ON WIELANDT'S INEQUALITY AND ITS APPLICATION TO THE ASYMPTOTIC DISTRIBUTION OF THE EIGENVALUES OF A RANDOM SYMMETRIC MATRIX

By Morris L. Eaton[1] and David E. Tyler[2]

University of Minnesota and Rutgers University

A relatively obscure eigenvalue inequality due to Wielandt is used to give a simple derivation of the asymptotic distribution of the eigenvalues of a random symmetric matrix. The asymptotic distributions are obtained under a fairly general setting. An application of the general theory to the bootstrap distribution of the eigenvalues of the sample covariance matrix is given.

1. Introduction and summary. The derivation of the asymptotic distribution of the eigenvalues of a random symmetric matrix arises in many papers in multivariate analysis. Although the main idea behind most of the derivations is quite basic, i.e., the expansion of the sample roots about the population roots, the derivations themselves are often quite involved. These complications are primarily due to the mathematical rather than the statistical nature of the eigenvalue problem.

One of the main objectives of this paper is to introduce a simple method for obtaining the asymptotic distribution of the eigenvalue of random symmetric matrices. The method is based upon a relatively obscure eigenvalue inequality due to Wielandt (1967).

Obtaining the asymptotic distribution of eigenvalues by expanding the sample roots about the population roots becomes even more cumbersome when the population roots vary, e.g., see Tyler (1983). This case arises when considering local alternatives to hypotheses on population covariance matrices, and it also arises when considering the bootstrap distribution of eigenvalues, see Beran and Srivastava (1985). For this case, the use of Wielandt's eigenvalue inequality again provides a fairly simple method for obtaining the asymptotic distribution of the roots.

This paper is organized as follows. Wielandt's eigenvalue inequality is stated and discussed in Section 2. General results on the asymptotic distribution of eigenvalues of random symmetric matrices are presented in Section 3. The case when the population roots vary is treated in Section 4. The results of Sections 3 and 4 are applied in Section 5 to obtain results on the asymptotic behavior of the bootstrap distribution of the eigenvalues of the sample covariance matrix. The results on the bootstrap extend the work of Beran and

260

Srivastava (1985, 1987). Some concluding historical remarks and comparisons are made in Section 6.

2. Wielandt's eigenvalue inequality.   Consider a symmetric matrix

$$(2.1) \qquad\qquad A = \begin{pmatrix} B & C \\ C' & D \end{pmatrix},$$

where A is $p \times p$, B is $q \times q$ and D is $r \times r$. Let $\rho^2(C)$ denote the largest eigenvalue of CC' and let $\alpha_1 \geq \cdots \geq \alpha_p$, $\beta_1 \geq \cdots \geq \beta_q$ and $\delta_1 \geq \cdots \geq \delta_r$ be the ordered eigenvalues of A, B and D, respectively.

THEOREM 2.1 (Wielandt).   *If* $\beta_q > \delta_1$, *then*

$$(2.2) \qquad 0 \leq \alpha_j - \beta_j \leq \rho^2(C)/(\beta_j - \delta_1), \qquad j = 1, \ldots, q$$

and

$$(2.3) \qquad 0 \leq \delta_{r-i} - \alpha_{p-i} \leq \rho^2(C)/(\beta_q - \delta_{r-i}), \qquad i = 0, \ldots, r - 1.$$

The first set of inequalities (2.2) is given in Wielandt's (1967) lecture notes on page 120, but only when A is positive definite. The inequalities follow immediately for any symmetric A by replacing A by $A + \delta I$, where $\delta > -\delta_r$ and noting that $A + \delta I$ is positive definite and the δ term cancels in (2.2). The second set of inequalities (2.3) follow from the first by multiplying A by -1.

The first inequality in (2.2) is simply a partial restatement of result (1f.2.13) in Rao (1973) which he refers to as a Sturmian separation theorem, and which is referred to by Wielandt [(1967), page 117] as the "interlacement theorem." The second inequality in (2.2) is apparently a novel result of Wielandt's. Two interesting features of this inequality deserve to be noted. First, note that the matrix A can be viewed as a perturbation of a block diagonal matrix, namely $A = A_0 + E$, where

$$(2.4) \qquad\qquad A_0 = \begin{pmatrix} B & 0 \\ 0 & D \end{pmatrix} \quad \text{and} \quad E = \begin{pmatrix} 0 & C \\ C' & 0 \end{pmatrix}.$$

By Wielandt's inequality, the eigenvalues of A_0 are perturbed quadratically in E when A_0 is perturbed linearly in E. It is well known that in general, eigenvalues are only perturbed linearly when the matrix is perturbed linearly. The quadratic perturbation obtained in Wielandt's inequality is due to the special structure of E relative to A_0. This quadratic perturbation result can also be obtained in a somewhat more cumbersome way by using perturbation techniques, for example as described in Chapter 2 of Kato (1980), and by observing that due to the special structure of E relative to A_0, the linear term is zero.

