Some necessary uniform tests for spherical symmetry

Jiajuan Liang · Kai-Tai Fang · Fred J. Hickernell

Received: 30 August 2004 / Revised: 4 December 2006 / Published online: 25 August 2007 @ The Institute of Statistical Mathematics, Tokyo 2007

Abstract While spherical distributions have been used in many statistical models for high-dimensional data analysis, there are few easily implemented statistics for testing spherical symmetry for the underlying distribution of high-dimensional data. Many existing statistics for this purpose were constructed by the theory of empirical processes and turn out to converge slowly to their limiting distributions. Some existing statistics for the same purpose were given in the form of high-dimensional integrals that are not easily evaluated in numerical computation. In this paper, we develop some necessary tests for spherical symmetry based on both univariate and multivariate uniform statistics. These statistics are easily evaluated numerically and have simple limiting distributions. A Monte Carlo study is carried out to demonstrate the performance of the statistics on controlling type I error rates and power.

Keywords Goodness-of-fit · Monte Carlo study · Spherical symmetry · Uniformity

J. Liang (🖂)

K.-T. Fang Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong, China

F. J. Hickernell Department of Applied Mathematics, Illinois Institute of Technology, 10 West 32nd Street, Chicago, IL 60616-3793, USA

University of New Haven, College of Business, 300 Boston Post Road, West Haven, CT 06516, USA e-mail: jiang@newhaven.edu

1 Introduction

Spherically symmetric (or simply spherical) distributions (SSD for simplicity) are natural extensions to the multivariate standard normal $N_d(0, I_d)$ ($I_d: d \times d$ identity matrix). The SSD possess many desirable properties similar to those of $N_d(0, I_d)$, see the comprehensive studies on SSD given by Fang et al. (1990). A *d*-dimensional random vector \mathbf{x} is said to have a spherical distribution if \mathbf{x} has a stochastic representation

$$\boldsymbol{x} \stackrel{\mathrm{d}}{=} \boldsymbol{\Gamma} \boldsymbol{x},\tag{1}$$

where Γ is a $d \times d$ constant orthogonal matrix such that $\Gamma'\Gamma = \Gamma\Gamma' = I_d$ and the sign " $\stackrel{d}{=}$ " means that both sides in (1) have the same distribution. We denote by $\mathbf{x} \sim S_d(\phi)$ if \mathbf{x} satisfies (1), here $\phi(\cdot)$ is a scale function. If $\mathbf{x} \sim S_d(\phi)$, then the characteristic function (c.f.) of \mathbf{x} has the form $\phi(t't) = \phi(||t||^2)$ ($t \in \mathbb{R}^d$, the *d*-dimensional Euclidean space, $|| \cdot ||$ stands for the Euclidean norm). The SSD have been used as distributional assumptions associated with statistical models, (see, for example, Zellner, 1976; Lange et al., 1989). The problem when the SSD can be considered as the underlying distribution of the sampled data has been the long lasting interest to statisticians in the study of goodness-of-fit techniques. For example, Kariya and Eaton (1977) and Gupta and Kabe (1993) proposed some robust tests for spherical symmetry based on non-independent samples.

Testing spherical symmetry based on an i.i.d. (independently identically distributed) sample x_1, \ldots, x_n with a c.d.f. (cumulative distribution function) F(x) $(x \in R^d)$ is to test the null hypothesis

$$H_0: F(\mathbf{x})$$
 is the c.d.f. of a spherical distribution, (2)

versus the alternative hypothesis H_1 : $F(\mathbf{x})$ is non-spherical. Some existing approaches or statistics for testing spherical symmetry based on i.i.d. samples were summarized in Fang and Liang (1999). These are: (1) graphical methods (Li et al., 1997); (2) tests based on stochastic representation (Baringhaus, 1991); and (3) tests based on projection NT-type statistics (Fang et al., 1993; Zhu et al., 1995). Some other approaches to testing spherical symmetry have been proposed since the past few years (see, for example, Koltchinskii and Li, 1998; Liang and Fang, 2000).

The purpose of this paper is to develop some new necessary tests for spherical symmetry and point out its possible extension to testing elliptical symmetry by employing both univariate and multivariate uniform statistics. Here necessary tests have the same meaning as in Fang et al. (1993). That is, smaller (e.g., less than 5%) *p*-values of the tests indicate evidence of a departure from spherical symmetry while larger *p*-values (e.g., larger than 10%) imply insufficient information to draw a statistical conclusion on the null hypothesis from the sampled data. The univariate uniform statistics are chosen from the recommendation

in Quesenberry and Miller (1977) and Miller and Quesenberry (1979). The multivariate uniform statistics are chosen from Liang et al. (2001). The rest of the paper is arranged as follows. In Sect. 2 a brief review on the univariate and multivariate uniform statistics is given. The principle for testing spherical symmetry based on the uniform statistics is derived. Section 3 presents the empirical results (type I error rates and power against some selected alternatives) for the performance of the uniform tests by a Monte Carlo study. Some concluding remarks and a possible extension of the uniform tests to testing elliptical symmetry are given in the last section.

2 The uniform tests for spherical symmetry

2.1 A review of the uniform statistics

Univariate uniform statistics are those for testing uniformity in the unit interval (0, 1). They are usually constructed by measuring the discrepancy between an ordered sample $u_{(1)} \leq \cdots \leq u_{(n)}$ that is associated with an i.i.d. sample $\{u_1, \ldots, u_n\}$ in (0, 1) and a set of reference ordered points in (0, 1). The ordered points $\{(2i - 1)/(2n) : i = 1, \ldots, n\}$ are known to be uniformly scattered in (0, 1) in the sense of discrepancy in Fang and Wang (1994). There are a number of uniform statistics in the literature. Based on their Monte Carlo studies, Quesenberry and Miller (1977) and Miller and Quesenberry (1979) recommended using Watson's U^2 -statistic and Neyman's smooth test with the fourth degree polynomials as general choices for testing univariate uniformity in (0, 1). These two statistics are described as follows:

1. Watson's U^2 -statistic

Let $W^2 = 1/(12n) + \sum_{i=1}^{n} [(2i-1)/2n - u_{(i)}]^2$, Watson (1962) proposed the statistic

$$WU^2 = W^2 - n(\bar{u} - 0.5)^2 \tag{3}$$

for testing uniformity in (0, 1), where \bar{u} is the sample mean from an i.i.d. sample $\{u_1, \ldots, u_n\}$. Tables of critical values for WU^2 are usually given for the modified form of WU^2 :

$$MU^{2} = \left(WU^{2} - \frac{1}{10n} + \frac{1}{10n^{2}}\right) \left(1 + \frac{0.8}{n}\right).$$
 (4)

The critical values of MU^2 are found to be only slightly dependent on the sample size *n*, and they are 0.267 ($\alpha = 1\%$), 0.187 ($\alpha = 5\%$) and 0.152 ($\alpha = 10\%$) from Stephens (1970). Large values of MU^2 indicate evidence of non-uniformity of the sample. For example, if $MU^2 > 0.187$, one rejects the null hypothesis of uniformity in (0, 1) at the significance level $\alpha = 5\%$.

