A vectorial notion of skewness and its use in testing for multivariate symmetry

N. Balakrishnan[1], M. R. Brito[2] & A. J. Quiroz[3]

Abstract By modifying the statistic of Malkovich and Afifi (1973), we introduce and study the properties of a notion of multivariate skewness that provides both a magnitude and an overall direction for the skewness present in multivariate data. This notion leads to a test statistic for the null hypothesis of multivariate symmetry. Under mild assumptions, we find the asymptotic distribution of the test statistic and evaluate, by simulations, the convergence of the finite sample size quantiles to their limits, as well as the power of the statistic against some alternatives.
Keywords: Multivariate skewness, tests for symmetry, asymptotic distributions.

1 Introduction

When reporting the skewness of a univariate distribution, it is customary to indicate its direction by talking of skewness 'to the left' (negative) or 'to the right' (positive). It seems natural, that in the multivariate setting, one would like as well to indicate a direction for the skewness of a distribution or a data set. There seems to be a lack for a vectorial notion of skewness in the literature, even though various measures of multivariate skewness have been proposed (see Chapter 44 in Kotz, Balakrishnan and Johnson, 2000), including those based on the notion of median balls introduced by Avérous and Meste (1997). Perhaps closer to the spirit of the present work is the 'geometric notion of quantiles' presented by Chaudhuri (1996).

Malkovich and Afifi (1973) introduced one of the most popular measures of multivariate skewness, defined as follows for an i.i.d. sample $X_1, X_2, \dots, X_n$ of random points in $\mathbb{R}^d$. Let Ω_d denote the unit d-dimensional sphere, $\{x \in \mathbb{R}^d : \|x\| = 1\}$. For $u \in \Omega_d$, let $b_1(u)$ denote the measure of skewness of the sample, in the u direction,

[1]Dpt. of Mathematics and Statistics, McMaster University
[2]Dpto. de Matemáticas Puras y Aplicadas, Universidad Simón Bolivar
[3]CESMa & Dpto. de Cómputo Científico y Estadística, Universidad Simón Bolivar

given by

$$b_{1,n}(u) = \frac{n\left(\sum_{1\le i\le n}\left(u^t(X_i-\overline{X})\right)^3\right)^2}{\left[\sum_{1\le i\le n}\left(u^t(X_i-\overline{X})\right)^2\right]^3}, \tag{1}$$

where $\overline{X}$ is the sample mean. Malkovich and Afifi's multivariate measure of skewness is

$$b_{1,n}^* = \sup_{u\in\Omega_d} b_{1,n}(u). \tag{2}$$

Let $S = S_X$ denote the sample covariance matrix for the X_i sample, and let $S^{-1/2}$ be the inverse of its square root, that is, $S(S^{-1/2})^2 = (S^{-1/2})^2S = S^{-1/2}SS^{-1/2} = I_d$, where I_d is the $d\times d$ identity matrix. In what follows, we will assume that all square roots and square root inverses of positive definite real symmetric matrices are symmetric matrices computed in the (natural) way described in Section 4. Denote by Z_i, $1\le i\le n$, the standardized sample obtained through Mahalanobis's transformation, i.e., $Z_i = S^{-1/2}(X_i-\overline{X})$. Malkovich and Afifi point out that the computation of $b_{1,n}^*$ is equivalent to that of

$$\sup_{u\in\Omega_d}\left(\frac{1}{\sqrt{n}}\sum_{1\le i\le n}\left(u^tZ_i\right)^3\right)^2, \tag{3}$$

a fact which simplifies significantly the computation of the statistic in (2).

Each value $c_{1,n}(u) = \frac{1}{\sqrt{n}}\sum_{1\le i\le n}\left(u^tZ_i\right)^3$ in (3) can be seen as a signed measure of skewness of the standardized sample in the direction of u (if negative it will indicate skewness in the direction of $-u$). Thus, the vector $uc_{1,n}(u)$ provides a vectorial indication of skewness in the u (or $-u$) direction. Summing these vectors over u will give us an overall vectorial measure of skewness for the Z_i sample. Therefore, we want to consider the statistic

$$T_n = \int_{\Omega_d} uc_{1,n}(u)d\lambda(u), \tag{4}$$

where λ denotes the rotationally invariant probability measure on the unit sphere. It turns out that the computation of T_n is rather simple and, when the distribution of X is symmetric (in a sense to be specified below) it has, under some moment assumptions, a gaussian asymptotic distribution with a limiting covariance matrix, D, that can be consistently estimated from the Z_i sample. Then, if $\tilde{D}$ is the sample estimator for D, the quadratic form $Q_n = T_n^t\tilde{D}^{-1}T_n$ will have a limiting χ^2 distribution with d degrees of freedom, under the null hypothesis of symmetry, and can be used in testing the X_i sample for symmetry. Here, symmetry is taken to mean that for every

$u \in \Omega_d$, the distribution of the random variable $u^t X_i$ is symmetric around its center, in the usual univariate sense.

