Journal of Multivariate Analysis **71**, 1–23 (1999)
Article ID jmva.1999.1823, available online at http://www.idealibrary.com on IDEAL

Shortcomings of Generalized Affine Invariant Skewness Measures

Steffen Gutjahr, Norbert Henze and Martin Folkers

Universität Karlsruhe, Karlsruhe, Germany

Received February 2, 1996

This paper studies the asymptotic behavior of a generalization of Mardia's affine invariant measure of (sample) multivariate skewness. If the underlying distribution is elliptically symmetric, the limiting distribution is a finite sum of weighted independent χ^2-variates, and the weights are determined by three moments of the radial distribution of the corresponding spherically symmetric generator. If the population distribution has positive generalized skewness a normal limiting distribution occurs. The results clarify the shortcomings of generalized skewness measures when used as statistics for testing for multivariate normality. Loosely speaking, normality will be falsely accepted for a short-tailed non-normal elliptically symmetric distribution, and it will be correctly rejected for a long-tailed non-normal elliptically symmetric distribution. The wrong diagnosis in the latter case, however, would be rejection due to positive skewness.

AMS 1991 subject classifications: 62H10, 62H15, 60F05.

Key words and phrases: multivariate skewness; test for multivariate normality; affine invariance; elliptically symmetric distribution.

1. INTRODUCTION AND SUMMARY

Let $\{X_j\}_{j=1}^n$ be independent copies of a d-dimensional $(d \geqslant 2)$ random column vector X with expectation $EX = \mu$ and nonsingular covariance matrix $\Sigma = E(X-\mu)(X-\mu)'$, where the prime means transpose. Furthermore, let

$$\bar{X} = n^{-1} \sum_{j=1}^{n} X_j, \qquad S = n^{-1} \sum_{j=1}^{n} (X_j - \bar{X})(X_j - \bar{X})' \tag{1.1}$$

denote the mean vector and the sample covariance matrix of $X_1, ..., X_n$, respectively. Assume S is nonsingular with probability one. This will hold, for example, if X has a continuous distribution and $n \geqslant d+1$ (see Eaton and Perlman [5]).

0047-259X/99 $30.00

This paper analyzes the asymptotic behavior of $b_k^d = b_k^d(X_1, ..., X_n)$, where

$$b_k^d = \frac{1}{n^2} \sum_{i, j=1}^{n} \{(X_i - \bar{X})' S^{-1}(X_j - \bar{X})\}^{2k+1}$$

and k is a fixed integer. Note that b_1^d is Mardia's measure of multivariate skewness (Mardia [19, 20]) for which limiting distributions were obtained by Baringhaus and Henze [2]. The generalized skewness measure b_k^d emerges as the leading term of the $(2k+1)$th component of the smooth test of fit for multivariate normality (see Klar [17]). An appealing feature of b_k^d is its invariance with respect to full-rank affine transformations of $X_1, ..., X_n$. Obviously, the affine invariant population counterpart of b_k^d is

$$\beta_k^d = E\{(X_1 - \mu)' \Sigma^{-1}(X_2 - \mu)\}^{2k+1}.$$

Under the tacit standing assumption $E\,|X|^{4k+2} < \infty$, where $|\cdot|$ is the Euclidean norm in $\mathbb{R}^d$, β_k^d is well-defined, and mixed moments of sufficiently high order exist.

To give another expression for β_k^d, let

$$W = \Sigma^{-1/2}(X - \mu), \tag{1.2}$$

where $\Sigma^{-1/2}$ is the symmetric positive definite square root of Σ^{-1}, so that $EW = 0$ and $EWW' = I_d$, the unit matrix of order d. Writing $W = (W_1, ..., W_d)'$, we have

$$\beta_k^d = \sum_{t_1, ..., t_{2k+1}=1}^{d} \left(E \prod_{j=1}^{2k+1} W_{t_j} \right)^2,$$

which shows that β_k^d is nonnegative. Moreover, regarded as a functional on a set of probability distributions, β_k^d vanishes within the class $\mathcal{N}_d$ of nondegenerate d-variate normal laws. It is thus tempting to use b_k^d as a statistic for testing the hypothesis H_0 that the distribution of X belongs to $\mathcal{N}_d$ by rejecting H_0 for large values of b_k^d. This approach, however, has the drawback that β_k^d is zero not only in case of multivariate normality, but also within the much wider semiparametric class of elliptically symmetric distributions (see Section 2). In this respect, b_k^d shares a general property of components of smooth tests of fit that is crucial with regard to a proper "diagnostic" interpretation of the results of such tests (see Henze [10] and Henze and Klar [13]).

The main result is that, under elliptical symmetry,

$$n \cdot b_k^d \xrightarrow{\mathcal{D}} \sum_{j=0}^{k} \alpha_{j,k} \cdot \chi^2_{v(2k+1-2j)}, \tag{1.3}$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution. Here

$$v(q) = \binom{d-3+q}{d-2} + \binom{d-2+q}{d-2} \tag{1.4}$$

is the number of linearly independent surface harmonics of degree q in $\mathbb{R}^d$ (see Erdélyi *et al.* [6]), and $\chi^2_{v(1)}, \chi^2_{v(3)}, ..., \chi^2_{v(2k+1)}$ are independent χ^2-variates with degrees of freedom $v(1)$, $v(3)$, ..., $v(2k+1)$, respectively. The weights $\alpha_{0,k}, ..., \alpha_{k,k}$ depend only on d and m_{2k}, m_{2k+2}, m_{4k+2}, where $m_r = E\,|W|^r$, and W is given in (1.2). The basic idea to get (1.3) is to approximate b_k^d by a V-statistic. Since the resulting kernel is degenerate, the asymptotic distribution of nb_k^d is determined by the eigenvalues of a certain integral operator. In the present case this operator is finite-dimensional, and the pertinent integral equation may be solved explicitly (see Section 2). Some examples of elliptically symmetric distributions will be considered in Section 3.

In Section 4 we show that, if $\beta_k^d > 0$ and the support of the underlying distribution has positive Lebesgue measure,

$$\sqrt{n}\,(b_k^d - \beta_k^d) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$

where the variance σ^2 of the limiting normal distribution depends on the distribution of W in a way to be made explicit. As an example, we consider mixtures of normal distributions with equal covariance matrices.

The final section addresses the diagnostic limitations of generalized skewness measures when used as statistics for testing for multivariate normality.

2. THE LIMIT LAW OF b_k^d FOR ELLIPTICALLY SYMMETRIC DISTRIBUTIONS

Throughout this section, the distribution of X is elliptically symmetric, i.e., X is the affine image of a spherically distributed vector Y (see Fang *et al.* [7]). Since b_k^d and β_k^d are affine invariant, there is no loss of generality in assuming that the distribution of X is spherically symmetric (for short: spherical), i.e., X has the same distribution as HX for any orthogonal $d \times d$ matrix H. This implies $EX = \mathbf{0}$, where $\mathbf{0}$ is the origin in $\mathbb{R}^d$, and (without loss of generality) $EXX' = I_d$ so that $E\,|X|^2 = d$. A further standing assumption is $P(X = \mathbf{0}) = 0$. As a consequence, we have the decomposition

$$X = |X| \cdot U, \qquad U = \frac{X}{|X|}, \tag{2.1}$$

where $|X|$ and U are independent, and the distribution of U is uniform on the surface of the unit d-sphere (see Fang *et al.* [7, p. 72]). The mixed moments of $U = (U_1, ..., U_d)'$ are given by

$$E\left(\prod_{j=1}^{d} U_j^{r_j}\right) = \frac{1}{(d/2)^{[l]}} \prod_{j=1}^{d} \frac{(2l_j)!}{4^{l_j} l_j!} \tag{2.2}$$

if $r_j = 2l_j$ (say) are even $(1 \leqslant j \leqslant d)$, $l = \sum_{j=1}^{d} l_j$, and $x^{[l]} = x(x+1) \cdot \cdots \cdot (x+l-1)$ (see Fang *et al.* [7, p. 72]). Note that the left-hand side of (2.2) is zero if at least one of the r_j is odd.

