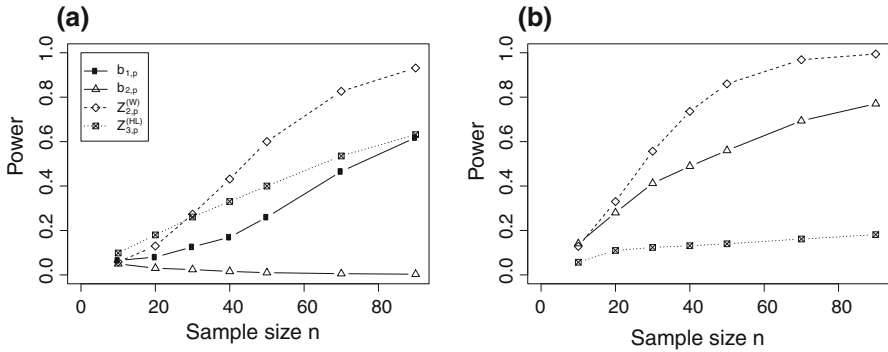
Fig. 1 **a** Power against dependent Beta(1,2) marginals, $p = 2$. **b** Power against dependent LogN(0,1) marginals, $p = 2$
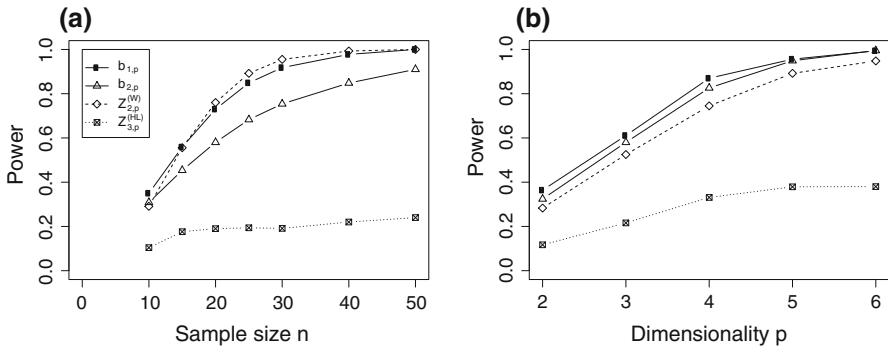

Fig. 2 **a** Power against AsL(3,I), $p = 2$. **b** Power against AsL(3,I), $n = 5p$

distributions, the $Z_{2,p}$ tests have higher power than the $b_{1,p}$ test for small n, while the relation is reversed for larger n.

In Fig. 2a, we see that against the $AL(\mathbf{3}, \Sigma_0)$ distribution, $b_{2,p}$ has slightly higher power than the $Z_{2,p}$ tests for small n, whereas the $Z_{2,p}$ tests have slightly higher power for larger n. In Fig. 2b however, when n/p is fixed and p is increased, the difference in power between the tests remains more or less unchanged.

8 Discussion

Based on the simulation results, our recommendations are that the $Z_{2,p}^{(max)}$ test should be used against asymmetric alternatives when $p = 2$. For higher p, $b_{1,p}$, $Z_{2,p}^{(W)}$ or $Z_{2,p}^{(PB)}$ should be used instead. Mardia's $b_{2,p}$ test should be used against heavy-tailed symmetric alternatives. For short-tailed symmetric alternatives, one of the $Z_{3,p}$ tests would be a better choice.

Most previous power studies for multivariate tests for normality have focused on alternatives with independent marginal distributions. Such distributions are likely to be rare in practice, and as is shown by the two distributions with $Laplace(0, 1)$ marginals

Table 5 Power of tests for normality against asymmetric alternatives, $\alpha = 0.05$, $p = 3$