The other interesting feature of Wielandt's inequality is that it not only shows that the perturbation of the eigenvalues are of quadratic order, but it also gives a bound which shows how the perturbation is related to the separation of the eigenvalues of B and D. Most perturbation techniques, such as Taylor series expansions, give only an order of perturbation in E rather

than bounds. Bounds of quadratic order on the perturbed eigenvalues can be obtained by using perturbation techniques described in Chapter 2 of Kato (1980), as is done in Section 6 of Tyler (1983). In light of Wielandt's inequality, this approach is unnecessary here, especially since it is more complicated and gives weaker bounds. Kato's (1980) perturbation technique, though, is useful if one wishes to obtain higher order approximations for the perturbed eigenvalues.

Because of its central role in this paper and its relative unavailability in the literature, a brief but complete proof of Wielandt's inequality is given below. The proof relies heavily on the following lemma which is given in Rao [(1973), Problem 1.9, page 68]. Wielandt (1967) refers to this lemma as Weyl's theorem (page 114) and refers to the corollary stated after the lemma as Aronszajn's theorem (page 119). The simple proof of the corollary is due to Wielandt.

LEMMA 2.1 (Weyl). *Let $T = R + S$, where R and S are $k \times k$ symmetric matrices, and let $t_1 \geq \cdots \geq t_k$, $r_1 \geq \cdots \geq r_k$ and $s_1 \geq \cdots \geq s_k$ be the ordered eigenvalues of T, R and S, respectively, then*

$$(2.5) \qquad t_j \leq \begin{cases} r_j + s_1 \\ \quad \vdots \\ r_1 + s_j \end{cases} \quad and \quad t_j \geq \begin{cases} r_j + s_k \\ \quad \vdots \\ r_k + s_j \end{cases}.$$

COROLLARY 2.1. Suppose in (2.1) that $A \geq 0$, i.e., A is positive semidefinite. Then

$$\beta_i \leq \alpha_i \leq \begin{cases} \beta_i + \delta_1 \\ \quad \vdots \\ \beta_1 + \delta_i \end{cases},$$

for $i = 1, \ldots, p$, where $\beta_i = 0$, for $i > q$ and $\delta_i = 0$, for $i > r$.

PROOF OF COROLLARY. Since A is symmetric positive semidefinite, it has a symmetric positive semidefinite square root $A^{1/2}$. Let $A^{1/2} = [F \quad G]$, where F is $p \times q$ and G is $p \times r$, and so A can be expressed as

$$A = \begin{bmatrix} F'F & F'G \\ G'F & G'G \end{bmatrix} \quad and \quad A = FF' + GG'.$$

The corollary follows from the lemma by noting that FF' and $F'F = B$ have the same eigenvalues apart from zeros. □

PROOF OF THEOREM 2.1. As noted previously, it only remains to show the second inequality in (2.2) holds. Since the result is invariant under the transformation $A \rightarrow A + \gamma I$ and under the transformation $B \rightarrow P'BP$, $D \rightarrow Q'DQ$ and $C \rightarrow P'CQ$ for orthogonal matrices P and Q, it can be assumed w.l.o.g. that $B = \text{diag}\{\beta_1, \ldots, \beta_q\}$, $D = \text{diag}\{\delta_1, \ldots, \delta_q\}$ and $\delta_1 = 0$. Note that the first inequality in (2.2) then implies $\alpha_i \geq \beta_i > 0$, for $i = 1, \ldots, q$.

Let $\tilde{B}$ be the $q \times q$ matrix $\tilde{B} = \text{diag}\{\beta_1, \ldots, \beta_j, \ldots, \beta_j\}$, for a fixed j. Also, let

$$\tilde{A} = \begin{pmatrix} \tilde{B} & C \\ C' & 0 \end{pmatrix} \quad \text{and hence} \quad \tilde{A}^2 = \begin{pmatrix} \tilde{B}^2 + CC' & \tilde{B}C \\ C'\tilde{B} & C'C \end{pmatrix}.$$

Let $\tilde{\alpha}_1 \geq \cdots \geq \tilde{\alpha}_p$ be the ordered eigenvalues of $\tilde{A}$. Now since $D \leq 0$ and $\tilde{B} \geq B$, it follows that $\tilde{A} \geq A$ and hence $\tilde{\alpha}_i \geq \alpha_i$, $i = 1, \ldots, p$. (The notation $M_1 \geq M_2$ means $M_1 - M_2$ is positive semidefinite.) Let $\pi_1 \geq \cdots \geq \pi_p$ be the ordered eigenvalues of $\tilde{A}^2$ and so they represent the ordered values of $\tilde{\alpha}_i^2$, $i = 1, \ldots, p$. Note that π_i is not necessarily equal to $\tilde{\alpha}_i^2$ since $\tilde{A}$ is not positive semidefinite. However, since $\tilde{\alpha}_i \geq \alpha_i > 0$, for $i = 1, \ldots, q$, it follows that $\pi_i \geq \tilde{\alpha}_i^2 \geq \alpha_i^2$, for $i = 1, \ldots, q$, and in particular $\pi_j \geq \alpha_j^2$.

