ASYMPTOTIC THEORY FOR PRINCIPAL COMPONENT ANALYSIS[1]

By T. W. Anderson

Columbia University

0. Summary. The asymptotic distribution of the characteristic roots and (normalized) vectors of a sample covariance matrix is given when the observations are from a multivariate normal distribution whose covariance matrix has characteristic roots of arbitrary multiplicity. The elements of each characteristic vector are the coefficients of a principal component (with sum of squares of coefficients being unity), and the corresponding characteristic root is the variance of the principal component. Tests of hypotheses of equality of population roots are treated, and confidence intervals for assumed equal roots are given; these are useful in assessing the importance of principal components. A similar study for correlation matrices is considered.

1. Introduction. Let x be a p-component random vector[2] with mean vector $\mathcal{E}x = \mu$ and covariance matrix $\mathcal{E}(x - \mu)(x - \mu)' = \Sigma$ (where the prime denotes the transpose of the vector). The variance of a linear combination $\gamma' x$ is

$$\mathcal{E}(\gamma' x - \mathcal{E}\gamma' x)^2 = \mathcal{E}[\gamma'(x - \mu)]^2 = \mathcal{E}\gamma'(x - \mu)(x - \mu)'\gamma = \gamma'\Sigma\gamma. \tag{1.1}$$

The linear combination normalized by $\gamma'\gamma = 1$ which has maximum variance may be called the first principal component of x. The linear combination uncorrelated with the first principal component and similarly normalized which has maximum variance may be called the second principal component. The other $p - 2$ principal components are similarly defined. (See [2], Chapter 11, for more detail; Hotelling [9] developed much of the theory.)

To give these linear combinations precisely we use the characteristic roots and vectors of Σ. Let $\delta_1 \geqq \cdots \geqq \delta_p > 0$ be the p characteristic roots of Σ (assumed to be positive definite). They are the roots of

$$|\Sigma - \delta I| = 0. \tag{1.2}$$

Let $\gamma_1, \cdots, \gamma_p$ be the corresponding normalized characteristic vectors; they satisfy

$$\Sigma\gamma_i = \delta_i\gamma_i, \tag{1.3}$$

$$\gamma_i'\gamma_i = 1. \tag{1.4}$$

Received June 26, 1961; revised July 19, 1962.

[1] This research was sponsored by the Office of Naval Research under Contract Number Nonr 3279(00) and Nonr 266(33). Reproduction in whole or in part is permitted for any purpose of the United States Government. Some of the research in this paper was conducted at Stanford University.

[2] Latin letters denote random variables and running variables, and Greek letters denote parameters. Scalars and vectors are indicated by lower case letters, and matrices are indicated by capital letters. To the extent of convenience the estimate of a parameter is indicated by a Latin letter corresponding to the Greek letter for the parameter.

If the characteristic roots are different,

$$\gamma_i'\gamma_j = 0, \qquad i \neq j; \tag{1.5}$$

if several characteristic roots are equal, the corresponding vectors have some indeterminancy, but they can be taken so that (1.5) holds. The linear combinations $v_i = \gamma_i'x$ are the principal components of x. They have the properties

$$\begin{aligned} \operatorname{Var} v_i &= \mathcal{E}(v_i - \mathcal{E}v_i)^2 = \gamma_i'\Sigma\gamma_i = \delta_i, \\ \operatorname{Cov}(v_i, v_j) &= \mathcal{E}(v_i - \mathcal{E}v_i)(v_j - \mathcal{E}v_j) = \gamma_i'\Sigma\gamma_j = 0, \qquad i \neq j. \end{aligned} \tag{1.6}$$

A geometric interpretation of this algebra may be helpful. If x is normally distributed, the contours of equal density are ellipsoids; a specific contour is $(x - \mu)'\Sigma^{-1}(x - \mu) = 1$, which is an ellipsoid with center at μ. If we ask for a point on this contour that is at a maximum distance from μ, that is, that maximizes $(x - \mu)'(x - \mu)$ we find the solution to be $\mu + \delta_1^{\frac{1}{2}}\gamma_1$ and the maximum distance to be $\delta_1^{\frac{1}{2}}$. If we ask for a point on the intersection of the contour and the hyperplane through μ orthogonal to γ_1 that is at a maximum distance from μ, we obtain $\mu + \delta_2^{\frac{1}{2}}\gamma_2$ at a distance of $\delta_2^{\frac{1}{2}}$.