Distribution	$b_{1,p}$	$b_{2,p}$	T	$Z_{2,p}^{(HL)}$	$Z_{2,p}^{(W)}$	$Z_{2,p}^{(PB)}$	$Z_{2,p}^{(max)}$	$Z_{3,p}^{(HL)}$	$Z_{3,p}^{(W)}$
Indep. $Exp(1)$									
$n = 20$	0.82	0.61	0.62	**0.84**	**0.86**	0.84	0.77	0.29	0.29
$n = 50$	**1.00**	0.93	0.91	**1.00**	**1.00**	**1.00**	**1.00**	0.59	0.59
$LogN(0, 2)$									
$n = 20$	**0.97**	0.89	0.88	**0.98**	**0.98**	**0.98**	0.96	0.61	0.63
$n = 50$	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.95	0.95
$LogN(0, 1)$									
$n = 20$	**0.41**	0.33	0.36	0.28	0.30	0.30	0.29	0.13	0.12
$n = 50$	**0.89**	0.69	0.67	0.87	**0.88**	**0.88**	0.84	0.22	0.21
$Beta(1, 2)$									
$n = 20$	0.08	0.05	0.06	0.09	0.10	0.10	0.11	**0.16**	**0.16**
$n = 50$	0.21	0.02	0.03	0.45	0.49	**0.52**	**0.56**	0.45	0.46
χ_2^2									
$n = 20$	**0.89**	0.79	0.76	0.88	**0.90**	**0.90**	0.86	0.35	0.34
$n = 50$	**1.00**	**0.99**	0.97	**1.00**	**1.00**	**1.00**	**1.00**	0.71	0.71
χ_8^2									
$n = 20$	**0.39**	0.31	0.32	0.31	0.33	0.32	0.29	0.10	0.09
$n = 50$	**0.87**	0.62	0.59	0.87	**0.88**	0.87	0.82	0.18	0.18
$AL(3, \Sigma_0)$									
$n = 20$	**0.78**	0.71	0.70	0.64	0.72	0.74	0.73	0.28	0.27
$n = 50$	**1.00**	0.98	0.96	**1.00**	**1.00**	**1.00**	**1.00**	0.39	0.39
$AL(1, \Sigma_{0.5})$									
$n = 20$	**0.72**	**0.71**	0.66	0.55	0.61	0.62	0.61	0.20	0.19
$n = 50$	0.98	0.98	0.94	0.97	**0.98**	**0.98**	**0.98**	0.31	0.30
$\frac{9}{10}N(0, \Sigma_0) + \frac{1}{10}N(1, \Sigma_0)$									
$n = 20$	0.06	0.06	0.06	0.05	0.05	0.05	0.05	0.05	0.05
$n = 50$	0.07	0.07	0.07	0.06	0.07	0.07	**0.07**	0.05	0.05
$\frac{9}{10}N(0, \Sigma_0) + \frac{1}{10}N(2, \Sigma_0)$									
$n = 20$	**0.15**	0.13	0.13	0.11	0.12	0.12	0.12	0.06	0.06
$n = 50$	0.39	0.21	0.19	0.27	0.32	0.36	**0.42**	0.07	0.07
$\frac{9}{10}N(0, \Sigma_0) + \frac{1}{10}N(1, \Sigma_{0.5})$									
$n = 20$	**0.16**	0.11	0.11	0.14	0.14	0.14	0.14	0.05	0.05
$n = 50$	**0.15**	0.12	0.12	0.11	0.12	0.12	0.12	0.07	0.07
$\frac{9}{10}N(0, \Sigma_0) + \frac{1}{10}N(2, \Sigma_{0.5})$									
$n = 20$	**0.74**	0.32	0.27	0.68	0.69	0.70	**0.71**	0.14	0.13
$n = 50$	**0.54**	0.25	0.25	0.40	0.44	0.48	**0.51**	0.14	0.14
$\frac{3}{4}N(0, \Sigma_0) + \frac{1}{4}N(1, \Sigma_0)$									
$n = 20$	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
$n = 50$	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.05	0.05

Table 5 continued

Distribution	$b_{1,p}$	$b_{2,p}$	T	$Z_{2,p}^{(HL)}$	$Z_{2,p}^{(W)}$	$Z_{2,p}^{(PB)}$	$Z_{2,p}^{(max)}$	$Z_{3,p}^{(HL)}$	$Z_{3,p}^{(W)}$
$\frac{3}{4}N(\mathbf{0}, \mathbf{\Sigma_0}) + \frac{1}{4}N(\mathbf{2}, \mathbf{\Sigma_0})$									
$n = 20$	0.07	0.05	0.05	0.09	0.09	0.09	0.09	0.06	0.06
$n = 50$	0.12	0.04	0.04	0.21	0.24	0.27	**0.32**	0.06	0.06
$\frac{3}{4}N(\mathbf{0}, \mathbf{\Sigma_0}) + \frac{1}{4}N(\mathbf{1}, \mathbf{\Sigma_{0.5}})$									
$n = 20$	0.08	0.07	0.07	0.09	0.08	0.08	0.08	0.05	0.05
$n = 50$	0.18	0.10	0.09	**0.19**	**0.19**	**0.20**	0.17	0.06	0.06
$\frac{3}{4}N(\mathbf{0}, \mathbf{\Sigma_0}) + \frac{1}{4}N(\mathbf{2}, \mathbf{\Sigma_{0.5}})$									
$n = 20$	0.11	0.07	0.07	**0.15**	**0.15**	**0.15**	0.13	0.06	0.06
$n = 50$	0.35	0.06	0.06	0.51	0.54	**0.57**	**0.56**	0.08	0.08