An upper bound on π_j can be obtained by first applying Corollary 2.1 to $\tilde{A}^2$. This gives $\pi_j \leq \tilde{\beta}_j^2 + \rho^2(C)$, where $\tilde{\beta}_j^2$ is the jth largest root of $\tilde{B}^2 + CC'$. Application of Lemma 2.1 to $\tilde{B}^2 + CC'$ then gives $\tilde{\beta}_j^2 \leq \beta_j^2 + \rho^2(C)$ and so $\pi_j \leq \beta_j^2 + 2\rho^2(C)$. Putting the two inequalities for π_j together yields $\alpha_j^2 \leq \beta_j^2 + 2\rho^2(C)$ or $(\alpha_j - \beta_j)(\alpha_j + \beta_j) = \alpha_j^2 - \beta_j^2 \leq 2\rho^2(C)$. However, since $\alpha_j \geq \beta_j$, the desired result $\alpha_j - \beta_j \leq \rho^2(C)/\beta_j$ follows. $\square$

3. Asymptotic results for eigenvalues of random symmetric matrices. Consider a sequence of random matrices S_n, $n = 1, 2, \ldots$, in $\mathscr{S}_p$, the set of $p \times p$ real symmetric matrices, and assume that

$$(3.1) \qquad W_n = n^{1/2}(S_n - \Sigma) \to_d W,$$

with $\Sigma \in \mathscr{S}_p$ and hence $W \in \mathscr{S}_p$. Given $M \in \mathscr{S}_p$, let the vector of ordered eigenvalues of M be $\varphi(M) = (\varphi_1(M), \ldots, \varphi_p(M))$. The dependence of φ on the dimension parameter is suppressed and the same symbol φ is used for the vector of ordered eigenvalues of symmetric matrices of different dimensions. In this section, the asymptotic distribution of

$$(3.2) \qquad X_n = n^{1/2}\{\varphi(S_n) - \varphi(\Sigma)\} \in \mathbb{R}^p$$

is studied. Without loss of generality, Σ is taken to be diagonal, in particular $\text{diag}\{\varphi(\Sigma)\}$. In what follows, the choice of norms in $\mathbb{R}^p$ and in $\mathscr{S}_p$ is irrelevant. The notation $T_n = O_p(b_n^{-1})$ implies for any sequence of positive numbers a_n with $a_n \to 0$, $a_n b_n \|T_n\| \to 0$ in probability. The notation $T_n = o_p(b_n^{-1})$ implies $b_n \|T_n\| \to 0$ in probability.

3.1. *A basic lemma.* Partition Σ and S_n as

$$(3.3) \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \quad \text{and} \quad S_n = \begin{pmatrix} T_n & U_n \\ U_n' & V_n \end{pmatrix},$$

where Σ_{11} is the $q \times q$ diagonal matrix $\Sigma_{11} = \text{diag}\{\varphi_1(\Sigma), \ldots, \varphi_q(\Sigma)\}$, Σ_{22} is the $r \times r$ diagonal matrix $\Sigma_{22} = \text{diag}\{\varphi_{q+1}(\Sigma), \ldots, \varphi_p(\Sigma)\}$ and $p = q + r$. The matrices T_n, V_n and U_n are $q \times q$, $r \times r$ and $r \times q$, respectively.

LEMMA 3.1.   *If $\varphi_q(\Sigma) > \varphi_{q+1}(\Sigma)$, then*

$$Y_n = \varphi(S_n) - \begin{bmatrix} \varphi(T_n) \\ \varphi(V_n) \end{bmatrix} \quad is \quad O_p(n^{-1}).$$

PROOF.   Let $A_n = \{\varphi_q(T_n) > \varphi_1(V_n)\}$. Since φ is a continuous function and from (3.1), $T_n \to_p \Sigma_{11}$ and $V_n \to_p \Sigma_{22}$, it follows that $\varphi_q(T_n) \to_p \varphi_q(\Sigma_{11}) = \varphi_q(\Sigma)$ and $\varphi_1(V_n) \to_p \varphi_1(\Sigma_{22}) = \varphi_{q+1}(\Sigma)$. Thus, $\text{Prob}(A_n) \to 1$, and so attention can be restricted to A_n, $n = 1, 2, \ldots$. For $S_n \in A_n$, Wielandt's Theorem (Theorem 2.1) implies for $1 \le i \le q$,

$$(3.4) \qquad |\varphi_i(S_n) - \varphi_i(T_n)| < \rho^2(U_n)/\{\varphi_q(T_n) - \varphi_1(V_n)\}.$$

Now, by (3.1), $U_n = O_p(n^{-1/2})$ and since ρ is continuous it follows that $\rho^2(U_n) = O_p(n^{-1})$. The top part of the lemma then follows from (3.4) since $\varphi_q(T_n) - \varphi_1(V_n) \to_p \varphi_q(\Sigma) - \varphi_{q+1}(\Sigma) > 0$. The proof of the bottom part is analogous to the top. □

