TESTING FOR AFFINE EQUIVALENCE OF ELLIPTICALLY SYMMETRIC DISTRIBUTIONS [1] [2]

Arjun K. Gupta, Norbert Henze[3] and Bernhard Klar

Department of Mathematics & Statistics
Bowling Green State University, Ohio, USA
and
Institut für Mathematische Stochastik, Universität Karlsruhe,
Karlsruhe, Germany

Abstract. Let X and Y be d-dimensional random vectors having elliptically symmetric distributions. Call X and Y affinely equivalent if Y has the same distribution as $AX + b$ for some nonsingular $d \times d$-matrix A and some $b \in \mathbb{R}^d$. This paper studies a class of affine invariant tests for affine equivalence under certain moment restrictions. The test statistics are measures of discrepancy between the empirical distributions of the norm of suitably standardized data.

Keywords. Multivariate two-sample problem; elliptically symmetric distribution; affine equivalence; affine invariance; empirical characteristic function.

1 Introduction

Let $X_1, \ldots, X_m, \ldots$ be independent copies of a random (column) vector X and, independently of the X_j, let $Y_1, \ldots, Y_n, \ldots$ be independent copies of a random vector Y. The distributions of X and Y are assumed to be continuous and elliptically symmetric, i.e., X and Y are some full rank affine transformations of spherically symmetric distributions (see, e.g. Fang, Kotz and Ng (1990), p. 31). Moreover, we assume $E\|X\|^4 < \infty$

[1] AMS 1991 subject classification numbers: 62H15, 62E20

[2] Proposed running head: TESTS FOR EQUIVALENCE OF ELLIPTICAL DISTRIBUTIONS

[3] Research started while N. Henze was a distinguished Lukacs professor visiting Bowling Green State University

and $E\|Y\|^4 < \infty$, where $\|\cdot\|$ denotes the Euclidean norm. Writing '$\sim$' for equality in distribution and Gl_d for the group of nonsingular matrices of order d, this paper deals with testing the hypothesis

$$(1.1) \qquad H_0: \; Y \sim AX + b \quad \text{for some} \quad A \in Gl_d \quad \text{and some} \quad b \in \mathbb{R}^d,$$

against general alternatives. Put in other words, H_0 means that X and Y have a density of the form

$$|\Sigma|^{-1/2}\, g\left((x-\mu)'\Sigma^{-1}(x-\mu)\right)$$

with the same (unspecified) density generator g (see Fang, Kotz and Ng (1990), p.35), but with possibly different values of the location vector μ and the positive definite matrix Σ. Here, $|\Sigma|$ denotes the determinant of Σ, and the prime stands for transpose of vectors and matrices.

Notice that H_0 includes the special case that both X and Y have nondegenerate normal distributions. The idea of testing H_0 is to test the equivalent hypothesis that the standardized distributions coincide. Since these distributions are spherically symmetric and thus uniquely determined by the distribution of their radial parts, the test statistic is a suitable measure of discrepancy between the empirical distributions of the radial part of the standardized samples. The paper is organized as follows. Section 2 introduces the new test statistics, and Section 3 is devoted to asymptotic distribution theory. Section 4 introduces a resampling procedure that renders the test asymptotically of a prespecified level. The final section presents the results of a Monte Carlo study.

2 The test statistics

We first transform (1.1) into an equivalent testing problem. If $Y \sim AX + b$ for $A \in Gl_d$ and $b \in \mathbb{R}^d$, then $\nu = A\mu + b$ and $T = A\Sigma A'$, where $\mu = EX$, $\nu = EY$, $S = E(X - \mu)(X-\mu)'$ and $T = E(Y-\nu)(Y-\nu)'$. Let $S^{-1/2}$ denote the symmetric positive definite square root of S^{-1}, and likewise define $T^{-1/2}$. Furthermore, put $\tilde{X} = S^{-1/2}(X-\mu)$, $\tilde{Y} = T^{-1/2}(Y-\nu)$. The distributional equality $Y \sim AX + b$ then implies

$$\tilde{Y} \sim T^{-1/2}AS^{-1/2}\tilde{X}.$$

Since $\tilde{X}$ and $\tilde{Y}$ have spherically symmetric distributions and the matrix $T^{-1/2}AS^{-1/2}$ is orthogonal, we have $\tilde{Y} \sim \tilde{X}$ and thus

$$(2.1) \qquad \|\tilde{Y}\|^2 \;\sim\; \|\tilde{X}\|^2.$$

To show that (2.1) entails (1.1), notice that $\tilde{Y} \sim U \cdot \|\tilde{Y}\|$ and $\tilde{X} \sim V \cdot \|\tilde{X}\|$, where $U, V, \|\tilde{X}\|$ and $\|\tilde{Y}\|$ are independent, and the distributions of U and V are uniform over the surface of the unit d-sphere (see Fang, Kotz and Ng (1990), p. 30). Therefore, (2.1) implies $\tilde{Y} \sim \tilde{X}$, from which (1.1) readily follows.

To test (2.1), let $\bar{X}_m = m^{-1}\sum_{j=1}^m X_j$, $\bar{Y}_n = n^{-1}\sum_{k=1}^n Y_k$ denote the sample means, and write $S_m = m^{-1}\sum_{j=1}^m (X_j - \bar{X}_m)(X_j - \bar{X}_m)'$, $T_n = n^{-1}\sum_{k=1}^n (Y_k - \bar{Y}_n)(Y_k - \bar{Y}_n)'$ for the sample covariance matrices. We assume that $m > d$ and $n > d$, thus ensuring the almost sure invertibility of S_m and T_n (see Eaton and Perlman (1973)). Define the standardized data

$$\tilde{X}_j = S_m^{-1/2}(X_j - \bar{X}_m), \qquad \tilde{Y}_k = T_n^{-1/2}(Y_k - \bar{Y}_n)$$

$(1 \le j \le m, 1 \le k \le n)$. For short, put

$$D_j = \|\tilde{X}_j\|^2 \quad (1 \le j \le m), \qquad \Delta_k = \|\tilde{Y}_k\|^2 \quad (1 \le k \le n).$$

Our measure of discrepancy between the empirical distributions of $D_1, \ldots, D_m$ and $\Delta_1, \ldots, \Delta_n$ is based on the empirical characteristic functions

$$\varphi_m(t) = \frac{1}{m}\sum_{j=1}^m \exp(itD_j),$$

$$\psi_n(t) = \frac{1}{n}\sum_{k=1}^n \exp(it\Delta_k)$$

of these samples. In the spirit of a class of tests for univariate and multivariate normality (Epps and Pulley (1983), Henze and Wagner (1997)), the test statistic is

$$(2.2) \qquad U_{m,n,a} \;=\; \int_{-\infty}^{\infty} \left|\varphi_m(t) - \psi_n(t)\right|^2 \exp(-at^2)\,dt,$$

where $a > 0$ is a constant the role of which will be discussed later. It is readily seen that, as $m, n \to \infty$,

$$U_{m,n,a} \;\longrightarrow\; \int_{-\infty}^{\infty} |\varphi(t) - \psi(t)|^2 \exp(-at^2)\,dt$$

almost surely, where $\varphi(t) = E\exp(it\|\tilde{X}\|^2)$ and $\psi(t) = E\exp(it\|\tilde{Y}\|^2)$. Thus, rejecting H_0 for large values of $U_{m,n,a}$ should give a reasonable test of H_0.

Using

$$\int_{-\infty}^{\infty} \cos(tc)\exp(-at^2)dt \;=\; \left(\frac{\pi}{a}\right)^{1/2} \exp\left(-\frac{c^2}{4a}\right),$$

some algebra yields the alternative representation

$$(2.3)\quad U_{m,n,a} \;=\; \sqrt{\frac{\pi}{a}}\left[\frac{1}{m^2}\sum_{j,k=1}^{m}\exp\left(-\frac{1}{4a}(D_j - D_k)^2\right) + \frac{1}{n^2}\sum_{j,k=1}^{n}\exp\left(-\frac{1}{4a}(\Delta_j - \Delta_k)^2\right) - \frac{2}{mn}\sum_{j=1}^{m}\sum_{k=1}^{n}\exp\left(-\frac{1}{4a}(D_j - \Delta_k)^2\right)\right].$$

This shows that a computer routine implementing $U_{m,n,a}$ is readily available. Moreover, since $D_j = (X_j - \bar{X}_m)'S_m^{-1}(X_j - \bar{X}_m)$ and $\Delta_k = (Y_k - \bar{Y}_n)'T_n^{-1}(Y_k - \bar{Y}_n)$, not even the computation of the square roots of S_m^{-1} and T_n^{-1} is needed.