The principal components are frequently defined alternatively as $\delta_i^{\frac{1}{2}}\gamma_i'x$. Then $(\delta_i^{\frac{1}{2}}\gamma_i)'(\delta_i^{\frac{1}{2}}\gamma_i) = \delta_i$ and $(\delta_i^{\frac{1}{2}}\gamma_i)'\Sigma^{-1}(\delta_i^{\frac{1}{2}}\gamma_i) = 1$. We shall find it more convenient to use the first normalization.

Let $x_1, \cdots, x_N$ be observations on x. The usual unbiased estimate of Σ is $S = [1/(N - 1)]A$, where

$$A = \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(x_\alpha - \bar{x})' \tag{1.7}$$

and $\bar{x} = (1/N)\sum_\alpha x_\alpha$. One estimates $\delta_1, \cdots, \delta_p$ by the characteristic roots, $d_1 \geqq \cdots \geqq d_p$, of S, that is, the roots of

$$|S - dI| = |[1/(N - 1)]A - dI| = 0, \tag{1.8}$$

and one estimates $\gamma_1, \cdots, \gamma_p$ by the normalized characteristic vectors, $c_1, \cdots, c_p$, of S, that is, the vectors satisfying

$$Sc_i = d_ic_i, \tag{1.9}$$

$$c_i'c_j = \delta_{ij}, \tag{1.10}$$

where δ_{ij} is the Kronecker delta. If $N - 1 \geqq p$, the roots of S are different with probability one and (1.9) for $i \neq j$ is automatically satisfied when $c_i'c_i = 1$.

In this paper we consider the asymptotic distribution of $d_1, \cdots, d_p, c_1, \cdots, c_p$ when the distribution of x is multivariate normal. If the characteristic roots of Σ are different, the deviations of $d_1, \cdots, d_p, c_1, \cdots, c_p$ from the corresponding population quantities are asymptotically normally distributed. When some of the roots of Σ are equal, the asymptotic distribution cannot be described so simply. A major purpose of this paper is to give the asymptotic distribution of the sample roots and vectors when the population roots are equal in sets.

In some exploratory work, an investigator wishes to study the variation in x; he may wish to consider the principal components with considerable variances and may wish to ignore the principal components with small variances. On the basis of a sample he may want to be able to infer that some of the smaller roots of Σ are small enough to be neglected. In this paper we consider statistical procedures for making such inferences.

In other situations the investigator may consider his measurements $x = x^{(1)} + x^{(2)}$ as made up of "systematic" or "true" values $x^{(1)}$ and errors of measurement $x^{(2)}$. If the error of measurement in each component is due to error in a measuring device applied independently to obtain each component of x (for example, due to the inherent inaccuracy of a micrometer in certain anthropological measurements), one can assume that the components of $x^{(2)}$ are uncorrelated and have the same variance; that is, $x^{(2)}$ has the covariance matrix $\sigma^2 I$. The components of $x^{(1)}$ may vary in a q-dimensional space and have a covariance matrix Ψ of rank $q(<p)$. Then $\Sigma = \Psi + \sigma^2 I$, and the $p - q$ smallest roots of Σ are σ^2. We consider testing this hypothesis, namely that the $p - q$ smallest roots are equal. We treat the likelihood ratio criterion for testing equality of any set of roots.

The principal component analysis can also be applied to correlation matrices. Let $\bar{R} = (\rho_{ij})$, where $\rho_{ij} = \sigma_{ij}/(\sigma_{ii}\sigma_{jj})^{\frac{1}{2}}$, and let $R = (r_{ij})$, where

$$r_{ij} = s_{ij}/(s_{ii}s_{jj})^{\frac{1}{2}} = a_{ij}/(a_{ii}a_{jj})^{\frac{1}{2}}.$$

Then $\bar{R}$ takes the place of Σ and R takes the place of S in the above discussion. We also consider the asymptotic theory in this case, but it is too complicated to be given in generality. The mathematical difficulties seem to reflect the difficulties of interpretation for correlations.

The methods used in this paper are similar to those used in a previous paper [1], though they are simpler in the present situation.

