Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT®
Journal of Multivariate Analysis 93 (2005) 58–80

Journal of Multivariate Analysis

http://www.elsevier.com/locate/jmva

A new test for multivariate normality

Gábor J. Székely[a,*,1] and Maria L. Rizzo[b]

[a] Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH, 43403-0221, USA

[b] Department of Mathematics, Ohio University, Athens, OH 45701, USA

Received 14 June 2002

Abstract

We propose a new class of rotation invariant and consistent goodness-of-fit tests for multivariate distributions based on Euclidean distance between sample elements. The proposed test applies to any multivariate distribution with finite second moments. In this article we apply the new method for testing multivariate normality when parameters are estimated. The resulting test is affine invariant and consistent against all fixed alternatives. A comparative Monte Carlo study suggests that our test is a powerful competitor to existing tests, and is very sensitive against heavy tailed alternatives.

AMS 2000 subject classifications: primary 62G10; secondary 62H15

Keywords: Goodness-of-fit; Strictly negative definite; BHEP test; Henze–Zirkler test; Multivariate skewness; Multivariate kurtosis; Projection pursuit

1. Introduction

We propose a new class of consistent tests for comparing multivariate distributions based on Euclidean distance between sample elements. Applications include one-sample goodness-of-fit tests for discrete or continuous multivariate distributions in arbitrary dimension $d \geqslant 1$. The new tests can be applied to assess distributional assumptions for many classical procedures in multivariate analysis. In this article we present the general form of the multivariate goodness-of-fit statistic,

*Corresponding author. Fax: +419-372-6092.

E-mail addresses: gabors@bgnet.bgsu.edu (G.J. Székely), rizzo@math.ohiou.edu (M.L. Rizzo).

[1] This work was partially supported by NSA Grant MDA904-02-1-0091.

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2003.12.002

and show that the corresponding tests are consistent against all fixed alternatives. We implement the proposed test of multivariate normality and prove consistency under the composite hypothesis when population parameters are estimated from the sample. A comparative Monte Carlo power study is presented to assess the empirical power performance of the new test of multivariate normality.

Recent new approaches to testing multivariate normality have been proposed by Baringhaus and Henze [2], Bowman and Foster [5], Henze and Wagner [11], Henze and Zirkler [12], Romeu and Ozturk [19]. The most widely applied tests of multivariate normality are based on Mardia's multivariate generalization of skewness and kurtosis [16]. Malkovich and Afifi [15] proposed tests of multivariate normality based on one dimensional skewness and kurtosis statistics. Projection pursuit approaches to testing multinormality have applied the same idea, and also apply this idea with Cramér–von Mises and Kolmogorov–Smirnov goodness-of-fit statistics, and sample entropy [6,28,29].

The family of d-variate normal distributions is closed with respect to all nonsingular linear transformations, called affine transformations, so it is desirable that a goodness-of-fit test of multinormality is invariant under such transformations. A test statistic T_n defined on a d-dimensional sample $X_1, \dots, X_n$ is *affine invariant* if $T_n(\mathscr{A}(X_1), \dots, \mathscr{A}(X_n)) = T_n(X_1, \dots, X_n)$ for every affine transformation $\mathscr{A}$ of R^d. A goodness-of-fit test of $H_0 : F \in \mathscr{F}$ vs. $H_1 : F \notin \mathscr{F}$ is *consistent* against all fixed alternatives if the probability of rejecting the null hypothesis approaches one as sample size tends to infinity, whenever the distribution of the sampled population is not a member of $\mathscr{F}$.

Mardia [16] proposed multivariate generalizations of skewness and kurtosis measures. The class of elliptically symmetric distributions, including d-variate normal, have population skewness zero. Baringhaus and Henze [3] obtained the nonnull limiting distribution of Mardia's skewness statistic and showed that Mardia's skewness test is consistent if and only if the sampled population is not elliptically symmetric. The kurtosis of a d-variate normal distribution as defined by Mardia is $d(d+2)$. Henze [10] showed that Mardia's kurtosis test of d-variate normality is consistent if and only if the kurtosis of the sampled population differs from $d(d+2)$.

Affine invariance and consistency are desirable properties for a test of multivariate normality, and certainly it is desirable that such a test is applicable for arbitrary dimension and sample size. Although there are several affine invariant tests of multivariate normality in the literature, not many of these tests are known to satisfy all these properties.

The BHEP class of tests [2,7,11], including the Henze–Zirkler test [12], are based on the empirical characteristic function. The BHEP tests of multivariate normality are affine invariant and consistent against all fixed alternatives.

Our proposed test of multivariate normality has all the desirable properties mentioned above, and it is practical to apply. Results of a Monte Carlo power study suggest that the new test we propose is a powerful competitor to other existing affine invariant tests, including Mardia's multivariate skewness and kurtosis tests, and the BHEP class of tests.

Section 2 presents the theoretical foundation of the test, and the corresponding goodness-of-fit statistics are developed in Section 3. Consistency is proved in Section 4. Section 5 presents a Monte Carlo comparative power study. Asymptotic behavior of the test is discussed in Section 6, followed by a Summary in Section 7.

2. The basic inequality

Suppose $X_1, \dots, X_n$ is a d-dimensional random sample, $d \geqslant 1$. Let X' denote an independent copy of the random variable X. Consider the V-statistic with kernel

$$h(x, y) = E\|x - Y\| + E\|y - Y\| - E\|Y - Y'\| - \|x - y\|, \tag{1}$$

where $\|x\| = (x^T x)^{1/2}$ is the Euclidean norm of x, and Y is a d-dimensional random vector. Define

$$\mathscr{E}_n = \frac{1}{n} \sum_{j,k=1}^{n} h(X_j, X_k) = n \frac{1}{n^2} \sum_{j,k=1}^{n} h(X_j, X_k).$$

Then $\mathscr{E}_n/n$ is a V-statistic with kernel $h(x, y)$. Since $E[h(x, X)] = 0$ for every x, $\mathscr{E}_n/n$ is a degenerate kernel V-statistic. The existence of the limiting distribution of $n(\mathscr{E}_n/n) = \mathscr{E}_n$ follows from U-statistics limit theorems of Hoeffding [13], provided $E[h^2(X, X')] < \infty$ and $X \stackrel{d}{=} Y$, where $\stackrel{d}{=}$ means X and Y are identically distributed (see Section 6).

We will show that if X, X', Y, Y' are independent, $E\|X\|$ and $E\|Y\|$ are finite, then

$$E[h(X, Y)] = 2E\|X - Y\| - E\|X - X'\| - E\|Y - Y'\| \geqslant 0 \tag{2}$$

with equality if and only if X and Y are identically distributed. An elementary proof of inequality (2) follows from the property of strict negative definiteness, defined below, of Euclidean distance.

Let S be an arbitrary nonempty set. A symmetric, real valued function $\gamma(x, y)$ defined on $S \times S$ is negative definite if for every positive integer n,

$$\sum_{j,k=1}^{n} \gamma(x_j, y_k) r_j r_k \leqslant 0 \tag{3}$$

for every set $\{(x_j, x_k) \in S \times S; j, k = 1, \dots, n\}$, and every n-tuple of real numbers $r_1, \dots, r_n$ such that $\sum_{j=1}^{n} r_j = 0$. The function $\gamma(x, y)$ is strictly negative definite if the above inequalities (3) are strict whenever $x_1, \dots, x_n$ are distinct and at least one of the $r_1, \dots, r_n$ does not vanish.

In case S is a metric space and γ is continuous on $S \times S$ we have the following equivalent definition. The function $\gamma(x, y)$ is strictly negative definite if

$$\int_S \int_S \gamma(x, y) r(x) r(y)\, dQ(x)\, dQ(y) \leqslant 0$$

whenever $\int_S r(x)\,dQ(x) = 0$ for some probability measure Q, and equality holds if and only if $r(x) = 0$ a.s. Q.

Proposition 1 (Strict negative definiteness of $||x - y||$). *Let $\gamma(x, y) : R^d \times R^d \to R$ be defined as the Euclidean distance between vectors x and y. Then $\gamma(x, y) = ||x - y||$ is strictly negative definite.*

Proposition 1 is proved in the appendix.