For later purposes, we note that, by analogy with (2.2), $U_{m,n,a}$ may be written in the form

$$(2.4)\quad U_{m,n,a} \;=\; \int_{-\infty}^{\infty} \left(\varphi_m^*(t) - \psi_n^*(t)\right)^2 \exp(-at^2)\,dt,$$

where

$$(2.5)\quad \varphi_m^*(t) = \frac{1}{m}\sum_{j=1}^{m}\left[\cos(tD_j) + \sin(tD_j)\right],$$

$$(2.6)\quad \psi_n^*(t) = \frac{1}{n}\sum_{k=1}^{n}\left[\cos(t\Delta_k) + \sin(t\Delta_k)\right]$$

are the empirical cosine-sine-transforms of $D_1, \ldots, D_m$ and $\Delta_1, \ldots, \Delta_n$, respectively. Representation (2.4) follows readily from symmetry arguments and the trigonometric formula $\cos(u - v) = \cos u \cos v + \sin u \sin v$.

An important property of $U_{m,n,a}$ is its invariance with repect to affine transformations $X_j \mapsto BX_j + \beta$, $Y_k \mapsto CY_k + \gamma$ $(B, C \in Gl_g; \beta, \gamma \in \mathbb{R}^d)$ of the data. Restriction to affine invariant test statistics is crucial since the testing problem (1.1) is affine invariant in the sense that H_0 holds for X and Y if, and only if, it holds for $BX+\beta$ and $CY+\gamma$ for any choice of $B, C \in Gl_d$ and $\beta, \gamma \in \mathbb{R}^d$. Consequently, a decision in favor or against H_0 should be the same for X_j, Y_k $(1 \le j \le m, 1 \le k \le n)$ and the transformed data $BX_j+\beta$, $CY_k+\gamma$ $(1 \le j \le m, 1 \le k \le n)$, a goal that is achieved by affine invariant test statistics. Since $U_{m,n,a}$ is affine invariant, its null distribution does not depend on the matrix A and the vector b figuring in (1.1). Under H_0 we thus may assume without loss of generality that X and Y have the same distribution.

In what follows, we discuss the role of the weight function $\exp(-at^2)$ figuring in (2.2). Our first result shows that $U_{m,n,a}$ has an alternative representation in terms of an L^2-distance between two nonparametric density estimators.

Proposition 2.1 *We have*

$$U_{m,n,a} \;=\; 2\pi \int_{-\infty}^{\infty} \left(\hat{f}_m(x) - \hat{g}_n(x)\right)^2 \, dx,$$

where

$$\hat{f}_m(x) \;=\; \frac{1}{m}\sum_{j=1}^{m} \frac{1}{\sqrt{2\pi a}} \, \exp\left(-\frac{(x-D_j)^2}{2a}\right),$$

$$\hat{g}_n(x) \;=\; \frac{1}{n}\sum_{k=1}^{n} \frac{1}{\sqrt{2\pi a}} \, \exp\left(-\frac{(x-\Delta_j)^2}{2a}\right).$$

PROOF. Let

$$\tilde{u}(x) = \int_{-\infty}^{\infty} \exp(itx) u(t)\, dt$$

be the Fourier transform of a square integrable complex-valued function u defined on $\mathbb{R}$. By Plancherel's theorem, we have

$$(2.7) \qquad \int_{-\infty}^{\infty} |\tilde{u}(x)|^2 \, dx \;=\; 2\pi \int_{-\infty}^{\infty} |u(t)|^2 \, dt.$$

Notice that $U_{m,n,a} = \int_{-\infty}^{\infty} |\tilde{u}(x)|^2 dx$, where

$$\tilde{u}(x) \;=\; \frac{1}{m}\sum_{j=1}^{m} \exp\left(itD_j - \frac{a}{2}t^2\right) \;-\; \frac{1}{n}\sum_{k=1}^{n} \exp\left(it\Delta_k - \frac{a}{2}t^2\right).$$

Write $\mathcal{P}_m$ for the empirical distribution of $D_1, \ldots, D_m$, and let $\mathcal{Q}_n$ be the empirical distribution of $\Delta_1, \ldots, \Delta_n$. The function $m^{-1} \sum_{j=1}^{m} \exp(itD_j - at^2/2)$ is the Fourier transform of the convolution $\mathcal{P}_m \star \mathcal{N}(0, a)$, and $n^{-1} \sum_{k=1}^{n} \exp(it\Delta_j - at^2/2)$ is the Fourier transform of the convolution $\mathcal{Q}_n \star \mathcal{N}(0, a)$. Since $\mathcal{P}_m \star \mathcal{N}(0, a)$ and $\mathcal{Q}_n \star \mathcal{N}(0, a)$ have densities $\hat{f}_m(x)$ and $\hat{g}_n(x)$, respectively, the assertion follows from (2.7). ■

Since $\hat{f}_m(x)$ and $\hat{g}_n(x)$ are nonparametric kernel density estimators with Gaussian kernel $(2\pi)^{-1/2} \exp(-t^2/2)$ and bandwidth $\sqrt{a}$, applied to $D_1, \ldots, D_m$ and $\Delta_1, \ldots, \Delta_n$, repectively, the role of a in the definition of $U_{m,n,a}$ ist that of a smoothing parameter. From the viewpoint of density estimation, the bandwidth must tend to zero as the sample size increases in order to obtain consistent estimates. However, we keep a fixed in what follows in order to be able to discriminate between alternatives that approach each other at the rate $1/\sqrt{m+n}$, where m and n are assumed to be of the same order of magnitude (as for this point, see Anderson, Hall, and Titterington (1994)).

Our next result shows that, in the limit as $a \to \infty$ (according to the above discussion, this case corresponds to 'infinite smoothing'), a rescaled version of $U_{m,n,a}$ approaches a limit statistic that may be of independent interest.

Proposition 2.2 *We have*

$$\lim_{a\to\infty} \frac{16a^{5/2}}{3\sqrt{\pi}} U_{m,n,a} = \left(\frac{1}{m} \sum_{j=1}^{m} D_j^2 - \frac{1}{n} \sum_{k=1}^{n} \Delta_k^2 \right)^2 . \tag{2.8}$$

Proof. By expanding the exponential terms in (2.3), we have

$$\begin{aligned} \sqrt{\frac{a}{\pi}}\, U_{m,n,a} &= \frac{1}{m^2} \sum_{j,k=1}^{m} \left[1 - \frac{1}{4a}(D_j - D_k)^2 + \frac{1}{32a^2}(D_j - D_k)^4 \right] \\ &\quad + \frac{1}{n^2} \sum_{j,k=1}^{n} \left[1 - \frac{1}{4a}(\Delta_j - \Delta_k)^2 + \frac{1}{32a^2}(\Delta_j - \Delta_k)^4 \right] \\ &\quad - \frac{2}{mn} \sum_{j=1}^{m} \sum_{k=1}^{n} \left[1 - \frac{1}{4a}(D_j - \Delta_k)^2 + \frac{1}{32a^2}(D_j - \Delta_k)^4 \right] + O\left(\frac{1}{a^3}\right) \end{aligned}$$

as $a \to \infty$. Since $\sum_{j=1}^{m} D_j = md$ and $\sum_{k=1}^{n} \Delta_k = nd$, the result follows by tedious but straightforward algebra. ■

Notice that the right-hand side of (2.8) is an estimator of $(E\|\tilde{X}\|^4 - E\|\tilde{Y}\|^4)^2$. Thus, for large a, $U_{m,n,a}$ is essentially a measure of discrepancy between the fourth moments of the norm of the radial part of the underlying standardized distributions.

3 Asymptotic distribution theory

In this section, we study the limit distribution of $U_{m,n,a}$ under H_0. Since $U_{m,n,a}$ is affine invariant, we assume without loss of generality that $X \sim Y$, and that X has a spherically symmetric distribution satisfying $EX = 0$ and $EXX' = I_d$, where I_d is the unit matrix of order d.