In the case of all the roots of Σ being different, Girshick [8] has given the asymptotic variances and covariances of $d_1, \cdots, d_p$, and of $d_1^{\frac{1}{2}}c_1, \cdots, d_p^{\frac{1}{2}}c_p$; in the case of all the roots of $\bar{R}$ being different, he has given the asymptotic variances and covariances of the roots of R. In the case of the smallest $p - q$ roots of Σ being equal and the others different, Lawley [10] has given the asymptotic variances and covariances of $d_1^{\frac{1}{2}}c_1, \cdots, d_q^{\frac{1}{2}}c_q$ (the sample vectors corresponding to the different roots of the population matrix), but he did not argue asymptotic normality. In this paper we obtain the asymptotic distribution of $d_1, \cdots, d_p, c_1, \cdots, c_p$ when the roots of Σ have any multiplicities; as a special case, the vectors Lawley treats are shown to be asymptotically normal (a result not following directly and easily from the usual theory of maximum likelihood estimates).

Bartlett in [4], [5] and [6] (summarized in [7]) has considered testing the hypothesis that $p - q$ roots of Σ or of R are equal and suggests the use of a criterion discussed in Section 3 of this paper. He recommends using χ^2-distributions as approximate distributions and justifies his recommendations by finding the asymptotic distributions in some cases. Lawley [11] has investigated more thoroughly the tests of equality of $p - q$ smallest roots of Σ, evaluating the expectation of the criterion to terms of order $1/(N-1)^2$ when the null hypothesis is true, but he does not consider explicitly the distribution of the roots involved. Besides a generalization of this work, another contribution in the present paper is a new proposal for confidence intervals and test procedures for the smallest roots. The approach here also gives some information about the similar analysis of correlation matrices. A discussion of the procedures available and their purposes indicates their usefulness (Section 6).

2. The asymptotic distribution of the characteristic roots and vectors of a sample covariance matrix. The matrix A is distributed as $\sum_1^n x_\alpha x_\alpha'$, where $n = N - 1$ and x_α is distributed according to $N(0, \Sigma)$. The first two moments of the elements of A are

$$\mathcal{E}a_{ij} = n\sigma_{ij}, \tag{2.1}$$

$$\mathcal{E}(a_{ij} - n\sigma_{ij})(a_{gh} - n\sigma_{gh}) = n(\sigma_{ih}\sigma_{jg} + \sigma_{ig}\sigma_{jh}). \tag{2.2}$$

By the multivariate central limit theorem $(1/n^{\frac{1}{2}})(A - n\Sigma)$ is asymptotically normally distributed with mean 0 and covariances given by (2.2) divided by n.

Let $(\gamma_1 \cdots \gamma_p) = \Gamma$. Then (1.3) can be summarized as

$$\Sigma\Gamma = \Gamma\Delta \tag{2.3}$$

where Δ is a diagonal matrix with $\delta_1, \cdots, \delta_p$ as the diagonal elements. From $\Gamma'\Gamma = I$ and (2.3), we find

$$\Gamma'\Sigma\Gamma = \Delta. \tag{2.4}$$

It will be convenient to carry out our analysis in terms of the transformed variates $\Gamma'x$. The characteristic roots of the transform of A, namely $\Gamma'A\Gamma$ are the roots of A, and the characteristic vectors of $\Gamma'A\Gamma$ are $\Gamma'c_i = e_i$, say. A set of characteristic vectors for Δ are $\{\epsilon_i\}$, where $(\epsilon_1 \cdots \epsilon_p) = I$. We can, therefore, study the asymptotic distribution of $(d_1, \cdots, d_p, e_1, \cdots, e_p)$. We know that $(1/n^{\frac{1}{2}})(\Gamma'A\Gamma - n\Delta) = U$, say, has a limiting normal distribution with mean 0 and covariances $\mathcal{E}u_{ij}u_{gh} = \delta_i\delta_j(\delta_{ih}\delta_{jg} + \delta_{ig}\delta_{jh})$, where δ_{ij} is the Kronecker delta. Because U is symmetric, $u_{ij} = u_{ji}$. The functionally independent u_{ij} are uncorrelated, and $\mathcal{E}u_{ii}^2 = 2\delta_i^2$, $\mathcal{E}u_{ij}^2 = \delta_i\delta_j$ for $i \neq j$.