2. Neyman's smooth test

Let

$$\begin{aligned} \pi_0(y) &= 1, \\ \pi_1(y) &= \sqrt{12}(y - 1/2), \\ \pi_2(y) &= \sqrt{5}[6(y - 1/2)^2 - 1/2], \end{aligned}$$

$$\begin{aligned} \pi_3(y) &= \sqrt{7}[20(y - 1/2)^3 - 3(y - 1/2)], \\ \pi_4(y) &= 210(y - 1/2)^4 - 45(y - 1/2)^2 + 9/8, \end{aligned}$$

which are Legendre polynomials, $y \in [0, 1]$. Denote by

$$t_r = \sum_{i=1}^n \pi_r(u_i), \quad r = 1, 2, 3, 4,$$
 (5)

where $\{u_1, \ldots, u_n\}$ is an i.i.d. sample in (0, 1). Neyman's smooth test (Neyman, 1937) with the fourth degree polynomials is defined by

$$P_4^2 = \frac{1}{n} \sum_{r=1}^4 t_r^2.$$
 (6)

Large values of P_4^2 indicate evidence of non-uniformity of the sample. Critical values for P_4^2 for some small sample size *n* and for large *n* ($n = \infty$) were provided by Miller and Quesenberry (1979). For example, for n > 50, the critical values for P_4^2 were given as 13.28 ($\alpha = 1\%$), 9.49 ($\alpha = 5\%$) and 7.78 ($\alpha = 10\%$).

Testing multi-dimensional (multivariate) uniformity is to test whether an i.i.d. *d*-dimensional sample $\{z_1, \ldots, z_n\}$ can be considered from the uniform distribution in the unit hypercube $\bar{C}^d = [0, 1]^d$. The hypothesis for uniformity of $\{z_1, \ldots, z_n\}$ can be set up as

$$H_0: z_1, \dots, z_n$$
 are uniformly distributed in C^d . (7)

The alternative hypothesis H_1 implies rejection for H_0 in (7). Liang et al. (2001) proposed two types of multivariate uniform statistics for testing uniformity in \bar{C}^d . The two types of multivariate uniform statistics are defined as follows (see Liang et al., 2001 for details):

Type 1. Approximate N(0, 1)-statistics

$$A_n = \sqrt{n} \left[(U_1 - M^d) + 2(U_2 - M^d) \right] / (5\sqrt{\zeta_1}) \xrightarrow{\mathcal{D}} N(0, 1) \quad (n \to \infty)$$
(8)

under H_0 in (7), where " $\xrightarrow{\mathcal{D}}$ " means convergence in probability distribution. There are three choices for A_n according to the three measures of discrepancy: symmetric, centered, and star (Hickernell, 1998).

Type 2. Approximate χ^2 -statistics

$$T_n = n[(U_1 - M^d), (U_2 - M^d)] \Sigma_n^{-1}[(U_1 - M^d), (U_2 - M^d)]' \xrightarrow{\mathcal{D}} \chi^2(2), \quad (n \to \infty)$$
(9)

under H_0 in (7), where

$$\Sigma_n = \begin{pmatrix} \zeta_1 & 2\zeta_1 \\ 2\zeta_1 & \frac{4(n-2)}{n-1}\zeta_1 + \frac{2}{n-1}\zeta_2 \end{pmatrix},$$
(10)

and ζ_1 and ζ_2 are calculated differently according to the three measures of discrepancies given as follows. There are also three choices for T_n .

The calculation of A_n in (8) and that of T_n in (9) are obtained according to any of the following three measures of discrepancy. From an i.i.d. *d*-dimensional sample $\{z_1, \ldots, z_n\}$ in \overline{C}^d , let $z_k = (z_{k1}, \ldots, z_{kd})'$ $(k = 1, \ldots, n)$.

1. The symmetric discrepancy gives

$$U_{1} = \frac{1}{n} \sum_{k=1}^{n} \prod_{j=1}^{d} (1 + 2z_{kj} - 2z_{kj}^{2}),$$

$$U_{2} = \frac{2^{d+1}}{n(n-1)} \sum_{k
(11)$$

with M = 4/3, $\zeta_1 = (9/5)^d - (6/9)^d$ and $\zeta_2 = 2^d - (16/9)^d$; 2. The centered discrepancy gives

$$U_{1} = \frac{1}{n} \sum_{k=1}^{n} \prod_{j=1}^{d} \left(1 + \frac{1}{2} \Big| z_{kj} - \frac{1}{2} \Big| - \frac{1}{2} \Big| z_{kj} - \frac{1}{2} \Big|^{2} \right),$$

$$U_{2} = \frac{2}{n(n-1)} \sum_{k
(12)$$

with M = 13/12, $\zeta_1 = (47/40)^d - (13/12)^{2d}$ and $\zeta_2 = (57/48)^d - (13/12)^{2d}$; 3. The star discrepancy gives

$$U_{1} = \frac{1}{n} \sum_{k=1}^{n} \prod_{j=1}^{d} \left(\frac{3 - z_{kj}}{2} \right),$$

$$U_{2} = \frac{2}{n(n-1)} \sum_{k
(13)$$

with M = 4/3, $\zeta_1 = (9/5)^d - (16/9)^d$ and $\zeta_2 = (11/6)^d - (16/9)^d$.