A convenient setting for asymptotics is the separable Hilbert space L^2 of measurable real-valued functions on $\mathbb{R}$ that are square-integrable with respect to the measure $\exp(-at^2)dt$. The norm in L^2 will be denoted by

$$\|u\|_{L^2} \; = \; \left(\int_{-\infty}^{\infty} u^2(t)\exp(-at^2)\;dt\right)^{1/2}.$$

The notation $\xrightarrow{\mathcal{D}}$ means weak convergence of random elements of L^2 and random variables, and $O_P(1)$ stands for a sequence of random variables that is bounded in probability. Likewise, $o_P(1)$ denotes a sequence of random variables that converges to 0 in probability. The first result of this section is as follows.

Theorem 3.1 *Let X have a spherically symmetric distribution satisfying $EXX' = I_d$ and $E\|X\|^4 < \infty$, and put*

$$\text{(3.1)} \qquad \varphi^*(t) \; = \; E\left[\cos(t\|X\|^2) + \sin(t\|X\|^2)\right],$$

$$\rho(t) \; = \; E\left[\|X\|^2\left(\sin(t\|X\|^2) - \cos(t\|X\|^2)\right)\right],$$

$$\text{(3.2)} \qquad g(t,x) \; = \; \cos(t\|x\|^2) + \sin(t\|x\|^2) - \varphi^*(t) + t\rho(t)\left(\frac{1}{d}\|x\|^2 - 1\right),$$

$t \in \mathbb{R}$. If $Y \sim X$, there exists a centered Gaussian process $\mathcal{W}(\cdot)$ on L^2 having covariance kernel

$$\text{(3.3)} \qquad K(s,t) \; = \; E[g(s,X)g(t,X)]$$

such that, as $m,n \to \infty$,

$$\text{(3.4)} \qquad \frac{mn}{m+n}\,U_{m,n,a} \; \xrightarrow{\mathcal{D}} \; \int_{-\infty}^{\infty} \mathcal{W}^2(t)\exp(-at^2)\,dt.$$

Proof. Put

$$\begin{aligned} \mathcal{U}_m(t) &= \sqrt{m}\,(\varphi_m^*(t) - \varphi^*(t))\,, \\ \mathcal{V}_n(t) &= \sqrt{n}\,(\psi_n^*(t) - \varphi^*(t))\,, \end{aligned}$$

where φ_m^*, ψ_n^* and φ^* are defined in (2.5), (2.6) and (3.1), respectively. From (2.4), we then have

$$\frac{mn}{m+n}\,U_{m,n,a} = \int_{-\infty}^{\infty} \left(\sqrt{\frac{n}{m+n}}\,\mathcal{U}_m(t) - \sqrt{\frac{m}{m+n}}\,\mathcal{V}_n(t) \right)^2 \exp(-at^2)dt.$$

We will prove

$$\mathcal{U}_m(\cdot) \xrightarrow{\mathcal{D}} \mathcal{U}(\cdot) \ \text{ as } m \to \infty \tag{3.5}$$

and

$$\mathcal{V}_n(\cdot) \xrightarrow{\mathcal{D}} \mathcal{V}(\cdot) \ \text{ as } n \to \infty \tag{3.6}$$

in L^2, where $\mathcal{U}(\cdot)$ and $\mathcal{V}(\cdot)$ are independent centered Gaussian processes on L^2 having covariance kernel $K(s,t)$. Since, for any choice of $\kappa, \lambda \in \mathbb{R}$ satisfying $\kappa^2 + \lambda^2 = 1$, the process $\kappa\,\mathcal{U}(\cdot) + \lambda\,\mathcal{V}(\cdot)$ is centered Gaussian with covariance kernel $K(s,t)$, it follows from (3.5) and (3.6) that

$$\sqrt{\frac{n}{m+n}}\,\mathcal{U}_m(\cdot) \ - \ \sqrt{\frac{m}{m+n}}\,\mathcal{V}_n(\cdot) \xrightarrow{\mathcal{D}} \mathcal{W}(\cdot),$$

where $\mathcal{W}(\cdot)$ has the properties stated in Theorem 3.1. Assertion (3.4) is then a consequence of the continuous mapping theorem.

Clearly, given their existence, the processes $\mathcal{U}(\cdot)$ and $\mathcal{V}(\cdot)$ are independent because the two samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ have this property. Moreover, since (3.5) and (3.6) are equivalent, only (3.5) needs to be proved.

To show (3.5), notice that $\mathcal{U}_m(t)$ is a sum of functions of the random variables

$$D_j = \|\tilde{X}_j\|^2 = (X_j - \bar{X}_m)' S_m^{-1} (X_j - \bar{X}_m) \qquad (1 \le j \le m),$$

which are not independent. We will decompose $\mathcal{U}_m(t)$ according to

$$\mathcal{U}_m(t) = \mathcal{U}_m^*(t) + \mathcal{R}_m(t), \tag{3.7}$$

where

$$\mathcal{U}_m^*(t) \;=\; \frac{1}{\sqrt{m}}\sum_{j=1}^{m} g(t, X_j),$$

$\|\mathcal{R}_m\|_{L^2} = o_P(1)$ as $m \to \infty$, and $g(t,x)$ is defined in (3.2). Since $\sqrt{m}\,\mathcal{U}_m^*(\cdot)$ is a sum of i.i.d. centered L^2-valued random elements, a standard Hilbert space central limit theorem yields $\mathcal{U}_m^*(\cdot) \xrightarrow{\mathcal{D}} \mathcal{U}(\cdot)$, where $\mathcal{U}(\cdot)$ has the properties stated above. Since $\|\mathcal{R}_m\|_{L^2} = o_P(1)$, (3.5) then follows from (3.7) and Slutzky's lemma.

To prove (3.7), start with

$$D_j \;=\; \|X_j\|^2 + \varepsilon_j,$$

where

$$\varepsilon_j \;=\; X_j'(S_m^{-1} - I_d)X_j - 2X_j' S_m^{-1}\bar{X}_m + \bar{X}_m' S_m^{-1}\bar{X}_m. \tag{3.8}$$

A Taylor expansion gives

$$\cos(tD_j) \;=\; \cos(t\|X_j\|^2) - t\,\varepsilon_j \sin(t\|X_j\|^2) + \frac{1}{2}\,t^2\,\varepsilon_j^2\,\xi_j, \tag{3.9}$$

$$\sin(tD_j) \;=\; \sin(t\|X_j\|^2) + t\,\varepsilon_j \cos(t\|X_j\|^2) + \frac{1}{2}t^2\,\varepsilon_j^2\,\eta_j, \tag{3.10}$$

where $|\xi_j| \le 1$ and $|\eta_j| \le 1$. We first assert

$$\frac{1}{\sqrt{m}}\left|\sum_{j=1}^{m} \varepsilon_j^2(\xi_j + \eta_j)\right| \;=\; o_P(1), \tag{3.11}$$

thus showing that the contribution of the rightmost terms in (3.9) and (3.10) is asymptotically negligible. To this end, notice that

$$\begin{aligned}
\frac{1}{\sqrt{m}}\left|\sum_{j=1}^{m} \varepsilon_j^2(\xi_j + \eta_j)\right| \;&\le\; \frac{2}{\sqrt{m}}\sum_{j=1}^{m}\varepsilon_j^2 \\
&\le\; \frac{24}{\sqrt{m}}\sum_{j=1}^{m}\left(X_j' S_m^{-1}\bar{X}_m\right)^2 + \frac{6}{\sqrt{m}}\sum_{j=1}^{m}\left(X_j'(S_m^{-1} - I_d)X_j\right)^2 \\
&\qquad + 6\sqrt{m}\left(\bar{X}_m' S_m^{-1}\bar{X}_m\right)^2.
\end{aligned}$$

