The Annals of Statistics,
1983, Vol. 11, No. 4, 1232–1242

THE ASYMPTOTIC DISTRIBUTION OF PRINCIPAL COMPONENT ROOTS UNDER LOCAL ALTERNATIVES TO MULTIPLE ROOTS

BY DAVID E. TYLER

Old Dominion University

The asymptotic distribution for the principal component roots under local alternatives to multiple population roots is derived. The asymptotic theory assumes the estimate of the population covariance or scatter matrix to be asymptotically normal and to possess certain invariance properties. These assumptions are satisfied for the affine-invariant M-estimates of scatter for an elliptical distribution. The local alternative framework is used in deriving a local power function for the test for subsphericity.

1. Introduction. The asymptotic theory for the distribution of the roots of a sample covariance matrix has been studied extensively. In these studies, two different approaches have been used. One of these approaches has been surveyed recently by Muirhead (1978). This approach assumes a normal population. It involves the use of asymptotic representations for the hypergeometric function which appears in the exact joint density of the sample roots. The exact joint density of the roots was derived by James (1960). The use of asymptotic representations for the hypergeometric function in the exact density was first introduced by G. A. Anderson (1965).

Another approach is based upon expanding the sample roots about the population covariance matrix. Using this approach, Girshick (1939) and T. W. Anderson (1963) derive the asymptotic distribution of the sample principal component roots assuming a normal population. For this case, Sugiura (1976) derives subasymptotic approximation to the distribution of the sample roots. This approach has also been applied in finding the asymptotic distribution of the roots of the sample covariance matrix taken from non-normal populations by Waternaux (1976) and by Davis (1976). Subasymptotic approximations for the non-normal population case have recently been given by Fujikoshi (1980). For the non-normal population problem, only the case when all the population roots are distinct has been studied.

In this paper, the asymptotic behavior of the principal component roots for a general class of estimates of a covariance or scatter matrix is studied. The estimates are assumed to be asymptotically normal and to possess certain invariance properties. This class of estimates include the affine-invariant M-estimates for the scatter matrix when sampling from an elliptical population. Special cases of these M-estimates are the sample covariance matrix and the maximum likelihood estimates of the scatter matrix for a specific elliptical population. The class of estimates is given in Section 2, along with a review of some distribution theory for eigenvalues of spherically invariant random symmetric matrices. The affine-invariant M-estimates are reviewed in Section 3.

The asymptotic distribution of the estimates of the principal component roots is derived under a sequence of local alternatives to multiple roots. This is done in Section 4. As a special case of the local alternative framework, one obtains the asymptotic distribution of the roots of the affine invariant M-estimates of scatter for elliptical populations under arbitrary multiplicities of the population roots. The local alternative framework is novel

Received April 1981; revised April 1983.

AMS 1970 *subject classifications*. Primary 62E20, 62H25; secondary 62H10.

Key words and phrases. Affine-invariant M-estimates of scatter, elliptical distributions, local alternatives to multiple roots, principal component roots, spherically invariant random matrices, test for subsphericity.

even when applied to the roots of the sample covariance matrix from a multivariate normal sample. It answers the following question: at what rates must two neighboring population roots approach each other with respect to "sample size" in order for the asymptotic theory to treat them as distinct roots, as multiple roots or as a mixture? As an example, the results in Section 4 are applied in Section 5 to derive a local power function for the test for subsphericity.

Although the asymptotic results under local alternatives to multiple roots appear to be somewhat obvious, the crucial step in the derivation is not straightforward. For clarity, this step is proven independently in Section 6. The method of proof involves expansions for eigenprojections and a study of their truncation bounds. These expansions and bounds are reviewed within Section 6.

2. Spherically invariant random symmetric matrices. Let Σ_n be a sequence of nonrandom symmetric positive definite matrices of order p, and let Γ_n be any sequence of $p \times p$ nonrandom matrices such that $\Gamma_n'\Gamma_n = \Sigma_n^{-1}$. Let S_n be a sequence of symmetric positive definite random matrices of order p such that $Z_n = n^{1/2}\Gamma_n(S_n - \Sigma_n)\Gamma_n'$ satisfies the following assumptions.

ASSUMPTION 2.1.

(i) The distribution of Z_n is invariant under the transformation $Z_n \to QZ_nQ'$ for any orthogonal Q.

(ii) $Z_n \to Z$ in distribution, where Z is multivariate normal.

Assumption 2.1 states that the distribution of Z_n is "spherically invariant". It readily follows that the distribution of Z is also spherically invariant. Since Z is multivariate normal, its distribution can be characterized by its first two moments. Since Z is spherically invariant, its first two moments can be characterized by three parameters (Tyler, 1982, Theorem 1).

LEMMA 2.1. *There exists constants* η, σ_1, σ_2 *with* $\sigma_1 \geq 0$ *and* $\sigma_2 \geq -2\sigma_1/p$ *such that* $E(Z) = \eta I$ *and* $\text{var}\{\text{vec}(Z)\} = \sigma_1(I + K_{p,p}) + \sigma_2\text{vec}(I)\{\text{vec}(I)\}'$.

The notation used in the lemma is as follows. If B is a $b \times t$ matrix, then vec(B) is the transformation of B into the bt-dimensional vector formed by stacking the columns of B. If B is $b \times t$ and C is $c \times u$, then the Kronecker product of B and C is the $bc \times tu$ partitioned matrix $B \otimes C = [b_{jk}C]$. The commutation matrix is the $ab \times ab$ matrix $K_{a,b} = \sum_{i=1}^{a} \sum_{j=1}^{b} J_{ij} \otimes J_{ij}'$ where J_{ij} is an $a \times b$ matrix with a one in the (i, j) position and zeroes elsewhere. This notation is to be used further in Section 5. For a good overview of the algebraic properties of the "vec" transformation, the Kronecker product and the commutation matrix, see Magnus and Neudecker (1979).

The form of var{vec(Z)} given in Lemma 2.1 states that the distinct off-diagonal elements of Z are uncorrelated with each other and with the diagonal elements. Each off-diagonal element has variance σ_1. The diagonal elements have variance $2\sigma_1 + \sigma_2$ with the covariance between any two diagonal elements being σ_2. The matrix Z has density

$$f(Z) = k_0\exp[k_1\text{tr}\{(Z - \eta I)^2\} + k_2\{\text{tr}(Z - \eta I)\}^2] \tag{2.1}$$

in $\frac{1}{2}p(p + 1)$ dimensional space, where $k_1 = -(4\sigma_1)^{-1}$, $k_2 = \sigma_2(2\sigma_1 + p\sigma_2)^{-1}(4\sigma_1)^{-1}$, and $k_0 = 2^{-(1/2)(p-1)}(2 + p\sigma_2/\sigma_1)^{-1/2}(2\pi\sigma_1)^{-(1/4)p(p+1)}$. The density function of spherically invariant symmetric matrices can be expressed in terms of the eigenvalues of the matrices. In particular, (2.1) can be expressed as

$$f(Z) = k_0\exp[k_1 \textstyle\sum_{i=1}^{p} (z_i - \eta)^2 + k_2\{\sum_{i=1}^{p} (z_i - \eta)\}^2], \tag{2.2}$$

where $z_1 \geq z_2 \geq \cdots \geq z_p$ are the eigenvalues of Z.