The empirical finite-sample percentiles of A_n and T_n under the above three discrepancies were provided in Liang et al. (2001) for some selected sample sizes (n = 25, n = 50, n = 100 and n = 200). From the Monte Carlo study on the type I error rates of A_n and T_n in Liang et al. (2001) by using the corresponding critical values of N(0, 1) (for A_n) and $\chi^2(2)$ (for T_n), it can be concluded that the approximation of A_n by N(0, 1), and T_n by $\chi^2(2)$ is good enough for

sample sizes *n* as small as 25. A large value of $|A_n|$ or T_n indicates evidence of non-uniformity for the underlying distribution of a sample from \bar{C}^d .

2.2 Uniform tests for spherical symmetry

In the univariate case, the uniform test for general goodness-of-fit problem is based on the principle: a random variable X has a c.d.f. F(x) that is strictly increasing if and only if the random variable U = F(X) has a uniform distribution U(0, 1). F(X) is called the uniform transformation of X. In the multivariate case, different uniform transformations from a random vector x can be defined. These transformations are based on some characterization of the underlying distribution of the random vector. Theorem 1 is an application of Lemma 1 and it gives the necessary characterization for the spherical distribution. Theorem 1 provides the principle for the necessary uniform tests for spherical symmetry.

Lemma 1 (*Rosenblatt*, 1952) Let random vector $\mathbf{x} = (x_1, ..., x_d)'$ have a probability density function (p.d.f.) $f(x_1, ..., x_d)$. Define the following random variables by the conditional probability distributions:

for k = 2, ..., d. Then the random variables $x_1, x_{2|1}, ..., x_{k|1...,k-1}$ (k = 2, ..., d) are mutually independent.

The assertion in Lemma 1 follows from the definition for the conditional p.d.f. and the property

$$f(x_1,\ldots,x_d) = f_{x_1}(x_1)f_{2|1}(x_2|x_1)\ldots f_{d|1\ldots d-1}(x_d|x_1,\ldots,x_{d-1}),$$

where f_{x_1} stands for the marginal p.d.f. of x_1 , $f_{2|1}(x_2|x_1)$ for the conditional p.d.f. of $x_{2|1}$, and in general, $f_{k|1...k-1}(x_k|x_1,...,x_{k-1})$ for the conditional p.d.f. of $x_{k|1...,k-1}$ for k = 2,...,d. The conditional probability transformation in (14) is called the *Rosenblatt transformation*.

Theorem 1 Let random vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be i.i.d. with a spherical distribution $S(\phi)$ and $P(\mathbf{x}_i = \mathbf{0}) = 0$ $(i = 1, \ldots, n)$. Denote by

$$u_1 = x_1 / ||x_1||, \ldots, u_n = x_n / ||x_n||,$$
 (15)

and $\mathbf{u}_i = (u_{i1}, \dots, u_{id})'$ for $i = 1, \dots, n$. Then the following assertions on the conditional distributions are true:

$$B_{1}(i) \stackrel{d}{=} u_{i1}^{2} \sim \beta(1/2, (d-1)/2),$$

$$B_{2}(i) \stackrel{d}{=} \{(1-u_{i1}^{2})^{-1}u_{i2}^{2}|u_{i1}\} \sim \beta(1/2, (d-2)/2),$$

$$\vdots : :$$

$$B_{k}(i) \stackrel{d}{=} \{(1-\sum_{j=1}^{k-1}u_{ij}^{2})^{-1}u_{ik}^{2}|(u_{i1}, \dots, u_{i,k-1})\}$$

$$\sim \beta(1/2, (d-k)/2),$$

$$i = 1, \dots, n; \quad k = 2, \dots, d-1$$

$$(16)$$

where $\beta(1/2, (d - j)/2)$ denotes the univariate beta distribution with parameters 1/2 and (d - j)/2 (j = 1, ..., d - 1). Moreover, the random variables $\{B_1(i), ..., B_{d-1}(i) : i = 1, ..., n\}$ are mutually independent.

Proof Under the spherical assumption on the x_i 's, the random vectors u_i 's given by (15) are independent and the u_i 's have a uniform distribution on the surface of the unit sphere in \mathbb{R}^d . By Chap. 2 of Fang et al. (1990), we have

$$\boldsymbol{u}_i \stackrel{\mathrm{d}}{=} \boldsymbol{z}_0 / \|\boldsymbol{z}_0\|, \quad \boldsymbol{z}_0 \sim N_d(\boldsymbol{0}, \boldsymbol{I}_d).$$
(17)

Then we can obtain the joint density function for any k < d components of the u_i in (15) by a direct calculation:

$$\frac{\Gamma(d/2)}{\Gamma[(d-k)/2]\pi^{k/2}} \left(1 - \sum_{j=1}^{k} u_j^2\right)^{(d-k)/2 - 1}, \quad \text{for } \sum_{j=1}^{k} u_j^2 < 1, \tag{18}$$

where k = 1, ..., d - 1. The p.d.f given by (18) exists in the unit sphere $S_k = \{u : u \in \mathbb{R}^k, \|u\|^2 < 1\}$ of \mathbb{R}^k and it diminishes (zero value) outside S_k (k = 1, ..., d - 1). Then a direct calculation for a conditional density leads to the assertions given by (16). The independence of the random variables $B_1(i), ..., B_{d-1}(i)$ (i = 1, ..., n) results from Lemma 1. This completes the proof.

Denote by $F_{b_j}(\cdot)$ the c.d.f. of $\beta(1/2, (d-j)/2)$ (j = 1, ..., d-1). Let $v_i = (v_{i1}, ..., v_{i,d-1})'$ be given by

$$v_{ij} = F_{b_i}(B_j(i)), \quad i = 1, \dots, n; \ j = 1, \dots, d-1.$$
 (19)

Based on Theorem 1, a necessary test for spherical symmetry can be constructed. Let x_1, \ldots, x_n be i.i.d. observations. Perform the transformation

$$\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n : \text{ in } R^d\} \Rightarrow \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n : \text{ in } R^d\}$$

$$\Rightarrow \{B_1(i), \dots, B_{d-1}(i) : i = 1, \dots, n; \text{ in } R^{d-1}\}$$

$$\Rightarrow \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n : \text{ in } R^{d-1}\},$$
(20)