In the sequel, $Z_n = O_P(a_n)$ and $Z_n = o_P(a_n)$ denote tightness and convergence in probability to zero of Z_n/a_n, respectively, for a sequence Z_n of random vectors and positive real numbers a_n. For example (recall S from (1.1)), the multivariate Central Limit Theorem (CLT) gives

$$S - I_d - n^{-1/2} A_n = O_P(n^{-1}), \tag{2.3}$$

where

$$A_n = n^{-1/2} \sum_{j=1}^{n} (X_j X_j' - I_d), \tag{2.4}$$

and thus

$$S^{-1} = I_d + O_P(n^{-1/2}). \tag{2.5}$$

Writing $X_j = (X_j^{(1)}, ..., X_j^{(d)})'$, $1 \leqslant j < \infty$, and $X = (X^{(1)}, ..., X^{(d)})'$, we state two auxiliary results which will be used repeatedly. The proofs follow by straightforward algebra from (2.1), (2.2) and the law of large numbers.

LEMMA 2.1. *For any $l \in \{1, 2, ..., 4k+2\}$ and $i_1, ..., i_l \in \{1, 2, ..., d\}$, we have*

(a) $$\frac{1}{n} \sum_{j=1}^{n} \prod_{\nu=1}^{l} X_j^{(i_\nu)} = E\left(\prod_{\nu=1}^{l} X^{(i_\nu)}\right) + o_P(1) = O_P(1).$$

(b) *In particular,*

$$\frac{1}{n} \sum_{j=1}^{n} \prod_{\nu=1}^{l} X_j^{(i_\nu)} = o_P(1)$$

if the distribution of X is spherical and l is odd.

LEMMA 2.2. *Let l_1, l_2, l_3 be nonnegative integers such that $l_1+l_2+l_3 \leqslant 2k+1$. If the distribution of X is spherical, then*

$$\frac{1}{n^2} \sum_{i,j=1}^{n} (X_i'X_j)^{l_1} (X_j'\bar{X})^{l_2} (X_i'\bar{X})^{l_3} = \begin{cases} o_P(1), & \text{if } \; l_2+l_3>0 \text{ or } l_1 \text{ odd} \\ O_P(1), & \text{otherwise.} \end{cases}$$

To highlight the key idea of the proof, let

$$V_{n,k}(X_1, ..., X_n) = \frac{1}{n^2} \sum_{i,j=1}^{n} h_k(X_i, X_j) \tag{2.6}$$

denote the V-statistic with kernel

$$h_k(x, y) = (x'y)^{2k+1} \qquad (x, y \in \mathbb{R}^d), \tag{2.7}$$

and put

$$h_{1,k}(x) = Eh_k(x, X) \qquad (x \in \mathbb{R}^d). \tag{2.8}$$

The fact that $h_{1,k} \equiv 0$ for spherically distributed X implies a nondegenerate limiting distribution for $nV_{n,k}$ (see, e.g., Gregory [8]). Since

$$b_k^d = V_{n,k}(S^{-1/2}(X_1 - \bar{X}), ..., S^{-1/2}(X_n - \bar{X})),$$

where $S^{-1/2}$ is the positive definite symmetric square root of S^{-1}, is a V-statistic with estimated parameters, the idea is to perform an asymptotic expansion in order to replace b_k^d by an asymptotically equivalent V-statistic without estimated parameters. This will be accomplished at the cost of a replacement of the kernel h_k by a more complicated kernel h_k^* given below.

PROPOSITION 2.3. *If the distribution of X is spherically symmetric with unit covariance matrix, then*

$$n \cdot b_k^d = \frac{1}{n} \sum_{i,j=1}^{n} h_k^*(X_i, X_j) + o_P(1), \tag{2.9}$$

where

$$h_k^*(x, y) = (x'y)^{2k+1} - \frac{(2k+1)!\, m_{2k}}{k!\, (d/2)^{[k]}\, 4^k} (|x|^{2k} + |y|^{2k}) \cdot x'y$$
$$+ \frac{(d+2k)(2k+1)!\, m_{2k}^2}{k!\, d(d/2)^{[k]}\, 4^k} \cdot x'y$$

and $m_r = E\,|X|^r$, $1 \leqslant r \leqslant 4k+2$. *Moreover,*

$$Eh_k^*(x, X) = 0, \qquad x \in \mathbb{R}^d. \tag{2.10}$$

Proof. The first step is to use (2.5) in order to get

$$n \cdot b_k^d = \frac{1}{n} \sum_{i,\, j=1}^{n} \{(X_i - \bar{X})'\,(X_j - \bar{X})\}^{2k+1} + o_P(1). \tag{2.11}$$

To this end, letting $Z_i = X_i - \bar{X} = (Z_i^{(1)}, ..., Z_i^{(d)})'$, (2.5) gives

$$n \cdot b_k^d = \frac{1}{n} \sum_{i,\, j=1}^{n} (Z_i' Z_j)^{2k+1} + \sum_{r=0}^{2k} \binom{2k+1}{r} D_r,$$

where $D_r = n^{-1} \sum_{i,\, j=1}^{n} (Z_i' Z_j)^r\,(Z_i' O_P(n^{-1/2})\,Z_j)^{2k+1-r}$. Since $D_{2k} = o_P(1)$ implies $D_r = o_P(1)$ for each $r \leqslant 2k$, we have to show $D_{2k} = o_P(1)$. Now, $D_{2k} = tr(O_P(n^{-1/2})\,R_{2k})$, where *tr* denotes trace and

$$\begin{aligned} R_{2k} &= \frac{1}{n} \sum_{i,\, j=1}^{n} Z_i (Z_i' Z_j)^{2k}\, Z_j' \\ &= \sum_{t_1, ..., t_{2k}=1}^{d} \left(n^{-1/2} \sum_{i=1}^{n} Z_i^{(t_1)} \cdots Z_i^{(t_{2k})} Z_i \right) \\ &\quad \times \left(n^{-1/2} \sum_{j=1}^{n} Z_j^{(t_1)} \cdots Z_j^{(t_{2k})} Z_j \right)'. \end{aligned}$$

Noting that $E(Z_1^{(t_1)} \cdots Z_1^{(t_{2k})} Z_1) = 0$ by spherical symmetry, the CLT and Slutzky's lemma yield $R_{2k} = O_P(1)$ and hence $D_{2k} = o_P(1)$, as was to be shown.

From (2.11) we obtain

$$n \cdot b_k^d = \sum \frac{(2k+1)!\,(-1)^{l_2+l_3}}{l_1!\, l_2!\, l_3!\, l_4!}\, T_n(l_1, l_2, l_3, l_4) = o_P(1), \tag{2.12}$$

where $T_n(l_1, l_2, l_3, l_4) = n^{-1} \sum_{i,\, j=1}^{n} (X_i' X_j)^{l_1}\,(X_j' \bar{X})^{l_2}\,(X_i' \bar{X})^{l_3}\,(\bar{X}' \bar{X})^{l_4}$ and the sum in (2.12) is over all nonnegative integers $l_1, ..., l_4$ satisfying $\sum_{v=1}^{4} l_v = 2k+1$. Invoking Lemma 2.2, a careful analysis shows that $T_n(l_1, l_2, l_3, l_4)$ is asymptotically negligible except for the cases

(1) $l_1 = 2k+1$,

(2) $l_1 = 2k$, $l_2 = 1$,

(3) $l_1 = 2k$, $l_3 = 1$,

(4) $l_1 = 2k$, $l_4 = 1$,

(5) $l_1 = 2k-1$, $l_2 = l_3 = 1$.