The distribution of the eigenvalues of spherically invariant random symmetric matrices has been well studied. For random symmetric matrices in general, a useful technique is to average the density of the matrix over the orthogonal group. That is, suppose Y is a $p \times p$ random symmetric matrix with density $f(Y)$ which is absolutely continuous in $\frac{1}{2}p(p+1)$ dimensional real space. The distribution of the eigenvalues of Y is the same as the distribution of the eigenvalues of the spherically invariant random symmetric matrix V with density $f_0(V) = \int_{O(p)} f(HVH')(dH)$. The integral is with respect to the invariant measure (dH) on the group $O(p)$ of orthogonal $p \times p$ matrices normalized so that the measure of the whole group is unity. In particular, the distribution of the roots of $Y = Z + M$, where Z has density of the form given by (2.1) and M is a fixed $p \times p$ symmetric matrix is the same as the distribution of the roots of W with density

$$f_0(W) = k_0 \exp[k_1 \textstyle\sum_{i=1}^{p} (w_i^2 + m_i^2) + k_2\{\sum_{i=1}^{p} (w_i - m_i)\}^2 {}_0F_0^{(p)}(-2k_1 W_0, M_0)], \tag{2.3}$$

where $w_1 \geq w_2 \geq \cdots \geq w_p$ are the roots of $W - \eta I$ and $m_1 \geq m_2 \geq \cdots \geq m_p$ are the roots of M. The constants η, k_0, k_1 and k_2 are the same as in (2.1). The hypergeometric function ${}_0F_0^{(p)}$ can be represented by ${}_0F_0^{(p)}(-2k_1 W_0, M_0) = \int_{O(p)} \operatorname{etr}\{-2k_1 W_0 H' M_0 H\}(dH)$, where $W_0 = \operatorname{diag}(w_1, w_2, \cdots, w_p)$ and $M_0 = \operatorname{diag}(m_1, m_2, \cdots, m_p)$. The role of hypergeometric functions of matrix arguments in multivariate distribution theory and their properties are treated in detail by James (1964) and Muirhead (1978).

The following lemmas are to be applied in Section 4. The first lemma follows from the continuity property of eigenvalues of symmetric matrices. The second lemma is an obvious generalization of Theorem 13.3.1 in Anderson (1958).

LEMMA 2.2. *Let $Z_{m,n}$ be a sequence with respect to n of $q(m) \times q(m)$ random symmetric matrices, and let $\hat{\lambda}_m = (\hat{\lambda}_{1,m}, \hat{\lambda}_{2,m}, \cdots, \hat{\lambda}_{q(m),m})$ where $\hat{\lambda}_{1,m} \geq \hat{\lambda}_{2,m} \geq \cdots \geq \hat{\lambda}_{q(m),m}$ are the eigenvalues of $Z_{m,n}$. If $(Z_{1,n}, Z_{2,n}, \cdots, Z_{k,n}) \to (Z_1, Z_2, \cdots, Z_k)$ in distribution, then $(\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_k) \to (\lambda_1, \lambda_2, \cdots, \lambda_k)$ in distribution, where $\lambda_m = (\lambda_{1,m}, \lambda_{2,m}, \cdots, \lambda_{q(m),m})$ and $\lambda_{1,m} \geq \lambda_{2,m} \geq \cdots \geq \lambda_{q(m),m}$ are the eigenvalues of Z_m.*

LEMMA 2.3. *Let $\{Z_m\}_{m=1,2,\cdots,k}$ be a set of random symmetric matrices with the order of Z_m being $q(m)$ and with joint density $f(Z_1, Z_2, \cdots, Z_k)$ which is absolutely continuous in $q_0 = \frac{1}{2} \sum_{m=1}^{k} q(m)\{q(m) + 1\}$ dimensional real space. Furthermore, assume that $f(Q_1 Z_1 Q_1', Q_2 Z_2 Q_2', \cdots, Q_k Z_k Q_k') = f(Z_1, Z_2, \cdots, Z_k)$ for any set of orthogonal matrices $Q_1, Q_2, \cdots, Q_k$. The joint density of $(\lambda_1, \lambda_2, \cdots, \lambda_k)$, where $\lambda_m = (\lambda_{1,m}, \lambda_{2,m}, \cdots, \lambda_{q(m),m})$ and $\lambda_{1,m} \geq \lambda_{2,m} \geq \cdots \geq \lambda_{q(m),m}$ are the roots of Z, is*

$$h(\lambda_1, \lambda_2, \cdots, \lambda_m) = h_0 f(\Delta_1, \Delta_2, \cdots, \Delta_k) \textstyle\prod_{m=1}^{k} \prod_{j,t=1, j<t}^{q(m)} (\lambda_{j,m} - \lambda_{t,m}),$$

where $h_0^{-1} = \pi^{-(1/2)q_0} \prod_{m=1}^{k} \prod_{j=1}^{q(m)} \Gamma[\frac{1}{2}\{q(m) - j + 1\}]$ *and* $\Delta_m = \operatorname{diag}(\lambda_{1,m}, \lambda_{2,m}, \cdots, \lambda_{q(m),m})$.

3. *M*-estimates of scatter. A p-dimensional random vector $\mathbf{Y}$ has an elliptical distribution if its density is of the form $f_g(\mathbf{y}; \mu, \Omega) = |\Omega|^{-1/2} g\{(\mathbf{y} - \mu)'\Omega^{-1}(\mathbf{y} - \mu)\}$ for some positive definite symmetric matrix Ω and some nonnegative function g, where g is independent of μ and Ω. Properties of elliptical distributions have been studied by Kelker (1971). The variable $T = (\mathbf{Y} - \mu)'\Omega^{-1}(\mathbf{Y} - \mu)$ has density

$$f_T(t) = \{\pi^{(1/2)p}/\Gamma(\tfrac{1}{2}p)\} t^{(1/2)p-1} g(t). \tag{3.1}$$

If the second moments of $\mathbf{Y}$ exist, then its covariance matrix is $c\Omega$, where $c = E(T)/p$.

Let $\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_n$ be a random sample from the elliptically distributed random vector $\mathbf{Y}$. Maronna (1976) defines affine-invariant M estimates of location and scatter to be solutions to a system of equations of the form $n^{-1} \sum_i u_1(d_i)(\mathbf{Y}_i - \mu_n) = \mathbf{0}$, and $n^{-1} \sum_i u_2(d_i^2)(\mathbf{Y}_i - \mu_n)(\mathbf{Y}_i - \mu_n)' = S_n$, where $d_i^2 = (\mathbf{Y}_i - \mu_n)' S_n^{-1}(\mathbf{Y}_i - \mu_n)$. The functions u_1 and u_2 must satisfy a set of general assumptions given in Section 2 of Maronna's paper. The solutions (μ_n, S_n) are estimates for the parameters (μ, Σ) where $\Sigma = \sigma^{-1}\Omega$. The parameter σ is the solution to the equation $E\{\psi(\sigma T)\} = p$, where $\psi(t) = tu_2(t)$ and T has density (3.1). Maronna (1976) shows that $n^{1/2}(S_n - \Sigma) \to N$ in distribution, where N is a multivariate normal matrix with zero mean. If the parameter Ω depends on n, then $(S_n, \Sigma_n = \sigma^{-1}\Omega_n)$ satisfies Assumptions 2.1.i and 2.1.ii. This follows from the affine-invariant properties of the M-estimates, and of the family of densities $\{f_g\}$ for a fixed g. The parameter σ does not depend on n. For the M-estimates, the values of η, σ_1 and σ_2 in Lemma 2.1 are $\eta = 0$, $\sigma_1 = (p+2)^2\psi_1/(2\psi_1 + p)^2$ and $\sigma_2 = \psi_2^{-2}[(\psi_1 - 1) - 2(\psi_2 - 1)\psi_1\{p + (p+4)\psi_2\}/(2\psi_2 + p)^2]$, where $\psi_1 = E\{\psi_1^2(\sigma T)\}/\{p(p+2)\}$ and $\psi_2 = E\{\sigma T\psi'(\sigma T)\}/p$ (Tyler, 1982, Example 3). The values of σ, σ_1 and σ_2 for a variety of M-estimates and a variety of elliptical distributions have been tabulated by the author (Tyler, 1983, Table 1).