Deringer

given by (15), (16) and (19). If the null hypothesis H_0 in (2) is true for the underlying distribution $F(\mathbf{x})$ of the i.i.d. sample $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, the random variables v_{ij} 's given by (19) are mutually independent and the v_{ij} has a uniform distribution U(0, 1). The random points $\mathbf{v}_i = (v_{i1}, \ldots, v_{i,d-1})'$ are located in the hypercube $[0, 1]^{d-1}$. A test for spherical symmetry (2) can be substituted by a test for multivariate uniformity

$$H_0$$
: the v_i 's are uniformly distributed in $[0, 1]^{d-1}$, (21)

versus H_1 that implies that H_0 in (21) is not true. Or a test for spherical symmetry (2) can also be substituted by a test for univariate uniformity

$$H_0$$
: the v_{ii} 's are uniformly distributed in (0, 1), (22)

versus H_1 that implies that H_0 in (22) is not true. It is obvious that both of the tests for (21) and (22) are necessary tests for spherical symmetry (2). That is, if H_0 in (21) or H_0 in (22) is rejected at some significance level, the H_0 in (2) is also rejected at the same level, and the underlying distribution of the sample $\{x_1, \ldots, x_n\}$ can be considered as non-spherical. The statistics A_n in (8) and T_n in (9) can be employed to test (21), and the statistics MU^2 in (4) and P_4^2 in (6) can be employed to test (22). Each of these four statistics can be used as a necessary test for hypothesis (2).

It should be pointed out that acceptance of H_0 in (21) can usually lead to acceptance of H_0 in (22) but the contrary is usually not true. This implies that univariate uniformity of all one-dimensional marginal distributions does not automatically lead to multivariate uniformity for the joint distribution. Figure 1 shows the extreme cases (dimension for the \mathbf{x}_i 's: d = 3) that the transformed sample { $\mathbf{v}_1, \ldots, \mathbf{v}_n$ } (dimension for the \mathbf{v}_i 's: d - 1 = 2) given by (20) has univariate uniformity on the horizontal axis (H_0 in (22) is true) but does not have multivariate uniformity in R^2 (H_0 in (21) is not true).

3 Monte carlo study

In this section we carry out a limited Monte Carlo study on the performance of the uniform tests A_n in (8), T_n in (9), MU^2 in (4) and P_4^2 in (6) for testing spherical symmetry. The type I error rates for MU^2 and P_4^2 , and the type I error rates for A_n and T_n under the three discrepancies (symmetric, centered and star) and the power of all uniform statistics will be studied. In calculating the type I error rates and power, the corresponding percentiles for the multivariate unform statistics A_n and T_n are chosen as those of their limiting distributions, respectively. That is, we consider $A_n \sim N(0, 1)$ and $T_n \sim \chi^2(2)$ for all sample sizes *n*. The percentiles of MU^2 and P_4^2 are given in Sect. 2.1. That is, for MU^2 , they are 0.267 $(1 - \alpha = 99\%)$, 0.187 $(1 - \alpha = 95\%)$ and 0.152 $(1 - \alpha = 90\%)$; for P_4^2 , they are 13.28 $(1 - \alpha = 99\%)$, 9.49 $(1 - \alpha = 95\%)$ and 7.78 $(1 - \alpha = 90\%)$.

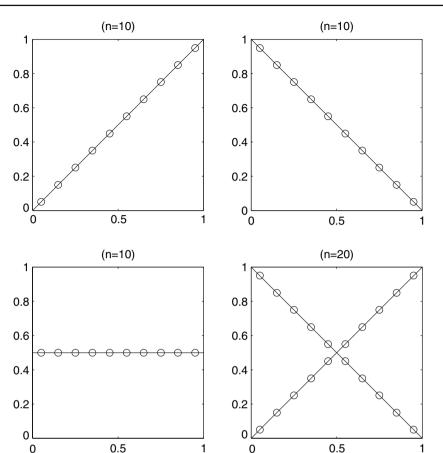


Fig. 1 Extreme cases of univariate uniformity but no multivariate uniformity in R^2

3.1 Monte Carlo study on type I error rates

In the study on type I error rates of the tests, six spherical distributions as discussed in detail in Chap. 3 of Fang et al. (1990) are selected as the null distributions. These spherical distributions are: (1) the standard normal distribution $N_d(\mathbf{0}, \mathbf{I}_d)$; (2) the multivariate *t*-distribution with degrees of freedom m = 5; (3) the Kotz type distribution with N = 2, r = 1 and s = 0.5; (4) the Pearson type VII distribution (PVII) with N = 10 and m = 2; (5) the Pearson type II distribution (PII) with m = 3/2; and (6) the Cauchy distribution. The so-called TFWW algorithm (Tashiro, 1977; Fang and Wang, 1994, pp. 166–170) is employed to generate empirical samples from the selected non-normal spherical distributions. The normal samples are generated by the MATLAB internal function. The empirical type I error rates are computed by

Type I error rate =
$$\frac{\text{Number of rejections}}{\text{Number of replications}}$$
. (23)

Deringer

Statistics	Discrepancy	Dimension $d = 5$							
		n	Normal	Mult <i>t</i>	Kotz	PVII	PII	Cauchy	
T _n	Symmetric	25	0.0530	0.0495	0.0530	0.0605	0.0545	0.0595	
		50	0.0515	0.0550	0.0465	0.0475	0.0425	0.0595	
		100	0.0550	0.0505	0.0495	0.0540	0.0500	0.0425	
		200	0.0500	0.0595	0.0520	0.0520	0.0530	0.0515	
	Centered	25	0.0480	0.0490	0.0535	0.0565	0.0540	0.0610	
		50	0.0515	0.0600	0.0480	0.0535	0.0485	0.0610	
		100	0.0555	0.0470	0.0450	0.0525	0.0495	0.0520	
		200	0.0590	0.0555	0.0515	0.0555	0.0515	0.0490	
	Star	25	0.0570	0.0580	0.0575	0.0550	0.0595	0.0630	
		50	0.0605	0.0675	0.0650	0.0575	0.0630	0.0630	
		100	0.0615	0.0575	0.0630	0.0635	0.0640	0.0635	
		200	0.0610	0.0660	0.0600	0.0595	0.0635	0.0630	
An	Symmetric	25	0.0645	0.0670	0.0650	0.0740	0.0750	0.0650	
		50	0.0580	0.0595	0.0545	0.0505	0.0535	0.0650	
		100	0.0540	0.0490	0.0470	0.0515	0.0575	0.0535	
		200	0.0510	0.0590	0.0480	0.0530	0.0570	0.0520	
	Centered	25	0.0590	0.0625	0.0685	0.0570	0.0720	0.0665	
		50	0.0620	0.0630	0.0640	0.0555	0.0435	0.0665	
		100	0.0580	0.0490	0.0515	0.0500	0.0555	0.0520	
		200	0.0555	0.0545	0.0455	0.0550	0.0505	0.0490	
	Star	25	0.0510	0.0500	0.0475	0.0445	0.0465	0.0550	
		50	0.0575	0.0545	0.0485	0.0500	0.0510	0.0550	
		100	0.0530	0.0455	0.0475	0.0510	0.0470	0.0525	
		200	0.0445	0.0490	0.0520	0.0495	0.0505	0.0545	