Bold figures mark tests with high power

Table 6 Critical values of the new tests

	$Z_{2,p}^{(HL)}$	$Z_{2,p}^{(W)}$	$Z_{2,p}^{(PB)}$	$Z_{2,p}^{(max)}$	$Z_{3,p}^{(HL)}$	$Z_{3,p}^{(W)}$
$p = 2, n = 20$						
$\alpha = 0.05$	0.85	0.31	1.75	0.58	1.25	0.13
$\alpha = 0.01$	1.06	0.21	2.65	0.67	1.40	0.08
$p = 2, n = 50$						
$\alpha = 0.05$	0.43	0.61	0.58	0.31	0.80	0.33
$\alpha = 0.01$	0.55	0.51	0.83	0.39	0.98	0.24
$p = 3, n = 20$						
$\alpha = 0.05$	1.62	0.07	4.98	0.78	2.30	0.01
$\alpha = 0.01$	1.83	0.04	6.84	0.82	2.40	0.01
$p = 3, n = 50$						
$\alpha = 0.05$	0.83	0.36	1.29	0.45	1.67	0.07
$\alpha = 0.01$	0.98	0.29	1.67	0.52	1.84	0.05

used in our study, multivariate dependence structures can greatly affect the power of tests for normality.

To complicate matters further, some of the results in the tables highlight the fact that what holds true for one combination of p and n can be false for a different combination. For instance, when $p = 2$, $Z_{2,p}^{(max)}$ had higher power than $b_{1,p}$ for the $AL(\mathbf{1}, \mathbf{\Sigma_0})$ and the multivariate χ_8^2 alternatives, but when $p = 3$, $Z_{2,p}^{(max)}$ had lower power than $b_{1,p}$. This phenomenon merits further investigation, as it implies that power studies performed for low values of p can be misleading when choosing between tests to use for higher-dimensional data. Further examples of this phenomenon are given in Figs. 1 and 2.

In recent years, several authors have studied robust testing for normality, i.e. normality test designed to be robust against outliers. See Stehlík et al. (2012) and Cerioli

et al. (2013) for examples. Stehlík et al. (2014) proposed a robustified version of the univariate $Z_{2,1}$ test. A robustified version of the multivariate $Z_{2,p}$ test will appear in a future paper by the author.

Looking at the normal mixtures, which can be viewed as contaminated normal distributions, we see that $Z_{2,p}^{(max)}$ and $b_{1,p}$ were on a par for the mildy polluted mixtures (with a 9:1 mixing ratio) and that $Z_{2,p}^{(max)}$ in general had higher power for the heavily polluted mixtures (with a 3:1 mixing ratio). This suggests the use of the $Z_{2,p}^{(max)}$ statistic for a test for outliers, an idea that perhaps could be investigated further.

Implementations of the $Z_{2,p}$ and $Z_{3,p}$ in R are available from the author. Some critical values for the new tests are given in Table 6.

Appendix: proofs and tables

For the proof of Theorems 3 and Theorems 4 we need some basic properties of the Kronecker product $\otimes$ and vech and vec operators from Henderson and Searle (1979). See also Kollo and von Rosen (2005) and Kollo (2002) for more on these tools from matrix algebra.

For a $p \times q$ matrix $A = \{a_{ij}\}$ and an $r \times s$ matrix B, the Kronecker product $A \otimes B$ is the $pr \times qs$ matrix $\{a_{ij}B\}$, $i = 1, \ldots, p$, $j = 1, \ldots, q$. The vec operator stacks the columns of a matrix underneath eachother, forming a single vector. If the columns of the $p \times q$ matrix A are denoted $a_1, \ldots, a_q$ then $\text{vec}(A) = (a_1', \ldots, a_q')'$ is a vector of length pq.