The estimate S_n corresponds to the sample covariance matrix whenever $u_1(d) = 1$ and $u_2(t) = 1$. For this case, $\sigma = 1$, $\sigma_1 = 1 + \kappa$ and $\sigma_2 = \kappa$ provided the fourth moments of the elliptical population exist. The parameter κ is a kurtosis parameter defined such that $3\kappa = E\{(Y_i - \mu_i)^4\}/(\Omega_{ii})^2 - 3$. If $\mathbf{Y}$ is multivariate normal, then $\kappa = 0$.

For a specific family of elliptical densities $\{f_g\}$ with g fixed, the maximum likelihood estimates are M-estimates with $u_1(d) = u_2(d^2)$ and $u_2(t) = -2g'(t)/g(t)$. Under the assumed family of densities, $\sigma = 1$, $\sigma_1 = \{\frac{1}{4}p(p+2)\}/E\{h^2(T)\}$ and $\sigma_2 = -2\sigma_1(1-\sigma_1)/\{2 + p(1-\sigma_1)\}$, where $h(t) = tg'(t)/g(t)$ (Tyler, 1982, Example 2). The multivariate t distribution on f degrees of freedom is an elliptical distribution with $g(t)\ \alpha\ (1 + t/f)^{-(1/2)(p+f)}$. For this distribution, $\sigma_1 = 1 + 2/(p+f)$ and $\sigma_2 = 2\sigma_1/f$. For $f > 2$, the covariance matrix of the multivariate t is $c\Omega$ with $c = f/(f-2)$. For $f > 4$, the kurtosis parameter is given by $1 + \kappa = (f-2)/(f-4)$.

4. The asymptotic distribution of the roots. Let the sequence (S_n, Σ_n) satisfy the conditions of assumption 2.1. In addition, for any $p \times p$ symmetric matrix M, let $\lambda_1(M) \geq \lambda_2(M) \geq \cdots \geq \lambda_p(M)$ represent the eigenvalues of M.

In this section, the asymptotic distribution of the roots of S_n are found under the following sequence of local alternatives to multiple roots:

$$\text{(4.1)} \qquad \frac{n^{1/2}\{\lambda_i(\Sigma_n) - \gamma_{m,n}\}}{\gamma_{m,n}} \to \begin{cases} \infty & i \in \mathscr{I}_r, \quad r < m \\ d_i & i \in \mathscr{I}_m \\ -\infty & i \in \mathscr{I}_r, \quad r > m \end{cases}$$

where $\mathscr{I}_1, \mathscr{I}_2, \cdots, \mathscr{I}_k$ is a partition of the set $\{1, 2, \cdots, p\}$ with $\mathscr{I}_m = \{i(m), i(m)+1, \cdots, i(m) + q(m) - 1\}$ and where $\gamma_{m,n}$ is the average of $\lambda_i(\Sigma_n)$ over $i \in \mathscr{I}_m$. Condition (4.1) implies that $\lambda_i(\Sigma_n)/\lambda_j(\Sigma_n) \to 1$ if $i \in \mathscr{I}_m$ and $j \in \mathscr{I}_m$, and that $\sum_{i\in\mathscr{I}_m} d_i = 0$. The framework (4.1) also includes the case whenever $\Sigma_n = \Sigma$ does not depend on n. For such cases, $d_i = 0$.

To obtain an asymptotic distribution, the roots of S_n are first "standardized" by defining $X_{i,n} = n^{1/2}\{\lambda_i(S_n) - \gamma_{m,n}\}/\gamma_{m,n}$. An asymptotic representation for the limiting joint distribution of $\mathbf{X}_n' = (X_{1,n}, X_{2,n}, \cdots, X_{p,n})$ is given in Theorem 4.1. This theorem shows that the asymptotic behavior of $\lambda_i(S_n)$ is not influenced by the eigenvalues of Σ_n which differ from $\lambda_n(\Sigma_n)$ by more than $O(n^{-(1/2)})$. Theorem 4.2 gives the density of the limiting distribution of $\mathbf{X}_n$.

The critical step in the proof of Theorem 4.1 is statement (4.3). As noted in the introduction, the proof of this step is technically involved. Therefore, rather than prove this step within the proof of Theorem 4.1, its proof is given separately in Section 6. For the case $p = 2$, it is possible to give a simple proof for (4.3) by using the explicit expression for the roots of a 2×2 symmetric matrix. That is, $B_{2\times 2} = (b_{ij})$ has roots $\frac{1}{2}(b_{11} + b_{22}) \pm \frac{1}{2}\{(b_{11} - b_{22})^2 + 4b_{12}^2\}^{1/2}$. However, the necessary generalization for $p > 2$ is not straightforward.

Before presenting Theorem 4.1, additional notation is needed. Let D_m be the $q(m) \times q(m)$ diagonal matrix with diagonal entries $d_{i(m)}, d_{i(m)+1}, \cdots, d_{i(m)+q(m)-1}$. Let Z be defined as in Assumption 2.1 and let

$$(4.2)\qquad Z = \begin{pmatrix} Z_{(1,1)} & Z_{(1,2)} & \cdots & Z_{(1,k)} \\ Z_{(2,1)} & Z_{(2,2)} & \cdots & Z_{(2,k)} \\ & & & \vdots \\ Z_{(k,1)} & Z_{(k,2)} & \cdots & Z_{(k,k)} \end{pmatrix}$$

represent a partitioning of the rows and columns of Z in accordance with the partition $\mathscr{I}_1, \mathscr{I}_2, \cdots, \mathscr{I}_k$. For any $p \times p$ matrix M, define $M_{(r,t)}$ in an analogous manner. Finally, for any p-dimensional vector $\mathbf{a}$, let $\mathbf{a}_{(m)}$ be the $q(m)$ dimensional vector $\mathbf{a}'_{(m)} = (\mathbf{a}_{i(m)}, \cdots, \mathbf{a}_{i(m)+q(m)-1})$. Note that $\mathbf{a}' = (\mathbf{a}'_{(1)}\ \mathbf{a}'_{(2)} \cdots \mathbf{a}'_{(k)})$.

THEOREM 4.1. *The standardized roots* $\mathbf{X}_n \to \mathbf{X}$ *in distribution, where the distribution of* $\mathbf{X}$ *is characterized by* $\mathbf{X}_{(m)} = \{\lambda_1(W_m), \lambda_2(W_m), \cdots, \lambda_{q(m)}(W_m)\}$ *with* $W_m = Z_{(m,m)} + D_m$.

PROOF. First observe that for $h \in \mathscr{I}_m$, $X_{h,n}$ can be expressed as $\lambda_h(M_{m,n})$ where $M_{m,n} = n^{1/2}(\gamma_{m,n}^{-1} S_n - I)$. Since $\lambda_h(M_{m,n}) = \lambda_h(QM_{m,n}Q')$ for any orthogonal Q, it then follows from Assumption 2.1.i that one can assume without loss of generality $\Sigma_n = \Delta_n = \operatorname{diag}\{\lambda_1(\Sigma_n), \lambda_2(\Sigma_n), \cdots, \lambda_p(\Sigma_n)\}$. By expressing $S_n = \Delta_n + n^{-1/2}\Delta_n^{1/2} Z_n \Delta_n^{1/2}$, the matrix $M_{m,n}$ can be rewritten as $M_{m,n} = D_{m,n} + T_{m,n}$, where $D_{m,n} = n^{1/2}(\gamma_{m,n}^{-1}\Delta_n - I)$ and $T_{m,n} = \gamma_{m,n}^{-1}\Delta_n^{1/2} Z_n \Delta_n^{1/2}$.