Table 1 Type I error rates of A_n and T_n for testing spherical symmetry ($\alpha = 0.05$, no. of replications = 2,000)

Tables 1 and 2, respectively, present the simulation results on the type I error rates of A_n and T_n , and those of MU^2 and P_4^2 , when testing spherical symmetry with the significance level $\alpha = 0.05$. Similar simulation results were also obtained for the significance levels $\alpha = 0.01$ and 0.10, but these are not presented to save space. The following empirical conclusions can be summarized:

- (1) When using the limiting distributions instead of the finite-sample distributions, the two multivariate uniform statistics A_n and T_n , and the two univariate uniform statistics MU^2 and P_4^2 can maintain feasible control of the type I error rates for the sample size as small as n = 25;
- (2) The dimension of the sample seems to have little influence on the type I error rates of the statistics A_n , T_n , MU^2 and P_4^2 when testing spherical symmetry. This is a good indication in testing goodness-of-fit in the sense of avoiding curse of dimensionality.

Table 1 conti	nued
---------------	------

Statistics	Discrepancy	Dimension $d = 10$							
		n	Normal	Multit	Kotz	PVII	PII	Cauchy	
T _n	Symmetric	25	0.0620	0.0540	0.0470	0.0510	0.0440	0.0500	
		50	0.0565	0.0445	0.0630	0.0625	0.0495	0.0520	
		100	0.0435	0.0515	0.0455	0.0540	0.0465	0.0455	
		200	0.0565	0.0490	0.0470	0.0465	0.0425	0.0505	
	Centered	25	0.0565	0.0480	0.0505	0.0605	0.0495	0.0430	
		50	0.0525	0.0520	0.0580	0.0585	0.0560	0.0480	
		100	0.0465	0.0490	0.0545	0.0590	0.0595	0.0585	
		200	0.0470	0.0455	0.0510	0.0465	0.0530	0.0525	
	Star	25	0.0570	0.0565	0.0615	0.0660	0.0540	0.0580	
		50	0.0560	0.0655	0.0570	0.0635	0.0650	0.0600	
		100	0.0500	0.0660	0.0595	0.0545	0.0705	0.0515	
		200	0.0560	0.0570	0.0605	0.0565	0.0630	0.0550	
A _n	Symmetric	25	0.0740	0.0750	0.0545	0.0720	0.0585	0.0765	
		50	0.0665	0.0480	0.0515	0.0685	0.0560	0.0565	
		100	0.0560	0.0590	0.0540	0.0625	0.0485	0.0480	
		200	0.0590	0.0605	0.0550	0.0505	0.0505	0.0445	
	Centered	25	0.0680	0.0645	0.0680	0.0705	0.0625	0.0635	
		50	0.0550	0.0495	0.0555	0.0495	0.0605	0.0625	
		100	0.0480	0.0595	0.0530	0.0675	0.0620	0.0630	
		200	0.0535	0.0490	0.0530	0.0505	0.0520	0.0435	
	Star	25	0.0505	0.0450	0.0430	0.0495	0.0415	0.0460	
		50	0.0495	0.0575	0.0475	0.0500	0.0560	0.0545	
		100	0.0400	0.0570	0.0510	0.0460	0.0555	0.0460	
		200	0.0485	0.0595	0.0440	0.0475	0.0520	0.0430	

3.2 Monte Carlo study on power

The empirical power of the tests is computed by (23) when choosing five nonspherical alternative distributions. Five meta-type normal distributions are chosen as the alternative distributions for studying the rejection rate or power of each of the tests for testing spherical symmetry. The theory on general meta-type distributions is given in Fang et al. (2002). The idea for constructing the meta-type normal distribution is as follows. Let $\mathbf{x} = (X_1, \ldots, X_d)'$ have a continuous c.d.f. $F(\mathbf{x})$ ($\mathbf{x} \in \mathbb{R}^d$) with a density function $f(\mathbf{x}) = f(x_1, \ldots, x_d)$. Denote by $f_i(x_i)$ the marginal density function of X_i . Define the random vector $\mathbf{y} = (Y_1, \ldots, Y_d)'$ by

$$Y_i = \Phi^{-1}(F_i(X_i)), \quad i = 1, \dots, d,$$
 (24)

where $\Phi(\cdot)$ is the c.d.f. of N(0,1) and $F_i(\cdot)$ the marginal c.d.f. of X_i . It is obvious that each Y_i has a normal distribution N(0,1) but the joint distribution

Statistics	n	Dimension $d = 5$							
		Normal	Multit	Kotz	PVII	PII	Cauchy		
MU^2	25	0.0470	0.0560	0.0465	0.0505	0.0530	0.0565		
	50	0.0515	0.0585	0.0530	0.0455	0.0415	0.0515		
	100	0.0550	0.0580	0.0470	0.0545	0.0510	0.0480		
	200	0.0525	0.0470	0.0475	0.0525	0.0575	0.0560		
P_{4}^{2}	25	0.0430	0.0475	0.0410	0.0445	0.0510	0.0630		
	50	0.0520	0.0505	0.0510	0.0510	0.0450	0.0480		
	100	0.0530	0.0545	0.0535	0.0525	0.0580	0.0490		
	200	0.0520	0.0495	0.0450	0.0530	0.0590	0.0530		
		Dimension $d = 10$							
MU^2	25	0.0605	0.0445	0.0595	0.0570	0.0605	0.0450		
	50	0.0500	0.0420	0.0500	0.0480	0.0500	0.0455		
	100	0.0480	0.0490	0.0465	0.0500	0.0500	0.0510		
	200	0.0485	0.0490	0.0410	0.0520	0.0625	0.0520		
P_{4}^{2}	25	0.0535	0.0525	0.0495	0.0480	0.0580	0.0500		
	50	0.0455	0.0500	0.0450	0.0435	0.0490	0.0465		
	100	0.0450	0.0545	0.0475	0.0520	0.0550	0.0550		
	200	0.0500	0.0475	0.0470	0.0520	0.0520	0.0515		