We claim that

$$T_n(2k, 1, 0, 0) = \frac{(2k)!\, m_{2k}}{k!\,(d/2)^{[k]}\, 4^k} \cdot \frac{1}{n} \sum_{i,\, j=1}^{n} |X_j|^{2k}\, X_i' X_j + o_P(1), \tag{2.13}$$

$$T_n(2k, 0, 0, 1) = \frac{(2k)!\, m_{2k}^2}{k!\,(d/2)^{[k]}\, 4^k} \cdot \frac{1}{n} \sum_{i,\, j=1}^{n} X_i' X_j + o_P(1), \tag{2.14}$$

$$T_n(2k-1, 1, 1, 0) = \frac{(2k)!\, m_{2k}^2}{k!\, d(d/2)^{[k]}\, 4^k} \cdot \frac{1}{n} \sum_{i,\, j=1}^{n} X_i' X_j + o_P(1) \tag{2.15}$$

(these terms correspond to the cases (2), (4), and (5), respectively). Note that cases (2) and (3) are symmetrical with respect to each other. On combining (2.12)–(2.15), (2.9) follows. To show (2.13), observe that $T_n(2k, 1, 0, 0)$ equals

$$\frac{1}{n} \sum_{j,\, l=1}^{n} \left[\sum_{p_1, \ldots, p_{2k}=1}^{d} E\left(\prod_{\nu=1}^{2k} X^{(p_\nu)} \right) \prod_{\nu=1}^{2k} X_j^{(p_\nu)} \right] X_j' X_l + o_P(1). \tag{2.16}$$

For $i = 1, \ldots, d$, let $n_i = n_i(p_1, \ldots, p_{2k}) = \#\{ j: 1 \leqslant j \leqslant 2k,\ p_j = 1\}$. Note that the expectation in (2.16) vanishes if at least one of the n_i is odd. Putting $n_i = 2l_i$ for $1 \leqslant i \leqslant d$, the squared-bracket term in (2.16) is

$$\sum_{\substack{l_1, \ldots, l_d \geqslant 0 \\ l_1 + \cdots + l_d = k}} \frac{(2k)!}{(2l_1)! \cdots (2l_d)!} \cdot E\left(\prod_{\nu=1}^{d} X^{(\nu)2l_\nu} \right) \prod_{\nu=1}^{d} X_j^{(\nu)2l_\nu}, \tag{2.17}$$

where, using (2.1), (2.2), and the definition of m_r,

$$E\left(\prod_{\nu=1}^{d} X^{(\nu)2l_\nu} \right) = \frac{m_{2k}(2l_1)! \cdots (2l_d)!}{(d/2)^{[k]}\, 4^k l_1! \cdots l_d!}.$$

Plugging this into (2.17) and using the multinomial theorem, (2.13) follows.

To show (2.14), note that, up to terms of order $o_P(1)$, $T_n(2k, 0, 0, 1)$ equals $E(X_1' X_2)^{2k} \cdot n^{-1} \cdot \sum_{i,\, j=1}^{n} X_i' X_j$. Putting

$$U_j = X_j / |X_j| = (U_j^{(1)}, \ldots, U_j^{(d)})', \qquad j = 1, 2, \tag{2.18}$$

and noting that $U_1' U_2$ and $U_2^{(1)}$ have the same distribution, (2.1) and (2.2) yield $E(X_1' X_2)^{2k} = E\,|X_1|^{2k}\, E\,|X_2|^{2k}\, E(U_2^{(1)^{2k}}) = m_{2k}^2 (2k)!/(4^k k!\,(d/2)^{[k]})$, from which (2.14) follows. To prove (2.15), observe that

$$T_n(2k-1, 1, 1, 0) = \frac{1}{n} \sum_{l,\, m=1}^{n} X_l' \left[\frac{1}{n^2} \sum_{i,\, j=1}^{n} (X_i' X_j)^{2k-1} \cdot X_i X_j' \right] X_m,$$

where the matrix within squared brackets equals

$$E[(X_1' X_2)^{2k-1} X_1 X_2'] + o_P(1). \tag{2.19}$$

With U_j defined in (2.18), we have

$$E[(X_1' X_2)^{2k-1} X_1 X_2'] = m_{2k}^2 \cdot E[(U_1' U_2)^{2k-1} U_1 U_2'].$$

Since $E[(U_1' U_2)^{2k-1} U_1^{(i)} U_2^{(j)}] = 0$ if $i \neq j$, and

$$E[(U_1' U_2)^{2k-1} U_1^{(i)} U_2^{(i)}] = \frac{1}{d} \cdot E(U_1' U_2)^{2k} = \frac{1}{d} \cdot E(U_2^{(1)^{2k}})$$

by symmetry, the reasoning given above shows that the matrix of expectations in (2.19) is $c \cdot I_d$, where $c = m_{2k}^2 \cdot (2k)!/(d \cdot (d/2)^{[k]} 4^k k!)$. This yields the assertion and completes the proof of (2.9). Since (2.10) follows from the spherical symmetry of X, the proof of Proposition 2.3 is completed. ∎

THEOREM 2.4. *Suppose X has an elliptically symmetric distribution with expectation μ and nonsingular covariance matrix Σ. If $E\{(X-\mu)' \Sigma^{-1}(X-\mu)\}^{2k+1} < \infty$ and $P(X=\mu)=0$, then*

$$n b_k^d(X_1, ..., X_n) \xrightarrow{\mathscr{D}} \sum_{j=0}^{k} \alpha_{j,k} \chi^2_{\nu(2k+1-2j)},$$

where $\chi^2_{\nu(1)}, \chi^2_{\nu(3)}, ..., \chi^2_{\nu(2k+1)}$ are independent χ^2-variates with degrees of freedom $\nu(1), ..., \nu(2k+1)$, respectively, and $\nu(q)$ is defined in (1.4). The weights $\alpha_{0,k}, ..., \alpha_{k,k}$ are given by

$$\alpha_{j,k} = \frac{(2k+1)!\, m_{4k+2}}{2j!\, (d/2)^{[2k-j+1]}\, 4^k} \qquad (j = 0, 1, ..., k-1)$$

and

$$\alpha_{k,k} = \frac{(2k+1)!}{k!\, d(d/2)^{[k]}\, 4^k} \cdot \left(\frac{d m_{4k+2}}{d+2k} - 2 m_{2k} m_{2k+2} + (d+2k)\, m_{2k}^2 \right),$$

where $m_{2j} = E\{(X-\mu)' \Sigma^{-1}(X-\mu)\}^j$ $(1 \leqslant j \leqslant 2k+1)$.