In the next two sections, we will discuss some of the properties of T_n and Q_n, including their calculation, asymptotics and behaviour on some examples. Some results will be stated for which proofs will be presented in Section 4.

2 Computation and properties of T_n and Q_n

In what follows, we will need the integrals of some monomials over the unit sphere Ω_d. Let us write x_j for the j-th coordinate of a point $x \in \Omega_d$. The values of the following integrals are obtained using Theorem 3.3 of Fang, Kotz and Ng (1990):

$$J_4 = \int_{\Omega_d} x_j^4 d\lambda(x) = \frac{3}{d(d+2)},\ J_2 = \int_{\Omega_d} x_j^2 d\lambda(x) = \frac{1}{d}$$
$$\text{and } J_{2,2} = \int_{\Omega_d} x_j^2 x_l^2 d\lambda(x) = \frac{1}{d(d+2)}, \tag{5}$$

for $j \neq l$, $1 \leq j, l \leq d$. The integrals do not depend on the particular choices of j and l. Let us write $Z_{i,j}$ for the j-th coordinate of Z_i. Direct calculation shows that the r-th coordinate of T_n is given by

$$T_{n,r} = \sqrt{n} J_4 \frac{1}{n} \sum_{i \leq n} Z_{i,r}^3 + 3\sqrt{n} \sum_{j \neq r} J_{2,2} \frac{1}{n} \sum_{i \leq n} Z_{i,j}^2 Z_{i,r}. \tag{6}$$

The following proposition expresses the type of invariance enjoyed by T_n. If an affine transformation $Y_i = MX_i + c$, for some non-singular M, is applied to the original data, the value of T_n will be rotated, by an orthonormal matrix which, in case M is a rotation, will coincide with M. All proofs are deferred to Section 4.

Proposition 2.1 Let T_n be the statistic given by (4) calculated on the sample $X_1, \ldots, X_n$. Denote by S_X the sample covariance matrix of the X_i's and by $S_X^{1/2}$ its square root, computed as indicated in Section 4. We assume that S_X is non-singular. Let M and c be, respectively, a non-singular $d \times d$ matrix and a vector in $\mathbb{R}^d$. Let T_n' be the statistic given by (4), computed on the sample $Y_1, \ldots, Y_n$, where $Y_i = MX_i + c$. Then, $T_n' = UT_n$, where $U = (MS_X M^t)^{-1/2} M S_X^{1/2}$ is an orthonormal matrix. If the eigenvalues of S_X are all different and $M \in O(d)$ (the orthogonal group), then $U = M$.

Note: When the sample comes from a distribution with a density with respect to Lebesgue measure, the eigenvalues of S_X will all be different, with probability 1, as can be easily checked.

In order to illustrate the behaviour of T_n on skewed data, consider the following bivariate example. Using the R Language, we generated a sample of 200 data points with independent coordinates, each coordinate with the exp(1) distribution. This data set, after Mahalanobis's standardization, is shown in Figure 1. In this Figure, as well as in the original data (not shown) a marked skewness to the 'north' and to the 'east' is evident, so that one should expect an overall indicator of skewness to point in the 'northeast' direction. When T_n is computed for this data set, its value turns out to be $(10.367, 9.041)$, a vector pointing, approximately, in the 41^{O} direction. Calculation of the quadratic form Q_n produces the value 13.909, that can be interpreted as strong evidence of asymmetry for this data set, when compared to the quantiles of the χ^2 distribution with 2 degrees of freedom.

For symmetric data, T_n has a limiting gaussian distribution, as the following Theorem states. Let X be a random vector with the same distribution of $X_1, X_2, \ldots$, and let $\mu = \mathbf{E}X$. We will call the distribution of X symmetric, when for every $u \in \Omega_d$ and $t > 0$, $\Pr(u^t(X-\mu) \geq t) = \Pr(u^t(X-\mu) \leq -t)$ (the one-dimensional projections are symmetric in the usual univariate sense).