We will use that

$$(A \otimes B)(C \otimes D) = AC \otimes BD, \qquad (A \otimes B)' = A' \otimes B'$$

and that if A is a $p \times p$ matrix and B a $q \times q$ matrix,

$$\det(A \otimes B) = \det(A)^q \det(A)^p.$$

The vech operator works as the vec operator, except that it only contains each distinct element of the matrix once. For a symmetric matrix A, $\text{vech}(A)$ thus contains only the diagonal and the elements above the diagonal, whereas $\text{vec}(A)$ contains the diagonal elements and the off-diagonal elements twice.

We have the following relationship between the vec operator and the Kronecker product:

$$\text{vec}(ABC) = (C' \otimes A)\text{vec}(B).$$

Furthermore, for a given symmetric $p \times p$ matrix A there exists a $p(p+1)/2 \times p^2$ matrix H and a $p^2 \times p(p+1)/2$ matrix G such that

$$\text{vech}(A) = H\text{vec}(A) \qquad \text{and} \qquad \text{vec}(A) = G\text{vech}(A).$$

As a preparation for the proof of Theorem 3, we prove the following auxiliary lemma.

Lemma 1 *Assume that $X, X_1, \ldots, X_n$ are i.i.d. p-variate random variables fulfilling the conditions of Theorem 1. Let $S_{ij} = (n-1)^{-1} \sum_{k=1}^{n} (X_{k,i} - \bar{X}_i)(X_{k,j} - \bar{X}_j)$ be the elements of the sample covariance matrix S.*

$$u_X = (S_{11}, S_{12}, \ldots, S_{1p}, S_{22}, S_{23}, \ldots, S_{2p}, S_{33}, \ldots, S_{p-1,p}, S_{pp})' = \text{vech}(S)$$

is a vector with $q = p(p+1)/2$ distinct elements. Denote its covariance matrix $\text{Cov}(u_X) = \Lambda_{22}$.

Let A be a nonsingular $p \times p$ matrix and let b be a p-dimensional vector. Then there exists a nonsingular $q \times q$ matrix D such that

(i) *the sample variances and covariances of $Y = AX + b$ are given by $u_Y = Du_X$,*
(ii) *$\text{Cov}(u_Y) = D\Lambda_{22}D'$ and*
(iii) *$\det(D) = \det(A)^{p+1}$,*

Proof The transformed sample $AX + b$ has sample covariance matrix ASA', so we wish to study $\text{vech}(ASA')$. We have

$$\text{vec}(ASA') = (A \otimes A)\text{vec}(S).$$

Moreover, since S is symmetric there exist nonsingular matrices G and H such that

$$\text{vec}(S) = G\text{vech}(S) \qquad \text{and} \qquad \text{vech}(S) = H\text{vec}(S).$$

Thus

$$u_Y = \text{vech}(ASA') = H(A \otimes A)G\text{vech}(S) =: Du_X,$$

which establishes the existence of D. From Section 4.2 of Henderson and Searle (1979) we have

$$\det(D) = \det(H(A \otimes A)G) = \det(A)^{p+1}$$

which is nonzero, since A is nonsingular. D is hence also nonsingular. In conclusion, we have established the existence and nonsingularity of D as well as (i) and (iii). Finally, (ii) follows immediately from (i). □

We now have the tools necessary to tackle Theorem 3.

Proof of Theorem 3 (i) From Theorem 10.2.4 in Mardia et al. (1979) we have that the canonical correlations between the random vectors $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are invariant under the nonsingular linear transformations $\boldsymbol{A}\boldsymbol{Y} + \boldsymbol{b}$ and $\boldsymbol{C}\boldsymbol{Z} + \boldsymbol{d}$. Clearly all five statistics are invariant under changes in location, since $\boldsymbol{S_{11}}$, $\boldsymbol{S_{22}}$, $\boldsymbol{S_{12}}$ and $\boldsymbol{S_{21}}$ all share that invariance property. It therefore suffices to show that the nonsingular linear transformation $\boldsymbol{A}\boldsymbol{X}$ induces nonsingular linear transformations $\boldsymbol{C}\bar{\boldsymbol{X}}$ and $\boldsymbol{D}\boldsymbol{u}$. $\boldsymbol{C} = \boldsymbol{A}$ is immediate and the existence of $\boldsymbol{D}$ is given by Lemma 1.