Now, the last term is $o_P(1)$ since $S_m^{-1} = O_P(1)$ and $\sqrt{m}\bar{X}_m = O_P(1)$. Using $tr(AB) = tr(BA)$, where AB and BA are square matrices and $tr(\cdot)$ denotes trace, we obtain

$$\begin{aligned}\frac{1}{\sqrt{m}}\sum_{j=1}^m \left(X_j' S_m^{-1}\bar{X}_m\right)^2 &= \frac{1}{\sqrt{m}}\sum_{j=1}^m tr\left(X_j' S_m^{-1}\bar{X}_m \bar{X}_m' S_m^{-1} X_j\right) \\ &= \frac{1}{\sqrt{m}}\, tr\left(S_m^{-1}\sqrt{m}\bar{X}_m\sqrt{m}\bar{X}_m' S_m^{-1}\frac{1}{m}\sum_{j=1}^m X_j X_j'\right),\end{aligned}$$

which is $o_P(1)$ since $m^{-1}\sum_{j=1}^m X_j X_j' = O_P(1)$. Finally,

$$\begin{aligned}\frac{1}{\sqrt{m}}\sum_{j=1}^m \left(X_j'(S_m^{-1} - I_d)X_j\right)^2 &\le \frac{1}{\sqrt{m}}\sum_{j=1}^m \|X_j\|^2 \|(S_m^{-1} - I_d)X_j\|^2 \\ &= \frac{1}{\sqrt{m}}\, tr\left[\left(\sqrt{m}(S_m^{-1} - I_d)\right)^2 \cdot \frac{1}{m}\sum_{j=1}^m X_j X_j' \|X_j\|^2\right].\end{aligned}$$

Since, in view of $E\|X\|^4 < \infty$, both factors within squared brackets are $O_P(1)$, (3.11) is proved.

We next approximate $m^{-1/2}\sum_{j=1}^m \varepsilon_j \sin(t\|X_j\|^2)$, up to terms that are asymptotically negligible, by a sum of i.i.d. random variables. To this end, notice that

$$\begin{aligned}\frac{1}{\sqrt{m}}\sum_{j=1}^m \varepsilon_j \sin(t\|X_j\|^2) &= \frac{1}{\sqrt{m}}\sum_{j=1}^m \left[X_j'(S_m^{-1} - I_d)X_j\right]\sin(t\|X_j\|^2) \\ &\quad -\frac{2}{\sqrt{m}}\sum_{j=1}^m X_j' S_m^{-1}\bar{X}_m \sin(t\|X_j\|^2) \\ &\quad +\frac{1}{\sqrt{m}}\sum_{j=1}^m \bar{X}_m' S_m^{-1}\bar{X}_m \sin(t\|X_j\|^2) \\ &= Z_{m,1}(t) \;-\; 2Z_{m,2}(t) \;+\; Z_{m,3}(t) \quad \text{(say)}.\end{aligned}$$

Putting $B_m = \sqrt{m}(S_m^{-1} - I_d)$, $A(t) = E[XX'\sin(t\|X\|^2)]$ and

$$\Delta_m(t) = \frac{1}{m}\sum_{j=1}^m X_j X_j' \sin(t\|X_j\|^2) \;-\; A(t),$$

we have

$$Z_{m,1}(t) = tr(B_m A(t)) \;+\; tr(B_m \Delta_m(t)). \tag{3.12}$$

Letting $B_m = (B_{m,k,l})_{1\le k,l\le d}$ and $\Delta_m(t) = (\Delta_{m,k,l}(t))_{1\le k,l\le d}$, and using the Cauchy-Schwarz inequality, it follows that

$$\left|tr(B_m\Delta_m(t))\right| \;\le\; \left(\sum_{k,l=1}^d B_{m,k,l}^2\right)^{1/2} \cdot \left(\sum_{k,l=1}^d \Delta_{m,k,l}^2(t)\right)^{1/2} \tag{3.13}$$

and thus

$$\int_{-\infty}^{\infty} (tr(B_m\Delta_m(t)))^2 \exp(-at^2)dt \;\leq\; O_P(1) \int_{-\infty}^{\infty} \sum_{k,l=1}^{d} \Delta_{m,k,l}^2(t) \exp(-at^2)dt,$$

since the first factor on the right-hand side of (3.13) is $O_P(1)$. Use Fubini's theorem to conclude that the expectation of the last integral converges to zero as $m \to \infty$. Consequently, $\|tr(B_m\Delta_m(\cdot)\|_{L^2} = o(1)$ which shows that the second term on the right-hand side of (3.12) is asymptotically negligible. In view of

$$\sqrt{m}(S_m^{-1} - I_d) \;=\; -\frac{1}{\sqrt{m}} \sum_{j=1}^{m} (X_j X_j' - I_d) \;+\; O_P(m^{-1/2}),$$

it follows readily that

$$Z_{m,1}(t) = -\frac{1}{\sqrt{m}} \sum_{j=1}^{m} tr\left((X_j X_j' - I_d)A(t)\right) \;+\; R_{m,1}(t)$$

for some L^2-valued random element $R_{m,1}(\cdot)$ satisfying $\|R_{m,1}(\cdot)\|_{L^2} = o(1)$. Using Theorem 3.3 of Fang, Kotz and Ng (1990) together with the decomposition $X \sim U \cdot \|X\|$, where U and $\|X\|$ are independent, and the distributions of U is uniform over the surface of the unit d-sphere, we obtain $A(t) = d^{-1}E[\|X\|^2 \sin(t\|X\|^2)]\, I_d$ and thus

$$Z_{m,1}(t) = -\frac{1}{\sqrt{m}} \sum_{j=1}^{m} E\left[\|X\|^2 \sin(t\|X\|^2)\right] \left(d^{-1}\|X_j\|^2 - 1\right) + o_P(1).$$

As for $Z_{m,2}(t)$, we have

$$Z_{m,2}(t) \;=\; \left(\frac{1}{m} \sum_{j=1}^{m} X_j \sin(t\|X_j\|^2)\right)' \sqrt{m}\left(S_m^{-1} - I_d\right) \bar{X}_m$$
$$+ \left(\sqrt{m}\bar{X}_m\right)' \frac{1}{m} \sum_{j=1}^{m} X_j \sin(t\|X_j\|^2).$$

Use the fact that $E[X \sin(t\|X\|^2)] = 0$ by spherical symmetry and thus $m^{-1}\sum_{j=1}^{m} X_j$ $\sin(t\|X_j\|^2) = o_P(1)$ to conclude $\|Z_{m,2}(\cdot)\|_{L^2} = o(1)$. Since also $\|Z_{m,3}(\cdot)\|_{L^2} = o(1)$, it follows that

$$\frac{1}{\sqrt{m}} \sum_{j=1}^{m} \varepsilon_j \sin(t\|X_j\|^2) \;=\; -\frac{1}{\sqrt{m}} \sum_{j=1}^{m} E\left[\|X\|^2 \sin(t\|X\|^2)\right]\left(\frac{\|X_j\|^2}{d} - 1\right) + o_P(1).$$

Likewise, we have

$$\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\varepsilon_j\cos(t\|X_j\|^2) \;=\; -\frac{1}{\sqrt{m}}\sum_{j=1}^{m}E\big[\|X\|^2\cos(t\|X\|^2)\big]\left(\frac{\|X_j\|^2}{d}-1\right)+o_P(1),$$

and (3.7) follows by straightforward algebra. ■

4 A resampling procedure

To perform the test based on $U_{m,n,a}$, we suggest the use of the following resampling procedure. Pool the values $D_1,\ldots,D_m,\Delta_1,\ldots,\Delta_n$ into a sample of size $N=m+n$ and draw a random sample $\mathcal{D}^*=\{D_1^*,\ldots,D_m^*,\Delta_1^*,\ldots,\Delta_n^*\}$ with replacement from the combined sample.

Independently of $\mathcal{D}^*$, generate independent random vectors $V_1,\ldots,V_N$, uniformly distributed over the surface of the unit d-sphere, and put

$$\begin{aligned} Z_{N,j} &= V_j\cdot\sqrt{D_j^*}, && j=1,\ldots,m,\\ Z_{N,m+k} &= V_{m+k}\cdot\sqrt{\Delta_k^*}, && k=1,\ldots,n. \end{aligned}$$

Note that, conditionally on $D_1,\ldots,D_m,\Delta_1,\ldots,\Delta_n$, the random vectors $Z_{N,1},\ldots,Z_{N,N}$ have the same spherically symmetric distribution P_N and distribution function F_N (say). The (conditional) distribution function of $\|Z_{N,1}\|^2$ is the empirical distribution function of $D_1,\ldots,D_m$, $\Delta_1,\ldots,\Delta_n$. Writing again $\tilde{Z}_{N,j}$ for the standardization of $Z_{N,j}$, let $D_{N,j}=\|\tilde{Z}_{N,j}\|^2$, $\Delta_{N,k}=\|\tilde{Z}_{N,m+k}\|^2$ $(1\le j\le m, 1\le k\le n)$.