Proof. Although Theorem 2.4 has been stated in a general form, assume (recall affine invariance) as before that the distribution of X is spherical with unit covariance matrix, so that $m_{2j} = E|X|^{2j}$. From Proposition 2.3 and the general theory of V-statistics (see, e.g., Gregory [8]), it follows that $n b_k^d \xrightarrow{\mathscr{D}} \sum_{l \geqslant 1} \lambda_l N_l^2$, where $(N_l)_{l \geqslant 1}$ are independent standard normal random variables, and $\{\lambda_l : l \geqslant 1\}$ are the nonzero eigenvalues corresponding to the operator $g \mapsto Ag(x) = \int h_k^*(x, y)\, g(y)\, P^X(dy)$ acting on the Hilbert space $L_2(P^X)$ of measurable functions on $\mathbb{R}^d$ that are square-integrable with respect to P^X. To determine λ_l, we have to solve the equation

$$\int h_k^*(x, y) \cdot g_l(y)\, P^X(dy) = \lambda_l \cdot g_l(x) \qquad P^X\text{-a.s.}, \tag{2.20}$$

where $\{g_l : l \geqslant 1\}$ is the set of associated orthonormal eigenfunctions, i.e., $E(g_l(X)\, g_m(X)) = 1$ if $l = m$ and is zero, otherwise. Generally, there is little hope to solve an integral equation like (2.20) even in the univariate case and for a "simple" kernel. In the present case, however, there is an explicit solution in terms of spherical harmonics (see Baringhaus and Henze [2] for the case $k = 1$).

To this end, use the decomposition $X = R \cdot U$, where $R = |X|$ is independent of $U = X/|X|$. Let G be the uniform distribution over $\mathcal{S}^{d-1}$, the surface of the unit d-sphere, and write F for the distribution of R. Denoting the left-hand side of (2.20) by I_l, Fubini's theorem gives

$$I_l = \iint h_k^*(x, sw)\, g_l(sw)\, dG(w)\, dF(s).$$

Putting $x = rz$, where $r > 0$ and $z \in \mathcal{S}^{d-1}$, and inserting the expression for h_k^*, it follows that

$$\begin{aligned} I_l = \iint [&r^{2k+1} s^{2k+1} (z'w)^{2k+1} - a_k (r^{2k} + s^{2k})\, rsz'w \\ &+ b_k rsz'w]\, g_l(sw)\, dG(w)\, dF(s), \end{aligned} \tag{2.21}$$

where

$$a_k = \frac{(2k+1)!\, m_{2k}}{k!\, (d/2)^{[k]}\, 4^k}, \qquad b_k = \frac{(d+2k)(2k+1)!\, m_{2k}^2}{k!\, d (d/2)^{[k]}\, 4^k}.$$

To solve the equation $I_l = \lambda_l \cdot g_l(rz)$ with I_l in (2.21), consider the case $d \geqslant 3$ first. Let $C_q^\gamma(t)$ be the Gegenbauer polynomial of degree q and order $\gamma = (d-2)/2$ (see Erdélyi *et al.* [6, p. 174]). Inserting the expression

$$t^{2l+1} = \frac{(2l+1)!}{2^{2l+1}} \cdot \sum_{m=0}^{l} \frac{2l+1+\gamma-2m}{m!\, \gamma^{[2l+2-m]}}\, C_{2l+1-2m}^\gamma(t) \tag{2.22}$$

(see Magnus *et al.* [18, p. 227]) for $t = z'w$ and $t^{2k+1} = (z'w)^{2k+1}$ into (2.21), some algebra gives

$$\begin{aligned}
I_l = {} & \frac{(2k+1)!}{2^{2k+1}} r^{2k+1} \sum_{m=0}^{k-1} \frac{2k+1+\gamma-2m}{m!\,\gamma^{[2k+2-m]}} \\
& \times \left[\int s^{2k+1} \int C^{\gamma}_{2k+1-2m}(z'w) \right] g_l(sw)\, dG(w)\, dF(s) \\
& + \frac{(2k+1)!}{k!\,(d-2)(d/2)^{[k]}\, 4^k} \iint \left[\frac{dr^{2k+1}s^{2k+1}}{d+2k} - m_{2k}(r^{2k}+s^{2k})\, rs \right. \\
& \left. + \frac{(d+2k)\, m_{2k}^2}{d} rs \right] C^{\gamma}_1(z'w)\, gl(sw)\, dG(w)\, dF(s).
\end{aligned}$$

The crucial point now is that there is a complete system $\{\varphi_{q,k}\colon k = 1, 2, ..., v(q);\ q \geqslant 0\}$ of orthonormal continuous functions $\varphi_{q,k} \in \mathscr{L}_2(G)$ such that

$$C^{\gamma}_q(z'w) = \frac{\gamma}{\gamma+q} \sum_{k=1}^{v(q)} \varphi_{q,k}(z)\, \varphi_{q,k}(w) \qquad (z, w \in \mathscr{S}^{d-1}), \tag{2.23}$$

where $v(q)$ is given in (1.4) (see Erdélyi *et al.* [6, p. 243]). This equation enables us to separate the variables z and w. Inserting (2.23) into the last expression for I_l yields

$$\begin{aligned}
I_l = {} & \frac{(2k+1)!}{2^{2k+1}} r^{2k+1} \sum_{m=0}^{k-1} \frac{\gamma}{m!\,\gamma^{[2k+1-m]}} \sum_{i=1}^{v(2k+1-2m)} \\
& \times \left[\int s^{2k+1} \int \varphi_{2k+1-2m,i}(z)\, \varphi_{2k+1-2m,i}(w) \right] g_l(sw)\, dG(w)\, dF(s) \\
& + \frac{(2k+1)!}{k!\, d(d/2)^{[k]}\, 4^k} \sum_{i=1}^{v(1)} \iint \left[\frac{dr^{2k+1}s^{2k+1}}{d+2k} - m_{2k}(r^{2k}+s^{2k})\, rs \right. \\
& \left. + \frac{(d+2k)\, m_{2k}^2}{d} rs \right] \varphi_{1,i}(z)\, \varphi_{1,i}(w)\, g_l(sw)\, dG(w)\, dF(s).
\end{aligned}$$

It is now easily verified that for each $m \in \{0, 1, ..., k-1\}$, the functions

$$g_{m,i}(sw) = \frac{s^{2k+1}\varphi_{2k+1-2m,i}(w)}{\sqrt{m_{4k+2}}} \qquad (i = 1, 2, ..., v(2k+1-2m))$$

are orthonormal eigenfunctions of (2.20) with associated eigenvalue

$$\lambda_{2k+1-2m} = \frac{(2k+1)!\, m_{4k+2}}{2m!\,(d/2)^{[2k-m+1]}\, 4^k},$$

where the multiplicity of $\lambda_{2k+1-2m}$ is $v(2k+1-2m)$. Moreover,

$$g_{k,i}(sw)=\frac{(s^{2k+1}-((d+2k)/d)\,m_{2k}s)\,\varphi_{1,i}(w)}{(m_{4k+2}-2((d+2k)/d)\,m_{2k}m_{2k+2}+(d+2k)^2\,m_{2k})^{1/2}}$$

$(i=1, ..., v(1))$ are orthonormal eigenfunctions of (2.20) with associated eigenvalue

$$\lambda_1=\frac{(2k+1)!}{k!\,d(d/2)^{[k]}\,4^k}\left(\frac{dm_{4k+2}}{d+2k}-2m_{2k}m_{2k+2}+(d+2k)\,m_{2k}^2\right).$$

To be sure that all nonzero eigenvalues of h_k^* have been obtained, it remains to prove

$$\sum_{m=0}^{k} v(2k+1-2m)\,\lambda_{2k+1-2m}=Eh_k^*(X,X) \tag{2.24}$$

(see Serfling [24, p. 226]). Using (1.4), straightforward algebra shows that (2.24) is equivalent to the combinatorial identity

$$\sum_{m=0}^{k}\frac{(d-2+2k-2m)!\,(d+4k-4m)}{m!\,2^m(1+2k-2m)!\,\prod_{v=0}^{2k-m}(d+2v)}=\frac{(d-2)!}{(2k+1)!}, \tag{2.25}$$

which may be proved by induction over d. If $d=2$ or $d=4$, (2.25) follows from Exercise 10 of Riordan [23, p. 34]. To show (2.25) for $d=3$, note that, for each $l=0, 1, ..., k$,

$$\sum_{m=0}^{l}\frac{3+4k-4m}{m!\,2^m\prod_{v=0}^{2k-m}(3+2v)}=\left[l!\,2^l\prod_{v=0}^{2k-l-1}(3+2v)\right]^{-1}$$