(ii) By part (ii) of Theorem 1, $\mu_{ijk} = 0$ for all i, j, k implies that $\boldsymbol{\Lambda}_{12} = \boldsymbol{0}$. But then $\boldsymbol{\Lambda_{11}}^{-1}\boldsymbol{\Lambda_{12}}\boldsymbol{\Lambda_{22}}^{-1}\boldsymbol{\Lambda_{21}} = \boldsymbol{0}$ and all canonical correlations are 0. If $\mu_{ijk} \neq 0$ then $\rho(\bar{X}_i, S_{jk}) \neq 0$. Thus the linear combinations $\boldsymbol{a}'\bar{\boldsymbol{X}} = \bar{X}_i$ and $\boldsymbol{b}'\boldsymbol{u} = S_{jk}$ have nonzero correlation. λ_1 must therefore be greater than 0.

(iii) Follows from the fact that the statistics are continuous function of sample moments that converge almost surely. □

The proofs of parts (ii) and (iii) of Theorem 4 are analog to the previous proof. The proof for part (i) is however slightly different as we omit to explicitly give a matrix that gives a nonsingular linear transformation of $\boldsymbol{v}_X$.

Proof of Theorem 4 (i) Let the third order central moment of a multivariate random variable $\boldsymbol{Z}$ be

$$\bar{m}_3(\boldsymbol{Z}) = \mathrm{E}\left[(\boldsymbol{Z} - \mathrm{E}\boldsymbol{Z}) \otimes (\boldsymbol{Z} - \mathrm{E}\boldsymbol{Z})' \otimes (\boldsymbol{Z} - \mathrm{E}\boldsymbol{Z})\right]'$$
$$= \mathrm{E}\left[(\boldsymbol{Z} - \mathrm{E}\boldsymbol{Z})\left((\boldsymbol{Z} - \mathrm{E}\boldsymbol{Z}) \otimes (\boldsymbol{Z} - \mathrm{E}\boldsymbol{Z})\right)'\right].$$

Given a sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_p$, let $S_{ijk} = \frac{n}{(n-1)(n-2)} \sum_{r=1}^{n}(X_{r,i}-\bar{X}_i)(X_{r,j}-\bar{X}_j)(X_{r,k}-\bar{X}_k)$. When the distribution of $\boldsymbol{Z}$ is the empirical distribution of said sample,

$$\boldsymbol{v}_X = (S_{111}, S_{112}, \ldots, S_{pp(p-1)}, S_{ppp})' = \frac{n^2}{(n-1)(n-2)}\mathrm{vech}\left(\bar{m}_3(\boldsymbol{Z})\right).$$

Similarly $\mathrm{vec}\left(\bar{m}_3(\boldsymbol{Z})\right)$ stacks the elements of $\bar{m}_3(\boldsymbol{Z})$ in a vector that simply is $\mathrm{vech}\left(\bar{m}_3(\boldsymbol{Z})\right)$ with a few repetitions:

$$\boldsymbol{w}_X = (S_{111}, S_{112}, \ldots, S_{112} \ldots, S_{pp(p-1)}, S_{ppp})' = \frac{n^2}{(n-1)(n-2)}\mathrm{vec}\left(\bar{m}_3(\boldsymbol{Z})\right).$$

Thus, for each linear combination $\boldsymbol{a}'\boldsymbol{w}_X$ there exists a $\boldsymbol{b}$ so that $\boldsymbol{b}'\boldsymbol{v}_X = \boldsymbol{a}'\boldsymbol{w}_X$ and therefore, by the definition of canonical correlations, the (sample) canonical correlations between $\bar{\boldsymbol{X}}$ and $\boldsymbol{v}_X$ are the same as those between $\bar{\boldsymbol{X}}$ and $\boldsymbol{w}_X$.

Writing $\boldsymbol{Y} = \boldsymbol{Z} - \mathrm{E}\boldsymbol{Z}$, we have $\bar{m}_3(\boldsymbol{Z}) = \mathrm{E}\left(\boldsymbol{Y}(\boldsymbol{Y} \otimes \boldsymbol{Y})'\right)$ and

$$\bar{m}_3(\boldsymbol{A}\boldsymbol{Z}) = \mathrm{E}\left(\boldsymbol{A}\boldsymbol{Y}(\boldsymbol{A}\boldsymbol{Y} \otimes \boldsymbol{A}\boldsymbol{Y})'\right) = \mathrm{E}\left(\boldsymbol{A}\boldsymbol{Y}(\boldsymbol{Y} \otimes \boldsymbol{Y})'(\boldsymbol{A} \otimes \boldsymbol{A})'\right)$$
$$= \boldsymbol{A}\bar{m}_3(\boldsymbol{Z})(\boldsymbol{A} \otimes \boldsymbol{A})'.$$