For $j \in \mathscr{I}_r$, $(D_{m,m})_{jj} \to \infty$ if $r < m$, $(D_{m,n})_{jj} \to -\infty$ if $r > m$. The submatrix $(D_{m,n})_{(m,m)} \to D_m$. The random matrix $T_{m,n}$ has the following convergence properties. The submatrix $(T_{m,n})_{(m,m)} \to Z_{(m,m)}$ in distribution since $Z_n \to Z$ in distribution and $\lambda_h(\Sigma_n)/\gamma_{m,n} \to 1$ for $h \in \mathscr{I}_m$. The elements of $(T_{m,n})_{(r,t)}$ for $r \neq m$ or $t \neq m$ may or may not converge depending on whether $\lambda_j(\Sigma_n)/\gamma_{m,n}$ converges for $j \notin \mathscr{I}_m$. The convergence of this ratio is not implied by (4.1).

Loosely interpreted, the block-diagonal matrices $(M_{m,n})_{(r,r)}$ "diverge" to $+\infty$ for $r < m$ and to $-\infty$ for $r > m$, and the block diagonal matrix $(M_{m,n})_{(m,m)}$ does not diverge. The off-diagonal matrices $(M_{m,n})_{(s,t)}$ are "negligible" with respect to $(M_{m,n})_{(r,r)}$ for $r \neq m$. Thus, one might conjecture that $X_{i(m)+j-1,n}$ is asymptotically equivalent to $\lambda_j\{(M_{m,n})_{(m,m)}\}$. More specifically, it is proven in Section 6 that

$$(4.3)\qquad X_{i(m)+j-1,n} - \lambda_j\{(M_{m,n(m,m)}\} \to 0 \quad \text{in probability}$$

for $j = 1, 2, \cdots, q(m)$ and $m = 1, 2, \cdots, k$. Theorem 4.1 then follows from Lemma 2.1 since $(M_{1,n}, M_{2,n}, \cdots, M_{k,n}) \to (W_1, W_2, \cdots, W_k)$ in distribution. □

It is possible to show that Theorem 4.1 is still valid if Assumption 2.1.i is replaced by the assumption that the distribution of Z is spherically invariant. The normality assumption for Z also is not essential. However, the normality assumption is needed in order to express the density function of $\mathbf{X}$ in a well-studied form.

THEOREM 4.2. *The joint density of the limiting distribution* $\mathbf{X}$ *in Theorem* 4.1 *is*

$$h(\mathbf{x}) = h_0 f_0(\mathbf{x})\{\textstyle\prod_{m=1}^{k}\prod_{j,t=1,j<t}^{q(m)}(x_{i(m)+j-1} - x_{i(m)+t-1})\}$$

on the domain $\{x_{i(m)} \geq x_{i(m)+1} \geq \cdots \geq x_{i(m)+q(m)-1};\ m = 1, 2, \cdots, k\}$, *where* h_0 *is defined in Lemma* 2.3 *and*

$$f_0(\mathbf{x}) = K_0 \exp[k_1 \textstyle\sum_{m=1}^{k} \operatorname{tr}\{(X_m - \eta I)^2 + D_m^2\} + k_2\{\sum_{m=1}^{k}\operatorname{tr}(X_m - \eta I)\}^2]$$
$$\times \textstyle\prod_{m=1}^{k}\, {}_0F_0^{[q(m)]}\{-2k_1(X_m - \eta I);\ D_m\}$$

with $k_1 = -(4\sigma_1)^{-1}$, $k_2 = \sigma_2(2\sigma_1 + p\sigma_2)^{-1}(4\sigma_1)^{-1}$ *and* $K_0 = 2^{-(1/2)(p-1)}(2 + p\sigma_2/\sigma_1)^{-(1/2)}(2\pi\sigma_1)^{-(1/2)q_0}$ *where* $q_0 = \frac{1}{2}\sum_{m=1}^{k} q(m)\{q(m) + 1\}$. *The parameters* η, σ_1 *and* σ_2

are defined in Lemma 2.1. *The matrix X_m is a $d(m) \times d(m)$ diagonal matrix with entries* $x_{i(m)}, x_{i(m)+1}, \cdots, x_{i(m)+q(m)-1}$.

PROOF. Without loss of generality, assume $\eta = 0$. The joint density $(Z_{(1,1)}, Z_{(2,2)}, \cdots, Z_{(k,k)})$ in q_0 dimensional space is

$$f_1(Z_1, Z_2, \cdots, Z_k) = K_0 \exp\{k_1 \textstyle\sum_{m=1}^k \operatorname{tr}(Z_m^2) + k_2(\sum_{m=1}^k \operatorname{tr} Z_m)^2\}, \tag{4.4}$$

and the density of $W = (W_1, W_2, \cdots, W_k)$ is $f_2(W_1, \cdots, W_k) = f_1(W_1 - D_1, \cdots, W_k - D_k)$. The joint density of the k set of eigenvalues of $\{W_m\}_{m=1,2,\cdots,k}$ is the same as the joint density of the k set of eigenvalues of $\{V_m\}_{m=1,2,\cdots,k}$ where by recalling that $\operatorname{tr} D_m = 0$, $V = (V_1, V_2, \cdots, V_k)$ has density

$$\begin{aligned} f_3(V) &= \int_{O[q(1)]} \cdots \int_{O[q(m)]} f_2(H_1 V_1 H_1', H_2 V_2 H_2', \cdots, H_k V_k H_k')(dH_1)(dH_2) \cdots (dH_k), \\ &= K_0 \exp[k_1 \textstyle\sum_{m=1}^k \operatorname{tr}(V_m^2 + D_m^2) + k_2\{\sum_{m=1}^k \operatorname{tr} V_m\}^2] \\ &\quad \times \textstyle\prod_{m=1}^k \left\{\displaystyle\int_{O[q(m)]} \operatorname{etr}(-2k_1 V_m H_m' D_m H_m)(dH_m)\right\}. \end{aligned} \tag{4.5}$$

The integral in the last expression is ${}_0F_0^{[q(m)]}(-2k_1 V_m, D_m)$. Since $f_3(V)$ depends on V only through the k set of roots of $\{V_m\}_{m=1,2,\cdots,k}$, the joint density of these roots can be found by applying Lemma 2.3. □

By examining $h(\mathbf{x})$, one can note that the variates $\mathbf{X}_{(1)}, \mathbf{X}_{(2)}, \cdots, \mathbf{X}_{(k)}$ are mutually independent if and only if $\sigma_2 = 0$. For sample covariances from elliptical populations, this implies that the "kurtosis" parameter $\kappa = 0$, see Section 3. Waternaux (1976) makes this observation for the sample covariance from an elliptical distribution whenever the covariance matrix Σ is not dependent on n and has all roots distinct. The joint marginal density of $\mathbf{X}_{(m)}$ is given by

$$\begin{aligned} h_m(\mathbf{x}_m) &= k_{0,m} \exp[k_1 \operatorname{tr}\{(X_m - \eta I)^2 + D_m^2\} + k_2\{\operatorname{tr}(X_m - \eta I)\}^2] \\ &\quad \times {}_0F_0^{[q(m)]}\{-2k_1(X_m - \eta I); D_m\} \textstyle\prod_{j,t=1, j<t}^{q(m)} (x_{i(m)+j-1} - x_{i(m)+t-1}), \end{aligned} \tag{4.6}$$

where

$$k_{0,m}^{-1} = 2^{(1/2)\{q(m)-1\}}(2 + q(m)\sigma_2/\sigma_1)^{1/2}(2\sigma_1)^{(1/4)q(m)\{q(m)+1\}} \textstyle\prod_{j=1}^{q(m)} \Gamma[\frac{1}{2}\{q(m) - j + 1\}].$$

Whenever $\Sigma_n = \Sigma$ does not depend on n with Σ having roots with multiplicities $q(1), q(2), \cdots, q(k)$, the densities given in Theorem 4.2 and in (4.6) simplify since $D_m = 0$ for $m = 1, 2, \cdots, k$ and the hypergeometric term ${}_0F_0^{[d(m)]}\{-2k_1(X_m - \eta I); 0\} = 1$. For this special case, the method of proof for Theorem 4.1 is not necessary. The result can be obtained by generalizing the arguments in Anderson's (1963) derivation for the principal component roots of a sample covariance matrix from a multivariate normal population.