Theorem 2.2 Suppose the X_i's form an i.i.d. sample from a symmetric distribution, with $\mathbf{E}\,\|X_1\|^6 < \infty$. Denote by Σ the covariance matrix of the X_i's, which we assume to be non-singular, and let $W_i = \Sigma^{-1/2}(X_i - \mu)$. The W_i are symmetric with mean 0 and covariance matrix the identity, I_d. Consider the following moments of coordinates of W_i. For different indices $r, j, k \leq d$, let

$$\begin{aligned} &m_r^4 = \mathbf{E}W_{i,r}^4,\ \ m_r^6 = \mathbf{E}W_{i,r}^6,\ \ m_{j,r}^{2,2} = \mathbf{E}W_{i,j}^2 W_{i,r}^2, \\ &m_{j,r}^{2,4} = \mathbf{E}W_{i,j}^2 W_{i,r}^4 \text{ and } m_{j,k,r}^{2,2,2} = \mathbf{E}W_{i,j}^2\,W_{i,k}^2\,W_{i,r}^2. \end{aligned} \tag{7}$$

Then, as $n \to \infty$, the distribution of T_n converges to a mean zero, d-dimensional gaussian distribution, with diagonal covariance matrix D=diag$(\sigma_{1,1}, \ldots, \sigma_{d,d})$, where

$$\begin{aligned} \sigma_{r,r} = J_4^2 m_r^6 + 9J_{2,2}^2 \left[\sum_{j\neq r} m_{r,j}^{2,4} + 2 \sum_{j<k;j,k\neq r} m_{j,k,r}^{2,2,2}\right] + \\ 9J_2^2 + 6J_4 J_{2,2} \sum_{j\neq r} m_{j,r}^{2,4} - 6J_4 J_2 m_r^4 - 18 J_2 J_{2,2} \sum_{j\neq r} m_{j,r}^{2,2}. \end{aligned} \tag{8}$$

The reading of Theorem 2.2 must be cautious in the sense that it is not saying that the limiting distribution of T_n is independent of Σ. The variable W_i has covariance matrix equal to the identity, but it is not necessarily rotationally invariant, and the moments in (7) might depend on the particular standardization being carried out to produce the W_i's, which depends on Σ (see also Proposition 2.1). Still, our next result does give us an asymptotic distribution for Q_n which is independent of Σ and other parameters of the distribution of X_1. It is also worth mentioning that the theoretical analysis leading to Theorem 2.2 parallels, to a certain extent, the analysis carried out by Baringhaus and Henze (1991) in their study of properties of $b^*_{1,n}$ under the hypothesis of elliptical symmetry.

Theorem 2.3 Under the assumptions of Theorem 2.2, the moments in (7) are consistently estimated by their sample counterparts computed on the standardized Z_i sample. For example, $m^{2,2}_{j,r}$ is consistently estimated by

$$\tilde{m}^{2,2}_{j,r} = \frac{1}{n} \sum_{1 \leq i \leq n} Z^2_{i,j} Z^2_{i,r}.$$

Substituting these estimated moments in formula (8), we obtain a consistent estimator $\tilde{D}$ of D, and the quadratic form $Q_n = T_n \tilde{D}^{-1} T_n$ has a limiting χ^2 distribution with d degrees of freedom.

In view of Theorems 2.2 and 2.3, we expect to use Q_n as a test statistic for the null hypothesis of multivariate symmetry and, when the value of Q_n suggests strong evidence of asymmetry in the data, T_n will indicate the overall direction of the skewness present in the standardized Z_i sample. In the following section, we evaluate, via Monte Carlo simulations, the convergence of quantiles of Q_n to their limiting values in the bivariate context for various distributions satisfying the conditions of Theorem 2.3, and evaluate the power of Q_n, as a statistic for testing the null hypothesis of bivariate symmetry, against some alternatives.

3 Simulation analysis of the behaviour of Q_n

In order to evaluate the performance of Q_n, we carried out a bivariate simulation analysis. We analyzed first the behavior of quantiles of Q_n, when the conditions of Theorem 2.3 hold. The following distributions were considered, which offer a variety of 'shapes' within the family of symmetric distributions:

- Uniform: The Uniform distribution on the unit square $[0, 1]^2$.

- Double exponential: The distribution with independent coordinates, each coordinate with the dexp(1) (double exponential) distribution.
- Standard Gaussian: The standard gaussian distribution.
- Symmetric Mixture: A fifty-fifty mixture of the standard gaussian distribution and the $N(\mu, I_2)$ distribution, for $\mu = (3, 3)^t$.