Let $T = \Gamma'S\Gamma = (1/n)\Gamma'A\Gamma$. Then $U = n^{\frac{1}{2}}(T - \Delta)$. The equations defining e_i and d_i can be written

$$TE = ED, \tag{2.5}$$

$$E'E = I, \tag{2.6}$$

where D is a diagonal matrix with $d_1, \cdots, d_p$ as diagonal elements and $E =$

$(e_1 \cdots e_p)$. Equation (2.5) can be replaced by

$$T = EDE'. \tag{2.7}$$

Equations (2.6) and (2.7), when $d_1 > \cdots > d_p$, define D and E uniquely except that e_i can be replaced by $-e_i$. We shall therefore require that $e_{ii} > 0$ ($e_{ii} \neq 0$ with probability 1).

Let us first consider the case of all characteristic roots of Σ equal, that is, $\delta_1 = \cdots = \delta_p = \lambda$, say. Then $\Delta = \lambda I$. Since the exact distribution of D and E can be given simply (Theorems 13.3.2 and 13.3.3 of [2], for example), the asymptotic distribution is derived here because the demonstration introduces methods and results to be used later. Let $n^{\frac{1}{2}}(D - \lambda I) = H$. We have

$$T = \lambda I + (1/n^{\frac{1}{2}})U, \tag{2.8}$$

$$\begin{aligned} EDE' &= E(\lambda I + (1/n^{\frac{1}{2}})H)E' \\ &= \lambda I + (1/n^{\frac{1}{2}})EHE'. \end{aligned} \tag{2.9}$$

Thus

$$U = EHE'. \tag{2.10}$$

This (with the ordering of the diagonal elements of H and the requirement that the diagonal elements of E be positive) defines E and H uniquely as a continuous function of U except for a set of probability 0. It follows that the limiting distribution of H and E is the distribution derived from (2.10) according to the limiting distribution of U (justified by the theorem of Section 7, for example). The limiting distribution of $U = U'$ is normal with mean 0 and $\mathcal{E}u_{ii}^2 = 2\lambda^2$, $\mathcal{E}u_{ij}^2 = \lambda^2$, $i \neq j$, and functionally independent elements are uncorrelated. The density of this distribution is proportional to

$$\exp[-\operatorname{tr} UU'/(4\lambda^2)] = \exp[-\operatorname{tr} U^2/(4\lambda^2)]. \tag{2.11}$$

The density of the limiting distribution of $h_1, \cdots, h_p$ (computed from (2.10)) is

$$\frac{K(p)}{\lambda^{p(p+1)/2}} e^{-\Sigma_i h_i^2/(4\lambda^2)} \prod_{i<j} (h_i - h_j) = f(h_1, \cdots, h_p; \lambda, p), \tag{2.12}$$

where $h_1 > \cdots > h_p$ and 0 otherwise and

$$1/K(p) = 2^{p(p+3)/4} \prod_{i=1}^{p} \Gamma[\tfrac{1}{2}(p + 1 - i)]. \tag{2.13}$$

This density is derived as Theorem 2 of [1] and Theorem 13.3.5 of [2].

The distribution of T (where $\Sigma = \lambda I$) is the same as the distribution of $P'TP$ for any orthogonal matrix P. Hence the distribution of EP is the same as that of E, except for the effect of requiring $e_{ii} > 0$. The normalized Haar invariant measure is the unique probability distribution on the group of orthogonal matrices that has the property that the distribution of EP is the same as that of P. The distribution involved here is 2^p times the Haar measure over orthogonal matrices with positive diagonal elements. We shall call this the conditional Haar invariant distribution. Thus E has the conditional Haar invariant distribution. This distribution is discussed in more detail in Section 13.3 of [2]. It can be considered as a uniform distribution of rotations; for instance, for $p = 2$, a rotation is specified by one angle and it has a rectangular distribution. The limiting distribution of E is, of course, this same Haar invariant distribution. This also follows from the density of the limiting distribution of U, which is invariant under the transformation $P'TP$.

We now proceed to the general case in which some of δ_i are equal; that is, the multiplicities of the characteristic roots may be arbitrary. In particular, this includes the case where the roots are all different, that is, of multiplicity 1; in that case all the submatrices below are scalars.