5. Test for subsphericity. Let $\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_n$ be a random sample from a $\text{Normal}_p(\mu, \Sigma)$ population. The likelihood ratio criterion for testing the sphericity hypothesis $H_0: \Sigma = \lambda I$ with λ unknown is $L_n = |\hat{\Sigma}|^{(1/2)n}/(p^{-1} \operatorname{tr} \hat{\Sigma})^{-np}$, where $\hat{\Sigma}$ represents the sample covariance matrix. Under H_0, it is well known that $-2 \ln L_n \to \chi^2_{(1/2)(p+2)(p-1)}$ in distribution. Under the sequence of alternatives $\Sigma_n = \lambda(I + n^{-1/2}D)$, Nagao (1970) shows that $-2 \ln L_n \to \chi^2_{(1/2)(p+2)(p-1)}(\xi)$ in distribution, where the noncentrality parameter $\xi = \frac{1}{2}\{\operatorname{tr} D^2 - p^{-1}(\operatorname{tr} D)^2\}$.

Consider the subsphericity hypothesis for testing the equality of the $q = p - i + 1$ ($q =$

$p - i + 1)$ smallest roots, that is $H_{0,q}: \lambda_i(\Sigma) = \cdots = \lambda_p(\Sigma)$ and assuming $\lambda_{i-1}(\Sigma) \neq \lambda_i(\Sigma)$. Anderson (1963) shows the likelihood ratio criterion for $H_{0,q}$ to be $L_{q,n} = \{\prod_{j=1}^{p} \lambda_j(\hat{\Sigma})\}^{(1/2)n}/\{q^{-1} \sum_{j=1}^{p} \lambda_j(\hat{\Sigma})\}^{(1/2)nq}$, and the limiting null distribution of $-2 \ln L_{q,n}$ to be $\chi^2_{(1/2)(q+2)(q-1)}$. The limiting distribution of $-2 \ln L_{q,n}$ under the sequence of alternatives $\lambda_j(\Sigma_n) = \lambda(1 + n^{-(1/2)}a_j)$ for $j \geq i$ is shown by Waternaux (1977) to be $\chi^2_{(1/2)(q+2)(q-1)}(\xi_q')$ where $\xi_q' = \frac{1}{2}\{\sum_{j=i}^{p} a_j^2 - q^{-1}(\sum_{j=i}^{p} a_j)^2\}$ provided the roots $\{\lambda_j(\Sigma_n), j < i\}$ are distinct and not dependent on n. The next theorem shows that this limiting distribution is valid no matter how $\lambda_j(\Sigma_n)$ behaves for $j < i$ provided they separate from λ faster than $O(n^{-1/2})$. Before stating the theorem, the following lemma concerning quadratic forms in normal variables is needed for the proof of the theorem, see Srivastava and Khatri (1979), Corollary 2.11.1.

LEMMA 5.1. *Let* $\mathbf{Z} \sim \text{Normal}_r(\theta, \Gamma)$, *and let* A *be an* $r \times r$ *symmetric matrix. If* $A\Gamma A = A$ *and* $t = \text{tr}(\Gamma A)$, *then* $Z'AZ \sim \chi_t^2(\xi)$ *with* $\xi = \theta' A \theta$.

THEOREM 5.1. *Let* Σ_n *be a sequence of alternatives with* $n^{1/2}\{\lambda_j(\Sigma_n) - \bar{\lambda}_{q,n}\}/\bar{\lambda}_{q,n} \to d_j$ *for* $j \geq i$, *and diverging to infinity for* $j < i$, *where* $\bar{\lambda}_{q,n} = q^{-1} \sum_{j=i}^{p} \lambda_j(\Sigma_n)$. *Under this sequence of alternatives,*

$$-2 \ln L_{n,q} \to \chi^2_{(1/2)(q+2)(q-1)}(\xi_q) \quad \textit{in distribution},$$

where $\xi_q = \frac{1}{2} \sum_{j=i}^{p} d_j^2$.

PROOF. By Theorem 4.1, the joint limiting distribution of $\{X_{i,n}, \cdots, X_{p,n}\}$ where $X_{j,n} = n^{1/2}\{\lambda_j(\hat{\Sigma}) - \bar{\lambda}_{q,n}\}/\bar{\lambda}_{q,n}$ is $\{\lambda_1(Z_{(q)} + D_{(q)}), \cdots, \lambda_q(Z_{(q)} + D_{(q)})\}$. The $q \times q$ diagonal matrix $D_{(q)}$ has entries $d_i, d_{i+1}, \cdots, d_p$. The random symmetric matrix $Z_{(q)}$ is multivariate normal with mean 0 and $\text{var}\{\text{vec}(Z_{(q)})\} = I + K_{q,q}$ since $\sigma_1 = 1$ and $\sigma_2 = 0$. This limiting distribution is valid no matter how $\lambda_j(\Sigma_n)$ behaves for $j < i$, provided it separates from $\bar{\lambda}_{q,n}$ at the rate indicated in the theorem.

The test statistic can be written in terms of $X_{j,n}$ as follows,

$$\begin{aligned}-2 \ln L_{q,n} &= -n \ln\{\textstyle\prod_{j=i}^{p} \lambda_j(\hat{\Sigma})\} + nq \ln\{q^{-1} \textstyle\sum_{j=i}^{p} \lambda_i(\hat{\Sigma})\} \\ &= -n \ln\{\textstyle\prod_{j=i}^{p} (1 + n^{-(1/2)}X_{j,n})\} + nq \ln\{\textstyle\sum_{j=i}^{p} q^{-1}(1 + n^{-(1/2)}X_{j,n})\}.\end{aligned}$$

By expanding $\ln(1 + x) = x - \frac{1}{2}x^2 + O(x^3)$, one obtains

$$-2 \ln L_{q,n} = \tfrac{1}{2}\{\textstyle\sum_{j=i}^{p} X_{j,n}^2 - q^{-1}(\sum_{j=i}^{p} X_{j,n})^2\} + O_p(n^{-(1/2)})$$

which converges in distribution to the quadratic form $Q(Z_{(q)})$, where

$$\begin{aligned}Q(Z_{(q)}) &= \tfrac{1}{2}[\textstyle\sum_{j=i}^{p} \{\lambda_j(Z_{(q)} + D_{(q)})\}^2 - q^{-1}\{\sum_{j=i}^{p} \lambda_j(Z_{(q)} + D_{(q)})\}^2] \\ &= \tfrac{1}{2}[\{\text{tr}(Z_{(q)} - D_{(q)})^2\} - q^{-1}\{\text{tr}(Z_{(q)} - D_{(q)})\}^2].\end{aligned}$$