(use induction over l). Putting $l=k$ gives (2.25) for the case $d=3$. Upon writing $j=k-m$ in (2.25), we see that, for $d\geqslant 2$, (2.25) is equivalent to the condition

$$A_{d,k}=\frac{(d-1)!}{(2k+1)!}\,2^k k!\prod_{v=0}^{k-1}(d+2v),$$

where

$$A_{d,k}=\sum_{j=0}^{k}\frac{(d-2+2j)!}{(1+2j)!}\binom{k}{j}2^j j!\,\frac{(d+4j)(d-1)}{\prod_{v=k}^{k+j}(d+2v)}.$$

Using the fact that $(d+4j)(d-1)=(d+2j)(d-1+2j)-(1+2j)\,2j$, straightforward algebra yields the recursion formula $A_{d+2,k}=(d+2k)(d+1)\,A_{d,k}$ $(d\geqslant 2)$ which, together with the cases $d=2$ and $d=3$, completes the proof of (2.25).

For the case $d=2$, (2.20) may be solved by using Chebyshev polynomials $C_r(t)=\cos(r \cdot \arccos t)$ instead of Gegenbauer polynomials (see Baringhaus and Henze [2, p. 1985], for the case $k=1$). The details are omitted. ▮

3. EXAMPLES

This section illustrates the general result of Theorem 2.4 by giving explicit formulae for the weights $\alpha_{0,k}, ..., \alpha_{k,k}$ for several parametric families of elliptically symmetric distributions and the case $k=1$ and $k=2$. Under the conditions of Theorem 2.4, it follows that

$$n \cdot b_1^d \xrightarrow{\mathscr{D}} \alpha_{0,1}\chi^2_{(d-1)\,d(d+4)/6} + \alpha_{1,1}\chi^2_d,$$

where

$$\alpha_{0,1} = \frac{6 \cdot m_6}{d(d+2)(d+4)}, \qquad \alpha_{1,1} = \frac{3}{d}\left(\frac{m_6}{d+2} - 2m_4 + d(d+2)\right)$$

(see also Theorem 2.2 of Baringhaus and Henze [2]). For the case $k=2$, Theorem 2.4 yields

$$n \cdot b_2^d \xrightarrow{\mathscr{D}} \alpha_{0,2}\chi^2_{(d-1)\,d(d+1)(d+2)(d+8)/120} + \alpha_{1,2}\chi^2_{(d-1)\,d(d+4)/6} + \alpha_{2,2}\chi^2_d,$$

where

$$\alpha_{0,2} = \frac{120 \cdot m_{10}}{d(d+2)(d+4)(d+6)(d+8)}, \qquad \alpha_{1,2} = \frac{60 \cdot m_{10}}{d(d+2)(d+4)(d+6)},$$

$$\alpha_{2,2} = \frac{15}{d^2(d+2)}\left(\frac{dm_{10}}{d+4} - 2m_4 m_6 + (d+4)\, m_4^2\right).$$

For notational simplicity, elliptically symmetric families will be given in their (reduced) spherically symmetric form with unit covariance matrix.

3.1. *The Normal Distribution*

In case of normality, we have

$$m_4 = d(d+2), \qquad m_6 = d(d+2)(d+4),$$

$$m_{10} = d(d+2)(d+4)(d+6)(d+8)$$

and thus

$$\alpha_{0,1} = \alpha_{1,1} = 6,$$

$$\alpha_{0,2} = 120, \qquad \alpha_{1,2} = 60(d+8), \qquad \alpha_{2,2} = 120(d+5).$$

Hence Mardia's result [19] $nb_1^d \xrightarrow{\mathscr{D}} 6\chi^2_{d(d+1)(d+2)/6}$ involving only one limiting χ^2-variate does not generalize to the case $k \geqslant 2$.

3.2. *Symmetric Pearson Type II Distribution*

The *symmetric multivariate Pearson Type II distribution* (see Fang *et al.*, [7, Sect. 3.4]), in its reduced form, has the density

$$f(x) = c(\kappa) \cdot \left(1 - \frac{|x|^2}{d+2\kappa+2}\right)^{\kappa} \cdot \mathbf{1}\{|x|^2 \leqslant d+2\kappa+2\},$$

where $\kappa > -1$, $c(\kappa)$ is a norming constant, and $\mathbf{1}\{\cdot\}$ denotes the indicator function. The normal distribution is approached in the limit as $\kappa \to \infty$. For the Pearson Type II distribution, we have

$$m_4 = d(d+2) \cdot \frac{d+2+2\kappa}{d+4+2\kappa},$$

$$m_6 = d(d+2)(d+4) \cdot \frac{(d+2+2\kappa)^2}{(d+4+2\kappa)(d+6+2\kappa)},$$

$$m_{10} = \frac{d(d+2)(d+4)(d+6)(d+8) \cdot (d+2+2\kappa)^4}{(d+4+2\kappa)(d+6+2\kappa)(d+8+2\kappa)(d+10+2\kappa)}$$

and thus

$$\alpha_{0,1} = 6 \cdot \frac{(d+2+2\kappa)^2}{(d+4+2\kappa)(d+6+2\kappa)},$$

$$\alpha_{1,1} = 6 - \frac{6[d(d+6+2\kappa)+16(\kappa+1)]}{(d+4+2\kappa)(d+6+2\kappa)},$$

$$\alpha_{0,2} = 120 \cdot \frac{(d+2+2\kappa)^4}{\prod_{\nu=2}^{5}(d+2\nu+2\kappa)},$$

$$\alpha_{1,2} = 60(d+8) \cdot \frac{(d+2+2\kappa)^k}{\prod_{\nu=2}^{5}(d+2\nu+2\kappa)},$$

$$\alpha_{2,2} = \frac{240 \cdot (d+2\kappa+2)^2 \cdot A}{(d+4+2\kappa)\prod_{\nu=2}^{5}(d+2\nu+2\kappa)},$$

where

$$A = (\kappa+2)\,d^3 + (4\kappa^2+16\kappa+28)\,d^2 + (4\kappa^3+34\kappa^2+66\kappa+112)\,d$$
$$+\,20\kappa^3+64\kappa^2+76\kappa+128.$$

The weights $\alpha_{i,\,j}$ are increasing functions of κ and converge to the "normal" values given in 3.1 as $\kappa \to \infty$. Hence, for the Pearson Type II family, the limiting distributions of nb_1^d and nb_2^d are stochastically increasing with increasing κ and stochastically bounded from above by the corresponding limits under normality. It may thus be expected that, for each distribution of the Pearson Type II family, the asymptotic power of a test of the hypothesis H_0 of multivariate normality that rejects H_0 for large values of b_1^d or b_2^d will be below a chosen nominal level. These theoretical findings are fully corroborated by the empirical results given in Section 5.