Let the multiplicities of the roots of Σ be $q_1, q_2, \cdots, q_r$. Let

$$\begin{aligned} \delta_1 &= \cdots = \delta_{q_1} = \lambda_1, \\ \delta_{q_1+1} &= \cdots = \delta_{q_1+q_2} = \lambda_2, \\ &\vdots \\ \delta_{p-q_r+1} &= \cdots = \delta_p = \lambda_r, \end{aligned} \tag{2.14}$$

where $\lambda_1 > \lambda_2 > \cdots > \lambda_r > 0$. We partition the matrices into submatrices with $q_1, \cdots, q_r$ rows and columns

$$\Delta = \begin{pmatrix} \Delta_1 & 0 & \cdots & 0 \\ 0 & \Delta_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \Delta_r \end{pmatrix} = \begin{pmatrix} \lambda_1 I & 0 & \cdots & 0 \\ 0 & \lambda_2 I & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_r I \end{pmatrix}, \tag{2.15}$$

$$U = \begin{pmatrix} U_{11} & U_{12} & \cdots & U_{1r} \\ U_{21} & U_{22} & \cdots & U_{2r} \\ \vdots & \vdots & & \vdots \\ U_{r1} & U_{r2} & \cdots & U_{rr} \end{pmatrix}, \tag{2.16}$$

$$D = \begin{pmatrix} D_1 & 0 & \cdots & 0 \\ 0 & D_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & D_r \end{pmatrix}, \qquad H = \begin{pmatrix} H_1 & 0 & \cdots & 0 \\ 0 & H_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & H_r \end{pmatrix}, \tag{2.17}$$

$$E = \begin{pmatrix} E_{11} & E_{12} & \cdots & E_{1r} \\ E_{21} & E_{22} & \cdots & E_{2r} \\ \vdots & \vdots & & \vdots \\ E_{r1} & E_{r2} & \cdots & E_{rr} \end{pmatrix}. \tag{2.18}$$

We shall now give a heuristic derivation of the general limiting distribution. A rigorous proof is completed in Section 7. Let $H_k = n^{\frac{1}{2}}(D_k - \lambda_k I)$ and $n^{\frac{1}{2}}E_{kl} = F_{kl}$, $k \neq l$. Then

$$
\begin{aligned}
T &= \Delta + \frac{1}{\sqrt{n}} U \\
&= \begin{pmatrix} \lambda_1 I & 0 & \cdots & 0 \\ 0 & \lambda_2 I & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_r I \end{pmatrix} + \frac{1}{\sqrt{n}} \begin{pmatrix} U_{11} & U_{12} & \cdots & U_{1r} \\ U_{21} & U_{22} & \cdots & U_{2r} \\ \vdots & \vdots & & \vdots \\ U_{r1} & U_{r2} & \cdots & U_{rr} \end{pmatrix} \\
&= \begin{pmatrix} E_{11} & n^{-\frac{1}{2}} F_{12} & \cdots & n^{-\frac{1}{2}} F_{1r} \\ n^{-\frac{1}{2}} F_{21} & E_{22} & \cdots & n^{-\frac{1}{2}} F_{2r} \\ \vdots & \vdots & & \vdots \\ n^{-\frac{1}{2}} F_{r1} & n^{-\frac{1}{2}} F_{r2} & \cdots & E_{rr} \end{pmatrix} \\
&\quad \cdot \begin{pmatrix} \lambda_1 I + n^{-\frac{1}{2}} H_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 I + n^{-\frac{1}{2}} H_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_r I + n^{-\frac{1}{2}} H_r \end{pmatrix} \\
&\quad \cdot \begin{pmatrix} E'_{11} & n^{-\frac{1}{2}} F'_{21} & \cdots & n^{-\frac{1}{2}} F'_{r1} \\ n^{-\frac{1}{2}} F'_{12} & E'_{22} & \cdots & n^{-\frac{1}{2}} F'_{r2} \\ \vdots & \vdots & & \vdots \\ n^{-\frac{1}{2}} F'_{1r} & n^{-\frac{1}{2}} F'_{2r} & \cdots & E'_{rr} \end{pmatrix} \\
&= \begin{pmatrix} \lambda_1 E_{11} E'_{11} & 0 & \cdots & 0 \\ 0 & \lambda_2 E_{22} E'_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_r E_{rr} E'_{rr} \end{pmatrix} \\
&\quad + \frac{1}{\sqrt{n}} \begin{pmatrix} E_{11} H_1 E'_{11} & \lambda_1 E_{11} F'_{21} + \lambda_2 F_{12} E'_{22} & \cdots & \lambda_1 E_{11} F'_{r1} + \lambda_r F_{1r} E'_{rr} \\ \lambda_1 F_{21} E'_{11} + \lambda_2 E_{22} F'_{12} & E_{22} H_2 E'_{22} & \cdots & \lambda_2 E_{22} F'_{r2} + \lambda_r F_{2r} E'_{rr} \\ \vdots & \vdots & & \vdots \\ \lambda_1 F_{r1} E'_{11} + \lambda_r E_{rr} F'_{1r} & \lambda_2 F_{r2} E'_{22} + \lambda_r E_{rr} F'_{2r} & \cdots & E_{rr} H_r E'_{rr} \end{pmatrix} \\
&\quad + (1/n) M,
\end{aligned} \tag{2.19}
$$