Theorem 1. *Let S be an arbitrary metric space, and let $\gamma(x, y)$ be a continuous, symmetric, real-valued function on $S \times S$. Suppose X, X', Y, and Y' are independent S-valued random variables, X and X' are identically distributed, and Y and Y' are identically distributed. Suppose $\gamma(X, X')$, $\gamma(Y, Y')$, and $\gamma(X, Y)$ have finite expected values. Then*

$$2E\gamma(X, Y) - E\gamma(X, X') - E\gamma(Y, Y') \geqslant 0 \tag{4}$$

if and only if γ is negative definite. If γ is strictly negative definite then equality holds in (4) *if and only if X and Y are identically distributed.*

Proof. Let μ and ν denote the distributions of X and Y respectively, let Q be an arbitrary probability measure such that Q dominates both μ and ν, and define

$$r(x) = \frac{d\mu(x)}{dQ} - \frac{d\nu(x)}{dQ}.$$

Then $\int_S r(x)\,dQ(x) = 0$, and by negative definiteness of $\gamma(x, y)$ we have

$$\begin{aligned}
&2E\gamma(X, Y) - E\gamma(X, X') - E\gamma(Y, Y') \\
&\quad = -\int_S\int_S \gamma(x, y)\,d\mu(x)\,d\mu(y) + \int_S\int_S \gamma(x, y)\,d\mu(x)\,d\nu(y) \\
&\qquad + \int_S\int_S \gamma(x, y)\,d\mu(y)\,d\nu(x) - \int_S\int_S \gamma(x, y)\,d\nu(x)\,d\nu(y) \\
&\quad = -\int_S\int_S \gamma(x, y) r(x) r(y)\,dQ(x)\,dQ(y) \geqslant 0.
\end{aligned}$$

If $\gamma(x, y)$ is strictly negative definite, equality holds if and only if $r(x) = 0$ a.s. Q; that is, if and only if X and Y are identically distributed. □

Corollary 1. *Suppose X, X', Y, and Y' are independent random vectors in R^d. If X and X' are identically distributed, Y and Y' are identically distributed, $E||X||$ and $E||Y||$ are finite, then*

$$2E||X - Y|| - E||X - X'|| - E||Y - Y'|| \geqslant 0, \tag{5}$$

and equality holds if and only if X and Y are identically distributed.

The following special case of Corollary 1 is proved by Morgenstern [17, p. 347].

For equal numbers of black and white points in euclidean space the sum of the pairwise distances between points of equal color is less than or equal to the sum of the pairwise distances between points of different color. Equality holds only in the case when black and white points coincide.

The special case $Y = -X$ (in distribution) of Corollary 1 is proved in [23, p. 458–459]. This special case was also associated with testing for diagonal symmetry in [24,25].

3. Goodness-of-fit tests

Suppose $X_1, \ldots, X_n$ is a random sample from a d-variate population with distribution F, and $x_1, \ldots, x_n$ are the observed values of the random sample. The proposed statistic for testing H_0: $F = F_0$ vs. H_1: $F \neq F_0$ is

$$\mathcal{E}_{n,d} = n\left(\frac{2}{n}\sum_{j=1}^{n} E||x_j - X|| - E||X - X'|| - \frac{1}{n^2}\sum_{j,k=1}^{n} ||x_j - x_k||\right),$$

where X and X' are independent and identically distributed (iid) with distribution F_0.

If the hypothesized distribution is d-variate normal with mean vector μ and nonsingular covariance matrix Σ, denoted $N_d(\mu, \Sigma)$, consider the transformed sample $y_j = \Sigma^{-1/2}(x_j - \mu)$, $i = 1, \ldots, n$. The test statistic for d-variate normality is

$$\mathcal{E}_{n,d} = n\left(\frac{2}{n}\sum_{j=1}^{n} E||y_j - Z|| - E||Z - Z'|| - \frac{1}{n^2}\sum_{j,k=1}^{n} ||y_j - y_k||\right), \tag{6}$$

where Z and Z' denote iid $N_d(0, I)$ random variables, and I is the $d \times d$ identity matrix. The first component of the test statistic involves computing $E||a - Z||$ where $a \in R^d$ is fixed. In the univariate case $E||a - Z|| = 2a\Phi(a) + 2\phi(a) - a$, where $\Phi(x)$ and $\phi(x)$ are the $N(0,1)$ cumulative distribution and density functions. This mean can also be expressed a series. Substituting the Taylor expansion of $\Phi(x)$,

$$\begin{aligned}
E|a - Z| - E|Z| &= -a + 2\int_0^a \Phi(x)\,dx \\
&= -a + 2\int_0^a \left[\frac{1}{2} + \frac{1}{\sqrt{2\pi}}\sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{2^k k!(2k+1)}\right] dx \\
&= \sqrt{\frac{2}{\pi}}\sum_{k=0}^{\infty} \frac{(-1)^k |a|^{2k+2}}{2^k k!(2k+1)(2k+2)}.
\end{aligned}$$

For $d \geqslant 1$ the following formula holds:

$$E||a-Z|| - E||Z|| = \sqrt{\frac{2}{\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!2^k} \frac{||a||^{2k+2}}{(2k+1)(2k+2)} \frac{\Gamma(\frac{d+1}{2})\Gamma(k+\frac{3}{2})}{\Gamma(k+\frac{d}{2}+1)}.$$

Then since

$$E||Z-Z'|| = \sqrt{2}E||Z|| = 2\frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})}, \tag{7}$$

the computing formula for the d-variate normality test statistic is given by

$$\mathscr{E}_{n,d} = n\left(\frac{2}{n}\sum_{j=1}^{n} E||y_j - Z|| - 2\frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} - \frac{1}{n^2}\sum_{j,k=1}^{n} ||y_j - y_k||\right), \tag{8}$$

where

$$E||a-Z|| = \frac{\sqrt{2}\,\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} + \sqrt{\frac{2}{\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\,2^k} \frac{||a||^{2k+2}}{(2k+1)(2k+2)} \frac{\Gamma(\frac{d+1}{2})\Gamma(k+\frac{3}{2})}{\Gamma(k+\frac{d}{2}+1)}.$$

A test of the simple hypothesis of d-variate normality, $d \geqslant 1$, rejects the null hypothesis for large values of $\mathscr{E}_{n,d}$. The test is affine invariant, and consistency is an immediate consequence of Theorem 1.

In practice, however, the parameters of the hypothesized normal distribution are usually unknown. If $\mathscr{N}_d(\mu, \Sigma)$ denotes the family of d-variate normal distributions with mean vector μ and covariance matrix $\Sigma > 0$, and F_X is the distribution of a d-dimensional random vector X, the problem is to test if $F_X \in \mathscr{N}_d(\mu, \Sigma)$. In this case, parameters μ and Σ are estimated from the sample mean vector $\bar{X}$ and sample covariance matrix $S = (n-1)^{-1}\sum_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X})^T$, and the standardized sample vectors are $Y_j = S^{-1/2}(X_j - \bar{X})$, $j = 1, \ldots, n$. The joint distribution of $Y_1, \ldots, Y_n$ does not depend on unknown parameters, but $Y_1, \ldots, Y_n$ are dependent. As a first step, we will ignore the dependence of the standardized sample, and use the computing formula given for testing the simple hypothesis. Denote by $\hat{\mathscr{E}}_{n,d}$ the version of the test statistic (6) and computing formula (8) obtained by standardizing the sample with estimated parameters. The test rejects the hypothesis of multivariate normality for large values of $\hat{\mathscr{E}}_{n,d}$. Percentiles of the finite sample null distribution of $\hat{\mathscr{E}}_{n,d}$ can be estimated by simulation (see Table 1). The test based on $\hat{\mathscr{E}}_{n,d}$ is clearly affine invariant, but since the standardized sample is dependent, the hypotheses of Theorem 1 do not hold.

Table 1
Empirical percentiles of $\hat{\mathscr{E}}_{n,d}$

	$n=25$		$n=50$		$n=100$	
d	0.90	0.95	0.90	0.95	0.90	0.95
1	0.686	0.819	0.695	0.832	0.701	0.840
2	0.856	0.944	0.872	0.960	0.879	0.969
3	0.983	1.047	1.002	1.066	1.011	1.077
4	1.098	1.147	1.119	1.167	1.124	1.174
5	1.204	1.241	1.223	1.263	1.234	1.277
10	1.644	1.659	1.671	1.690	1.685	1.705

4. Consistency

In this section a proof of consistency is presented for testing the composite hypothesis of multivariate normality using the test statistic $\hat{\mathscr{E}}_{n,d}$.