3.3. *Symmetric Pearson Type VII Distribution*

The *symmetric multivariate Pearson Type VII distribution*, in its reduced form, has the density

$$f(x) = c(a)\cdot\left(1+\frac{|x|^2}{2a-d-2}\right)^{-a},$$

where $a > 1 + d/2$ and $c(a)$ is a norming constant (see Fang *et al.* [7, Sect. 3.3]). This class consists of distributions with longer tails compared to those of the normal distribution, which is obtained in the limit as $a \to \infty$. Note that the standing moment condition $m_{4k+2} < \infty$ holds if $a > 2k + 1 + d/2$. For the Pearson Type VII distribution, we have

$$m_4 = d(d+2)\cdot\frac{d+2-2a}{d+4-2a},$$

$$m_6 = d(d+2)(d+4)\cdot\frac{(d+2-2a)^2}{(d+4-2a)(d+6-2a)},$$

$$m_{10} = \frac{d(d+2)(d+4)(d+6)(d+8)(d+2-2a)^4}{(d+4-2a)(d+6-2a)(d+8-2a)(d+10-2a)}.$$

Obviously, these moments agree with the corresponding moments of the Pearson Type II distribution (see 3.2) by letting $a = -\kappa$. Hence the formulae for the weights $\alpha_{i,\,j}$ given in Section 3.2 carry over to the present case by putting $\kappa = -a$. In view of the discussion at the end of Section 3.2, note that, for the long-tailed Pearson Type VII family, the limiting distributions of nb_1^d and nb_2^d are stochastically decreasing as a increases. Furthermore, the corresponding limiting distributions under normality constitute stochastic lower bounds. Consequently, although not being consistent against those distributions from the symmetric Pearson Type VII distribution that satisfy the moment condition $a > 2k + 1 + d/2$, the tests for multivariate normality based on b_1^d or b_2^d exhibit an asymptotic power that is above the nominal level, for such alternatives. These findings are in complete accordance with the finite-sample simulation results presented in Section 5.

3.4. *Symmetric Kotz Type Distribution*

As a final example, consider the density

$$f(x) = c(\rho) \cdot |x|^{2(\rho-1)} \exp\left(-\frac{2\rho + d - 2}{2d} |x|^2 \right),$$

where $\rho > 1 - d/2$ and $c(\rho)$ is a norming constant. This class of distributions is a subclass of the *symmetric Kotz type distributions* (see Fang *et al.* [7, Sect. 3.2]) and includes the normal distribution for $\rho = 1$.

Straightforward calculations yield

$$m_4 = \frac{d^2(d+2\rho)}{d+2\rho-2}, \qquad m_6 = \frac{d^3(d+2\rho)(d+2\rho+2)}{(d+2\rho-2)^2},$$

$$m_{10} = \frac{d^5(d+2\rho+6)(d+2\rho+4)(d+2\rho+2)(d+2\rho)}{(d+2\rho-2)^4}$$

and thus

$$\alpha_{0,1} = \frac{6d^2(2\rho+d)(2\rho+d+2)}{(d+2)(d+4)(2\rho+d-2)^2},$$

$$\alpha_{1,1} = 6 \cdot \frac{d^2(d+2\rho) + 8(\rho-1)^2}{(d+2)(d+2\rho-2)^2},$$

$$\alpha_{0,2} = \frac{120 \cdot d^4 \prod_{j=0}^{3} (2\rho+d+2j)}{(2\rho+d-2)^4 \prod_{j=1}^{4} (d+2j)},$$

$$\alpha_{1,2} = \frac{60 \cdot d^4 \prod_{j=0}^{3} (2\rho+d+2j)}{(2\rho+d-2)^4 \prod_{j=1}^{3} (d+2j)},$$

$$\alpha_{2,2} = \frac{120 \cdot d^2(d+2\rho)[d^2(d+2\rho+2)(d+2\rho+3) + 8(d+2\rho)(\rho-1)^2]}{(d+2)(d+4)(d+2\rho-2)^4}.$$

Since the weights $\alpha_{i,j}$ are decreasing functions of ρ, the limiting distributions of nb_1^d and nb_2^d are stochastically decreasing as ρ increases. As a consequence, the asymptotic power of tests for normality based on b_1^d or b_2^d against the symmetric Kotz distribution with parameters ρ is a decreasing function of ρ, and it attains the nominal level for the value $\rho = 1$. As before, these findings are in complete accordance with the results of a Monte Carlo study (see Section 5).

4. THE LIMITING DISTRIBUTION OF b_k^d IN THE NONDEGENERATE CASE

We now treat the case that the kernel h_k in (2.7) is nondegenerate, that is,

$$\operatorname{Var} h_{1,k}(X) > 0, \tag{4.1}$$

where $h_{1,k}$ is given in (2.8). As before, assume $EX = 0$ and $EXX' = I_d$ without loss of generality. Since

$$h_{1,k}(x_1, ..., x_d) = \sum_{t_1, ..., t_{2k+1}=1}^{d} E\left(\prod_{j=1}^{2k+1} X^{(t_j)}\right) x_{t_1} \cdot \cdots \cdot x_{t_{2k+1}},$$

the weak assumption that the support of the underlying distribution has positive d-dimensional Lebesgue measure implies the equivalence of (4.1) and the condition $\beta_k^d > 0$ (see Okamoto [21]). Let $T_{n,k} = \binom{n}{2}^{-1} \sum_{1 \leqslant i < j \leqslant n} h_k(X_i, X_j)$, and recall $V_{n,k}$ from (2.6). Since $E\,|X|^{4k+2} < \infty$, it follows that

$$V_{n,k} - T_{n,k} = o_P(n^{-1/2}). \tag{4.2}$$

In view of the asymptotic normality of U-statistics in the nondegenerate case (see, e.g., Serfling [24]), this implies a nondegenerate limiting normal distribution for $n^{1/2}(V_{n,k} - \beta_k^d)$.

PROPOSITION 4.1. *If* $EX = 0$, $EXX' = I_d$, $E\,|X|^{4k+2} < \infty$, *and* $\operatorname{Var} h_{1,k}(X) > 0$, *then*

$$b_k^d = V_{n,k} - \frac{2k+1}{\sqrt{n}}\, tr(A_n B_k) - (4k+2)\, a_k' \bar{X} + o_P(n^{-1/2}),$$

where

$$B_k = E[X_1 (X_1' X_2)^{2k} X_2'], \qquad a_k = E[(X_1' X_2)^{2k} X_2], \tag{4.3}$$

and A_n *is given in* (2.4).

Proof. Use (2.3) and Lemma 2.1 and proceed by analogy with the reasoning given in the proof of Proposition 2.3 (see also Lemma 3.1 of Baringhaus and Henze [2]). ∎

To state the limiting behavior of b_k^d, let u_k be the $(1+d^2+d)$-dimensional vector

$$u_k = (2, -\xi_k b_{11}, -\xi_k b_{12}, ..., -\xi_k b_{1d}, -\xi_k b_{21}, ..., -\xi_k b_{dd}, -2\xi_k a_k')', \tag{4.4}$$

where $\xi_k = 2k+1$, and $B_k = (b_{ij})_{1 \leqslant i, j \leqslant d}$ and a_k are defined in (4.3). Furthermore, let

$$\begin{aligned} Z_{i,k} = (&h_{1,k}(X_i) - \beta_k^d, X_i^{(1)^2} - 1, X_i^{(1)} X_i^{(2)}, ..., X_i^{(1)} X_i^{(d)}, \\ &X_i^{(2)} X_i^{(1)}, ..., X_i^{(d)^2} - 1, X_i')', \qquad i \geqslant 1, \end{aligned} \tag{4.5}$$

and write $Z_k = Z_{1,k}$.

THEOREM 4.2. *Let X have a d-variate distribution P^X with expectation μ and nonsingular covariance matrix Σ such that $E\{(X-\mu)' \Sigma^{-1}(X-\mu)\}^{2k+1} < \infty$. If $\beta_k^d > 0$ and the support of P^X has positive Lebesgue measure, then*

$$\sqrt{n}(b_k^d(X_1, ..., X_n) - \beta_k^d) \xrightarrow{\mathscr{D}} \mathscr{N}(0, \sigma^2),$$

where

$$\sigma^2 = u_k' E(Z_k Z_k') u_k, \tag{4.6}$$

and u_k, Z_k in (4.6) are computed from the standardized vector $\Sigma^{-1/2}(X-\mu)$.