where the submatrices of M are sums of products of E_{kk}, λ_k, H_k, F_{kl} and $1/n^{\frac{1}{2}}$. The orthogonality of E implies

$$
\begin{aligned}
\begin{pmatrix} I & 0 & \cdots & 0 \\ 0 & I & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & I \end{pmatrix} &= \begin{pmatrix} E_{11} E'_{11} & 0 & \cdots & 0 \\ 0 & E_{22} E'_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & E_{rr} E'_{rr} \end{pmatrix} \\
&\quad + \frac{1}{\sqrt{n}} \begin{pmatrix} 0 & E_{11} F'_{21} + F_{12} E'_{22} & \cdots & E_{11} F'_{r1} + F_{1r} E'_{rr} \\ F_{21} E'_{11} + E_{22} F'_{12} & 0 & \cdots & E_{22} F'_{r2} + F_{2r} E'_{rr} \\ \vdots & \vdots & & \vdots \\ F_{r1} E'_{11} + E_{rr} F'_{1r} & F_{r2} E'_{22} + E_{rr} F'_{2r} & \cdots & 0 \end{pmatrix}, \\
&\quad + (1/n) W,
\end{aligned} \tag{2.20}
$$

where the submatrices of W are sums of products of F_{kl}. From (2.20) we see that

$$
E_{kk} E'_{kk} = I - (1/n) W_{kk}. \tag{2.21}
$$

When this is inserted in (2.19), the first matrix on the right hand side is simply Δ plus $1/n$ times a matrix with diagonal blocks $-\lambda_k W_{kk}$. Then the other submatrix equations in (2.19) and (2.20) (multiplied by $n^{\frac{1}{2}}$) are

$$U_{kk} = E_{kk}H_kE'_{kk} + (1/n^{\frac{1}{2}})(M_{kk} - \lambda_k W_{kk}), \tag{2.22}$$

$$U_{kl} = \lambda_k E_{kk}F'_{lk} + \lambda_l F_{kl}E'_{ll} + (1/n^{\frac{1}{2}})M_{kl}, \quad k \neq l, \tag{2.23}$$

$$0 = E_{kk}F'_{lk} + F_{kl}E'_{ll} + (1/n^{\frac{1}{2}})W_{kl}, \quad k \neq l. \tag{2.24}$$

From (2.21) and (2.22) we argue (and prove in Section 7) that the limiting distribution of E_{kk} and H_k is the unique distribution such that E_{kk} is orthogonal and the distribution of $E_{kk}H_kE'_{kk}$ is the limiting distribution of U_{kk}; that is, it is the limiting distribution given before (with λ replaced by λ_k and p replaced by q_k). From (2.24) we see that $E_{kk}F'_{lk}$ and $-F_{kl}E'_{ll}$ are asymptotically equivalent, and from (2.23) we see that the limiting distribution of either is the limiting distribution of $[1/(\lambda_k - \lambda_l)]U_{kl}$. The limiting distribution of U_{kl} is normal and has density with exponent $-\frac{1}{2}\operatorname{tr} U_{kl}U'_{kl}/(\lambda_k\lambda_l)$; hence the density is the same for $PU_{kl}Q$ where P and Q are arbitrary orthogonal matrices. Since the limiting distributions of U_{kk} (which determines E_{kk}) and U_{kl} are independent, the limiting distribution of the elements of either F'_{lk} or $-F_{kl}$ is normal with means 0, variances $\lambda_k\lambda_l/(\lambda_k - \lambda_l)^2$ and correlations 0. Note that $n^{\frac{1}{2}}(E_{kk}E'_{kk} - I)$ converges stochastically to 0.