Let $c_{\alpha,n,d}$ denote the constant satisfying $P(\hat{\mathscr{E}}_{n,d} \geqslant c_{\alpha,n,d}) = \alpha$. The test is consistent against all fixed alternatives if whenever the sampled population is nonnormal and fixed, $\lim_{n\to\infty} P(\hat{\mathscr{E}}_{n,d} \geqslant c_{\alpha,n,d}) = 1$. To prove consistency it is necessary that the sequence of critical values $c_{\alpha,n,d}$ for any fixed α and d is bounded above by a finite constant $k_{\alpha,d}$ that does not depend on n. The existence of $k_{\alpha,d}$ follows from the existence of a finite upper bound for the expected value of $\hat{\mathscr{E}}_{n,d}$ that does not depend on n. The following inequalities are used to prove that such an upper bound exists.

Proposition 2. *If $Z_1, \ldots, Z_n$ is a univariate standard normal random sample, and $Y_1, \ldots, Y_n$ is the standardized sample, then*

$$E|Y_1 - Y_2| > E|Z_1 - Z_2|. \tag{9}$$

Proposition 3. *Let*

$$E_{\mathscr{S}}\left[\frac{1}{n}\sum_{j=1}^{n} E|Y_j - Z|\right]$$

denote the expected value of $\frac{1}{n}\sum_{j=1}^{n} E|y_j - Z|$, with respect to all standardized univariate normal samples $Y_1, \ldots, Y_n$, where $E|a - Z| = \int_{-\infty}^{\infty} |a - z|\phi(z)\,dz$. Then for $X_1, \ldots, X_n$ and Z iid $N(0,1)$,

$$E_{\mathscr{S}}\left[\frac{1}{n}\sum_{j=1}^{n} E|Y_j - Z|\right] \leqslant E\left[\frac{1}{n}\sum_{j=1}^{n} E|X_j - Z|\right]. \tag{10}$$

Proofs of Propositions 2 and 3 are given in the appendix.

Proposition 4. *Under the hypothesis of normality, for every fixed integer $d \geqslant 1$,*

1. $E[\hat{\mathscr{E}}_{n,d}]$ *is bounded above by a finite constant that depends only on* d.
2. *For every fixed* $\alpha \in (0, 1)$, *the sequence of critical values* $c_{\alpha,n,d}$ *of* $\hat{\mathscr{E}}_{n,d}$ *are bounded above by a finite constant* $k_{\alpha,d}$ *that depends only on* α *and* d.

Proof. Inequalities (9) and (10) imply that under univariate normality

$$\begin{aligned} E[\hat{\mathscr{E}}_{n,1}] &= E_{\mathscr{S}}\left[n\left(\frac{2}{n}\sum_{j=1}^{n} E|Y_j - Z| - E|Z - Z'| - \frac{1}{n^2}\sum_{j,k=1}^{n} |Y_j - Y_k|\right)\right] \\ &< 2nE|X - Z| - nE|Z - Z'| - (n-1)E|Z_1 - Z_2| \\ &= E|Z - Z'| = \frac{2}{\sqrt{\pi}}. \end{aligned}$$

It follows that the sequence of critical values $c_{\alpha,n}$ of $\hat{\mathscr{E}}_{n,1}$ for any fixed α are bounded above by some finite constant k_α. In the multivariate case, the same relationship holds: $E[\hat{\mathscr{E}}_{n,d}]$ is at most d times the corresponding univariate expected value, so $E[\hat{\mathscr{E}}_{n,d}]$ and the critical values of $\hat{\mathscr{E}}_{n,d}$ are bounded. □

Theorem 2 (Consistency). *The test of multivariate normality based on* $\hat{\mathscr{E}}_{n,d}$ *is consistent against all nonnormal fixed alternatives.*

Proof. Suppose X is nonnormal with $E[X] = \mu = (\mu_k),\ |\mu_k| < \infty, k = 1, \ldots, d$, and $\mathrm{Cov}(X) = \Sigma = (\sigma_{jk}),\ |\sigma_{j,k}| < \infty,\ j,k = 1, \ldots, d$. Without loss of generality, assume that $\mu = 0$ and $\Sigma = I$. Let $X_1, \ldots, X_n$ be a random sample from the distribution of X and $Y_j = S^{-1/2}(X_j - \bar{X}),\ j = 1, \ldots, n$. If S is singular, define $Y_j = (X_j - \bar{X})$. Then

$$\begin{aligned} \hat{\mathscr{E}}_{n,d} &= n\left(\frac{2}{n}\sum_{j=1}^{n} E||y_j - Z|| - E||Z - Z'|| - \frac{1}{n^2}\sum_{j,k=1}^{n} ||y_j - y_k||\right) \\ &= \sum_{j=1}^{n} R_j, \end{aligned}$$

where $R_1, \ldots, R_n$ are identically distributed as

$$R_1 = 2E||y_1 - Z|| - E||Z - Z'|| - \frac{1}{n}\sum_{j=1}^{n} ||Y_1 - Y_j||$$

and Z, Z' are iid $N_d(0, I)$. From (A.7) in the appendix

$$\lim_{n\to\infty} E_{\mathscr{S}}\left[\frac{1}{n}\sum_{j=1}^{n} E|Y_j - Z|\right] = E\left[\frac{1}{n}\sum_{j=1}^{n} E|X_j - Z|\right].$$

Hence

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} R_j = 2E\|X - Z\| - E\|Z - Z'\| - E\|X - X'\| \geqslant 0,$$

with equality if and only if $X \stackrel{d}{=} Z$, by Theorem 1. Since X is nonnormal, and $\lim_{n\to\infty} c_{\alpha,n} \leqslant k_{d,\alpha}$, there exists $\delta > 0$ such that

$$\begin{aligned}\lim_{n\to\infty} P(\hat{\mathscr{E}}_{n,d} > c_{\alpha,n,d}) &\geqslant \lim_{n\to\infty} P(\hat{\mathscr{E}}_{n,d} \geqslant k_{\alpha,d}) \\ &= \lim_{n\to\infty} P\left(\sum_{j=1}^{n} R_j \geqslant k_{\alpha,d}\right) \geqslant \lim_{n\to\infty} P(n\delta > k_{\alpha,d}) = 1.\end{aligned}$$

This proves consistency for all fixed nonnormal alternatives X with finite covariance. In the general case, consider the truncated distributions. The details of the truncation argument are given in the appendix. □

5. Empirical results

In order to assess the performance of the new test, we performed a comparative Monte Carlo power study of eight affine invariant tests of multivariate normality. We chose two categories of tests for comparison; tests based on skewness and kurtosis, and the BHEP tests based on the empirical characteristic function. Projection pursuit tests based on five univariate goodness-of-fit tests of normality were also compared.

5.1. Multivariate tests

The multivariate skewness test proposed by Mardia [16] is based on the sample skewness statistic defined

$$b_{1,d} = \frac{1}{n^2} \sum_{j,k=1}^{n} ((X_j - \bar{X})^T \hat{\Sigma}^{-1} (X_k - \bar{X}))^3, \tag{11}$$

where $\hat{\Sigma} = n^{-1} \sum_{j=1}^{n} (X_j - \bar{X})(X_j - \bar{X})^T$ denotes the maximum likelihood estimator of population covariance. Normality is rejected for large values of $b_{1,d}$. Mardia's multivariate kurtosis test is based on the sample kurtosis

$$b_{2,d} = \frac{1}{n} \sum_{j=1}^{n} ((X_j - \bar{X})^T \hat{\Sigma}^{-1} (X_j - \bar{X}))^2. \tag{12}$$

Large values of $|d(d+2) - b_{2,d}|$ are significant.

The BHEP statistics are defined as follows. Assume that the sample covariance matrix is nonsingular and $Y_k = \hat{\Sigma}^{-1/2}(X_k - \bar{X}),\ k = 1, \ldots, n,$ is the standardized sample. The statistic $T_{n,d}(\beta)$ is the weighted integral of the squared difference between the multivariate normal characteristic function and the empirical characteristic function $\Psi_n(t) = \frac{1}{n}\sum_{k=1}^n e^{it^T Y_k}$, where i is the complex unit. The test statistic is defined (see [11])

$$T_{n,d}(\beta) = \int_{R^d} ||\Psi_n(t) - e^{-||t||^2/2}||^2 \varphi_\beta(t)\,dt$$

where $\varphi_\beta(t)$ is a weighting function. When the weighting function is

$$\varphi_\beta(t) = (2\pi\beta^2)^{-d/2} e^{-||t||^2/(2\beta^2)}$$

for a fixed parameter $\beta > 0$, the BHEP test statistic is

$$T_{n,d}(\beta) = \frac{1}{n^2} \sum_{j,k=1}^n e^{-(\beta^2/2)||Y_j - Y_k||^2} - 2(1+\beta^2)^{-d/2} \frac{1}{n} \sum_{j=1}^n e^{-\beta^2 ||Y_j||^2)/(2(1+\beta^2)} + (1+2\beta^2)^{-d/2}. \quad (13)$$

Normality is rejected for large values of $T_{n,d}(\beta)$. Four versions of $T_{n,d}(\beta)$ are included in the power comparison, where $\beta = .1, .5, 1, 3$.