Putting $U_{m,n,a}^R=U_{m,n,a}(D_{N,1},\ldots,D_{N,m},\Delta_{N,1},\ldots,\Delta_{N,n})$, we have

$$\frac{mn}{m+n}\,U_{m,n,a}^R=\int_{-\infty}^{\infty}\left(\sqrt{\frac{n}{m+n}}\,\mathcal{U}_{N,m}(t)-\sqrt{\frac{m}{m+n}}\,\mathcal{V}_{N,n}(t)\right)^2\exp(-at^2)dt,$$

where

$$\begin{aligned} \mathcal{U}_{N,m}(t) &= \sqrt{m}\left(\varphi_{N,m}(t)-\varphi_N(t)\right), \quad \varphi_{N,m}(t)=\frac{1}{m}\sum_{j=1}^{m}\left[\cos(tD_{N,j})+\sin(tD_{N,j})\right],\\ \mathcal{V}_{N,n}(t) &= \sqrt{n}\left(\psi_{N,n}(t)-\varphi_N(t)\right), \quad \psi_{N,n}(t)=\frac{1}{n}\sum_{k=1}^{n}\left[\cos(t\Delta_{N,k})+\sin(t\Delta_{N,k})\right], \end{aligned}$$

and

$$(4.1) \qquad \varphi_N(t) = \int_{\mathbb{R}^d} \left(\cos(t\|z\|^2) + \sin(t\|z\|^2)\right) \, dF_N(z).$$

To prove the conditional convergence in distribution of the resampling process $\mathcal{U}_{N,m}$ to the Gaussian process $\mathcal{W}$ figuring in Theorem 3.1, we use the following Hilbert space Central Limit Theorem of Kundu et al. (Kundu, Majumdar, and Mukherjee (2000), Theorem 1.1). Therein, $\mathcal{H}$ denotes a real separable infinite-dimensional Hilbert space.

Lemma 4.1 *Let $\{e_k : k \geq 0\}$ be an orthonormal basis of $\mathcal{H}$. For each $N \geq 1$, let $W_{N1}, W_{N2}, \ldots, W_{NN}$ be a finite sequence of independent $\mathcal{H}$-valued random elements with zero means and finite second moments, and put $W_N = \sum_{j=1}^N W_{Nj}$. Let C_N be the covariance operator of W_N. Assume that the following conditions hold:*

a) $\lim_{N\to\infty} \langle C_N e_k, e_l\rangle = a_{kl}$ *(say) exists for all $k \geq 0$ and $l \geq 0$.*

b) $\lim_{N\to\infty} \sum_{k=0}^\infty \langle C_N e_k, e_k\rangle = \sum_{k=0}^\infty a_{kk} < \infty$.

c) $\lim_{N\to\infty} L_N(\varepsilon, e_k) = 0$ *for every $\varepsilon > 0$ and every $k \geq 0$, where, for $b \in \mathcal{H}$,* $L_N(\varepsilon, b) = \sum_{j=1}^N E\left(\langle W_{Nj}, b\rangle^2 \, \mathbf{1}\{|\langle W_{Nj}, b\rangle| > \varepsilon\}\right)$.

Then $W_N \xrightarrow{\mathcal{D}} \mathcal{N}(0, C)$ in $\mathcal{H}$, where $\mathcal{N}(0, C)$ is a centered Gaussian random element of $\mathcal{H}$ with covariance operator C characterized by $\langle Ch, e_l\rangle = \sum_{j=0}^\infty \langle h, e_j\rangle a_{jl}$, for every $l \geq 0$.

The main result of this section is as follows.

Theorem 4.2 *For almost all sample sequences $X_1(\omega), X_2(\omega), \ldots, Y_1(\omega), Y_2(\omega), \ldots$, we have under H_0*

$$\frac{mn}{m+n} U^R_{m,n,a} \xrightarrow{\mathcal{D}} \int_{-\infty}^{\infty} \mathcal{W}^2(t) \exp(-at^2) \, dt \quad (m, n \to \infty)$$

in L^2, where we use the notation and assume the conditions of Theorem 3.1.

Proof. Define

$$\rho_N(t) = E\left[\|Z_{N,1}\|^2 \left(\sin(t\|Z_{N,1}\|^2) - \cos(t\|Z_{N,1}\|^2)\right)\right],$$

$$g_N(t, x) = \cos(t\|x\|^2) + \sin(t\|x\|^2) - \varphi_N(t) + t\rho_N(t)\left(\frac{1}{d}\|x\|^2 - 1\right),$$

and

$$\mathcal{U}^*_{N,m}(t) = \frac{1}{\sqrt{m}} \sum_{j=1}^{m} g_N(t, Z_{N,j}).$$

We show

$$\mathcal{U}_{N,m}(t) = \mathcal{U}^*_{N,m}(t) \; + \; \mathcal{R}_{N,m}(t), \tag{4.2}$$

where $\|\mathcal{R}_{N,m}\|_{L^2} = o_{P_N}(1)$ as $N \to \infty$ and $\mathcal{U}^*_{N,m}(\cdot) \xrightarrow{\mathcal{D}} \mathcal{U}(\cdot)$. Then, a reasoning similar as in the proof of Theorem 3.1 yields

$$\mathcal{U}_{N,m}(\cdot) \xrightarrow{\mathcal{D}} \mathcal{U}(\cdot) \quad (m \to \infty), \qquad \mathcal{V}_{N,n}(\cdot) \xrightarrow{\mathcal{D}} \mathcal{V}(\cdot) \quad (n \to \infty),$$

and the assertion of the theorem.

To show (4.2), let D_1 be a countable dense subset of $\mathbb{R}$, and let Ω_1 be the set of all $\omega \in \Omega$ for which $\bar{X}_m \to 0$, $S_m \to I_d$, $m^{-1}\sum_{j=1}^m \|X_j\|^4 \to E\|X\|^4$, $m^{-\frac{1}{2}} \max_{1\le j\le m} \|Z_{N,j}\|^2 \to 0$, $m^{-1}\sum_{j=1}^m \cos(t\|X_j\|^2) \to E[\cos(t\|X\|^2)]$ and $m^{-1}\sum_{j=1}^m \sin(t\|X_j\|^2) \to E[\sin(t\|X\|^2)]$ as $m \to \infty$ for each $t \in D_1$. Clearly, Ω_1 has measure one, and it is readily seen that, for each fixed $\omega \in \Omega_1$,

$$\frac{1}{m}\sum_{j=1}^m \cos(t\|X_j\|^2) \to E[\cos(t\|X\|^2)], \quad \frac{1}{m}\sum_{j=1}^m \sin(t\|X_j\|^2) \to E[\sin(t\|X\|^2)]$$

for each $t \in \mathbb{R}$. Likewise, let Ω_2 be the set (of measure one) of all $\omega \in \Omega$ for which $\bar{Y}_n \to 0$, $T_n \to I_d$, $n^{-1}\sum_{k=1}^n \|Y_k\|^4 \to E\|Y\|^4$, $n^{-\frac{1}{2}} \max_{1\le k\le n} \|Z_{N,k}\|^2 \to 0$, $n^{-1}\sum_{k=1}^n \cos(t\|Y_k\|^2) \to E[\cos(t\|Y\|^2)]$ and $n^{-1}\sum_{k=1}^n \sin(t\|Y_k\|^2) \to E[\sin(t\|Y\|^2)]$ for each $t \in \mathbb{R}$. Putting $\Omega_0 = \Omega_1 \cap \Omega_2$, the following reasoning will be done for a fixed $\omega \in \Omega_0$.