THEOREM 1. *Let $d_1 > d_2 > \cdots > d_p$ be the characteristic roots and $e_1, \cdots, e_p$ the corresponding characteristic vectors (normalized by $e'_ie_i = 1$ and $e_{ii} > 0$) of a covariance matrix based on a sample of $n + 1$ from $N(\mu, \Delta)$ where Δ has the structure indicated by* (2.15) *and λ_k is of multiplicity q_k ($\lambda_1 > \lambda_2 > \cdots > \lambda_r$). Let the diagonal matrix with diagonal elements $d_1, \cdots, d_p$ be D and the matrix $(e_1 \cdots e_p)$ be E partitioned as in* (2.18). *Then $n^{\frac{1}{2}}(D_k - \lambda_k I) = H_k$, E_{kk}, and $n^{\frac{1}{2}}E_{kl} = F_{kl}$, $k < l$, are independent in the limiting distribution. The limiting distribution of the diagonal elements of H_k has density $f(x_1, \cdots, x_{q_k}; \lambda_k, q_k)$ given by* (2.10); *the limiting distribution of E_{kk} is the conditional Haar invariant distribution; and the limiting distribution of the elements of F_{kl}, $k < l$, is normal with means* 0, *variances $\lambda_k\lambda_l/(\lambda_k - \lambda_l)^2$, and correlations* 0. *$(E_{kk}F'_{lk} + F_{kl}E'_{ll})$ converges stochastically to* 0 *for $k \neq l$.*

Now consider these results in terms of the original coordinates. The characteristic roots are the same but the characteristic vectors make up the matrix $C = \Gamma E$. Let $C = (C_1C_2 \cdots C_r)$ and $\Gamma = (\Gamma_1\Gamma_2 \cdots \Gamma_r)$. Then

$$\begin{aligned}(C_1C_2 \cdots C_r) &= (\Gamma_1\Gamma_2 \cdots \Gamma_r)\begin{pmatrix} E_{11} & E_{12} & \cdots & E_{1r} \\ E_{21} & E_{22} & \cdots & E_{2r} \\ \vdots & \vdots & & \vdots \\ E_{r1} & E_{r2} & \cdots & E_{rr}\end{pmatrix} \\ &= \left(\sum_{l=1}^{r} \Gamma_l E_{l1} \sum_{l=1}^{r} \Gamma_l E_{l2} \cdots \sum_{l=1}^{r} \Gamma_l E_{lr}\right),\end{aligned} \tag{2.25}$$

$$C_k = \Gamma_k E_{kk} + n^{-\frac{1}{2}} \sum_{l \neq k} \Gamma_l F_{lk}. \tag{2.26}$$

If $q_k = 1$, then C_k is a vector, say c_i, the ith column of C. Then $E_{kk} = e_{ii}$ converges stochastically to 1; in fact $n^{\frac{1}{2}}(e_{ii} - 1)$ converges stochastically to 0. Then $n^{\frac{1}{2}}(c_i - \gamma_i)$ has a limiting normal distribution with mean 0 and the covariance between its gth and hth elements is

$$(2.27) \qquad \sum_{j \neq i} [\delta_j \delta_i / (\delta_j - \delta_i)^2] \gamma_{gj} \gamma_{hj} .$$

Note that the multiplicities of roots other than the ith one are irrelevant. If the ith and jth columns of Γ correspond to roots of multiplicity 1, the covariance between the gth element of the ith column of C and the hth element of the jth column is

$$(2.28) \qquad -[\delta_i \delta_j / (\delta_i - \delta_j)^2] \gamma_{gj} \gamma_{hi} , \; i \neq j.$$

If $q_k > 1$, E_{kk} is a random matrix with the invariant distribution of orthogonal matrices as its limiting distribution. (2.25) shows $E_{kk} = \Gamma_k' C_k$. From (2.26) we see that

$$(2.29) \qquad n^{\frac{1}{2}}(C_k - \Gamma_k \Gamma_k' C_k) = \sum_{l \neq k} \Gamma_l \Gamma_l' n^{\frac{1}{2}} C_k = \sum_{l \neq k} \Gamma_l F_{lk}$$

has a limiting (singular) multivariate normal distribution with means 0 and variances and covariances derived from the theorem.