3.2. *The main theorems.*   Let $d_1 > d_2 > \cdots > d_k$ represent the distinct eigenvalues of Σ with the multiplicity of d_i being p_i, $i = 1, \ldots, k$, and hence $p_1 + \cdots + p_k = p$. Let I_i be the $p_i \times p_i$ identity matrix and partition Σ and S_n as

$$\Sigma = \begin{bmatrix} d_1 I_1 & 0 & \cdots & 0 \\ 0 & d_2 I_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_k I_k \end{bmatrix},$$

(3.5)

$$S_n = \begin{bmatrix} S_{n,11} & S_{n,11} & \cdots & S_{n,1k} \\ S_{n,21} & S_{n,22} & \cdots & S_{n,2k} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n,k1} & S_{n,k2} & \cdots & S_{n,kk} \end{bmatrix},$$

where $S_{n,ij}$ is $p_i \times p_j$, $i, j = 1, \ldots, k$. By applying Lemma 3.1 $k - 1$ times, the following asymptotic equivalence result is obtained. The vector $e_i \in \mathbb{R}^{p_i}$ is the vector of ones, $i = 1, \ldots, k$.

THEOREM 3.1.   *In the notation above, $n^{1/2}\{\varphi(S_n) - \varphi(\Sigma)\} = Z_n + R_n$, where*

$$Z_n = n^{1/2} \begin{bmatrix} \varphi(S_{n,11}) - d_1 e_1 \\ \vdots \\ \varphi(S_{n,kk}) - d_k e_k \end{bmatrix}$$

and the remainder term R_n is $O_p(n^{-1/2})$.

The asymptotic distribution of the leading term Z_n can be readily obtained from (3.1). Analogous to the partitioning of S_n, let $W = \{W_{ij}\}$ represent the partitioning of W in blocks of order $p_i \times p_j$, and hence

$$(3.6) \qquad \tilde{W}_n = n^{1/2} \begin{bmatrix} S_{n,11} - d_1 I_1 \\ \vdots \\ S_{n,kk} - d_k I_k \end{bmatrix} \to_d \tilde{W} = \begin{bmatrix} W_{11} \\ \vdots \\ W_{kk} \end{bmatrix}.$$

Now, on the space $\mathscr{S}_{p_1} \oplus \mathscr{S}_{p_2} \oplus \cdots \oplus \mathscr{S}_{p_k}$, where $\tilde{W}$ takes its values, the function

$$(3.7) \qquad H(\tilde{W}) = \begin{bmatrix} \varphi(W_{11}) \\ \vdots \\ \varphi(W_{kk}) \end{bmatrix}$$

is continuous, and hence $H(\tilde{W}_n) \to_d H(\tilde{W})$. However, since $\varphi\{n^{1/2}(S_{n,ii} - d_i I_i)\} = n^{1/2}\{\varphi(S_{n,ii}) - d_i e_i\}$, $i = 1, \ldots, k$, the following theorem is obtained.

THEOREM 3.2. *In the above notation, $\sqrt{n}\{\varphi(S_n) - \varphi(\Sigma)\} = H(\tilde{W}_n) + R_n$, where R_n is $O_p(n^{-1/2})$ and $H(\tilde{W}_n) \to_d H(\tilde{W})$.*

Thus, the asymptotic distribution of the roots of S_n is found by calculating the distribution of $H(\tilde{W})$.

4. Asymptotic behavior under a more general setting. Consider now two sequences of random matrices S_n and Σ_n, $n = 1, 2, \ldots$, both in $\mathscr{S}_p$, and assume that

$$(4.1) \qquad W_n = n^{1/2}(S_n - \Sigma_n) = O_p(1).$$

As a special case, Σ_n may be a sequence of nonrandom matrices, and in particular if Σ_n does not depend on n and W_n converges in distribution, then this reduces to the setting in Section 3. In this section, the asymptotic behavior of

$$(4.2) \qquad X_n = n^{1/2}\{\varphi(S_n) - \varphi(\Sigma_n)\} \in \mathbb{R}^p$$

is studied. Using the spectral value decomposition, express $\Sigma_n = P_n' \Delta_n P_n$, where P_n is an orthogonal matrix and Δ_n is the diagonal matrix with diagonal entries $\varphi_1(\Sigma_n), \ldots, \varphi_p(\Sigma_n)$, respectively. Define $S_n^0 = P_n S_n P_n'$ and note that $X_n = n^{1/2}\{\varphi(S_n^0) - \varphi(\Delta_n)\}$.

4.1. *A basic lemma.* Partition Δ_n and S_n^0 as

$$(4.3) \qquad \Delta_n = \begin{pmatrix} \Delta_{11,n} & 0 \\ 0 & \Delta_{22,n} \end{pmatrix} \quad \text{and} \quad S_n^0 = \begin{pmatrix} T_n & U_n \\ U_n' & V_n \end{pmatrix},$$

where the dimensions are the same as in (3.3).