Table 2 Type I error rates of MU^2 and P_4^2 for testing spherical symmetry ($\alpha = 0.05$, no. of replications = 2,000)

of $y = (Y_1, ..., Y_d)'$ may have a big difference from the multivariate normal distribution. A direct calculation gives the joint density function of y:

$$q(y_1, \dots, y_d) = f\left(F_1^{-1}(\Phi(y_1)), \dots, F_d^{-1}(\Phi(y_d))\right) \prod_{i=1}^d \{\phi(y_i) / f_i(F_i^{-1}(\Phi(y_i)))\}, \quad (25)$$

where $\phi(\cdot)$ is the density function of N(0, 1) and $(y_1, \ldots, y_d)' \in \mathbb{R}^d$. In particular, if the random variables X_1, \ldots, X_d are independent, then $\mathbf{y} = (Y_1, \ldots, Y_d)' \sim N_d(\mathbf{0}, \mathbf{I}_d)$. The five meta-type normal distributions are obtained as follows. All related spherical distributions with the corresponding parameters are referred to Chap. 3 of Fang et al. (1990).

- When the random vector x = (X₁,...,X_d)' ~multivariate *t*-distribution with m = 5 in (24), the meta-type normal distribution given by the distribution of y = (Y₁,...,Y_d)' in (24)-(25) is denoted by y ~ MTN;
- (2) when the random vector x = (X₁,...,X_d)' ∼Kotz type distribution with N = 2, r = 1 and s = 0.5, the meta-type normal distribution given by the distribution of y = (Y₁,...,Y_d)' in (24)–(25) is denoted by u ∼ MKN;
- (3) when the random vector $\mathbf{x} = (X_1, \dots, X_d)'$ ~Pearson type VII distribution with N = 10 and m = 2, the meta-type normal distribution given by the distribution of $\mathbf{y} = (Y_1, \dots, Y_d)'$ in (24)–(25) is denoted by $\mathbf{u} \sim MPVIIN$;
- (4) when the random vector $\mathbf{x} = (X_1, \dots, X_d)'$ ~Pearson type II distribution with m = 3/2, the meta-type normal distribution given by the distribution of $\mathbf{y} = (Y_1, \dots, Y_d)'$ in (24)–(25) is denoted by $\mathbf{u} \sim MPIIN$;

(5) when the random vector $\mathbf{x} = (X_1, \dots, X_d)'$ ~Cauchy distribution, the meta-type normal distribution given by the distribution of $\mathbf{y} = (Y_1, \dots, Y_d)'$ in (24)–(25) is denoted by $\mathbf{u} \sim MCN$.

In order to have a visual perception on the power performance of the four statistics A_n , T_n , MU^2 and P_4^2 , we plot their power values versus the above five nonspherical alternative distributions. The simulation for obtaining the empirical power of the four statistics was carried out with 2,000 replications. Figures 2 and 3, respectively, present the plots of the power values versus the five metatype normal distributions for dimension d = 5 and d = 10 for different sample sizes.

Figures 2 and 3 provide a quick view on the power performance of the four statistics A_n , T_n , MU^2 and P_4^2 in testing spherical symmetry against the selected meta-type normal distributions that are nonspherical. Based on Figs. 2 and 3, we can summarize our empirical conclusions on the power performance of the uniform statistics as follows.

- (1) Similar to the performance in their type I error rates, all four uniform statistics A_n , T_n , MU^2 and P_4^2 are not sensitive to the increase of sample dimension in testing spherical symmetry. For example, when the sample dimension *d* increases from d = 5 (Fig. 2) to d = 10 (Fig. 3), the four statistics have similar power performance for each of the selected alternative distributions;
- (2) For the two multivariate uniform statistics A_n and T_n , their power performance is different for different choices of discrepancy measures. The symmetric discrepancy tends to be the best, centered discrepancy the second, and the star discrepancy the worst in all cases;
- (3) The two multivariate uniform statistics A_n and T_n cannot outperform the two univariate uniform statistics MU^2 and P_4^2 in all cases and vice versa. So in practical applications, all of these statistics can be used together to give more confidence in drawing a conclusion.

4 Concluding remarks and a possible extension

In Sect. 3 we employ the two multivariate uniform statistics A_n and T_n to test spherical symmetry. Although these two statistics cannot outperform the traditional univariate uniform statistics MU^2 and P_4^2 in all cases as considered, they provide a new way to substitute a test for spherical symmetry by a test for uniformity. The statistics A_n and T_n are easy to compute and their percentiles can be approximately taken as those for the standard normal N(0, 1) and the chi-square $\chi^2(2)$, respectively, for the sample size as small as n = 25 based on the simulation results. In some extreme cases like those in Fig. 1, if the underlying distribution of the sample is not spherical, the univariate uniform statistics MU^2 and P_4^2 may fail to detect nonspherical symmetry but the multivariate uniform statistics A_n and T_n may still work well. This implies that the statistics A_n and T_n possess some good multivariate properties that the statistics MU^2 and P_4^2 do not have.

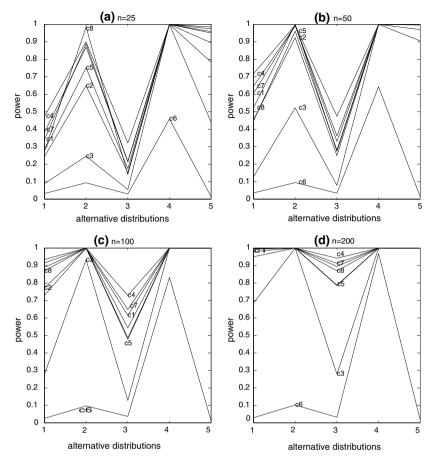


Fig. 2 Plots of power of the statistics A_n and T_n (under the three discrepancies: symmetric, centered and star), and MU^2 and P_4^2 in testing spherical symmetry against the meta-type normal distributions (d = 5, $\alpha = 5\%$). The curves c1–c8 stand for: (1) c1 for T_n (symmetric); (2) c2 for T_n (centered); (3) c3 for T_n (star); (4) c4 for A_n (symmetric); (5) c5 for A_n (centered); (6) c6 for A_n (star); (7) c7 for MU^2 ; and (8) c8 for P_4^2 . The alternative distributions are arranged in order: "1" for the distribution MTN; "2" for the distribution MKN; "3" for the distribution MPVIIN; "4" for the distribution MPIIN; and "5" for the distribution MCN

It is also a desirable property that the uniform tests for spherical symmetry are not sensitive to the increase of sample dimension. While the power of the uniform statistics is relatively high for some selected alternative distributions with suitable measures of discrepancy, it is very low for the selected alternative distributions "2" (MKN) and "3" (MPVII) by using A_n and T_n with the star discrepancy (the curves c3 and c6 in Figs. 2 and 3). This implies that choosing the symmetric discrepancy and the centered discrepancy for A_n and T_n may give better results in general.