By employing the "vec" notation, the quadratic form can be expressed as $Q(Z_{(q)}) = \{\text{vec}(Z_{(q)} + D_{(q)})\}' A_q \text{vec}(Z_{(q)} + D_{(q)})$, where $A_q = \frac{1}{2}\{\frac{1}{2}(I + K_{q,q}) - q^{-1}\text{vec}(I)\text{vec}(I)'\}$. This follows from the properties $\text{tr}(BC) = \{\text{vec}(B)\}'\{\text{vec}(C')\}$ and $K_{q,q}\text{vec}(B) = \text{vec}(B')$, see Magnus and Neudecker (1979). The theorem then follows from Lemma 5.1 by using the property $K_{q,q}^2 = I$ to note that $A_q(I + K_{q,q})A_q = A_q$, $\text{tr}\{(I + K_{q,q})A_q\} = 2 \text{ tr } A_q = \frac{1}{2}(q + 2)(q - 1)$, and $\xi_q = \{\text{vec}(D_{(q)}\}' A_q \{\text{vec}(D_{(q)})\} = \frac{1}{2}(\text{tr } D_{(q)})^2$ since $\text{tr } D_{(q)} = 0$. □

One might suspect that Theorem 5.1 could be proven by using the general theory for likelihood ratio tests. The difficulty in using the general theory lies in the apparent inability to express $H_{0,q}$ for $q < p$ in the form $\mathbf{h}(\Sigma) = \mathbf{0}$ for some continuously differentiable $\frac{1}{2}(q + 2)(q - 1)$ dimensional function $\mathbf{h}$. The proof of Theorem 5.1 can be easily extended to prove the following result for the class of estimates considered in Section 2. Waternaux (1977) gives a similar result for the sample covariance matrix from an elliptical population under the more restrictive conditions on the population roots described prior to Theorem 5.1.

THEOREM 5.2. *Let Assumption* 2.1 *and the local alternative framework* (4.1) *hold, and define*

$$\Lambda_{m,n} = \{\textstyle\prod_{j=i(m)}^{i(m)+q(m)-1} \lambda_j(S_n)\}^{(1/2)n}/\{q(m)^{-1} \sum_{j=i(m)}^{i(m)+q(m)-1} \lambda_j(S_n)\}^{(1/2)nq(m)}.$$

The limiting distribution of $-2\{\ln \Lambda_{m,n}\}/\sigma_1$ *is a noncentral chi-squared on* $\frac{1}{2}\{q(m) + 2\}\{q(m) - 1\}$ *degrees of freedom and noncentrality parameter* $\xi_{(m)} = \frac{1}{2}(\text{tr } D_m)^2/\sigma_1$. *Furthermore, if* $\sigma_2 = 0$ *then the limiting distributions are jointly independent for* $m = 1, 2, \cdots, k$.

PROOF. By arguments analogous to the proof to Theorem 5.1, $-2\{\ln \Lambda_{m,n}\}/\sigma_1 \to \sigma_1^{-1}\{\text{vec}(Z_{(m,m)} + D_m)\}' A_{q(m)}\{\text{vec}(Z_{(m,m)} + D_m)\}$ jointly in distribution for $m = 1, 2, \cdots, k$. The marginal distribution of $Z_{(m,m)}$ is multivariate normal with mean ηI and covariance matrix $\Gamma_m = \text{var}\{\text{vec}(Z_{(m,m)})\} = \sigma_1(I + K_{q(m),q(m)}) + \sigma_2 \text{vec}(I)\{\text{vec}(I)\}'$. The submatrices $Z_{(m,m)}$ are independent only if $\sigma_2 = 0$. After noting that $A_{q(m)}\text{vec}(I) = 0$, it can then be verified that $(\sigma_1^{-1}A_{q(m)})\Gamma_m(\sigma_1^{-1}A_{q(m)}) = \sigma_1^{-1}A_{q(m)}$ and $\text{tr}\{\sigma_1^{-1}A_{q(m)}\Gamma_m\} = \frac{1}{2}\{q(m) + 2\}\{q(m) - 1\}$. The theorem then follows from Lemma 5.1 since

$$\{\text{vec}(D_m - \eta I)\}'(\sigma_1^{-1}A_{q(m)})\text{vec}(D_m - \eta I) = \{\text{vec}(D_m)\}'(\sigma_1^{-1}A_{q(m)})\text{vec}(D_m) = \xi_{(m)}. \quad \square$$

6. Proof of statement (4.3). Before giving the proof for statement (4.3), a review of some concepts in spectral theory is presented. Let M be a $p \times p$ real symmetric matrix with eigenvalues $\lambda_1(M) \geq \lambda_2(M) \geq \cdots \geq \lambda_p(M)$, and let $s(M) = \{\lambda_1(M), \lambda_2(M), \cdots \lambda_p(M)\}$ denote the spectral set of M. As a function of M, $s(M)$ is uniformly continuous.

The eigenspace of M associated with $\lambda \in s(M)$ is $V_\lambda(M) = \{\mathbf{x} \in R^p \mid M\mathbf{x} = \lambda\mathbf{x}\}$ where R^p is the set of all p-dimensional real vectors. The eigenprojection of M associated with $\lambda \in s(M)$, denoted $P_\lambda(M)$, is the orthogonal projection operator onto $V_\lambda(M)$. The eigenprojections have the properties $P_\lambda(M) = \{P_\lambda(M)\}' = \{P_\lambda(M)\}^2$, $P_\lambda(M)P_\mu(M), = P_\mu(M)P_\lambda(M) = 0$ for $\lambda \neq \mu$, and $MP_\lambda(M) = P_\lambda(M)M = \lambda M$. The spectral decomposition of M is $M = \sum_{\lambda \in s(M)} \lambda P_\lambda(M)$. If ν is any subset of $s(M)$, then the total eigenprojection of M associated with ν is $P(\nu, M) = \sum_{\lambda \in \nu} P_\lambda(M)$.

Lemma 6.1 gives the Taylor series expansion and truncation term for the product $MP(\nu, M)$ for $\nu \subset s(M)$. This product is important in studying the eigenvalues of M. The non-zero eigenvalues of the product are the non-zero elements in ν. In addition, if $\lambda_{i-1}(M). \neq \lambda_i(M)$ and $\lambda_{i+q-1}(M) \neq \lambda_{i-q}(M)$, and if M_0 is a $p \times p$ symmetric matrix with $\lambda_{i-1}(M_0) \neq \lambda_i(M_0)$ and $\lambda_{i+q-1}(M_0) \neq \lambda_{i+q}(M_0)$, then

$$|\lambda_t(M) - \lambda_t(M_0)| \leq 2 \|MP(\nu, M) - M_0P(\nu_0, M_0)\| \tag{6.1}$$

for $t = i, i + 1, \cdots, i + q - 1$ where $\nu = \{\lambda_i(M), \cdots, \lambda_{i+q-1}(M)\}$ and $\nu_0 = \{\lambda_i(M_0), \cdots, \lambda_{i+q-1}(M_0)\}$. The norm, $\|\cdot\|$, on the set of all real $p \times p$ symmetric matrices is the maximum absolute value of the eigenvalues of the matrix argument. Statement (6.1) is obtained by appling the inequality $\lambda_p(A - B) \leq \lambda_j(A) - \lambda_j(B) \leq \lambda_1(A - B)$ for any two symmetric matrices A and B of order p. The matrices $MP(\nu, M)$ and $M_0P(\nu_0, M_0)$ are symmetric. The 2 factor in (6.1) is due to the possibility that $\lambda_t(M)$ may be matched with the zero root of $M_0P(\nu_0, M_0)$ rather than with $\lambda_t(M_0)$.