LEMMA 4.1. *If* $a_n\{\varphi_q(\Sigma_n) - \varphi_{q+1}(\Sigma_n)\} \to_p \infty$, *for some increasing sequence of positive numbers* $a_n \to \infty$ *with* $a_n^{-1} = O_p(n^{-1/2})$, *then*

$$Y_n = \varphi(S_n) - \begin{bmatrix} \varphi(T_n) \\ \varphi(V_n) \end{bmatrix} \quad \text{is } o_p(a_n/n).$$

PROOF. Let $A_n = \{\varphi_q(T_n) > \varphi_1(V_n)\}$ and $B_n = \{\varphi_q(\Delta_{11,n}) > \varphi_1(\Delta_{22,n})\}$. By the condition in the lemma, it readily follows that $\text{Prob}(B_n) \to 1$. Condition (4.1) implies $P_n W_n P_n' = O_p(1)$ and hence $n^{1/2}(T_n - \Delta_{11,n}) = O_p(1)$ and $n^{1/2}(V_n - \Delta_{22,n}) = O_p(1)$. Application of Lemma 2.1 thus gives $\varphi_q(T_n) = \varphi_q(\Delta_{11,n}) + O_p(n^{-1/2})$ and $\varphi_1(V_n) = \varphi_1(\Delta_{22,n}) + O_p(n^{-1/2})$, which implies $a_n\{\varphi_q(T_n) - \varphi_1(V_n)\} \to_p \infty$, and so $\text{Prob}(A_n) \to 1$.

Attention is now restricted to A_n, $n = 1, 2, \ldots$. For $S_n \in A_n$, Wielandt's Theorem and the identity $\varphi(S_n) = \varphi(S_n^0)$ gives

(4.4) $$|\varphi_i(S_n) - \varphi_i(T_n)| < \rho^2(U_n)/\{\varphi_q(T_n) - \varphi_1(V_n)\}.$$

The numerator is $O_p(n^{-1})$, since $P_n W_n P_n'$ is $O_p(n^{-1/2})$ and so U_n is $O_p(n^{-1/2})$. It already has been shown that $a_n\{\varphi_q(T_n) - \varphi_1(V_n)\} \to_p \infty$ and hence the right-hand side of (4.4) is $o_p(a_n/n)$. The proof of the bottom part of the theorem is analogous. □

4.2. *The main theorems.* Partition the matrices Δ_n and S_n^0, respectively, as

$$\Delta_n = \begin{bmatrix} \Delta_{n,1} & 0 & \cdots & 0 \\ 0 & \Delta_{n,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Delta_{n,k} \end{bmatrix} \quad \text{and}$$

(4.5)

$$S_n^0 = \begin{bmatrix} S_{n,11}^0 & S_{n,12}^0 & \cdots & S_{n,1k}^0 \\ S_{n,21}^0 & S_{n,22}^0 & \cdots & S_{n,2k}^0 \\ \vdots & \vdots & \ddots & \vdots \\ S_{n,k1}^0 & S_{n,k2}^0 & \cdots & S_{n,kk}^0 \end{bmatrix},$$

where the dimensions are analogous to those in (3.5). For example, $\Delta_{n,i}$ and $S_{n,ii}^0$ are $p_i \times p_i$. Application of Lemma 4.1 $k - 1$ times with $a_n = n^{1/2}$ gives the following result.

THEOREM 4.1. *If* $n^{1/2}\{\varphi_{p_i}(\Delta_{n,i}) - \varphi_1(\Delta_{n,i+1})\} \to_p \infty$, *for* $i = 1, 2, \ldots, k - 1$, *then* $X_n = n^{1/2}\{\varphi(S_n) - \varphi(\Sigma_n)\} = Z_n + R_n$, *where*

$$Z_n = n^{1/2} \begin{bmatrix} \varphi(S_{n,11}^0) - \varphi(\Delta_{n,1}) \\ \vdots \\ \varphi(S_{n,kk}^0) - \varphi(\Delta_{n,k}) \end{bmatrix}.$$

and the remainder term R_n *is* $o_p(1)$.

Thus, the asymptotic distribution of X_n, if it exists, is the same as the asymptotic distribution of Z_n. Even if the asymptotic distribution does not exist, Z_n represents a simpler asymptotically equivalent variate. The term Z_n can be reexpressed as follows. Let $d_{n,i}$ represent the average of the p_i eigenvalues in $\Delta_{n,i}$, define $D_{n,i} = n^{1/2}(\Delta_{n,i} - d_{n,i}I_i)$. Also, define $W_n^0 = P_n W_n P_n'$ with $W_n^0 = \{W_{n,ij}^0\}$ representing the partitioning of W_n^0 in blocks of order $p_i \times p_j$, and let

$$(4.6) \quad \tilde{W}_n^0 = n^{1/2}\begin{bmatrix} S_{n,11}^0 - \Delta_{n,1} \\ \vdots \\ S_{n,kk}^0 - \Delta_{n,k} \end{bmatrix} = \begin{bmatrix} W_{n,11}^0 \\ \vdots \\ W_{n,kk}^0 \end{bmatrix} \quad \text{and} \quad \tilde{D}_n = \begin{bmatrix} D_{n,1} \\ \vdots \\ D_{n,k} \end{bmatrix}.$$

Note that $n^{1/2}\{\varphi(S_{n,ii}^0) - d_{n,i}e_i\} = \varphi(n^{1/2}(S_{n,ii}^0 - d_{n,i}I_i)) = \varphi(W_{n,ii}^0 + D_{n,i})$ and $n^{1/2}\{\varphi(\Delta_{n,i}) - d_{n,i}e_i\} = \varphi(D_{n,i})$, and so using the function H defined in Section 3, the following result is obtained.