Henze and Zirkler [12] show that setting

$$\beta = \frac{1}{\sqrt{2}}\left[\frac{(2d+1)n}{4}\right]^{1/(d+4)} \quad (14)$$

corresponds to choosing optimal bandwidth for a multivariate nonparametric kernel density estimator with Gaussian kernel. This choice of β in (13) defines the Henze–Zirkler test, denoted HZ.

5.2. Projection pursuit tests

The projection pursuit (PP) approach to testing multivariate normality is based on the well known fact that a d-dimensional random variable X with mean vector μ and covariance matrix Σ has a multivariate normal $N_d(\mu, \Sigma)$ distribution if and only if the distribution of $a^T X$ is (univariate) $N(a^T\mu, a^T \Sigma a)$ for all vectors $a \in R^d$. The PP method tests the "worst" one dimensional projection of the multivariate data according to a univariate goodness-of-fit index. Suppose C_n is a statistic for testing univariate normality that rejects normality for large C_n. For a d-dimensional random sample $X_1, \ldots, X_n$, define

$$C_n^*(X_1, \ldots, X_n) = \sup_{a \in R^d, ||a||=1} \{C_n(a^T X_1, \ldots, a^T X_n)\}.$$

The PP test based on the index C_n rejects multinormality for large values of C_n^*. To implement the PP test based on C_n, we approximate C_n^* by finding the maximum C_n over a finite set of projections determined by a suitable "uniformly scattered" net of vectors in R^d. In $d = 2$ a suitable net of K points in R^2 is given by $\{a_k = (\cos(\theta_k), \sin(\theta_k)),\ k = 1, \ldots, K\}$, where $\theta_k = 2\pi(2k-1)/(2K)$. See Fang and Wang [8, Chapters 1 and 4] for details of implementation in $d > 2$.

We apply PP tests based on five univariate tests of normality: univariate statistic $\mathcal{E}$, modified Cramér–von Mises W^2, Kolmogorov–Smirnov D, univariate skewness $\sqrt{b_1}$, and univariate kurtosis b_2 (kurtosis is a two-tailed test). If $x_1, \dots, x_n$ is the observed d-dimensional random sample, and $a \in R^d$, let $y_{aj} = a^T S^{-1/2}(x_j - \bar{x}),\ j = 1, \dots, n$. Let $y_{a(1)}, \dots, y_{a(n)}$ denote the ordered standardized univariate sample corresponding to a. The univariate $\mathcal{E}$ statistic has a very simple form, implemented as

$$\mathcal{E}^a = n\left(\frac{2}{n}\sum_{j=1}^{n} [2y_{aj}\Phi(y_{aj}) + 2\phi(y_{aj})] - \frac{2}{\sqrt{\pi}} - \frac{2}{n^2}\sum_{j=1}^{n} (2j - 1 - n)y_{a(j)}\right).$$

Let $F_n(j) = n^{-1}\sum_{k=1}^{n} \mathbf{1}\{x_k \leqslant x_j\}$, where $\mathbf{1}(\cdot)$ is the indicator function, denote the empirical distribution function (EDF) of the sample, and let $F_n^a(\cdot)$ denote the EDF of the projected sample $y_{a1}, \dots, y_{an}$. The statistics $W^2, D, \sqrt{b_1}$, and b_2 corresponding to a are $W^{2^a} = n\sum_{j=1}^{n}(F_n^a(y_{aj}) - \Phi(y_{aj}))^2$, $D^a = \sqrt{n} \max_{1 \leqslant j \leqslant n} |F_n^a(y_{aj}) - \Phi(y_{aj})|$, $\sqrt{b_1^a} = n^{-1}\sum_{j=1}^{n}(y_{aj})^3$, and $b_2^a = n^{-1}\sum_{j=1}^{n}(y_{aj})^4$.

5.3. Simulation design

Since our sample sizes are not large, we used empirical critical values for all of the test statistics in our power study. The critical values of all multivariate and PP tests compared were estimated by Monte Carlo simulation of 20,000 standardized $N_d(0, I)$ samples (40,000 for $n = 25$). Approximate percentiles for $\hat{\mathcal{E}}_{n,d}$ are given in Table 1 for selected values of d and n.

Our Monte Carlo power study for multivariate normality compared $\hat{\mathcal{E}}_{n,d}$ with the seven multivariate tests described above for $d = 2, 3, 5, 10,\ n = 25, 50, 100,$ at significance level $\alpha = 0.05$. The five PP tests were compared for $d = 2,\ n = 25, 50, 100$, at significance level $\alpha = 0.05$. Power was estimated from a simulation of 2000 random samples from the alternative distribution. The PP tests were implemented by projecting each standardized sample in 15,000 directions: $\{a_k = (\cos(\theta_k), \sin(\theta_k)),\ k = 1, \dots, 15,000\}$, where $\theta_k = 2\pi(2k - 1)/(30,000)$.

A normal mixture is denoted $pN_d(\mu_1, \Sigma_1) + (1 - p)N_d(\mu_2, \Sigma_2)$, where the sampled populations is $N_d(\mu_1, \Sigma_1)$ with probability p, and $N_d(\mu_2, \Sigma_2)$ with probability $1 - p$. As the mixing parameter p and other parameters are varied, the multivariate normal mixtures have a wide variety of types of departures from normality. A 50% normal location mixture is symmetric with light tails, and a 90% normal location mixture is skewed with heavy tails. A normal location mixture with $p = 1 - \frac{1}{2}(1 - \frac{\sqrt{3}}{3}) = 0.7887$, provides an example of a skewed distribution with normal kurtosis [10]. The scale mixtures in the comparison are symmetric with heavier tails than normal.

Table 2
Percentage of significant tests of bivariate normality of 2000 Monte Carlo samples at $\alpha = 0.05$

Alternative	n	$\mathcal{E}$	b_1	b_2	$T_{0.1}$	$T_{0.5}$	T_1	T_3	HZ
$0.5N_d(0,I) + 0.5N_d(3,I)$	25	26	2	19	2	5	24	39	35
$0.79N_d(0,I) + 0.21N_d(3,I)$	25	50	24	9	28	39	54	34	53
$0.9N_d(0,I) + 0.1N_d(3,I)$	25	44	45	20	46	52	47	19	41
$0.5N_d(0,B) + 0.5N_d(0,I)$	25	19	19	14	19	19	18	14	18
$0.9N_d(0,B) + 0.1N_d(0,I)$	25	22	29	26	29	27	21	10	18
$0.5N_d(0,I) + 0.5N_d(3,I)$	50	71	1	34	2	6	64	79	82
$0.79N_d(0,I) + 0.21N_d(3,I)$	50	92	51	8	58	83	94	77	92
$0.9N_d(0,I) + 0.1N_d(3,I)$	50	81	83	36	86	89	82	48	74
$0.5N_d(0,B) + 0.5N_d(0,I)$	50	32	21	24	20	25	31	21	30
$0.9N_d(0,B) + 0.1N_d(0,I)$	50	39	46	51	46	48	38	15	29
$0.5N_d(0,I) + 0.5N_d(3,I)$	100	99	1	62	2	17	98	99	100
$0.79N_d(0,I) + 0.21N_d(3,I)$	100	100	91	7	93	100	100	99	100
$0.9N_d(0,I) + 0.1N_d(3,I)$	100	98	100	53	100	99	98	84	96
$0.5N_d(0,B) + 0.5N_d(0,I)$	100	65	23	40	23	43	62	50	64
$0.9N_d(0,B) + 0.1N_d(0,I)$	100	61	55	75	57	70	61	28	47

$\mu = 3$ denotes the mean vector $(3,3)^T$; $\Sigma = B$ denotes 1 on diagonal and 0.9 off diagonal.