For each distribution in this set, and each sample size $n = 50, 100, 200$, a total of $m = 10,000$ samples were produced, using the R Statistical Environment (for a description of R see Ihaka and Gentleman, 1996). For each sample, T_n and Q_n were computed, according to the procedure outlined in the previous sections, and from the m values of Q_n, approximate quantiles were obtained. The results are displayed in Table 1, where we can appreciate the (moderately fast) convergence of the quantiles to their limiting values for all the distributions considered. In all cases, the Monte Carlo quantiles are smaller than the limiting values, which appear in the last row of the Table, labelled 'symmetric', suggesting that the use of the limiting χ^2 quantiles will result in a conservative test for the null hypothesis of bivariate symmetry. In the same experiment just described, we computed, in each case, the empirical cummulative distribution function at the theoretical limiting quantiles, i.e., the percentage of values of Q_n which were less than or equal to the given quantile. The resulting probabilities, expressed as percentages, are shown in Table 2. In this Table we can see how the agreement between the MonteCarlo and nominal probabilities generally improves with sample size, being quite acceptable at $n = 200$.

In order to evaluate the power of the procedure proposed as a test for symmetry, we considered the following bivariate alternatives:

- Moonshape: The Uniform distribution on the region of the plane limited by the curves $y = 2x(1 - x)$ and $y = 4x(1 - x)$.
- Exponential: The distribution with independent coordinates, each coordinate with the exp(1) distribution.
- Contaminated Gaussian 1: The mixture defined by $0.9\, N(0, I_2) + 0.1\, N(\mu, I_2)$ distribution, for $\mu = (3, 3)^t$.
- Contaminated Gaussian 2: The mixture defined by $0.95\, N(0, I_2) + 0.05\, N(\mu, I_2)$ distribution, for $\mu = (2, 2)^t$.

These alternatives present different forms of departure from the null hypothesis of symmetry. The last two are included to evaluate the ability of Q_n to detect contamination on gaussian data, with the last one being a more difficult alternative, since it presents less contamination in terms of percentage (of data contaminated) and magnitude of the contamination. Power was estimated, at the 5% level, by generating 1000 samples from each distribution and each sample size ($n = 50, 100, 200$), calculating Q_n for each sample, and comparing with the 95% quantile of the asymptotic χ^2 distribution with 2 degrees of freedom (5.991). The resulting approximate power values are displayed, as percentages, in Table 3. We observe that Q_n offers very good power against the first three alternatives, even for sample size $n = 50$. The last alternative turns out to be a more difficult one, as expected, and Q_n shows very little power against it for $n = 50$, a case in which the expected number of contaminated points is only 2.5. Yet, against this alternative power does improve with sample size, and a reasonable amount of power is obtained for sample size $n = 200$.

4 Proofs of results

For a real symmetric positive definite $d \times d$ matrix A, let Γ be a matrix whose columns form an orthonormal basis of eigenvectors of A. Then, $E = \Gamma^t A \Gamma$ is a diagonal matrix, with the eigenvalues of A on its diagonal. We require Γ to be such that the eigenvalues of A appear in increasing order in the diagonal of E. Let $E^{1/2}$ and $E^{-1/2}$ be the matrices obtained by taking square root of the elements in the diagonal of E, and by taking the reciprocal of the square root of the diagonal elements of E, respectively. Then, as is well known, $\Gamma E^{1/2} \Gamma^t$ is a symmetric square root of A and $\Gamma E^{-1/2} \Gamma^t$ is a symmetric square root inverse of A. We assume in this Section, that square roots and inverse square roots are computed, when required, in this fashion.

Proof of Proposition 2.1 Let S_Y denote the sample covariance matrix of the Y_i sample. Then $S_Y = M S_X M^t$, and T_n' can be written in this case as

$$\begin{aligned} T_n' &= \int_{\Omega_d} \frac{1}{n} \sum_{1 \le i \le n} \left(u^t (M S_X M^t)^{-1/2} M S_X^{1/2} S_X^{-1/2} (X_i - \overline{X}) \right)^3 u du \\ &= \int_{\Omega_d} \frac{1}{n} \sum_{1 \le i \le n} \left(u^t U S_X^{-1/2} (X_i - \overline{X}) \right)^3 u du, \qquad (9) \end{aligned}$$