As a first step, we prove that $\lim_{N\to\infty} K_N(s,t) = K(s,t)$ pointwise on $\mathbb{R}^2$ for $\omega \in \Omega_0$, where $K_N(s,t) = E[g_N(s, Z_{N,1})\, g_N(t, Z_{N,1})]$. Using the Taylor expansion

$$\cos(tD_j) = \cos(t\|X_j\|^2) - t\,\varepsilon_j \xi_j, \qquad \sin(tD_j) = \sin(t\|X_j\|^2) + t\,\varepsilon_j \eta_j,$$

where $|\xi_j| \le 1$, $|\eta_j| \le 1$ and ε_j is defined in (3.8), we obtain

$$\begin{aligned}
\varphi_N(t) &= E\left[\cos(t\|Z_{N,1}\|^2) + \sin(t\|Z_{N,1}\|^2)\right] \\
&= E\left[\cos(tD_1^*) + \sin(tD_1^*)\right] \\
&= N^{-1}\left(\sum_{j=1}^{m}(\cos(tD_j) + \sin(tD_j)) + \sum_{k=1}^{n}(\cos(t\Delta_k) + \sin(t\Delta_k))\right) \\
&\longrightarrow E[\cos(t\|X\|^2) + \sin(t\|X\|^2)] = \varphi^*(t)
\end{aligned}$$

as $N \to \infty$ for $\omega \in \Omega_0$, where φ_N and φ^* are defined in (3.1) and (4.1), respectively. Using similar arguments, we obtain $\lim_{N\to\infty} \rho_N(t) = \rho(t)$ and, finally, $\lim_{N\to\infty} K_N(s,t) = K(s,t)$ for $\omega \in \Omega_0$.

Next, we verify conditions a) - c) of Lemma 4.1 for $W_{N1}, \ldots, W_{NN}$, where $W_{Nj}(t) = g_N(t, Z_{N,j})/\sqrt{m}, 1 \le j \le m$ and $W_{Nj}(t) = 0, m+1 \le j \le N$. To this end, let C_N be the covariance operator of $W_N = \sum_{j=1}^{N} W_{Nj}$ $(= \mathcal{U}_{N,m}^*)$ with kernel $E\left[W_N(s)W_N(t)\right] = K_N(s,t)$.

As complete orthonormal set $\{e_k\}$ in $\mathcal{L}_2$, one can choose products of univariate Hermite polynomials (see, e.g., Rayner and Best (1989), p. 100). Since, for $\omega \in \Omega_0$ and sufficiently large N, $|K_N(s,t)| \le c_1|st|$ for some constant c_1, dominated convergence yields

$$\begin{aligned}
\lim_{N\to\infty} \langle C_N e_k, e_l \rangle &= \lim_{N\to\infty} \int\int K_N(s,t)e_k(s)e_l(t)P_a(ds)P_a(dt) \\
&= \int\int K(s,t)e_k(s)e_l(t)P_a(ds)P_a(dt) \\
&= \langle Ce_k, e_l \rangle,
\end{aligned}$$

where $P_a(dt)$ is shorthand for $\exp(-a\|t\|^2)dt$, and C is the covariance operator of $\mathcal{W}$. Here and in what follows, an unspecified integral denotes integration over the whole space $\mathbb{R}$. Setting $a_{kl} = \langle Ce_k, e_l \rangle$, this proves condition a) of Lemma 4.1.

To verify condition b) of Lemma 4.1, use monotone convergence, Parseval's equality and dominated convergence to show

$$\begin{aligned}
\lim_{N\to\infty}\sum_{k=0}^{\infty}\langle C_N e_k, e_k\rangle &= \lim_{N\to\infty}\sum_{k=0}^{\infty}E\langle e_k, W_N\rangle^2 \\
&= \lim_{N\to\infty}E\|W_N\|^2_{\mathcal{L}^2} \\
&= \int \lim_{N\to\infty}K_N(t,t)P_a(dt) \\
&= \int K(t,t)P_a(dt) \\
&= E\|\mathcal{W}\|^2_{\mathcal{L}^2} \\
&= \sum_{k=0}^{\infty}a_{kk} < \infty.
\end{aligned}$$

To prove condition c) of Lemma 4.1, notice that

$$\begin{aligned}
|\langle W_{Nj}, e_k\rangle| &= m^{-\frac{1}{2}}\left|\int g_N(t, Z_{N,j})e_k(t)P_a(dt)\right| \\
&\leq m^{-\frac{1}{2}}\int |g_N(t, Z_{N,j})e_k(t)|\, P_a(dt) \\
&\leq m^{-\frac{1}{2}}\left(\int |g_N(t, Z_{N,j})|^2\, P_a(dt)\right)^{1/2}\|e_k\|_{\mathcal{L}^2}.
\end{aligned}$$

Using $|\rho_N(t)| \leq 2E\|Z_{N,j}\|^2 = 2(\sum_{j=1}^m D_j + \sum_{k=1}^n \Delta_k)/N \to 2E\|X\|^2$ as $m, n \to \infty$ for $\omega \in \Omega_0$, and $|g_N(t, Z_{N,j})| \leq 4 + |t|\,|\rho_N(t)|\max_{1\leq j\leq m}\|Z_{N,j}\|^2$, we obtain

$$|\langle W_{Nj}, e_k\rangle| \leq m^{-\frac{1}{2}}\left(c_2 + c_3 E\|Z_{N,j}\|^2 \max_{1\leq j\leq m}\|Z_{N,j}\|^2\right)$$

for some positive constants c_2, c_3, which converges to zero for $\omega \in \Omega_0$. Hence

$$E\left(\langle W_{Nj}, e_k\rangle^2\, \mathbf{1}\{|\langle W_{Nj}, e_k\rangle| > \varepsilon\}\right) = 0$$

for sufficiently large N, and thus $\lim_{N\to\infty} L_N(\varepsilon, e_k) = 0$. By Lemma 4.1, $W_N \Rightarrow \mathcal{N}(0, C)$ in $\mathcal{L}^2$.

Finally, $\|\mathcal{R}_m\|_{L^2} = o_{P_N}(1)$ as $N \to \infty$ can be proved as in Theorem 3.1. ■

5 Simulation results

To assess the actual level of the tests for affine equivalence based on $U_{m,n,a}$, a simulation study was performed for sample sizes $N = 50$ ($m = n = 25$) and $N = 100$ ($m = n = 50$)

and dimensions $d = 2$ and $d = 5$. Besides $U_{m,n,0.5}$, $U_{m,n,1}$, $U_{m,n,2}$ and $U_{m,n,5}$, we included the limit statistic of Proposition 2.2. We used the following distributions:

- MN_1: the d-variate standard Normal distribution $\mathcal{N}(0, I_d)$
- MN_2: the d-variate Normal distribution $\mathcal{N}(0, \Sigma_d^1)$, where $\Sigma_2^1 = diag(2, 4)$ and $\Sigma_5^1 = diag(2, 2, 2, 4, 4)$;
- MN_3: a d-variate normal distribution with mean zero, unit variances and equal correlation $\rho = 0.5$ between components; the covariance matrix is denoted by Σ_d^2;
- MT_1: the multivariate t distribution with 10 degrees of freedom $t_d(10; 0, I_d)$, generated as $U/\sqrt{V}$, where U and V are independent, and $U \sim \mathcal{N}(0, I_d)$, $V \sim \chi^2_{10}/10$;
- MT_2, MT_3: multivariate t distribution $t_d(10; 0, \Sigma_d^1)$ and $t_d(5; 0, \Sigma_d^2)$;
- MT_4: the multivariate t distribution with 5 degrees of freedom $t_d(5; 0, I_d)$
- MP_1: the multivariate Pearson Type II distribution with shape parameter -1/2
- MP_2: the multivariate Pearson Type II distribution with shape parameter 0
- MP_3: the multivariate Pearson Type II distribution with shape parameter 1

Using these distributions, we simulated data from the following H_0 cases:

- $X \sim MN_1$ and $Y \sim MN_k$ $(k = 1, 2, 3)$;
- $X \sim MT_1$ and $Y \sim MT_k$ $(k = 1, 2, 3)$;
- $X \sim MT_4$ and $Y \sim MT_4$;
- $X \sim MP_k$ and $Y \sim MP_k$ $(k = 1, 2, 3)$.