Besides using the two statistics A_n and T_n to construct the necessary tests for spherical symmetry, we can actually employ A_n and T_n to construct a class of

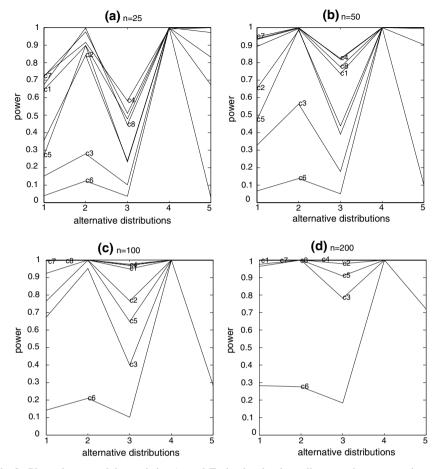


Fig. 3 Plots of power of the statistics A_n and T_n (under the three discrepancies: symmetric, centered and star), and MU^2 and P_4^2 in testing spherical symmetry against the meta-type normal distributions ($d = 10, \alpha = 5\%$). The curves c1–c8 stand for: (1) c1 for T_n (symmetric); (2) c2 for T_n (centered); (3) c3 for T_n (star); (4) c4 for A_n (symmetric); (5) c5 for A_n (centered); (6) c6 for A_n (star); (7) c7 for MU^2 ; and (8) c8 for P_4^2 . The alternative distributions are arranged in order: "1" for the distribution MTN; "2" for the distribution MKN; "3" for the distribution MPVIIN; "4" for the distribution MPIIN; and "5" the distribution MCN

necessary tests for a much bigger family of symmetric multivariate distributions. It was pointed out by Fang et al. (1990, p. 9) that the following approach is a common way to construct symmetric multivariate distributions. Let y have a symmetric multivariate distribution in some sense. One can define a class of symmetric multivariate distributions generated by y as follows:

$$\mathcal{F}(\mathbf{y}) = \{ \mathbf{x} \mid \mathbf{x} \stackrel{\mathrm{d}}{=} R\mathbf{y}, R \ge 0 \text{ is independent of } \mathbf{y} \}.$$
(26)

The following special cases give some subclasses of $\mathcal{F}(\mathbf{y})$:

- (1) If $\mathbf{y} \sim N_d(\mathbf{0}, \sigma^2 \mathbf{I}_d)$, the class $\mathcal{F}(\mathbf{y})$ in (26) is equivalent to the class of mixtures of the spherical normal distribution.
- (2) If y is uniformly distributed on the unit sphere in \mathbb{R}^d , $\mathcal{F}(\mathbf{y})$ in (26) is the class of spherical distributions.
- (3) If y is uniformly distributed on the l₁-norm unit sphere, F(y) in (26) is the class of the l₁-norm symmetric distributions (Fang and Fang, 1988, 1989; Fang et al., 1990, Chap. 5).
- (4) If y is uniformly distributed on the l_p -norm unit sphere, $\mathcal{F}(\mathbf{y})$ in (26) is the class of the l_p -norm symmetric distributions (Yue and Ma, 1995).
- (5) If y is uniformly distributed on the L_p -norm unit sphere, $\mathcal{F}(y)$ in (26) is the class of the L_p -norm spherical distributions (Osiewalski and Stel, 1993; Gupta and Song, 1997).

In this paper we actually develop some necessary goodness-of-fit tests for a subclass (the class of spherical distributions) of $\mathcal{F}(\mathbf{y})$ in (26). The same technique can be extended to constructing necessary goodness-of-fit tests for the distributions in the class $\mathcal{F}(\mathbf{y})$ defined by (26). Note that all distributions in the class $\mathcal{F}(\mathbf{y})$ have zero means and some distributions have covariance matrices of the form $c\mathbf{I}$ (c > 0 is a constant and \mathbf{I} is an identity matrix) when $\operatorname{cov}(\mathbf{y}) = \sigma^2 \mathbf{I}$ as for the case of an SSD for \mathbf{y} . An important family of distributions, called the elliptical (or elliptically contoured) distributions (ECD for simplicity), is obtained from a subclass of $\mathcal{F}(\mathbf{y})$ by linear transformations (Fang et al., 1990):

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{y}, \quad \boldsymbol{y} \sim SSD, \tag{27}$$

where μ is a constant vector and A is a constant full-rank matrix. x in (27) is said to have an ECD and denote by $x \sim ECD(\mu, \Sigma)$ ($\Sigma = AA'$). ECD has been used in statistical analysis in many areas (see, e.g., Anderson, 1993; Kariya and Sinha, 1989; and Wakaki, 1994). Testing whether an i.i.d. sample can be considered as an elliptical sample (the underlying distribution is an ECD) is called testing elliptical symmetry. A heuristic approach to extending tests for spherical symmetry to tests for elliptical symmetry could be carried out as follows. $x \sim ECD(\mu, \Sigma)$ can be represented as

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{y}, \quad \boldsymbol{y} \sim \text{SSD}, \tag{28}$$

where $\Sigma^{\frac{1}{2}}$ stands for the positively definite square root of $\Sigma > 0$ (assuming Σ is positively definite). For a given i.i.d. sample x_1, \ldots, x_n , by using the sample mean \bar{x} to estimate the unknown mean μ and the sample covariance matrix S to estimate the unknown covariance matrix Σ , we define a transformed "sample" (it is not an i.i.d. sample for finite sample sizes):

$$\mathbf{y}_{j} = \mathbf{S}^{-1/2}(\mathbf{x}_{j} - \bar{\mathbf{x}}), \quad i = 1, \dots, n,$$
 (29)

where $S^{-1/2} = (S^{1/2})^{-1}$ and $S^{1/2}$ is the positively definite square root of S (assuming large sample size). It is well known that the sample mean \bar{x} is a

strongly consistent estimate for μ , and the sample covariance matrix **S** is a strongly consistent estimate for Σ under minor conditions. Therefore, if $x_j \sim \text{ECD}(\mu, \Sigma)$, the distribution of y_j defined by (29) can be approximated by an SSD. Edgeworth expansions of the distribution of y_j have been studied by Wakaki (1994) and Fujikoshi (1997). Under a large sample size, it is possible to employ the results in Wakaki (1994) and Fujikoshi (1997), the relationship (29), and the uniform statistics to construct some necessary tests for elliptical symmetry. Due to limited space in one paper, we will study the theoretical justifications of this approach in our future research.