for $U = (M S_X M^t)^{-1/2} M S_X^{1/2}$. Now, $U U^t = (M S_X M^t)^{-1/2} M S_X M^t (M S_X M^t)^{-1/2}$ $= I_d$, showing that U is orthogonal. Then, the change of variables $v = U^t u$ in equation (9) finishes the proof of the first claim in Proposition 2.1. Suppose now that $MM^t = I_d$ and the eigenvalues of S_X are all distinct. We can write

$$(i)\quad MS_XM^t = \Gamma_1 E\,\Gamma_1^t \text{ and } (ii)\quad S_X = \Gamma_2 E\,\Gamma_2^t,$$

for orthogonal matrices Γ_1 and Γ_2, as described at the beginning of this section. The diagonal matrix E is the same in both expressions, since M is orthogonal. Now, S_X can be solved for in (i), yielding

$$S_X = M^t\Gamma_1 E\,\Gamma_1^t M = \Gamma_2 E\,\Gamma_2^t.$$

Since the eigenvalues of S_X are distinct, Γ_2 is unique, and it follows that $M^t\Gamma_1 = \Gamma_2$. Using this fact and our convention for computing square roots of matrices, we have

$$U = (MS_XM^t)^{-1/2}MS_X^{1/2} = \Gamma_1 E^{-1/2}\Gamma_1^t M\Gamma_2 E^{1/2}\Gamma_2^t = \Gamma_1\Gamma_2^t = M,$$

as we wanted to prove.

The next two Lemmas will be used in the proof of Theorem 2.2. Lemma 4.2 provides an approximation to the stochastic process $R_n(u) = n^{-1/2}\sum_{1\le i\le n}(u^tZ_i)^3$, considered as a process indexed on $u \in \Omega_d$. The approximation obtained is the same one obtained by Baringhaus and Henze (1991, Lemma 2.2) for the related process $\sqrt{nb_{1,n}(u)}$. Still, we decided to include a proof of this fact, since the process being approximated is a different one and we are working in more generality (we do not assume elliptical symmetry). In the statement of Lemma 4.1 and some of the proofs that follow, we use the language and results from the Theory of Empirical Processes. When necessary, the reader can consult the texts of van der Vaart and Wellner (1996), or Pollard (1984).

In what follows, we assume that $X_1, \dots, X_n$ form an i.i.d. sample from a symmetric distribution on $I\!R^d$, with mean μ and positive definite covariance matrix Σ. We write A for its symmetric square root, $\Sigma^{1/2}$. We consider the variables $W_i = A^{-1}(X_i - \overline{X})$, which are i.i.d. symmetric, with mean 0 and covariance matrix I_d. We call P the distribution of the W_i. We denote by S_X the sample covariance matrix of the X_i's and by S_W the sample covariance matrix of the W_i's. We assume that P has sixth moment, that is, $\mathbf{E}\,\|W_i\|^6 < \infty$.

Lemma 4.1 Let $\mathcal{K}$ denote a compact neighborhood of the identity matrix, I_d. We assume $\mathcal{K}$ small enough to guarantee that all the eigenvalues of matrices in $\mathcal{K}$ are positive and bounded away from 0. Then

(a) $\sup_{u\in\Omega, B\in\mathcal{K}} |\frac{1}{n}\sum_{i\leq n}(u^t BW_i)^2 - \mathbf{E}(u^t BW_1)^2| \to 0$, in probability, as $n\to\infty$.

(b) To $u\in\Omega$ and $B\in\mathcal{K}$ associate the function $g_{u,B}: I\!R^d \to I\!R$, defined by $g_{u,B}(w) = u^t Bw$. The collection of functions $\mathcal{G} = \{g_{u,B} : u\in\Omega,\ B\in\mathcal{K}\}$ is a P-Donsker class of functions.

(c) To $u\in\Omega$ and $B\in\mathcal{K}$ associate the function $h_{u,B}: I\!R^d \to I\!R$, defined by $h_{u,B}(w) = (u^t Bw)^3$. The collection of functions $\mathcal{H} = \{h_{u,B} : u\in\Omega,\ B\in\mathcal{K}\}$ is a P-Donsker class of functions.