Proof. In view of affine invariance, assume the conditions of Proposition 4.1. Note that $\hat{T}_{n,k} - \beta_k^d = 2n^{-1} \sum_{j=1}^n (h_{1,k}(X_j) - \beta_k^d)$, where $\hat{T}_{n,k} = \sum_{j=1}^n E(T_{n,k} \mid X_j) - (n-1)\beta_k^d$ is the Hajek projection of $T_{n,k}$. Since $E(T_{n,k} - \hat{T}_{n,k})^2 = O(n^{-2})$, (4.2) and Proposition 4.1 imply $\sqrt{n}(b_k^d - \beta_k^d) = u_k' n^{-1/2} \sum_{j=1}^n Z_{j,k} + o_P(1)$. The result then follows from the multivariate CLT and the continuous mapping theorem. ∎

As an example, consider a mixture $\mathscr{N}MIX_d(p, \mu_1, \mu_2, \Sigma)$ of normal distributions with equal covariance matrices, i.e., the distribution of $X = SY_1 + (1-S) Y_2$, where S, Y_1, Y_2 are independent, $P(S=1) = p = 1 - P(S=0)$, $0 < p < 1$, $p \neq 1/2$, $Y_j \sim \mathscr{N}_d(\mu_j, \Sigma)$, $j = 1, 2$, $\mu_1 \neq \mu_2$. Putting

$$\delta = \{(\mu_1 - \mu_2)' \Sigma^{-1}(\mu_1 - \mu_2)\}^{1/2}$$

and performing the transformation $x \mapsto H\Sigma^{-1/2}(x-\mu_2)$, where the orthogonal matrix H maps $\Sigma^{-1/2}(\mu_1-\mu_2)$ into $(\delta, 0, ..., 0)'$, we may assume (recall affine invariance) that X has the same distribution as $(T, W_2, W_3, ..., W_d)'$. Here,

$$T = \frac{S(W_1+\delta)+(1-S)\,W_1^* - p\delta}{(1+p(1-p)\,\delta^2)^{1/2}}, \tag{4.7}$$

and $W_1, W_2, ..., W_d, W_1^*$ are independent standard normal random variables, also independent of S. Note that $EX=0$ and $EXX'=I_d$.

In what follows, only the case $k=1$ will be treated. Computation of the limiting variance σ^2 requires knowledge of the moments $t_j=ET^j$ of the random variable T in (4.7). These are given by

$$t_3 = \frac{\zeta(1-2p)\,\delta^3}{(1+\zeta\delta^2)^{3/2}}, \qquad t_4 = 3 + \frac{(1-6\zeta)\,\delta^4}{(1+\zeta\delta^2)^2},$$

$$t_5 = \frac{\zeta(1-2p)[10+(1-2\zeta)\,\delta^2]\,\delta^3}{(1+\zeta\delta^2)^{5/2}} \tag{4.8}$$

and

$$t_6 = 15 + \frac{\zeta[15(1-6\zeta)\,\delta^4 + (1-5\zeta(1+2\zeta))\,\delta^6]}{(1+\zeta\delta^2)^3},$$

where $\zeta = p(1-p)$ (see also Henze [9]). Some easy calculations show that the matrix $B_1=(b_{ij})$ defined in (4.3) is given by $b_{11}=t_3^2$ and $b_{ij}=0$, otherwise. Moreover, a_1 of (4.3) takes the form $(t_3, 0, ..., 0)'$ which shows that u_1 of (4.4) is the vector $u_1=(2, -3t_3^2, 0, ..., 0, -6t_3, 0, ..., 0)'$ ending up with $d-1$ zeroes. Since $\beta_1^d=t_3^2$ and $h_{1,1}(X)=t_3T^3$, the vector $Z_1=Z_{1,1}$ of (4.5) is given by

$$Z_1 = (L, M_1, ..., M_d, X')',$$

where $L=h_{1,1}(X)-\beta_1^d$,

$$\begin{aligned} M_1 &= (T^2-1, TW_2, TW_3, ..., TW_d)', \\ M_2 &= (W_2T, W_2^2-1, W_2W_3, ..., W_2W_d)', \\ &\vdots \\ M_d &= (W_dT, W_dW_2, ..., W_dW_{d-1}, W_d^2-1)'. \end{aligned}$$

By some algebra, we get $E(LX) = (t_3 t_4, 0, ..., 0)'$, $E(LM_1) = (t_3 t_5 - t_3^2, 0, ..., 0)'$, $E(LM_j) = \mathbf{0}'$ $(j = 2, ..., d)$, $E(M_1 M_1') = \operatorname{diag}(t_4 - 1, 1, ..., 1)$, $E(M_1 M_j') = \Delta_{j1}$ $(j = 2, ..., d)$, $E(XM_1') = t_3 \Delta_{11}$ and $E(L^2) = t_3^2 t_6 - t_3^4$. Here, Δ_{lm} is shorthand for a $d \times d$ matrix having only one nonzero element $(=1)$ in its (l, m)th place. The evaluation of σ^2 in (4.6) is now straightforward, and we have the following result.

Corollary 4.3. *If the underlying distribution is the normal mixture* $\mathcal{N}MIX_d(p, \mu_1, \mu_2, \Sigma)$, $0 < p < 1$, $p \neq 1/2$, *then*

$$\sqrt{n}(b_1^d - t_3^2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$

where

$$\sigma^2 = t_3^2 \cdot (35 t_3^2 + 36 - 12 t_3 t_5 - 24 t_4 + 9 t_3^2 t_4 + 4 t_6),$$

and t_3, t_4, t_5, t_6 *are given in* (4.8).

Since $t_3 = 0$ if $p = 1/2$, the normal mixture with equal mixing probabilities provides an example of a non-elliptically symmetric distribution with zero skewness in the sense of Mardia. A careful analysis of the proof of Proposition 2.3 (to this end, observe that $t_5 = 0$ if $p = 1/2$) shows that, for this symmetric normal mixture, we have $n \cdot b_1^d = n^{-1} \sum_{i, j=1}^n \tilde{h}_1(X_i, X_j) + o_P(1)$, where $\tilde{h}_1(x, y) = (x'y)^3 - 3(|x|^2 + |y|^2)\, x'y + 3(d+2)\, x'y$ is a degenerate kernel. Hence the limiting distribution of nb_1^d is a weighted sum of independent χ^2-variates. It seems to be difficult, however, to obtain the weights (eigenvalues) in this case where the method of proof of Theorem 2.4 is not applicable.

5. CONCLUSIONS

Our interest in the generalized skewness statistic b_k^d arose from the problem of testing for multivariate normality. Since the class $\mathcal{N}_d$ of non-degenerate d-variate normal distributions is closed with respect to affine transformations, there is at least a "soft argument" for confining oneself to affine invariant statistics for this testing problem. While affine invariant and universally consistent tests for multivariate normality are available (Baringhaus and Henze [1], Csörgő [4], Henze and Zirkler [12], Henze and Wagner [14]), there is still a widespread belief that "directed" tests like, e.g., those based on measures of skewness or kurtosis, should have specific diagnostic properties. That is, directed tests are presumed to be able to assess that a distribution is non-normal and at the same time to indicate the kind of departure from normality. For example, Mardia [19, p. 523] claims that b_1^d can be used "to test $\beta_1^d = 0$ (i.e., the hypothesis of non-multivariate skewness) for large samples."