THEOREM 4.2. *In the above notation and under the conditions of Theorem 4.1, $n^{1/2}\{\varphi(S_n) - \varphi(\Sigma_n)\} = \{H(\tilde{W}_n^0 + \tilde{D}_n) - H(\tilde{D}_n)\} + R_n$, where R_n is $o_p(1)$.*

For nonrandom Σ_n, Theorem 4.2 can be used to obtain the asymptotic distribution of the roots of S_n under the sequence Σ_n. Suppose $\Sigma_n \to \Sigma$, which without loss of generality is taken as in (3.5). The sequence P_n can be chosen so that $P_n \to I$, and so if

$$(4.7) \qquad\qquad W_n = n^{1/2}(S_n - \Sigma_n) \to_d W,$$

then $W_n^0 \to_d W$ and hence $\tilde{W}_n^0 \to_d \tilde{W}$. Furthermore if $\tilde{D}_n \to D$, then

$$(4.8) \qquad n^{1/2}\{\varphi(S_n) - \varphi(\Sigma_n)\} \to_d H(\tilde{W} + D) - H(D).$$

Note that no condition on the rate at which $\Sigma_n \to \Sigma$ is made in obtaining (4.8). Only conditions on the rates at which the roots of Σ_n approach each other or diverge from each other are needed.

It is interesting to compare (4.8) to Theorem 3.2. For example, when $p = 2$, note that if $\varphi_1(\Sigma_n)$ and $\varphi_2(\Sigma_n)$ differ by $o(n^{-1/2})$, then the asymptotic theory treats them as a multiple root, if they differ by $O(a_n^{-1})$, where $a_n = o(n^{1/2})$, then the asymptotic theory treats them as two simple roots and if they differ by $O(n^{-1/2})$ then the asymptotic theory treats them as a mixture of the two cases.

The term $n^{1/2}$ in (3.1) and (4.1) is the most common rate arising in practice. The results of this section and Section 3, though, readily generalize if the rate $n^{1/2}$ in (3.1) and (4.2) is replaced by an increasing sequence $c_n \to \infty$. The resulting modification in all the statements and theorems is made by simply replacing n by c_n^2 (except, of course, when n is used as an index or subscript).

Wielandt's theorem is also valid when A in (2.1) has complex entries and is selfadjoint. Correspondingly, the results of this section and Section 3 can be

easily extended to the case when S_n, Σ and Σ_n have complex entries and are selfadjoint.

5. An application to bootstrapping eigenvalues.

Let $\{\mathbf{x}_i; 1 \leq i \leq n\}$ represent a random sample from a distribution with covariance matrix Σ and finite fourth moments. If S_n represents the sample covariance matrix, then

$$(5.1) \qquad W_n = n^{1/2}(S_n - \Sigma) \to_d W,$$

where W has a multivariate normal distribution. Without loss of generality, let $\Sigma = \Delta$ be diagonal and represented as in (3.5). Using the notation established in Section 3.2, application of Theorem 3.2 gives

$$(5.2) \qquad X_n = n^{1/2}\{\varphi(S_n) - \varphi(\Sigma)\} \to_d H(\tilde{W}).$$

Let F_n be the sample distribution function of $\{\mathbf{x}_i; 1 \leq i \leq n\}$. The covariance matrix associated with the distribution F_n is thus S_n. Consider now a random sample $\{\mathbf{x}_i^*; 1 \leq i \leq n\}$ from the distribution F_n and let S_n^* be the sample covariance matrix of this sample. The idea behind the bootstrap is to use the distribution of $W_n^* = n^{1/2}(S_n^* - S_n)$ under F_n, which is realizable, as a nonparametric estimate of the distribution of $W_n = n^{1/2}(S_n - \Sigma)$. Beran and Srivastava (1985) show that the bootstrap estimate is strongly consistent, that is

$$(5.3) \qquad W_n^* = n^{1/2}(S_n^* - S_n) \to_{d*} W \quad \text{a.s.}$$

The notation $\to_{d*}$ refers to the weak convergence of the distribution function of W_n^* under F_n to the distribution function of W. Under F_n, S_n is a fixed matrix and S_n^* is a random matrix. The almost sure statement refers to the underlying product measure on $\{\mathbf{x}_i; 1 \leq i < \infty\}$.