Proof: We will prove only (c), since the arguments leading to (a) and (b) are quite similar. Since $\mathcal{K}$ is compact, there is a $\kappa > 0$ such that, for all $h_{u,B}\in\mathcal{H}$, we have $|h_{u,B}(w)| \leq \kappa\|w\|^3$. Furthermore, by our moment assumption, $\kappa\|w\|^3$ is in $L^2(P)$. The subgraph of $h_{u,B}$ is the subset of $I\!R^d\times I\!R$ defined by

$$\mathrm{subg}(h_{u,B}) = \{(w,t)\in I\!R^d\times I\!R : t\leq h_{u,B}(w)\}$$

(see van der Vaart and Wellner, 1996). By the definition of $h_{u,B}$, $\mathrm{subg}(h_{u,B})$ can be written as the positivity set of a polynomial, $p_{u,B}(w,t)$, of degree 3, on the coordinates of w and the variable t. It follows from Theorem 3.1 in Wenocur and Dudley (1981), that $\mathcal{H}$ is a VC-subgraph class and, using Theorems 2.6.7 and 2.5.2 of van der Vaart and Wellner (1996), we get that $\mathcal{H}$ is a P-Donsker class.

Lemma 4.2 Let X_i and W_i, $1\leq i\leq n$ be as stated above. The process $R_n(u) = n^{-1/2}\sum_{1\leq i\leq n}(u^t Z_i)^3$, as a process indexed by $u\in\Omega_d$, is approximated as

$$R_n(u) = \frac{1}{\sqrt{n}}\sum_{1\leq i\leq n}(u^t W_i)^3 - \frac{3}{\sqrt{n}}\sum_{1\leq i\leq n} u^t W_i + o_{\mathrm{Pr}}(1),$$

where the $o_{\mathrm{Pr}}(1)$ term is uniform over Ω_d.

Proof: We have

$$u^t Z_i = u^t S_X^{-1/2}(X_i - \overline{X}) = u^t (AS_W A)^{-1/2} A(W_i - \overline{W}) = u^t B_n(W_i - \overline{W}), \tag{10}$$

where the matrix $B_n = (AS_W A)^{-1/2}A$ converges, in probability, to the identity I_d, by the consistency of S_W and the continuity of the 'square root inverse' operator over the set of symmetric positive definite real matrices. Then,

$$R_n(u) = \frac{1}{\sqrt{n}}\sum_{i\leq n}(u^t B_n W_i)^3 - 3\frac{1}{\sqrt{n}}\sum_{i\leq n}(u^t B_n W_i)^2(u^t B_n\overline{W})$$

$$+3\frac{1}{\sqrt{n}}\sum_{i\le n}(u^tB_nW_i)(u^tB_n\overline{W})^2-\sqrt{n}(u^tB_n\overline{W})^3$$

$$=(i)+(ii)+(iii)+(iv). \tag{11}$$

For term (iv) in (11) we have the bound

$$\sup_u|\sqrt{n}(u^tB_n\overline{W})^3|=K_n\sqrt{n}\|\overline{W}\|^3=\mathrm{O}_{\mathrm{Pr}}(n^{-1}), \tag{12}$$

where K_n is the largest eigenvalue of B_n (bounded in probability) and we have used that $\|\overline{W}\|=\mathrm{O}_{\mathrm{Pr}}(n^{-1/2})$. Now, from Lemma 4.1(b), we have that $\sup_u|\frac{1}{\sqrt{n}}\sum_{i\le n}(u^tB_nW_i)|=\mathrm{O}_{\mathrm{Pr}}(1)$, while $\sup_u(u^tB_n\overline{W})^2=\mathrm{O}_{\mathrm{Pr}}(n^{-1})$ by the same argument used for term (iv). Thus, we get the following bound for (iii):

$$\sup_u|\frac{3}{\sqrt{n}}\sum_{i\le n}(u^tB_nW_i)(u^tB_n\overline{W})^2|=\mathrm{O}_{\mathrm{Pr}}(n^{-1}). \tag{13}$$

The approximation

$$\sup_u|\frac{1}{\sqrt{n}}\sum_{i\le n}(u^tB_nW_i)^3-\frac{1}{\sqrt{n}}\sum_{i\le n}(u^tW_i)^3|=\mathrm{o}_{\mathrm{Pr}}(1) \tag{14}$$

for term (i) follows from the asymptotic equicontinuity condition implied by Lemma 4.1(c). Since $\mathbf{E}(u^tBW_i)^2$ is a uniformly continuous function of u and B, $B_n\to I_d$, in probability, and $\mathbf{E}(u^tW_i)^2\equiv 1$, for every $u\in\Omega_d$, we have, by Lemma 4.1(a),

$$\sup_u|\frac{1}{n}\sum_{i\le n}(u^tB_nW_i)^2-1|=\mathrm{o}_{\mathrm{Pr}}(1), \tag{15}$$