5.4. Results of simulations

Empirical results summarized in Table 2 for tests of bivariate normality suggest that the $\hat{\mathcal{E}}_{n,d}$ test is more powerful than Mardia's skewness or kurtosis tests against the symmetric mixtures with light or normal tails, and more powerful than the kurtosis test against the skewed heavy tailed mixtures. The Henze–Zirkler test is very sensitive against the alternatives with light or normal tails, but less powerful in this comparison against the heavy tail alternatives. Monte Carlo results in dimensions $d = 3, 5$, and 10 for $n = 50$ in Table 3 suggest that the relative performance of the eight tests is similar to the bivariate case, and that all eight tests are practical and effective in higher dimensions.

Fig. 1 shows empirical power across sample sizes $n = 25, 50, 100$ in $d = 5$ for a 90% normal location mixture, where the $\hat{\mathcal{E}}_{n,d}$ test was superior to the multivariate skewness, kurtosis, and HZ tests. In Fig. 2 power is compared in $d = 5$ for a 79% normal location mixture, which suggests that $\hat{\mathcal{E}}_{n,d}$ and HZ have comparable performance superior to the skewness and kurtosis tests. Results shown in Fig. 3 suggest that $\hat{\mathcal{E}}_{n,d}$ is at least as powerful as the HZ and skewness tests against this 50% normal location–scale mixture.

Empirical power performance of the multivariate skewness, kurtosis, and BHEP tests in our study was consistent with results of similar power studies of Henze and Zirkler [12], Romeu and Ozturk [19], Horswell and Looney [14], and others. The multivariate skewness test was one of the most sensitive tests against skewed heavy tailed alternatives, but relatively very poor at detecting nonnormality in symmetric

Table 3
Percentage of significant tests of multivariate normality of 2000 Monte Carlo Samples at $\alpha = 0.05$, $n = 50$

Alternative	d	$\mathcal{E}$	b_1	b_2	$T_{0.1}$	$T_{0.5}$	T_1	T_3	HZ
$0.5N_d(0,I)+0.5N_d(3,I)$	3	58	2	21	3	6	62	70	80
$0.79N_d(0,I)+0.21N_d(3,I)$	3	98	42	8	49	78	96	66	95
$0.9N_d(0,I)+0.1N_d(3,I)$	3	91	89	35	90	94	88	40	81
$0.5N_d(0,B)+0.5N_d(0,I)$	3	71	40	52	37	53	67	47	68
$0.9N_d(0,B)+0.1N_d(0,I)$	3	65	71	76	71	72	58	23	48
$0.5N_d(0,I)+0.5N_d(3,I)$	5	20	3	11	4	7	39	23	50
$0.79N_d(0,I)+0.21N_d(3,I)$	5	79	28	7	36	66	85	25	82
$0.9N_d(0,I)+0.1N_d(3,I)$	5	93	82	33	86	95	84	16	72
$0.5N_d(0,B)+0.5N_d(0,I)$	5	99	79	95	75	92	99	76	99
$0.9N_d(0,B)+0.1N_d(0,I)$	5	89	95	95	95	92	76	22	66
$0.5N_d(0,I)+0.5N_d(3,I)$	10	5	3	6	3	5	13	8	13
$0.79N_d(0,I)+0.21N_d(3,I)$	10	27	11	6	15	29	26	10	24
$0.9N_d(0,I)+0.1N_d(3,I)$	10	66	48	20	57	68	28	9	25
$0.5N_d(0,B)+0.5N_d(0,I)$	10	100	100	100	100	100	100	92	100
$0.9N_d(0,B)+0.1N_d(0,I)$	10	97	99	98	99	96	70	17	66

$\mu = 3$ denotes the $d \times 1$ mean vector $(3,\ldots,3)^T$; $\Sigma = B$ denotes 1 on diagonal and 0.9 off diagonal.

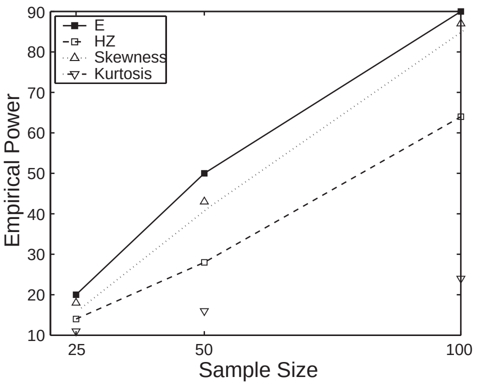

Fig. 1. Empirical power of tests of multivariate normality ($d = 5, n = 50$) against normal location mixture $0.9N_5(0,I) + 0.1N_5(2,I)$: percent of significant tests of 2000 Monte Carlo samples at $\alpha = 0.05$. E denotes $\hat{\mathcal{E}}_{n,5}$ and HZ denotes the Henze–Zirkler test.

light tailed distributions. The multivariate kurtosis test had good performance against symmetric distributions with nonnormal kurtosis. Performance of fixed β BHEP tests against different classes of alternatives depends on the parameter β. Henze and Wagner [11] give a graphical illustration and discussion of this dependence. Empirical results agree with conclusions of Henze and Zirkler [12] and Henze and Wagner [11] that small β is a good choice against symmetric heavy tailed alternatives, while large β is better against the symmetric light tailed

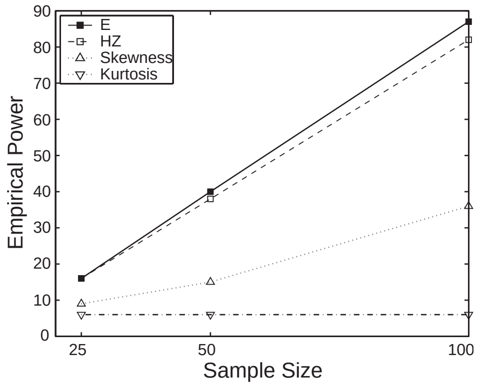

Fig. 2. Empirical power of tests of multivariate normality $(d = 5, n = 50)$ against normal location mixture $0.7887N_5(0, I) + 0.2113N_5(2, I)$: percent of significant tests of 2000 Monte Carlo samples at $\alpha = 0.05$. E denotes $\hat{\mathscr{E}}_{n,5}$ and HZ denotes the Henze–Zirkler test.

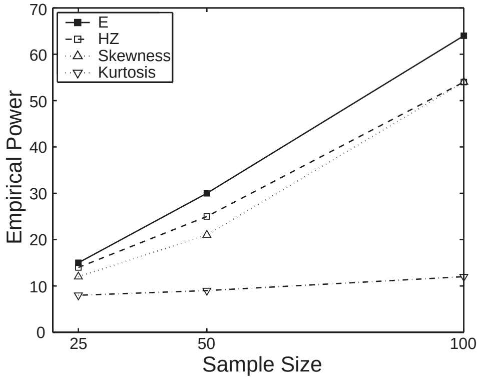

Fig. 3. Empirical power of tests of multivariate normality $(d = 5, n = 50)$ against normal mixture $0.5N_5(0, C) + 0.5N_5(2, I)$: percent of significant tests of 2000 Monte Carlo samples at $\alpha = 0.05$. Covariance matrix C has 1 on diagonal and 0.5 off diagonal. E denotes $\hat{\mathscr{E}}_{n,5}$ and HZ denotes the Henze–Zirkler test.

alternatives. The parameter β in the Henze–Zirkler test is between 1 and 2, and although not the optimal choice of β for most alternatives, provides a test that is effective against a wide class of alternatives.