For each fixed combination of N, d and underlying distributions as given above, the following procedure was replicated 5 000 times:

1. generate random samples $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$

2. compute D_j $(1 \le j \le m), \Delta_k$ $(1 \le k \le n)$ and $U_{m,n,a}(D_1, \ldots, D_m, \Delta_1, \ldots, \Delta_n)$

3. draw 500 samples $D_1^*, \ldots, D_m^*, \Delta_1^*, \ldots, \Delta_n^*$ with replacement from the pooled sample $D_1, \ldots, D_m, \Delta_1, \ldots, \Delta_n$; for each sample, generate random vectors $V_1, \ldots, V_N$, uniformly distributed over the surface of the unit d-sphere, and put

$$\begin{aligned} Z_{N,j} &= V_j \cdot \sqrt{D_j^*}, \quad j = 1, \ldots, m, \\ Z_{N,m+k} &= V_{m+k} \cdot \sqrt{\Delta_k^*}, \quad k = 1, \ldots, n \end{aligned}$$

4. calculate the corresponding 500 realizations $U_{m,n,a}^R(l),\ 1 \le l \le 500$, (say) of the resampling statistic $U_{m,n,a}^R$

5. reject H_0 if $U_{m,n,a}$, computed on $D_1, \ldots, D_m, \Delta_1, \ldots, \Delta_n$, exceeds the empirical 95%-quantile of $U_{m,n,a}^R(l),\ 1 \le l \le 500$.

Table 1 shows the percentage of the number of rejections of H_0 for sample sizes $n = m = 25$ and $n = m = 50$ and dimensions 2 and 5. Table 2 shows the percentage of the number of rejections of some H_0 cases for large sample size ($n = m = 100$ and $n = m = 200$) and dimension 2.

Notice that, for $a = 1$ and $a = 2$, the observed level is fairly close to the nominal level 5% even for samples of size $n = m = 25$; for the cases $a = 5.0$ and $a = \infty$, however, the actual level is sometimes far below or above the nominal level. Particularly for very long tailed distributions, the observed level of significance for $a = \infty$ seem to approach its nominal value 5% only very slowly with increasing sample size, as the simulation results in Table 2 indicate.

To assess the power of the different tests, we simulated data from the following distributions:

- MN_1 against MT_k $(k = 1, 4)$;
- MN_1 against MP_k $(k = 1, 2, 3)$;
- MT_1 against MT_2;

			$a = 0.5$	$a = 1.0$	$a = 2.0$	$a = 5.0$	$a = \infty$
$X \sim \mathcal{MN}_1$, $Y \sim \mathcal{MN}_1$	$n = m = 25$	$d = 2$	4.4	5.2	5.8	6.1	5.9
		$d = 5$	5.4	5.8	6.3	7.5	9.5
	$n = m = 50$	$d = 2$	4.4	4.7	5.3	5.4	5.0
		$d = 5$	5.4	5.0	6.0	6.1	7.8
$X \sim \mathcal{MN}_1$, $Y \sim \mathcal{MN}_2$	$n = m = 25$	$d = 2$	4.9	5.3	5.3	6.5	6.7
		$d = 5$	5.6	5.4	6.4	7.4	9.1
	$n = m = 50$	$d = 2$	4.0	5.3	4.9	5.7	5.5
		$d = 5$	4.5	5.2	6.3	6.6	8.3
$X \sim \mathcal{MN}_1$, $Y \sim \mathcal{MN}_3$	$n = m = 25$	$d = 2$	5.2	5.4	5.6	6.3	6.5
		$d = 5$	5.7	6.1	6.3	7.2	8.5
	$n = m = 50$	$d = 2$	4.6	5.1	5.2	5.1	5.1
		$d = 5$	5.4	5.6	5.8	6.7	7.9
$X \sim \mathcal{MT}_1$, $Y \sim \mathcal{MT}_1$	$n = m = 25$	$d = 2$	5.9	6.1	6.9	8.1	12.2
		$d = 5$	5.6	6.0	7.3	8.8	16.8
	$n = m = 50$	$d = 2$	5.2	5.6	5.8	7.1	11.2
		$d = 5$	5.8	5.7	6.0	6.9	18.1
$X \sim \mathcal{MT}_1$, $Y \sim \mathcal{MT}_2$	$n = m = 25$	$d = 2$	5.5	5.7	6.7	8.0	12.4
		$d = 5$	5.1	6.9	7.6	9.4	17.0
	$n = m = 50$	$d = 2$	4.6	5.8	6.0	7.1	11.4
		$d = 5$	5.4	5.3	6.5	7.6	18.7
$X \sim \mathcal{MT}_1$, $Y \sim \mathcal{MT}_3$	$n = m = 25$	$d = 2$	5.8	6.2	6.9	7.9	12.0
		$d = 5$	5.9	6.3	6.6	9.8	17.3
	$n = m = 50$	$d = 2$	5.2	5.7	6.0	7.2	11.6
		$d = 5$	5.3	5.6	5.6	7.2	19.1
$X \sim \mathcal{MT}_4$, $Y \sim \mathcal{MT}_4$	$n = m = 25$	$d = 2$	6.4	7.4	8.1	10.2	21.3
		$d = 5$	5.8	7.0	8.0	11.1	25.3
	$n = m = 50$	$d = 2$	6.4	6.9	7.4	9.0	22.5
		$d = 5$	6.3	7.0	7.1	8.4	31.2
$X \sim \mathcal{MP}_1$, $Y \sim \mathcal{MP}_1$	$n = m = 25$	$d = 2$	7.6	6.3	4.4	3.1	1.4
		$d = 5$	5.2	4.7	3.3	2.9	1.5
	$n = m = 50$	$d = 2$	9.2	7.1	5.9	4.4	2.5
		$d = 5$	7.0	5.4	4.6	3.5	1.7
$X \sim \mathcal{MP}_2$, $Y \sim \mathcal{MP}_2$	$n = m = 25$	$d = 2$	5.4	5.6	4.5	2.9	1.3
		$d = 5$	5.8	4.9	4.3	3.3	2.3
	$n = m = 50$	$d = 2$	6.5	5.4	4.6	4.2	2.8
		$d = 5$	5.8	5.4	5.2	3.9	2.5
$X \sim \mathcal{MP}_3$, $Y \sim \mathcal{MP}_3$	$n = m = 25$	$d = 2$	4.7	4.5	4.4	3.9	1.9
		$d = 5$	5.0	5.2	4.5	4.1	3.2
	$n = m = 50$	$d = 2$	5.0	5.6	5.4	4.3	2.9
		$d = 5$	5.4	5.3	5.5	4.8	3.0

Table 1: Estimated level for the bootstrap test (nominal level: 5%)

		$a=0.5$	$a=1.0$	$a=2.0$	$a=5.0$	$a=\infty$
$X \sim \mathcal{MN}_1$,	$n=m=100$	4.7	5.2	4.8	5.3	5.1
$Y \sim \mathcal{MN}_1$	$n=m=200$	4.9	4.9	4.9	4.8	4.9
$X \sim \mathcal{MT}_1$,	$n=m=100$	4.8	5.3	5.9	6.6	9.3
$Y \sim \mathcal{MT}_1$	$n=m=200$	5.0	5.0	5.5	6.0	7.3
$X \sim \mathcal{MT}_4$,	$n=m=100$	6.5	6.5	7.5	7.6	19.8
$Y \sim \mathcal{MT}_4$	$n=m=200$	6.2	6.4	6.2	7.1	15.3
$X \sim \mathcal{MP}_1$,	$n=m=100$	8.6	6.3	6.0	4.8	3.9
$Y \sim \mathcal{MP}_1$	$n=m=200$	6.6	6.4	5.1	4.7	4.9
$X \sim \mathcal{MP}_2$,	$n=m=100$	6.7	6.0	5.7	4.6	4.2
$Y \sim \mathcal{MP}_2$	$n=m=200$	5.6	5.9	5.4	4.7	3.9
$X \sim \mathcal{MP}_3$,	$n=m=100$	5.6	5.1	4.9	5.0	4.1
$Y \sim \mathcal{MP}_3$	$n=m=200$	5.4	5.4	5.2	4.5	4.6

Table 2: Estimated level for $d = 2$ and large sample size (nominal level: 5%)

- MT_k against MP_l $(k = 1, 4; l = 1, 2, 3)$;
- MP_k against MP_l $(k, l = 1, 2, 3; k < l)$;

Table 3 and Table 4 show the percentages of rejection of H_0. The main conclusions that can be drawn from the power study are the following:

1. In all cases, power increases with the sample size.
2. For the alternatives $\mathcal{MN}_1$ against $\mathcal{MT}_1$, $\mathcal{MN}_1$ against $\mathcal{MT}_4$ and $\mathcal{MT}_1$ against $\mathcal{MT}_4$, power is higher for $d = 5$ than for $d = 2$. In all other cases, power decreases with increasing dimension.
3. Power of $\mathcal{MN}_1$ against $\mathcal{MT}_4$ is higher than for $\mathcal{MN}_1$ against $\mathcal{MT}_1$. Power of $\mathcal{MN}_1$, $\mathcal{MT}_1$ and $\mathcal{MT}_4$ against $\mathcal{MP}_k$ increases with k.
4. In most cases, power depends heavily on the weight parameter a. Partly, power increases with increasing values a; in other cases power decreases with a. Therefore, no test is superior to the other tests in all cases.
5. If nothing is known about the alternative, the tests based on $U_{m,n,1}$ or $U_{m,n,2}$ can be recommended since they maintain their level quite closely; furthermore, they distribute their power more evenly over the range of alternatives.