The nonparametric bootstrap distribution of the sample roots is the distribution of $X_n^* = n^{1/2}\{\varphi(S_n^*) - \varphi(S_n)\}$ under F_n. This estimates the distribution of $X_n = n^{1/2}\{\varphi(S_n) - \varphi(\Sigma)\}$. It is easy to verify that the conditions of Theorem 4.2 are almost surely satisfied and so

$$(5.4) \qquad X_n^* - \{H(\tilde{W}_n^{*0} + \tilde{D}_n) - H(\tilde{D}_n)\} \to_{d*} 0 \quad \text{a.s.}$$

The notation in (5.4) is as follows. Let $S_n = P_n' \Delta_n P_n$ represent the spectral value decomposition of S_n with $\Delta_n = \text{diag}\{\varphi(S_n)\}$ and define $W_n^{*0} = P_n W_n^* P_n'$. Next, partition Δ_n as in (4.5), define $\tilde{D}_n$ as in (4.6) and then define $\tilde{W}_n^{*0}$ accordingly.

Since $\varphi_i(A + aI) = \varphi_i(A) + a$, expression (5.4) can be reexpressed as

$$(5.5) \qquad X_n^* - \{H(\tilde{W}_n^{*0} + \tilde{A}_n) - H(\tilde{A}_n)\} \to_{d*} 0 \quad \text{a.s.},$$

where $A_n = n^{1/2}(\Delta_n - \Delta)$ and $\tilde{A}_n$ is defined accordingly.

Now, since $S_n \to \Sigma = \Delta$ a.s., the sequence P_n can be chosen so that $P_n \to I$ a.s. and hence $\tilde{W}_n^{*0} \to_{d*} \tilde{W}$ a.s. The matrices $\tilde{A}_n$ are fixed matrices with respect to F_n and converge in distribution but not almost surely with respect

to the product measure on $\{x_i; 1 \le i < \infty\}$. More specifically,

$$(5.6) \qquad \tilde{A}_n = \begin{pmatrix} A_{n,1} \\ \vdots \\ A_{n,k} \end{pmatrix} \to_d A = \begin{pmatrix} A_1 \\ \vdots \\ A_k \end{pmatrix},$$

where $A_{n,i} = n^{1/2}(\Delta_{n,i} - d_i I)$ and so from (5.2) the joint distribution of $A_1, \ldots, A_k$ is the same as the joint distribution of $\text{diag}\{\varphi(W_{11})\}, \ldots, \text{diag}\{\varphi(W_{kk})\}$.

If all the eigenvalues of Σ are simple, then $H(\tilde{W}_n^{*0} + \tilde{A}_n) - H(\tilde{A}_n) = H(\tilde{W}^{*0})$, and so from (5.5) it follows that

$$(5.7) \qquad X_n^* = n^{1/2}\{\varphi(S_n^*) - \varphi(S_n)\} \to_{d*} H(\tilde{W}) \quad \text{a.s.}$$

Thus, for this case the bootstrap distribution for the sample roots is strongly consistent. If only some of the eigenvalues of Σ are simple, then by the same argument it can be shown that the joint marginal distribution of the bootstrap distribution associated with these roots are strongly consistent. However, since $\tilde{A}_n$ does not go to zero almost surely and does not cancel out in (5.5) when Σ has multiple roots, the marginal bootstrap distribution associated with a multiple root is not consistent. The consistency of the bootstrap for simple roots was proven by Beran and Srivastava (1985). They also showed the inconsistency of the bootstrap in the presence of multiple population roots for dimension $p = 2$, see Beran and Srivastava (1987). Their proof in the latter case makes use of the explicit form of the eigenvalues of a 2×2 matrix.

For the $p = 2$ dimensional case, Beran and Srivastava (1987) show that bootstrapping based upon samples of size m with $m/n \to 0$, gives a strongly consistent estimate of the limiting distribution of $n^{1/2}\{\varphi(S_n) - \varphi(\Sigma)\}$ regardless of the eigenvalue multiplicities. This approach works in general. That is, suppose $\{x_i^*; 1 \le i \le m\}$ represent a random sample of size m from the distribution F_n, with $m/n \to 0$. Let $S_{(m)}^*$ be the sample covariance matrix of this sample and let $X_{(m)}^* = m^{1/2}\{\varphi(S_{(m)}^*) - \varphi(S_n)\}$. The bootstrap estimate of the distribution is still strongly consistent. That is,

$$(5.8) \qquad W_{(m)}^* = m^{1/2}\left(S_{(m)}^* - S_n\right) \to_{d*} W \quad \text{a.s.}$$

Likewise, an analogous statement to (5.5) holds,

$$(5.9) \qquad X_{(m)}^* - \left\{H\left(\tilde{W}_{(m)}^{*0} + \tilde{A}_{(m)}\right) - H\left(\tilde{A}_{(m)}\right)\right\} \to_{d*} 0 \quad \text{a.s.},$$

where $W_{(m)}^{*0} = P_n W_{(m)}^* P_n'$ and hence $\tilde{W}_{(m)}^{*0}$ is defined accordingly. Also, $A_{(m)} = m^{1/2}(\Delta_n - \Delta)$ with $\tilde{A}_{(m)}$ defined accordingly. Now, $\tilde{W}_{(m)}^{*0} \to_{d*} \tilde{W}$ a.s. and $A_{(m)} = (m/n)^{1/2}A_n \to_{\text{a.s.}} 0$, which by (5.8) gives

$$(5.10) \qquad X_{(m)} = m^{1/2}\{\varphi(S_{(m)}^*) - \varphi(S_n)\} \to_{d*} H(\tilde{W}) \quad \text{a.s.}$$

Although bootstrapping a sample of size $o(n)$ gives consistent results, its asymptotic efficiency is zero with respect to bootstrapping a sample of size n when the roots are simple. Finding a consistent and efficient method of

bootstrapping eigenvalues which does not presuppose knowledge of the population eigenvalues' multiplicity is an open problem.