Such a "directed diagnosis," however, is not valid because the critical value of b_1^d is determined *under the distributional assumption of multivariate normality*. Consequently, rejection of H_0 due to a "too large" observed value of b_1^d casts doubt on H_0, but nothing more can be inferred, at least with regard to statistical significance. This point is of paramount importance in order to avoid typical "diagnostic pitfalls" in goodness-of-fit testing, and it has been commented on by many authors (see, e.g., Bera and John [3], Horswell and Looney [15, 16], Rayner, Best and Mathews [22], Henze and Klar [13], Henze [10, 11]).

The following results of a Monte Carlo study illustrate the diagnostic limitations of tests for multivariate normality based on generalized skewness. For the case $d = 5, n = 50$ and the nominal level of significance $\alpha = 0.05$, random samples of size n were generated from each of the families of spherically symmetric distributions considered in Section 3. The test statistics are b_1^d and b_2^d, and rejection of the hypothesis H_0 of 5-variate normality is for large values of b_1^d and b_2^d, respectively. Critical values for b_1^d and b_2^d were obtained by simulation; they are based on 50.000 Monte Carlo replications under H_0.

Each entry in Tables I through III is the number of rejections of H_0 out of 10.000 Monte Carlo replications. Table I gives the results for several distributions of the symmetric Pearson Type II family (see Subsection 3.2). Since about 500 significant cases are expected under H_0, the values exhibit very poor power of (generalized) skewness tests for this class of elliptically symmetric distributions. This behavior, of course, was anticipated from the fact that the limiting distributions of b_1^d and b_2^d are stochastically bounded from above by the corresponding limits under normality (see the discussion at the end of Subsection 3.2).

TABLE I

Number of Rejections (out of 10.000 Monte Carlo Replications) of H_0 for Symmetric Pearson II type Distributions

κ:	1	4	6	10	20	100
b_1^d:	1	3	12	43	149	408
b_2^d:	0	3	8	41	108	304

TABLE II

Number of Rejections (out of 10.000 Monte Carlo Replications) of H_0 for the Symmetric Pearson Type VII Distribution

a:	3	5	10	15	100
b_1^d:	10.000	8.675	3.014	1.636	502
b_2^d:	10.000	9.087	3.683	2.111	667

Likewise, Table II shows estimated power against several members of the symmetric Pearson Type VII family.

These results which, as compared with the values of Table I, might at first sight seem striking (note that both families are symmetric!), are in complete agreement with the discussion at the end of Subsection 3.3. The very high power for $a = 3$ and $a = 5$ is due to the fact that m_6 and m_{10} are infinite (for $a = 3$, m_4 is also infinite).

Finally, Table III shows the simulation results for the symmetric Kotz type distribution. Again, these fully corroborate the theoretical findings, as discussed at the end of Section 3.4. Note that b_2^d exhibits higher power than Mardia's measure b_2^d against Pearson Type VII and symmetric Kotz type distributions, but is inferior to b_1^d against alternatives from the Pearson Type II family.

To sum up, the generalized skewness test based on b_k^d does not reliably identify how a distribution departs from normality. Although consistent against a population distribution having positive generalized skewness (this follows from Theorem 4.2), its power against elliptically symmetric alternatives depends on three moments of the Euclidean norm of the standardized distribution (Theorem 2.4).

There will be false acceptance of H_0 for a short-tailed non-normal elliptically symmetric distribution. On the other hand, a long-tailed elliptically symmetric distribution will lead to the (correct) rejection of H_0. The wrong diagnosis in the latter case, however, would be rejection of H_0 due to positive skewness.

TABLE III

Number of Rejections (out of 10.000 Monte Carlo Replications) of H_0 for the Symmetric Kotz Type Distribution

ρ:	-0.25	0	1	2	3
b_1^d:	5.491	3.564	496	35	5
b_2^d:	5.844	3.914	512	76	23

ACKNOWLEDGMENTS

The authors are grateful to a referee for many constructive comments.

REFERENCES

1. L. Baringhaus and N. Henze, A consistent test for multivariate normality based on the empirical characteristic function, *Metrika* **35** (1988), 339–348.
2. L. Baringhaus and N. Henze, Limit distributions for Mardia's measure of multivariate skewness, *Ann. Statist.* **20** (1992), 1889–1902.
3. A. Bera and S. John, Tests for multivariate normality with Pearson alternatives, *Comm. Statist. A Theory Methods* **12** (1983), 103–107.
4. S. Csörgő, Consistency of some tests for multivariate normality, *Metrika* **36** (1989), 708–723.
5. M. L. Eaton and M. D. Perlman, The nonsingularity of generalized sample covariance matrices, *Ann. Statist.* **1** (1973), 710–717.
6. A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, "Higher Transcendental Functions," Vol. 2, McGraw–Hill, New York, 1953.
7. K.-T. Fang, S. Kotz, and K.-W. Ng, "Symmetric Multivariate and Related Distributions," Chapman & Hall, London, 1990.
8. G. G. Gregory, Large sample theory for U-statistics and tests of fit, *Ann. Statist.* **5** (1977), 110–123.
9. N. Henze, On Mardia's Kurtosis test for multivariate normality, *Comm. Statist. A Theory Methods* **23** (1994), 1031–1045.
10. N. Henze, Do components of smooth tests of fit have diagnostic properties?, *Metrika* **45** (1997), 121–130.
11. N. Henze, Limit laws for multivariate skewness in the sense of Móri, Rohatgi and Székely, *Statist. Probab. Lett.* **33** (1997), 299–307.
12. N. Henze and B. Zirkler, A class of invariant consistent tests for multivariate normality based on the empirical characteristic function, *Comm. Statist. A Theory Methods* **19** (1990), 3595–3617.
13. N. Henze and B. Klar, Properly rescaled components of smooth tests of fit are diagnostic, *Austral. J. Statist.* **38** (1996), 61–74.
14. N. Henze and Th. Wagner, A new approach to the BHEP tests for multivariate normality, *J. Multivariate Anal.* **62** (1997), 1–23.
15. R. L. Horswell and S. W. Looney, A comparison of tests for multivariate normality that are based on measures of skewness and Kurtosis, *J. Statist. Comput. Simulation* **42** (1992), 21–38.
16. R. L. Horswell and S. W. Looney, Diagnostic limitations of skewness coefficients in assessing departures from univariate and multivariate normality, *Comm. Statist. B Simul. Comp.* **22** (1993), 437–459.
17. B. Klar, "Classical and New Goodness-of-Fit Tests," Doctoral dissertation, University of Karlsruhe, 1998. [In German]
18. W. Magnus, F. Oberhettinger, and R. P. Soni, "Formulas and Theorems for the Special Functions of Mathematical Physics," Springer-Verlag, Berlin, 1966.
19. K. V. Mardia, Measures of multivariate skewness and Kurtosis with applications, *Biometrika* **57** (1970), 519–530.
20. K. V. Mardia, Assessment of multinormality and the robustness of hotelling's T^2-test, *J. Roy. Statist. Soc. Ser. C* **24** (1975), 163–171.

21. M. Okamoto, Distinctness of the eigenvalues of a quadratic form in a multivariate sample, *Ann. Statist.* **1** (1973), 763–765.
22. J. C. W. Rayner, D. J. Best, and K. L. Mathews, Interpreting the skewness coefficient, *Comm. Statist. A Theory Methods* **24** (1995), 593–600.
23. J. Riordan, "Combinatorial Identities," Wiley, New York, 1968.
24. R. Serfling, "Approximation Theorems of Mathematical Statistics," Wiley, New York, 1980.
25. F. G. Tricomi, "Vorlesungen über Orthogonalreihen," Springer-Verlag, Berlin, 1955.