			$a = 0.5$	$a = 1.0$	$a = 2.0$	$a = 5.0$	$a = \infty$
$X \sim \mathcal{MN}_1$, $Y \sim \mathcal{MT}_1$	$n = m = 25$	$d = 2$	6.2	7.0	8.0	9.4	12.9
		$d = 5$	9.2	9.9	12.2	14.7	26.4
	$n = m = 50$	$d = 2$	8.7	9.7	10.5	11.8	17.1
		$d = 5$	14.4	17.7	21.1	26.5	44.8
$X \sim \mathcal{MN}_1$, $Y \sim \mathcal{MT}_4$	$n = m = 25$	$d = 2$	11.5	14.1	15.1	18.6	27.3
		$d = 5$	17.3	21.9	27.2	34.7	54.6
	$n = m = 50$	$d = 2$	22.3	24.2	27.8	29.9	42.4
		$d = 5$	43.4	49.5	55.3	64.5	85.3
$X \sim \mathcal{MN}_1$, $Y \sim \mathcal{MP}_1$	$n = m = 25$	$d = 2$	34.0	25.6	18.0	9.9	4.4
		$d = 5$	2.4	1.9	1.6	0.9	0.1
	$n = m = 50$	$d = 2$	91.3	88.7	86.0	81.2	75.6
		$d = 5$	36.6	35.8	32.9	27.0	11.8
$X \sim \mathcal{MN}_1$, $Y \sim \mathcal{MP}_2$	$n = m = 25$	$d = 2$	70.2	65.3	60.0	48.5	32.1
		$d = 5$	14.1	13.6	11.7	8.7	3.0
	$n = m = 50$	$d = 2$	99.4	99.2	98.9	99.4	99.0
		$d = 5$	83.2	85.6	86.0	84.0	70.1
$X \sim \mathcal{MN}_1$, $Y \sim \mathcal{MP}_3$	$n = m = 25$	$d = 2$	86.8	84.4	81.8	78.2	69.9
		$d = 5$	36.0	38.9	38.4	35.3	22.3
	$n = m = 50$	$d = 2$	100.0	99.9	99.8	99.8	99.9
		$d = 5$	97.3	98.6	98.8	99.2	98.0
$X \sim \mathcal{MT}_1$, $Y \sim \mathcal{MT}_4$	$n = m = 25$	$d = 2$	7.8	8.8	9.8	12.4	20.3
		$d = 5$	8.3	9.3	12.4	15.2	31.2
	$n = m = 50$	$d = 2$	10.0	10.2	12.4	13.1	24.3
		$d = 5$	13.4	16.0	17.8	20.1	45.2

Table 3: Estimated power for the resampling test

References

Anderson, H.H., Hall, P. and Titterington, D.M. (1994). Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *J. Multiv. Anal.*, *50*, 41-54.

Eaton, M.L., and Perlman, M.D. (1973). The non–singularity of generalized sample covariance matrices. *Ann. Statist.*, *1*, 710–717.

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

Fang, K.T., Kotz, S. and Ng, K.-W. (1990). Symmetric multivariate and related distributions. *Monographs on Statistics and Applied Probablity 36, Chapman and Hall, London, New York.*

			$a=0.5$	$a=1.0$	$a=2.0$	$a=5.0$	$a=\infty$
$X \sim \mathcal{MT}_1$, $Y \sim \mathcal{MP}_1$	$n=m=25$	$d=2$	33.4	26.4	18.4	10.2	3.8
		$d=5$	2.4	1.8	1.5	0.7	0.1
	$n=m=50$	$d=2$	91.1	89.0	85.6	81.8	74.9
		$d=5$	36.0	36.7	36.0	27.4	12.6
$X \sim \mathcal{MT}_1$, $Y \sim \mathcal{MP}_2$	$n=m=25$	$d=2$	71.6	66.4	59.5	48.4	32.6
		$d=5$	14.3	13.5	12.4	8.9	3.6
	$n=m=50$	$d=2$	99.2	99.1	99.0	99.1	99.0
		$d=5$	82.7	86.2	85.3	84.5	71.2
$X \sim \mathcal{MT}_1$, $Y \sim \mathcal{MP}_3$	$n=m=25$	$d=2$	86.6	84.7	83.0	79.0	69.3
		$d=5$	35.3	38.9	37.9	36.5	22.3
	$n=m=50$	$d=2$	99.9	99.8	99.9	99.9	99.9
		$d=5$	97.5	98.5	98.8	99.0	97.8
$X \sim \mathcal{MT}_4$, $Y \sim \mathcal{MP}_1$	$n=m=25$	$d=2$	34.1	26.6	16.4	9.4	4.3
		$d=5$	2.6	2.1	1.2	0.7	0.1
	$n=m=50$	$d=2$	91.2	88.6	86.1	80.6	75.4
		$d=5$	37.1	36.3	33.8	27.3	11.6
$X \sim \mathcal{MT}_4$, $Y \sim \mathcal{MP}_2$	$n=m=25$	$d=2$	70.7	65.2	59.4	48.7	33.5
		$d=5$	14.7	13.5	12.2	8.8	2.9
	$n=m=50$	$d=2$	99.3	99.3	99.1	98.9	98.9
		$d=5$	82.2	85.3	85.9	82.6	71.0
$X \sim \mathcal{MT}_4$, $Y \sim \mathcal{MP}_3$	$n=m=25$	$d=2$	85.3	85.1	82.9	79.5	69.7
		$d=5$	35.7	37.7	39.5	34.7	21.6
	$n=m=50$	$d=2$	99.9	99.9	99.9	99.9	99.9
		$d=5$	97.6	98.5	98.6	99.1	97.5
$X \sim \mathcal{MP}_1$, $Y \sim \mathcal{MP}_2$	$n=m=25$	$d=2$	70.2	65.9	59.2	49.0	33.0
		$d=5$	14.1	14.9	11.4	8.4	2.8
	$n=m=50$	$d=2$	99.2	99.3	99.2	98.9	99.0
		$d=5$	82.6	85.5	86.6	84.4	70.8
$X \sim \mathcal{MP}_1$, $Y \sim \mathcal{MP}_3$	$n=m=25$	$d=2$	86.6	85.5	83.3	79.5	68.4
		$d=5$	36.3	37.8	39.7	35.5	19.9
	$n=m=50$	$d=2$	99.8	99.9	99.9	99.9	99.9
		$d=5$	97.4	98.6	98.9	99.3	98.0
$X \sim \mathcal{MP}_2$, $Y \sim \mathcal{MP}_3$	$n=m=25$	$d=2$	87.4	84.4	83.6	78.3	68.1
		$d=5$	35.1	37.4	39.4	36.2	21.7
	$n=m=50$	$d=2$	99.9	99.9	99.9	100.0	99.9
		$d=5$	97.6	98.0	98.7	99.2	97.7

Table 4: Estimated power for the resampling test

Henze, N., and Wagner, Th. (1997). A new approach to the BHEP tests for multivariate normality. *J. Multiv. Anal.*, *62*(1), 1–23.

Kundu, S., Majumdar, S., and Mukherjee, K. (2000). Central Limit Theorems revisited. *Statist.* & *Prob. Lett.*, *47*, 265–275.

Rayner, J., and Best, D. (1989). *Smooth tests of goodness of fit.* New York: Oxford University Press.