Empirical results comparing projection pursuit tests are presented in Table 4. The PP-$\mathscr{E}$ test was comparable to or better than the multivariate $\mathscr{E}$ depending on the alternative. Overall, the PP-$\mathscr{E}$ test was more powerful than the PP EDF tests W^2 and D, however for sample sizes $n \geqslant 50$ there was little or no difference in power between PP-$\mathscr{E}$ and PP-W^2. The multivariate skewness test was more powerful than the PP-

Table 4
Percentage of significant tests of bivariate normality, by projection pursuit method, of 2000 Monte Carlo samples at $\alpha = 0.05$

Alternative	n	$\mathcal{E}$	$\sqrt{b_1}$	b_2	W^2	D
$0.5N_d(0,I) + 0.5N_d(3,I)$	25	36	3	48	28	22
$0.79N_d(0,I) + 0.21N_d(3,I)$	25	60	24	10	60	48
$0.9N_d(0,I) + 0.1N_d(3,I)$	25	47	49	21	46	36
$0.5N_d(0,B) + 0.5N_d(0,I)$	25	20	17	15	19	15
$0.9N_d(0,B) + 0.1N_d(0,I)$	25	20	18	14	20	14
$0.5N_d(0,I) + 0.5N_d(3,I)$	50	88	2	88	84	61
$0.79N_d(0,I) + 0.21N_d(3,I)$	50	96	55	8	96	87
$0.9N_d(0,I) + 0.1N_d(3,I)$	50	82	87	35	80	67
$0.5N_d(0,B) + 0.5N_d(0,I)$	50	40	22	26	36	26
$0.9N_d(0,B) + 0.1N_d(0,I)$	50	40	20	25	37	25
$0.5N_d(0,I) + 0.5N_d(3,I)$	100	100	2	100	100	97
$0.79N_d(0,I) + 0.21N_d(3,I)$	100	100	94	7	100	100
$0.9N_d(0,I) + 0.1N_d(3,I)$	100	99	100	62	98	94
$0.5N_d(0,B) + 0.5N_d(0,I)$	100	75	24	53	73	54
$0.9N_d(0,B) + 0.1N_d(0,I)$	100	77	24	52	73	53

$\mu = 3$ denotes the mean vector $(3,3)^T$; $\Sigma = B$ denotes 1 on diagonal and 0.9 off diagonal.

skewness test against the 90% scale mixture alternative, but otherwise both versions of skewness tests were comparable in power. Against some alternatives, such as the symmetric location mixture, the PP-kurtosis test was considerably more sensitive than the multivariate version. Against other alternatives such as the 90% scale mixture, the PP-kurtosis was considerably weaker than the multivariate test. Based on these results, it does not appear that PP-skewness or PP-kurtosis offers a clear advantage over Mardia's tests, which are computationally much less intensive. The extra computational burden does, however, seem to correspond to increased power for the PP-$\mathcal{E}$ statistic against some alternatives, without sacrificing power against others.

The purpose of this power study was to assess the performance of the new test based on $\hat{\mathcal{E}}_{n,d}$. The $\hat{\mathcal{E}}_{n,d}$ statistic had impressive performance against all the symmetric and skewed heavy tailed alternatives considered.

6. On the asymptotic behavior of the test

We have shown that under the hypotheses of Corollary 1, the V-statistics $\mathcal{E}_n/n$ based on the kernel $h(x,y)$ defined in (1), satisfy $E[h(X,Y)] \geqslant 0$ with equality if and only if $X \overset{d}{=} Y$. If V_n is a degenerate V-statistic generated by square integrable kernel

h, then nV_n converges in distribution to

$$\sum_{k=0}^{\infty} \lambda_k Z_k^2,$$

where $\{Z_k\}$ is a sequence of independent standard normal random variables, and $\{\lambda_k\}$ are the eigenvalues determined by the integral equation

$$\int \psi(x)h(x,y)\, dF_X(x) = \lambda\psi(y). \tag{15}$$

See [21] or [27] for asymptotic distribution theory of degenerate kernel V-statistics. Since $\mathcal{E}_n/n$ is a degenerate kernel V-statistic, the distribution of $\mathcal{E}_n$ under $X \overset{d}{=} Y$ converges to a proper limiting distribution (the quadratic form $\sum_{k=0}^{\infty}\lambda_k Z_k^2$) provided the second moments of X are finite. If X and Y are not identically distributed, then $\lim_{n\to\infty} \mathcal{E}_n = \infty$ with probability 1. Hence the statistics $\mathcal{E}_n$ determine consistent goodness-of-fit tests for any multivariate distributions with finite second moments.

In the univariate case, there is an interesting connection between

$$E[h(X,Y)] = 2E|X-Y| - E|X-X'| - E|Y-Y'|$$

and the Cramér–von Mises distance given by

$$\int_{-\infty}^{\infty} (G(t) - F(t))^2 \psi(t)\, dF(t), \tag{16}$$

where F and G are the distribution functions of X and Y, respectively, and $\psi(t)$ is a weight function. It is easy to prove that in one dimension

$$2E|X-Y| - E|X-X'| - E|Y-Y'| = 2\int_{-\infty}^{\infty} (G(t) - F(t))^2\, dt. \tag{17}$$

(Equality (17) cannot hold in higher dimensions since the right-hand side is not rotation invariant.) Our test is a rotation invariant multivariate version of a Cramér–von Mises type test, and the power of our univariate test is similar to the power of a suitable Cramér–von Mises type test. From the right-hand side of (17) we get the classical "distribution free" Cramér–von Mises goodness-of-fit formula if dt is replaced by $dF(t)$ (and G is replaced by the empirical distribution F_n). When the weight function in (16) is $\psi(t) = [F(t)(1-F(t))]^{-1}$, we have the Anderson–Darling distance [1]. In case of standard normal null F, the shape of the curve $[\psi(t)]^{-1} = F(t)(1-F(t))$ is similar to the shape of the density $F'(t)$: their ratio is close to a constant c (empirically $c \approx 0.67$.) That is, if we replace $dF(t)/[F(t)(1-F(t))]$ by $c^{-1}\, dt$, we can see that our univariate test hardly differs from the powerful Anderson–Darling test of univariate normality. In fact our simulations show (see Table 5) that in one dimension the power of our test, even for small samples sizes, is almost the same as that of Anderson–Darling, which has extremely good 0.96 Bahadur local index for Gaussian null and location alternatives [18, p. 80].

Integral equations of the form (15) typically do not have nice analytical solutions. A remarkable exception is the Cramér–von Mises case when F is linear in [0,1] and

Table 5
Empirical power of $\mathcal{E}$ test of univariate normality compared with the Anderson–Darling A^2 test: percentage of significant tests of 5000 Monte Carlo samples, at $\alpha = 0.05$

	$n=10$		$n=25$		$n=50$		$n=100$	
Alternative	$\mathcal{E}$	A^2	$\mathcal{E}$	A^2	$\mathcal{E}$	A^2	$\mathcal{E}$	A^2
Student $t(4)$	13	14	26	27	40	41	64	65
Student $t(10)$	6	6	9	9	11	12	15	16
Uniform(0,1)	8	8	23	23	56	58	92	94
Logistic(0,1)	8	8	12	12	16	16	22	23
Laplace	16	16	31	31	55	55	83	83
$\chi^2(5)$	19	19	45	47	78	80	99	99
$\chi^2(10)$	11	11	25	26	46	48	78	80
$0.5N(0,1)+0.5N(3,1)$	8	8	19	19	45	44	81	81
$0.79N(0,1)+0.21N(3,1)$	13	13	34	34	65	65	93	93
$0.9N(0,1)+0.1N(3,1)$	13	13	28	29	49	50	80	82
Binomial(10,0.5)	20	20	51	50	99	98	100	100
Binomial(20,0.5)	11	11	19	19	50	48	100	100
Poisson(4)	16	16	37	37	86	86	100	100
Poisson(8)	10	9	16	16	33	33	86	84

$\psi(t)$ is identically 1. For other exceptions see [26]. So far we have not tried numerical approximations because (i) our test does not apply the concrete form of the limit distribution, only its existence, and (ii) according to Götze [9] and Bickel et al. [4], the rate of convergence is $o(n^{-1})$ for degenerate kernel U-statistics (and V-statistics), compared with $O(n^{-1/2})$ rate of convergence to the normal limit in the nondegenerate case. For this reason as Henze and Wagner [11] observed for their degenerate kernel BHEP V-statistics, "the test is practically sample size independent." We observed the same behavior for our degenerate kernel; see Table 1. This fact decreases the practical importance of large sample analysis.

7. Summary

We have proposed a new class of tests for comparing multivariate distributions, and applied a new test of multivariate normality that is affine invariant and consistent against all fixed alternatives. The new test of multivariate normality is practical to apply for arbitrary dimension and sample size, and our Monte Carlo power comparisons suggest that it is a powerful competitor to two important categories of affine invariant tests of multivariate normality, Mardia's skewness and kurtosis tests and the BHEP tests.