The proof of Lemma 6.1 uses well-known results and perturbation techniques in spectral theory. The reader should refer to Kato (1966), Chapter 2, for a good overview. It is necessary to introduce complex variables for the lemma and its proof, and to extend the norm defined previously to the set of all $p \times p$ matrices with complex entries. Let $\mathbf{i} = \sqrt{-1}$. Define the norm $\|A\| = \{\lambda_1(A^*A)\}^{1/2}$, where A^* represents the conjugate transpose of A, and $\lambda_1(\cdot)$ represents the maximum eigenvalue. This norm has the important property

$$\|AB\| \leq \|A\| \cdot \|B\|. \tag{6.2}$$

LEMMA 6.1. *Let D be a real $p \times p$ symmetric matrix with $\lambda_{i-1}(D) \neq \lambda_i(D)$ and $\lambda_{i+q-1}(D) \neq \lambda_{i+q}(D)$, and let $M = D + T$ be a real $p \times p$ symmetric matrix. Define $w = \{\lambda_i(D), \cdots, \lambda_{i+q-1}(D)\}$ and $\nu = \{\lambda_i(M), \cdots, \lambda_{i+q-1}(M)\}$. Let G be a circle in the complex plane with the roots of D in w inside G and the roots of D not in w outside G. Also, for any $p \times p$ matrix B and any $p \times p$ symmetric positive definite matrix Γ, define $r(G, B, \Gamma) = \max_{\alpha \in G} \| \Gamma^{-1}B(D - \alpha I)^{-1}\Gamma \|$. If $r(G, T, \Gamma) < 1$ for some Γ, then*

(i) *$\lambda_{i-1}(M) \neq \lambda_i(M)$ and $\lambda_{i+q-1}(M) \neq \lambda_{i+q}(M)$, and*

(ii) *$MP(\nu, M) = DP(w, D) + P(w, D)TP(w, D) + H(T) + E(T)$, where*

$$H(T) = \textstyle\sum_{\lambda \in w, \mu \notin w} \lambda(\mu - \lambda)^{-1}\{P_\lambda(D)TP_\mu(D) + P_\mu(D)TP_\lambda(D)\},$$

and

$$E(T) = -(2\pi \mathbf{i})^{-1} \textstyle\sum_{j=2}^{\infty} \displaystyle\int_G \alpha(D - \alpha I)^{-1}\{-T(D - \alpha I)^{-1}\}^j \, d\alpha.$$

NOTE. If $i = 1$, the property $\lambda_{i-1} \neq \lambda_i$ and $\lambda_{i+q-1} \neq \lambda_{i+q}$ is understood to read $\lambda_{i+q-1} \neq \lambda_{i+q}$. If $i + q - 1 = p$, the property is understood to read $\lambda_{i-1} \neq \lambda_i$.

PROOF. Let Γ be such that $r(G, T, \Gamma) < 1$. For any $\alpha \in G$ we can apply the geometric series to obtain

$$(M - \alpha I)^{-1} = (D - \alpha I)^{-1}\Gamma[\textstyle\sum_{j=0}^{\infty} \{\Gamma^{-1}(-T)(D - \alpha I)^{-1}\Gamma\}^j]\Gamma^{-1}. \tag{6.3}$$

For $\alpha \in G$, $(M - \alpha I)^{-1}$ exists and so no eigenvalues of M lie on G. By the continuity of eigenvalues of symmetric matrices, it then follows that the eigenvalues of M within G are $\lambda_i(M), \cdots, \lambda_{i+q-1}(M)$, and that the other eigenvalues of M are outside G. Thus, $\lambda_{i-1}(M) \neq \lambda_i(M)$ and $\lambda_{i+q-1}(M) \neq \lambda_{i+q}(M)$.

To show part (ii), the expansion (6.3) and the identity

$$P(\nu, M) = -(2\pi \mathbf{i})^{-1} \int_G (M - \alpha I)^{-1} \, d\alpha, \tag{6.4}$$

can be used to obtain

$$MP(\nu, M) = DP(w, D) + (2\pi \mathbf{i})^{-1} \int_G \alpha(D - \alpha I)^{-1}T(D - \alpha I)^{-1} \, d\alpha + E(T), \tag{6.5}$$

where $E(T)$ is defined in the statement of the lemma. The middle two terms in the expansion in the lemma follow by integrating the second term in (6.5). □

A bound on $\| E(T) \|$ can be constructed by using the "triangle inequality" on the integral involved, and by using property (6.2) and the geometric series. In particular, if Γ = diagonal $(\tau_1, \tau_2, \cdots, \tau_p)$, D = diagonal $(\delta_1, \delta_2, \cdots, \delta_p)$, and G has center (0, 0) and radius R, then

$$\| E(T) \| \le \{c_1(R, \Gamma)\}^2 \cdot \| \Gamma^{-1}T\Gamma^{-1} \| \cdot [r(G, T, \Gamma)/\{1 - r(G, T, \Gamma)\}], \tag{6.6}$$

where $c_1(R, \Gamma) = \max_j | \{\tau_j R/(| \delta_j | - R)\} |$. By applying (6.2), the term $r(G, T, \Gamma)$ can be bounded by

$$r(G, T, \Gamma) \le c_2(R, \Gamma) \cdot \| \Gamma^{-1}T\Gamma^{-1} \|, \tag{6.7}$$

where $c_2(G, \Gamma) = \max_j | \{\tau_j^2/(| \delta_j | - R)\} |$.

PROOF OF (4.3). Let Lemma 6.1 be applied with $D_{m,n}$ and $M_{m,n}$ replacing D and M respectively, and with $i(m)$ and $q(m)$ replacing i and q respectively. For any new notation, terms dependent on m will have the m suppressed.

It is to be shown that there exists a sequence of circles G_n satisfying the respective conditions on G in Lemma 6.1 such that

(6.8) $$r(G_n, T_{m,n}, \Gamma_n) \to 0 \quad \text{in probability}$$

where $\Gamma_n = \gamma_{m,n}^{-(1/2)} \Delta_n^{1/2}$. It is also to be shown that

(6.9) $$H(T_{m,n}) \to 0 \quad \text{in probability, and}$$

(6.10) $$E(T_{m,n}) \to 0 \quad \text{in probability.}$$

It then follows that

(6.11) $$M_{m,n}P(\nu, M_{m,n}) - P(w, D_{m,n})M_{m,n}P(w, D_{m,n}) \to 0 \quad \text{in probability}$$

after noting that $D_{m,n}P(w, D_{m,n}) = P(w, D_{m,n})D_{m,n}P(w, D_{m,n})$.

Let M_n^0 be defined by

$$\begin{aligned} (M_n^0)_{(m,m)} &= (M_{m,n})_{(m,m)}, \\ (M_n^0)_{(r,r)} &= [\lambda_1\{(M_{m,n})_{(m,m)}\} + 1] \times I \quad \text{for} \quad r < m, \\ (M_n^0)_{(r,r)} &= [\lambda_{q(m)}\{(M_{m,n})_{(m,m)}\} - 1] \times I \quad \text{for} \quad r > m, \quad \text{and} \\ (M_n^0)_{(r,t)} &= 0 \quad \text{for} \quad r \neq t. \end{aligned}$$

Since $D_{m,n}$ is diagonal, $\{P(w, D_{m,n})\}_{(m,m)} = I$ and $\{P(w, D_{m,n})\}_{(r,t)} = 0$ for $r \neq m$ or $t \neq m$. By noting that $P(w, D_{m,n}) = P(w, M_n^0)$, it then follows that $P(w, D_{m,n})M_{m,n}P(w, D_{m,n}) = M_n^0 P(w, M_n^0)$. Inserting this identity into (6.11) gives

(6.12) $$M_{m,n}P(\nu, M_{m,n}) - M_n^0 P(w, M_n^0) \to 0 \quad \text{in probability.}$$

Statement (4.3) is obtained by applying (6.1) with $M_{m,n}$ and M_n^0 replacing M and M_0, since $\lambda_{i(m)+j-1}(M_n^0) = \lambda_j\{(M_{m,n})_{(m,m)}\}$. Thus, to complete the proof, (6.8), (6.9) and (6.10) must be proven.