Hence

$$\mathrm{vec}\left(\bar{m}_3(\boldsymbol{A}\boldsymbol{Z})\right) = (\boldsymbol{A} \otimes \boldsymbol{A} \otimes \boldsymbol{A})\mathrm{vec}\left(\bar{m}_3(\boldsymbol{Z})\right).$$

Now, $\det(A \otimes A \otimes A) = \det(A \otimes A)^p \det(A)^{p^2} = \det(A)^{3p^2} > 0$, so $E := (A \otimes A \otimes A)$ is a nonsingular matrix such that $\bar{m}_3(AZ) = E\bar{m}_3(Z)$. Since canonical correlations are invariant under nonsingular linear transformations of the two sets of variables, this means that the canonical correlations between $\bar{X}$ and w_X remain unchanged under the transformation $AX + b$. Thus the canonical correlations between $\bar{X}$ and v_Y must also necessarily remain unchanged. This proves the affine invariance of the statistics. $\quad\square$

References

Bartlett MS (1939) A note on tests of significance in multivariate analysis. Math Proc Camb Philos Soc 35:180–185

Cerioli A, Farcomeni A, Riani M (2013) Robust distances for outlier-free goodness-of-fit testing. Comput Stat Data Anal 65:29–45

Doornik JA, Hansen H (2008) An omnibus test for univariate and multivariate normality. Oxf Bull Econ Stat 70:927–939

Dubkov AA, Malakhov AN (1976) Properties and interdependence of the cumulants of a random variable. Radiophys Quantum Electron 19:833–839

Fisher RA (1936) The use of multiple measurements in taxonomic problems. Ann Eugen 7:179–188

Henderson HV, Searle SR (1979) Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics. Can J Stat 7:65–81

Henze N (2002) Invariant tests for multivariate normality: a critical review. Stat Pap 43:467–506

Kankainen A, Taskinen S, Oja H (2007) Tests of multinormality based on location vectors and scatter matrices. Stat Methods Appl 16:357–359

Kaplan EL (1952) Tensor notation and the sampling cumulants of k-statistics. Biometrika 39:319–323

Kollo T (2002) Multivariate skewness and kurtosis measures with an application in ICA. J Multivar Anal 99:2328–2338

Kollo T, von Rosen D (2005) Advanced multivariate statistics with matrices. Springer, Berlin. ISBN 978-1-4020-3418-3

Kotz S, Kozubowski TJ, Podgórski K (2000) An asymmetric multivariate Laplace distribution, Technical Report No. 367, Department of Statistics and Applied Probability, University of California at Santa Barbara

Kshirsagar AM (1972) Multivariate analysis. Marcel Dekker, ISBN 0-8247-1386-9

Lin C-C, Mudholkar GS (1980) A simple test for normality against asymmetric alternatives. Biometrika 67:455–61

Mardia KV (1970) Measures of multivariate skewness and kurtosis with applications. Biometrika 57:519–530

Mardia KV (1974) Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. Sankhya Indian J Stat 36:115–128

Mardia KV, Kent JT (1991) Rao score tests for goodness of fit and independence. Biometrika 78:355–363

Mardia KV, Kent JT, Bibby JM (1979) Multivariate analysis. Academic Press, ISBN 0-12-471250-9

McCullagh P (1987) Tensor methods in statistics. University Press, ISBN 0-412-27480-9

Mecklin CJ, Mundfrom DJ (2004) An appraisal and bibliography of tests for multivariate normality. Int Stat Rev 72:123–128

Mecklin CJ, Mundfrom DJ (2005) A Monte Carlo comparison of the type I and type II error rates of tests of multivariate normality. J Stat Comput Simul 75:93–107

Mudholkar GS, Marchetti CE, Lin CT (2002) Independence characterizations and testing normality against restricted skewness–kurtosis alternatives. J Stat Plan Inference 104:485–501

Stehlík M, Fabián Z, Střelec L (2012) Small sample robust testing for normality against Pareto tails. Commun Stat Simul Comput 41:1167–1194

Stehlík M, Střelec L, Thulin M (2014) On robust testing for normality in chemometrics. Chemom Intell Lab Syst 130:98–109

Thulin M (2010) On two simple tests for normality with high power. Pre-print, arXiv:1008.5319