These empirical results illustrate that none of the tests are universally superior, but some general aspects of power performance are evident. Overall, the relative performance of $\hat{\mathcal{E}}_{n,d}$ was impressive against heavy tailed alternatives, and better than a skewness test against symmetric light tailed alternatives. Among the tests compared, $\hat{\mathcal{E}}_{n,d}$ was the only test that never ranked below all other tests in power. We conclude that $\hat{\mathcal{E}}_{n,d}$ provides a powerful *omnibus* test of multivariate normality, in the sense that it is consistent against all alternatives and has relatively good power against general alternatives compared with other tests. When applied as the index of fit in the projection pursuit test, univariate $\hat{\mathcal{E}}$ had impressive empirical power relative to other univariate measures of fit, and PP-$\hat{\mathcal{E}}$ outperformed multivariate $\hat{\mathcal{E}}_{n,d}$ against certain alternatives. Moreover, in view of the close relation between $\hat{\mathcal{E}}$ and Anderson–Darling, multivariate $\hat{\mathcal{E}}_{n,d}$ can be viewed as a computationally simple way to lift a powerful EDF test to arbitrarily high dimension.

Acknowledgments

The authors thank the referees for their constructive comments.

Appendix. Proofs of statements

Proof of Proposition 1. Let $x_1, \ldots, x_n$ be arbitrary distinct points in R^d. To prove that

$$\sum_{j,k=1}^{n} ||x_j - x_k|| r_j r_k \leqslant 0 \text{ whenever } \sum_{j=1}^{n} r_j = 0,$$

with equality if and only if $r_1 = \cdots = r_n = 0$, it is equivalent to prove the following: if $p_1, \ldots, p_n$ and $q_1, \ldots, q_n$ are probability distributions, and $x_1, \ldots, x_n$ are distinct arbitrary points in R^d, then

$$\sum_{j,k=1}^{n} ||x_j - x_k||(2p_j q_k - p_j p_k - q_j q_k) \geqslant 0, \tag{A.1}$$

and equality holds if and only if $p_1, \ldots, p_n$ and $q_1, \ldots, q_n$ are identical distributions.

Fix the points $x_1, \ldots, x_n$ in R^d and a hyperplane H, and suppose that there are m points on one side of H, $0 < m < n$. Select two points at random from $x_1, \ldots, x_n$ according to the p or q distribution and connect them if the points are on opposite sides of H. The connected pairs are called homogeneous pairs if both are selected by the same distribution, and called mixed pairs otherwise. If $p = \sum_{j=1}^{m} p_j$ and $q = \sum_{j=1}^{m} q_j$, then the expected number of mixed pairs minus the expected number of homogeneous pairs is

$$\begin{aligned} &p(1-q) + q(1-p) - p(1-p) - q(1-q) \\ &\quad = (p-q)^2 \geqslant 0, \end{aligned} \tag{A.2}$$

and equality holds if and only if $p = q$.

By a suitable randomization of H (see below for details), the sum in inequality (A.1) is determined by the difference of expected length of line segments connecting mixed pairs and expected length of line segments connecting homogenous pairs.

Suppose H is chosen randomly, and consider all possible partitions $P_m = \{S_{m_1}, S_{m_2}\}$ of the points $x_1, \dots, x_n$ determined by H. Then each partition defines probability distributions $\{p^{(m)}, 1-p^{(m)}\}$ and $\{q^{(m)}, 1-q^{(m)}\}$, such that $p^{(m)} = \sum_{j \in S_{m_1}} p_j$ and $q^{(m)} = \sum_{j \in S_{m_1}} q_j$.

For each of the r possible partitions $P_1, \dots, P_r$ let α_m denote the probability that H determines partition P_m. Then the expected number of line segments between mixed pairs minus the expected number of line segments between homogeneous pairs is

$$\sum_{m=1}^{r} \alpha_m (p^{(m)}(1-q^{(m)}) + q^{(m)}(1-p^{(m)}) \\ - p^{(m)}(1-p^{(m)}) - q^{(m)}(1-q^{(m)})) \\ = \sum_{m=1}^{r} \alpha_m (p^{(m)} - q^{(m)})^2 \geqslant 0.$$

The hyperplane H is randomized as follows. Select a ball B in R^d with center O that contains the points $x_1, \dots, x_n$, a point P uniformly distributed on the surface of B, and a point Q uniformly distributed on radius OP. Choose the hyperplane H such that H contains Q and H is perpendicular to OP. Then for each pair (x_j, x_k), the probability that H intersects the segment between x_j and x_k is proportional to $||x_j - x_k||$. This is a known fact in integral geometry; see e.g. [20] or [22].

If H is chosen in this way, the expected length of line segments between mixed pairs minus the expected length of line segments between homogeneous pairs is proportional to

$$\sum_{j,k=1}^{n} ||x_j - x_k||(2p_j q_k - p_j p_k - q_j q_k).$$

This sum is nonnegative and equals zero if and only if $(p^{(m)} - q^{(m)})^2 = 0$ for all m. For each j, there is an m such that $p^{(m)} = p_j$ and $q^{(m)} = q_j$, so equality holds in (A.1) if and only if $p_j = q_j,\ j = 1, 2, \dots, n$. $\square$

Proof of Proposition 2. Y is an ancillary statistic and independent of $(\bar{Z}, \mathrm{S})$ by Basu's Theorem, so

$$E|Z_1 - Z_2| = E[E[|Z_1 - Z_2|\,|S]] \\ = E[E[S|Y_1 - Y_2||S]] = E[S]E|Y_1 - Y_2|.$$

Hence

$$E|Y_1 - Y_2| = \frac{E|Z_1 - Z_2|}{E[S]} = \frac{2}{\sqrt{\pi}} \frac{1}{E[S]}$$
$$= \frac{2}{\sqrt{\pi}} \sqrt{\frac{2}{n-1}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}$$
$$> \frac{2}{\sqrt{\pi}} \sqrt{\frac{2}{n-1}} \sqrt{\frac{n-1}{2}} = E|Z_1 - Z_2|. \qquad \square$$

Proof of Proposition 3. Suppose $Y_1, \dots, Y_n$ is a standardized normal sample, and let $X_1, \dots, X_m$, and $Z_1, \dots, Z_n$ be iid $N(0,1)$, independent of $Y_1, \dots, Y_n$. Denote the combined ordered sample $X_1, \dots, X_m, Y_1, \dots, Y_n$ by $U_{(1)}, \dots, U_{(m+n)}$, and denote the combined ordered sample $X_1, \dots, X_m, Z_1, \dots, Z_n$ by $W_{(1)}, \dots, W_{(m+n)}$. If y_j is fixed, then the sample mean $m^{-1}\sum_{k=1}^{m}|y_j - X_k|$ is an unbiased estimator of

$$E|y_j - Z| = \int_{-\infty}^{\infty} |y_j - z|\phi(z)\, dz.$$

If $y_1, \dots, y_n$ are the observed values of a standardized normal random sample, an unbiased estimator of the expected value of $n^{-1}\sum_{j=1}^{n} E|y_j - Z|$ is

$$\frac{1}{mn} \sum_{k=1}^{m} \sum_{j=1}^{n} |y_j - X_k|, \tag{A.3}$$

and hence

$$E_{\mathscr{S}}\left[\frac{1}{n}\sum_{j=1}^{n} E|Y_j - Z|\right] = E\left[\frac{1}{mn}\sum_{j=1}^{m}\sum_{k=1}^{n} |Y_j - X_k|\right].$$

An L-statistic is a linear combination of the ordered sample. If $X_{(1)}, \dots, X_{(n)}$ denotes the ordered sample $X_1, \dots, X_n$, then the following L-statistics identities hold.

$$\sum_{j=1}^{n}\sum_{k=1}^{n} |X_j - X_k| = 2\sum_{k=1}^{n} ((2k-1) - n)X_{(k)}, \tag{A.4}$$

and

$$\sum_{j=1}^{n}\sum_{k=1}^{m} |X_j - Y_k| = \sum_{k=1}^{n+m} ((2k-1) - (n+m))U_{(k)} - \sum_{k=1}^{n} ((2k-1) - n)X_{(k)} - \sum_{k=1}^{m} ((2k-1) - m)Y_{(k)}. \tag{A.5}$$

An unbiased estimator of the expected value of $n^{-1}\sum_{j=1}^{n} E|x_j - Z|$ is

$$\frac{1}{mn} \sum_{k=1}^{m} \sum_{j=1}^{n} |Z_j - X_k|. \tag{A.6}$$

By identity (A.5) the difference in the Monte Carlo estimates (A.3) and (A.6) is $(mn)^{-1}$ times

$$\sum_{k=1}^{m}\sum_{j=1}^{n}|X_k-Y_j|-\sum_{k=1}^{m}\sum_{j=1}^{n}|X_k-Z_j|$$
$$=\sum_{k=1}^{n+m}((2k-1)-(n+m))(U_{(k)}-W_{(k)})$$
$$-\sum_{k=1}^{n}((2k-1)-m)(Y_{(k)}-Z_{(k)}). \tag{A.7}$$