while the approximation

$$\sup_u|\sqrt{n}(u^tB_n\overline{W})-\frac{1}{\sqrt{n}}\sum_{i\le n}(u^tW_i)|=\mathrm{o}_{\mathrm{Pr}}(1) \tag{16}$$

follows by Lemma 4.1(b). From (15) and (16) we have that

$$\sup_u|3\frac{1}{\sqrt{n}}\sum_{i\le n}(u^tB_nW_i)^2(u^tB_n\overline{W})-3\frac{1}{\sqrt{n}}\sum_{i\le n}(u^tW_i)|=\mathrm{o}_{\mathrm{Pr}}(1). \tag{17}$$

Putting together (14), (17), (13) and (12) finishes the proof of Lemma 4.2.

Proof of Theorem 2.2 By Lemma 4.2, we have

$$T_n=\int_{\Omega_d}u\frac{1}{\sqrt{n}}\sum_{i\le n}\Big((u^tW_i)^3-3u^tW_i\Big)\,d\lambda(u)+\mathrm{o}_{\mathrm{Pr}}(1). \tag{18}$$

¿From (18), direct calculation shows that the r-th coordinate of T_n is given by

$$T_{n,r} = \sqrt{n}J_4\,\frac{1}{n}\sum_{i\leq n} W_{i,r}^3 + 3\sqrt{n}\sum_{j\neq r} J_{2,2}\,\frac{1}{n}\sum_{i\leq n} W_{i,j}^2 W_{i,r} - 3\sqrt{n}J_2\,\frac{1}{n}\sum_{i\leq n} W_{i,r} + o_{\mathrm{Pr}}(1), \quad (19)$$

and Theorem 2.2 follows, after some algebra, by application of the Multivariate Central Limit Theorem.

Proof of Theorem 2.3 It suffices to show that the moments in (7) are consistently estimated by their sample counterparts computed on the standardized Z_i sample. Since all the proofs are similar, we will consider only the estimation of $m_{j,r}^{2,4} = \mathbf{E}W_{i,j}^2 W_{i,r}^4$ by $\tilde{m}_{j,r}^{2,4} = \frac{1}{n}\sum_{i\leq n} Z_{i,j}^2 Z_{i,r}^4$.

¿From (10), $\tilde{m}_{j,r}^{2,4}$ can be written as

$$\frac{1}{n}\sum_{i\leq n}(b_{n,j}^t(W_i - c_n))^2(b_{n,r}^t(W_i - c_n))^4,$$

where $b_{n,j}^t$ and $b_{n,r}^t$ are, respectively, the j-th and the r-th rows of matrix B_n, while $c_n = \overline{W}$. Recall that $b_{n,j} \to e_j$, in probability, where e_j is the vector with a 1 in the j-th position and zeroes in all other positions. Similarly, $b_{n,r} \to e_r$, while $c_n \to 0$, in probability. Let $\mathcal{N}_j$ (resp. $\mathcal{N}_r$) denote a compact neighborhood of e_j (resp. e_r), and let $\mathcal{V}$ denote a compact neighborhood of the vector $0 \in \mathbb{R}^d$. To each triplet $(b_j, b_r, c) \in \mathcal{N}_j \times \mathcal{N}_r \times \mathcal{V}$, associate a function $f_{b_j,b_r,c} : \mathbb{R}^d \to \mathbb{R}$, defined by $f_{b_j,b_r,c}(w) = (b_j^t(w - c))^2(b_r^t(W_i - c))^4$. Notice that $\mathbf{E}\,f_{e_j,e_r,0}(W_1) = m_{j,r}^{2,4}$. Consider the class of functions

$$\mathcal{F} = \{f_{b_j,b_r,c} : (b_j, b_r, c) \in \mathcal{N}_j \times \mathcal{N}_r \times \mathcal{V}\}.$$

By an application of the triangle inequality, it suffices to prove two things:

(a) $\sup_{f_{b_j,b_r,c}\in\mathcal{F}} |\frac{1}{n}\sum_{i\leq n} f_{b_j,b_r,c}(W_i) - \mathbf{E}\,f_{b_j,b_r,c}(W_1)| \to 0$, in probability, and

(b) $\mathbf{E}\,f_{b_j,b_r,c}(W_1)$ is continuous, as a function of the argument $(b_j, b_r, c) \in \mathcal{N}_j \times \mathcal{N}_r \times \mathcal{V}$, at $(b_j, b_r, c) = (e_j, e_r, 0)$.