It should be noted that the results of this section depend on the sample covariance matrix only through properties (5.1), (5.3) and (5.8). The results generalize to any symmetric estimate of Σ for which (5.1), (5.3) and (5.8) hold.

6. Concluding remarks. The asymptotic theory for the distribution of the roots of a sample covariance matrix has been studied extensively. In these studies, primarily two different approaches have been used. One approach assumes a normal population and involves the use of asymptotic representations for the hypergeometric function which appears in the exact joint density of the sample roots. This approach also applies to canonical correlations and to MANOVA roots when sampling from normal populations. An extensive survey has been given by Muirhead (1978).

Another approach essentially involves the delta method, i.e., expanding the sample roots about the population roots. For normal populations, Lawley (1956) and Anderson (1963) use this approach for deriving the asymptotic distribution of the roots of a sample covariance matrix for the simple population root case and for the general case, respectively. This approach is not dependent on the assumption of a normal population and can be applied to nonnormal populations, e.g., see Waternaux (1976) and Davis (1977). This approach is also not specific to the sample covariance matrix. It can be applied to random symmetric matrices in general, e.g., see Tyler (1983) and Amemiya (1986), or to canonical correlations and MANOVA roots under fairly general distributional assumptions, e.g., see Anderson (1951, 1987) and Amemiya (1986).

Using the delta or perturbation method for the simple population root case is fairly straightforward since eigenvalues are analytic about a simple root. However, although continuous, eigenvalues are not differentiable at points of multiple roots. Obtaining expansions for sample roots associated with a multiple population root is thus more complicated and usually requires expansions for the eigenvectors as well, see e.g., Anderson (1951, 1963). The use of Wielandt's inequality circumvents these complications.

The asymptotic distribution of eigenvalues when the population roots are allowed to vary have been previously considered by Tyler (1983), but under a less general setting on the random symmetric matrix. The arguments used to prove Theorem 4.1 in Tyler (1983) though can be applied with only slight modification to obtain result (4.8) of the present paper. These arguments, however, are technically rather cumbersome. They involve a study of the truncation error associated with the expansion of the sample roots about the population roots. On the other hand, the use of Wielandt's inequality in the varying population root case is essentially as straightforward as in the fixed root case. Amemiya (1986) has shown that a slight modification of Anderson's (1951, 1963) classical approach for the fixed root case can also be made to include the varying population root case. This approach works for general random symmetric matrices but under slightly more restrictive conditions on

how the population roots may vary. Amemiya's (1986) proof applies when neighboring population roots approach each other at a rate of $O(n^{-1/2})$. His arguments, though, do not apply when the neighboring roots approach each other more slowly, e.g., at a rate $O(n^{-\alpha})$ for $0 < \alpha < \frac{1}{2}$.

REFERENCES

AMEMIYA, Y. (1986). Limiting distribution of certain characteristic roots under general conditions and their applications. Technical Report 17, Econometric Workshop, Stanford Univ.

ANDERSON, T. W. (1951). The asymptotic distribution of certain characteristic roots and vectors. *Proc. Second Berkeley Symp. Math. Statist. Probab.* 103–130. Univ. California Press, Berkeley.

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

ANDERSON, T. W. (1987). The asymptotic distribution of characteristic roots and vectors in multivariate components of variance. Technical Report 23, Econometric Workshop, Stanford Univ.

BERAN, R. and SRIVASTAVA, M. S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix. *Ann. Statist.* **13** 95–115.

BERAN, R. and SRIVASTAVA, M. S. (1987). Correction: Bootstrap tests and confidence regions for functions of a covariance matrix. *Ann. Statist.* **15** 470–471.

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

KATO, T. (1980). *Perturbation Theory for Linear Operators*. Springer, Berlin.

LAWLEY, D. N. (1956). Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika* **43** 128–136.

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

RAO, C. R. (1973). *Linear Statistical Inference and its Applications*. Wiley, New York.

TYLER, D. E. (1983). The asymptotic distribution of principal component roots under local alternatives to multiple roots. *Ann. Statist.* **11** 1232–1242.

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

WIELANDT, H. (1967). Topics in the analytic theory of matrices. (Lecture notes prepared by R. R. Meyer.) Univ. Wisconsin Press, Madison.

SCHOOL OF STATISTICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MINNESOTA 55455

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903