Acknowledgment This research is supported in part by *Hong Kong Research Grants Council* grant HKBU/2007/03P, and a University of New Haven 2004 and 2005 Summer Faculty Fellowships and Research Award.

References

- Anderson, T. W. (1993). Nonnormal multivariate distributions: Inference based on elliptically contoured distributions. In C. R. Rao (Ed.), *Multivariate analysis: Future directions* (pp. 1–24). Elsevier: Amsterdam.
- Baringhaus, L. (1991). Testing for spherical symmetry of a multivariate distribution. The Annals of Statistics, 19, 899–917.
- Fang, H., Fang, K.-T., Kotz, S. (2002). The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis*, 82, 1–16.
- Fang, K.-T., Fang, B.-Q. (1988). Some families of multivariate symmetric distributions related to exponential distribution. *Journal of Multivariate Analysis*, 24, 109–122.
- Fang, K.-T., Fang, B.-Q. (1989). A characterization of multivariate l₁-norm symmetric distribution. Statistics and Probability Letters, 7, 297–299.
- Fang, K.-T., Kotz, S., Ng, K. W. (1990). *Symmetric multivariate and related distributions*. London: Chapman and Hall.
- Fang, K.-T., Liang, J. (1999). Testing spherical and elliptical symmetry. In S. Kotz, C. B. Read, D. L. Banks (Eds.), *Encyclopedia of statistical sciences* (Update). (Vol. 3, pp. 686–691). New York: Wiley.
- Fang, K.-T., Wang, Y. (1994). Number-theoretic methods in statistics. London: Chapman and Hall.
- Fang, K.-T., Zhu, L.-X., Bentler, P. M. (1993). A necessary test for sphericity of a high-dimensional distribution. *Journal of Multivariate Analysis*, 44, 34–55.
- Fujikoshi, Y. (1997). An asymptotic expansion for the distribution for the distribution of Hotelling T²-statistic. Journal of Multivariate Analysis, 61, 187–193.
- Gupta, A. K., Kabe, D. G. (1993). Multivariate robust tests for spherical symmetry with applications to multivariate least squares regression. *Journal of Applied Statistical Science*, 1(2), 159–168.
- Gupta, A. K., Song, D. (1997). L_p-norm spherical distributions. Journal of Statistical Planning and Inference, 60, 241–260.
- Hickernell, F. J. (1998). A generalized discrepancy and quadrature error bound. *Mathematics of Computation*, 67, 299–322.
- Kariya, T., Eaton, M. L. (1977). Robust tests for spherical symmetry. The Annals of Statistics, 5, 206–215.
- Kariya, T., Sinha, B. K. (1989). Robustness of statistical tests. New York: Academic Press.
- Koltchinskii, V. I., Li, L. (1998). Testing for spherical symmetry of a multivariate distribution. Journal of Multivariate Analysis, 65, 228–244.
- Lange, K. L., Little, R. J. A., Taylor, J. M. G. (1989). Robust statistical modeling using the t-distribution. Journal of the American Statistical Association, 84, 881–896.
- Li, R., Fang, K.-T., Zhu, L.-X. (1997). Some Q-Q probability plots to test spherical and elliptical symmetry. *Journal of Computational and Graphical Statistics*, 6, 435–450.
- Liang, J., Fang, K.-T. (2000). Some applications of Läuter's technique in tests for spherical symmetry. Biometrical Journal, 42, 923–936.

- Liang, J., Fang, K.-T., Hickernell, F. J., Li, R. (2001). Testing multivariate uniformity and its applications. *Mathematics of Computation*, 70, 337–355.
- Miller, F. L. Jr., Quesenberry, C. P. (1979). Power studies of tests for uniformity, II. Communications in Statistics –Simulation and Computation, B8(3), 271–290.
- Neyman, J. (1937). "Smooth" test for goodness of fit. Journal of the American Statistical Association, 20, 149–199.
- Osiewalski, J., Stel, M. F. J. (1993). Robust Bayesian inference on *lq*-spherical models. *Biometrika*, 80, 456–460.
- Quesenberry, C. P., Miller, F. L., Jr. (1977). Power studies of some tests for uniformity. *Journal of Statistical Computation and Simulation*, 5, 169–191.
- Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23, 470–472.
- Stephens, M. A. (1970). Use of the Kolmogorov Smirnov, Cramér-von Mises and related statistics without extensive tables. *Journal of the Royal Statistical Society (Series B)*, 32, 115–122.
- Tashiro, D. (1977). On methods for generating uniform points on the surface of a sphere. *The Annals of the Institute of Statistical Mathematics*, 29, 295–300.
- Wakaki, H. (1994). Discriminant analysis under elliptical distributions. *Hiroshima Mathematical Journal*, 24, 257–298.
- Watson, G. S. (1962). Goodness-of-fit tests on a circle. II. Biometrika, 49, 57-63.
- Yue, X., Ma, C. (1995). Multivariate l_p-norm symmetric distributions. Statistics and Probability Letters, 24, 281–288.
- Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-*t* error terms. *Journal of the American Statistical Association*, *71*, 400–405.
- Zhu, L.-X., Fang, K.-T., Zhang, J.-T. (1995). A projection NT-type test for spherical symmetry of a multivariate distribution. *New trends in probability and statistics* (Vol. 3, pp. 109–122). Utrecht, The Netherland, Tokyo, Japan: VSP and Uilnius, Lithuania: TEV.
- Zhu, L.-X., Fang, K.-T., Bhatti, M. I., Bentler, P. M. (1995). Testing sphericity of a high-dimensional distribution based on bootstrap approximation. *Pakistan Journal of Statistics*, 11(1), 49–65.