From identity (A.4) and inequality (9),

$$E\left[\sum_{k=1}^{n}((2k-1)-n)(Y_{(k)}-Z_{(k)})\right]$$
$$=\frac{n(n-1)}{2}(E[Y_1-Y_2]-E[Z_1-Z_2])>0.$$

Since $\sum_{k=1}^{n+m}((2k-1)-(n+m))(U_{(k)}-W_{(k)})$ has negative expected value,

$$\lim_{m\to\infty}\frac{1}{mn}\sum_{k=1}^{m}\sum_{j=1}^{n}|Y_j-X_k|-\lim_{m\to\infty}\frac{1}{mn}\sum_{k=1}^{m}\sum_{j=1}^{n}|Z_j-X_k|$$

cannot be positive for any finite sample size n. $\square$

Proof of Theorem 2. In Section 4, proof of consistency of the test of multivariate normality that is based on $\hat{\mathscr{E}}_{n,d}$ is given for the case when the alternative distribution has finite second moments. In the general case, we apply consistency in the special case of finite second moments to the truncated distribution.

For a fixed constant M define

$$\tilde{Y}=S^{-1/2}(\tilde{X}-\bar{X})$$

where

$$\tilde{X}=\begin{cases}X & \text{if } ||X||<M,\\ M\dfrac{X}{||X||} & \text{if } ||X||\geqslant M,\end{cases}$$

and $\bar{X}, S$ denote the sample mean vector and sample covariance matrix of $\tilde{X}$, and let

$$\hat{\mathscr{E}}^{M}_{n,d}=n\left(\frac{2}{n}\sum_{k=1}^{n}E||\tilde{y}_k-Z||-E||Z-Z'||-\frac{1}{n^2}\sum_{j,k=1}^{n}||\tilde{y}_j-\tilde{y}_k||\right).$$

To find a suitable truncation point M, first define a constant M_0 as follows. For arbitrary $\mu\in R^d$ and symmetric positive definite $\Sigma_{d\times d}$, define

$$m(\mu,\Sigma)=\sup\{m\geqslant 0:\ X\overset{d}{=}N_d(\mu,\Sigma)\ \text{a.e. on}\ B(0,m)\},$$

where $B(0,m)$ denotes the ball of radius m centered at 0, and let $M_0 = \sup_{\mu,\Sigma}\{m(\mu,\Sigma)\}$. Then $M_0 < \infty$ since X is nonnormal. If X is truncated at $M > M_0$, $\tilde{X}$ is nonnormal with finite covariance, so for any $c \in R$, $\lim_{n\to\infty} P(\hat{\mathscr{E}}^M_{n,d} > c) = 1$.

Let $\{p_n\} \subseteq [0,1]$ be a sequence such that $\lim_{n\to\infty} p_n^n = 1$. For each n, choose $M_n > M_0$ such that $P(X = \tilde{X}) > p_n$ if X is truncated at M_n, and $\lim_{n\to\infty} M_n = \infty$. Then $||y_j - y_k|| \geqslant ||\tilde{y}_j - \tilde{y}_k||$ and

$$\lim_{n\to\infty} P\left(\sum_{j,k=1}^{n} ||y_j - y_k|| > \sum_{j,k=1}^{n} ||\tilde{y}_j - \tilde{y}_k||\right)$$
$$= \lim_{n\to\infty} P(||x_j|| \geqslant M_n \text{ for some j}) = \lim_{n\to\infty} (1 - p_n^n) = 0.$$

Similarly, $E||y_j - Z|| \geqslant E||\tilde{y}_j - Z||$ and

$$\lim_{n\to\infty} P\left(\sum_{j=1}^{n} E||y_j - Z|| > \sum_{j=1}^{n} E||\tilde{y}_j - Z||\right)$$
$$= \lim_{n\to\infty} P(||x_j|| \geqslant M_n \text{ for some j}) = \lim_{n\to\infty} (1 - p_n^n) = 0.$$

Therefore $\lim_{n\to\infty} P(\hat{\mathscr{E}}_{n,d} > c) = \lim_{n\to\infty} P(\hat{\mathscr{E}}^{M_n}_{n,d} > c) = 1$. $\square$

References

[1] T. Anderson, D. Darling, A test of goodness-of-fit, J. Amer. Statist. Assoc. 49 (1954) 765–769.

[2] L. Baringhaus, N. Henze, A consistent test for multivariate normality based on the empirical characteristic function, Metrika 35 (1988) 339–348.

[3] L. Baringhaus, N. Henze, Limit distributions for Mardia's measure of multivariate skewness, Ann. Statist. 20 (1992) 1889–1902.

[4] P.J. Bickel, F. Götze, W.R. van Zwet, The Edgeworth expansion for U-statistics of degree two, Ann. Statist. 14 (4) (1986) 1463–1484.

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

[6] H. Cui, P. Cheng, The P-values of testing for multinormality based on the PP skewness and kurtosis, Progr. Nat. Sci. 6 (3) (1996) 277–283.

[7] T.W. Epps, L.B. Pulley, A test for normality based on the empirical characteristic function, Biometrika 70 (1983) 723–726.

[8] K.-T. Fang, Y. Wang, Number Theoretic Methods in Statistics, Chapman & Hall, New York, 1994.

[9] F. Götze, Asymptotic expansions for bivariate von Mises functionals, Z. Wahrsch. Verw. Gebiete 50 (1979) 333–355.

[10] N. Henze, On Mardia's kurtosis test for multivariate normality, Comm. Statist. Theory Methods 23 (1994) 1031–1045.

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

[12] N. Henze, B. Zirkler, A class of invariant and consistent tests for multivariate normality, Comm. Statist. Theory Methods 19 (1990) 3595–3617.

[13] W. Hoeffding, A class of statistics with asymptotically normal distribution, Ann. Math. Statist. 49 (1948) 293–325.

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

[15] J.F. Malkovich, A.A. Afifi, On tests for multivariate normality, J. Amer. Statist. Assoc. 68 (1973) 519–530.
[16] K.V. Mardia, Measures of multivariate skewness and kurtosis with applications, Biometrika 57 (1970) 519–530.
[17] D. Morgenstern, Proof of a conjecture by Walter Deuber concerning the distance between points of two types in R^d, Discrete Math. 226 (2001) 347–349.
[18] Y. Nikitin, Asymptotic Efficiency of Nonparametric Tests, Cambridge University Press, New York, 1995.
[19] J.L. Romeu, A. Ozturk, A comparative study of goodness-of-fit tests for multivariate normality, J. Multivariate Anal. 46 (1993) 309–334.
[20] L.A. Santaló, Integral Geometry and Geometric Probability, Addison-Wesley, Reading, MA, 1976.
[21] R.J. Serfling, Approximation Theorems of Mathematical Statistics, Wiley, New York, 1980.
[22] H. Solomon, Geometric Probability, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1978.
[23] G.J. Székely, Contests in Higher Mathematics, Springer, New York, 1996.
[24] G.J. Székely, Pre-limit and post-limit theorems for statistics, in: C.R. Rao, G.J. Székely (Eds.), Statistics for the 21st Century, Marcel Dekker, New York, 2001, pp. 411–422.
[25] G.J. Székely, T.F. Móri, A characteristic measure of asymmetry and its application for testing diagonal symmetry, Comm. Statist. Theory Methods 30 (8&9) (2001) 1633–1639.
[26] O. Thas, J.P. Ottoy, Some generalizations of the Anderson–Darling statistic, Statist. Probab. Lett. 64 (2003) 255–261.
[27] A.W. van der Vaart, Asymptotic Statistics, Cambridge University Press, New York, 1998.
[28] L.-X. Zhu, K.-T. Fang, M.I. Bhatti, On estimated projection pursuit-type Cramer–von Mises Statistics, J. Multivariate Anal. 63 (1997) 1–14.
[29] L.-X. Zhu, H.L. Wong, K.-T. Fang, A test for multivariate normality based on sample entropy and projection pursuit, J. Statist. Plann. Inference 45 (1995) 373–385.