For both (a) and (b), we want the class $\mathcal{F}$ to have an envelope function in $L^1(P)$. Since the neighborhoods considered are compact, it is not hard to see that there exists a $\kappa > 0$ such that, for all $f_{b_j,b_r,c} \in \mathcal{F}$ and $w \in \mathbb{R}^d$, $|f_{b_j,b_r,c}(w)| \leq \kappa(1 + \|w\|^6)$. Our moment assumption guarantees that the function $\kappa(1 + \|w\|^6)$ is in $L^1(P)$. Then, (b) is obtained by an application of the Dominated Convergence Theorem. To prove (a), one can proceed as in the proof of Lemma 4.1(c): We observe that the subgraph of a function $f_{b_j,b_r,c} \in \mathcal{F}$, given by

$$\operatorname{subg}(f_{b_j,b_r,c}) = \{(w,s) \in I\!R^d \times I\!R : s \leq f_{b_j,b_r,c}(w)\},$$

can be written as the positivity set of a polynomial, $p_{b_j,b_r,c}(w,s)$, of degree 6, on the coordinates of w and the variable s. Then, $\mathcal{F}$ is a VC-subgraph class by Theorem 3.1 in Wenocur and Dudley (1981), and the result follows.

References

Avérous, J. and Meste, M. (1997) Median Balls: An Extension of the Interquantile Intervals to Multivariate Distributions *Journal of Multivariate Analysis* **63** 222-241.

Baringhaus, L. and Henze, N. (1991) Limit Distributions for Measures of Multivariate Skewness and Kurtosis Based on Projections *Journal of Multivariate Analysis* **38**, 51-69.

Chaudhuri, P. (1996) On a Geometric Notion of Quantiles for Multivariate Data *Journal of the American Statistical Association* **91**, 862-872.

Fang, K. T., Kotz, S. and Ng, K. W. (1990) *Symmetric Multivariate and Related Distributions.* Monographs on Statistics and Applied Probability **36**, Chapman and Hall, London.

Ihaka, R. and Gentleman, R. (1996) R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* **5**, 299-314.

Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000) *Continuous Multivariate Distributions. Volume 1: Methods and Applications. Second Edition*, John Wiley and Sons, New York.

Malkovich, J. F. and Afifi, A. A. (1973) On Tests for Multivariate Normality *Journal of the American Statistical Association* **68**, 176-179.

Pollard, D. (1984) *Convergence of Stochastic Processes.* Springer-Verlag, New York.

van der Vaart, A. and Wellner, J. A. (1996) *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer, New York.

Wenocur, R. S. and Dudley, R. M. (1981) Some special Vapnik-Chervonenkis classes. *Discrete Mathematics* **33** 313-318.

Table 1: Monte Carlo quantiles for Q_n

Distribution	Sample size	90%	92.5%	95%
Uniform	50	3.741	4.181	4.798
	100	4.175	4.655	5.270
	200	4.279	4.740	5.542
Double exponential	50	4.532	4.954	5.536
	100	4.607	5.112	5.673
	200	4.503	4.939	5.582
Standard Gaussian	50	4.154	4.637	5.169
	100	4.391	4.805	5.414
	200	4.400	4.899	5.578
Symmetric Mixture	50	3.867	4.304	4.835
	100	4.249	4.685	5.311
	200	4.389	4.923	5.565
Symmetric	∞	4.605	5.181	5.991

Table 2: Monte Carlo cummulative probabilities for Q_n at limit quantiles

Distribution	Sample size	90%	92.5%	95%
Uniform	50	94.3	96.2	97.9
	100	92.3	94.7	96.7
	200	91.9	93.9	96.1
Double exponential	50	90.6	93.6	96.5
	100	90.0	92.9	96.1
	200	90.6	93.5	96.1
Standard Gaussian	50	92.2	95.1	97.5
	100	91.4	94.0	96.6
	200	91.0	93.6	96.1
Symmetric Mixture	50	94.0	96.1	97.9
	100	92.1	94.5	96.8
	200	91.1	93.5	96.0

Table 3: Approximate power against asymmetric alternatives

Alternative	$n = 50$	$n = 100$	$n = 200$
Moonshape	64.9	98.4	100
Exponential	72.8	92.7	99.0
Contam. Gaussian 1	57.5	95.4	100
Contam. Gaussian 2	6.7	20.6	44.1