To show (6.9), expanding $P(w, D_{m,n})$ gives

$$H(T_{m,n}) = \sum_{\lambda \in w, \mu \notin w} h_n(\lambda, \mu)\{P_\lambda(D_{m,n})Z_n P_\mu(D_{m,n}) + P_\mu(D_{m,n})Z_n P_\lambda(D_{m,n})\},$$

where $h_n(\lambda, \mu) = \lambda(\lambda - \mu)^{-1}(1 + n^{-(1/2)}\lambda)^{1/2}(1 + n^{-(1/2)}\mu)^{1/2}$. Recalling that $\lambda_t(D_{m,n}) \to d_t$ for $t = i(m), \cdots, i(m) + q(m) - 1$ and $|\lambda_t(D_{m,n})| \to \infty$ for $t < i(m)$ or $t > i(m) + q(m) - 1$, it follows that $\max_{\lambda \in w, \mu \notin w} |h_n(\lambda, \mu)| \to 0$. Thus, since $Z_n \to Z$ in distribution, (6.9) is true.

By definition of $D_{m,n}$, $\lambda_{i(m)}(D_{m,n}) \geq 0 \geq \lambda_{i(m)+q(m)-1}(D_{m,n})$. Let $a_n = \max\{|\lambda_{i(m)}(D_{m,n})|, |\lambda_{i(m)+q(m)-1}(D_{m,n})|\}$ and let $b_n = \min\{|\lambda_{i(m)-1}(D_{m,n})|, |\lambda_{i(m)+q(m)}(D_{m,n})|\}$. Define G_n to be the circle in the complex plane with center (0, 0) and radius $R_n = a_n + (b_n - a_n)^{1/2}$. Since $a_n \to \max\{d_{i(m)}, d_{i(m)+q(m)-1}\}$ and $b_n \to \infty$, for large n the roots $\lambda_j(D_{m,n})$ for $j = i(m), i(m) + 1, \cdots, i(m) + q(m) - 1$ are inside G_n, and the roots $\lambda_j(D_{m,n})$ for $j < i(m)$ or $j > i(m) + q(m) - 1$ are outside G_n.

To show (6.8), inequality (6.7) can be used to obtain $r(G_n, T_{m,n}, \Gamma_n) \leq c_{2,n} \| Z_n \|$ where

$$c_{2,n} = \max_j |\{1 + n^{-(1/2)}\lambda_j(D_{m,n})\}/\{|\lambda_j(D_{m,n})| - R_n\}|.$$

Evaluating the maximum for $j \in \mathscr{I}_m$ and $j \notin \mathscr{I}_m$ separately gives

$$c_{2,n} \leq \max[(1 + n^{-(1/2)}a_n)(b_n - a_n)^{-(1/2)}, (1 + n^{-(1/2)}b_n)\{(b_n - a_n) - (b_n - a_n)^{1/2}\}].$$

The right-hand side of the inequality goes to zero. Statement (6.8) is thus justified since $Z_n \to Z$ in distribution.

To show (6.10), inequality (6.6) can be used to obtain

$$\| E(T_{m,n}) \| \leq c_{1,n}^2 \| Z_n \| \cdot [r(G_n, T_{m,n}, \Gamma_n)/\{1 - r(G_n, T_{m,n}, \Gamma_n)\}],$$

where $c_{1,n} = \max_j |\{1 + n^{-1/2}\lambda_j(D_{m,n})\}^{1/2}R_n/\{|\lambda_j(D_{m,n})| - R_n\}|$. Since $Z_n \to Z$ in distribution and $r(G_n, T_{m,n}, \Gamma_n) \to 0$, it only needs to be shown that $c_{1,n}$ is bounded for large n. Evaluating the maximum for $j \in \mathscr{I}_m$ and $j \notin \mathscr{I}_m$ separately gives

$$c_{1,n} \leq \max[(1 + n^{-(1/2)}a_n)^{1/2}\{1 + a_n(b_n - a_n)^{-(1/2)}\},$$
$$(1 + n^{-(1/2)}b_n)^{1/2}\{1 + a_n(b_n - a_n)^{-(1/2)}\}\{(b_n - a_n)^{1/2} - 1\}].$$

The right-hand side of the inequality goes to 1. □

Acknowledgments. I am indebted to the referees for their comments which led to major changes in the presentation. I also wish to thank Professor G. S. Watson. Personal communications with him led to the inclusion of Section 5.

REFERENCES

ANDERSON, G. A. (1965). An asymptotic expansion for the distribution of the latent roots of the estimated covariance matrix. *Ann. Math. Statist.* **36** 1153–1173.

ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.

ANDERSON, T. W. (1963). Asymptotic theory for principal components analysis. *Ann. Math. Statist.* **34** 122–148.

DAVIS, A. W. (1977). Asymptotic theory for principal components analysis: non-normal case. *Aust. J. Statist.* **19** 206–212.

FUJIKOSHI, Y. (1980). Asymptotic expansions for the distribution of the sample roots under nonnormality. *Biometrika* **67** 45–51.

GIRSHICK, M. A. (1939). On the sampling distribution of the roots of determinantal equations. *Ann. Math. Statist.* **10** 203–224.

JAMES, A. T. (1960). The distribution of the latent roots of the covariance matrix. *Ann. Math. Statist.* **31** 151–158.

JAMES, A. T. (1964). Distribution of matrix variates and latent roots derived from normal samples. *Ann. Math. Statist.* **35** 475–501.

KATO, T. (1966). *Perturbation Theory for Linear Operators.* Springer, Berlin.

MAGNUS, J. R. and NEUDECKER, H. (1979). The commutation matrix: some properties and applications. *Ann. Statist.* **7** 381–394.

MUIRHEAD, R. J. (1978). Latent roots and matrix variates: a review of some asymptotic results. *Ann. Statist.* **6** 5–33.

NAGAO, H. (1970). Asymptotic expansions of some test criteria for homogeneity of variances and covariance matrices from normal populations. *J. Sci. Hiroshima University A1* **34** 153–247.

SRIVASTAVA, M. S. and KHATRI, C. G. (1979). *An Introduction to Multivariate Analysis.* North Holland, New York.

SUGUIRA, N. (1976). Asymptotic expansions of the distribution of the latent roots and latent vectors of the Wishart and multivariate F matrices. *J. Multivariate Anal.* **6** 500–525.

TYLER, D. E. (1982). Radial estimates and the test for sphericity. *Biometrika* **69** 429–436.

TYLER, D. E. (1983). Robustness and efficiency properties of scatter matrices. *Biometrika* **70** 411–420.

WATERNAUX, C. M. (1976). Asymptotic distribution of the sample roots for a nonnormal population. *Biometrika* **63** 639–645.

WATERNAUX, C. M. (1977). Principal components in the nonnormal case: the test for sphericity. *J. Multivariate Anal.* to appear.

DEPARTMENT OF MATHEMATICAL SCIENCES
OLD DOMINION UNIVERSITY
NORFOLK, VIRGINIA 23508