If the other normalization for the principal components is used, we treat $d_i^{\frac{1}{2}} c_{gi}$. If δ_i is a root of multiplicity 1, then $n^{\frac{1}{2}}[d_i^{\frac{1}{2}} c_{gi} - \delta_i^{\frac{1}{2}} \gamma_{gi}]$ is asymptotically distributed as

$$(2.30) \qquad n^{\frac{1}{2}} \delta_i^{\frac{1}{2}} (c_{gi} - \gamma_{gi}) + n^{\frac{1}{2}} \gamma_{gi} (d_i - \delta_i) / (2 \delta_i^{\frac{1}{2}})$$

and the two terms are asymptotically independent. The asymptotic variances and covariances can be calculated from the above results. Similarly we can treat the estimates of $\gamma_{gi}(\delta_i - \lambda_r)^{\frac{1}{2}}$, where the last root is λ_r, of multiplicity q_r, which was the normalization Lawley [10] used. The theory of this paper includes asymptotic normality, which Lawley's did not.

3. Statistical inference about characteristic roots of covariance matrices. The population covariance matrix Σ can be written as $\Gamma \Delta \Gamma'$, where Γ is an orthogonal matrix with $\gamma_{ii} > 0$ and Δ is a diagonal matrix with diagonal elements $\delta_1 \geqq \cdots \geqq \delta_p$; the representation is unique if the roots δ_i are all different and no γ_{ii} is 0. The sample convariance matrix S can similarly be represented as CDC', and the representation is unique with probability 1. If the population representation is assumed unique, C is the maximum likelihood estimate of Γ and $(n/N)D$ is the maximum likelihood estimate of Δ.

THEOREM 2. *If the characteristic roots of the population covariance matrix are $\lambda_1 > \cdots > \lambda_r$ with multiplicities $q_1, \cdots, q_r$, respectively, the maximum likelihood estimate of λ_k is*

$$(3.1) \qquad \hat{\lambda}_k = (1/q_k)(n/N) \sum_{j \varepsilon L_k} d_j , \qquad k = 1, \cdots, r,$$

where L_k is the set of integers $q_1 + \cdots + q_{k-1} + 1, \cdots, q_1 + \cdots + q_k$, and the *maximum likelihood estimate of* Γ_k *is* C_k *multiplied by any* $(q_k \times q_k)$ *orthogonal matrix on the right such that* $\hat{\gamma}_{jj} > 0$.

PROOF. The logarithm of the likelihood function after maximization with respect to μ is

$$\begin{aligned} -\tfrac{1}{2}N \log |\Sigma| - \tfrac{1}{2} \operatorname{tr} \Sigma^{-1} nS &= -\tfrac{1}{2}N \log |\Gamma\Delta\Gamma'| - \tfrac{1}{2}n \operatorname{tr} \Gamma\Delta^{-1}\Gamma' CDC' \\ &= -\tfrac{1}{2}N \log |\Delta| - \tfrac{1}{2}n \operatorname{tr} \Delta^{-1}P'DP, \end{aligned} \tag{3.2}$$

where $P = C'\Gamma$. The trace term can be written as

$$\sum_{i=1}^{p} \sum_{k=1}^{r} d_i q_{ik}/\lambda_k , \tag{3.3}$$

where $q_{ik} = \sum_j p_{ij}^2$ for $j \varepsilon L_k$. Since P is orthogonal, $\sum_{i=1}^{p} q_{ik} = q_k$ and

$$\sum_{k=1}^{r} q_{ik} = 1.$$

Then (3.3) is minimized by $q_{ik} = 1, i \varepsilon L_k$, and $q_{ik} = 0, i \not\varepsilon L_k$, and the minimum is $\sum_{k=1}^{r} \sum_{i \varepsilon L_k} d_i/\lambda_k$ because for any other set of q's the sum can be decreased. To demonstrate this fact let l be the smallest index k for which $q_{ik} < 1$ for some $i \varepsilon L_k$, and let j be the smallest of such indices i. Then $q_{ik} = 1, i \varepsilon L_k$, and $q_{ik} = 0, i \not\varepsilon L_k$, for $k < l$, because $\sum_i q_{ik} = q_k$; and $q_{ik} = 0$ for $i < j, i \not\varepsilon L_k$, because $\sum_k q_{ik} = 1$. Thus there is an index $k > l$ such that $q_{jk} > 0$ and there is an index $i > j$ $(i > q_1 + \cdots + q_l)$ such that $q_{il} > 0$. If Δq_{jl} is the smaller of q_{jk} and q_{il}, the sum can be decreased by replacing q_{jl}, q_{jk}, q_{il} and q_{ik} by $q_{jl} + \Delta q_{jl}$, $q_{jk} - \Delta q_{jl}$, $q_{il} - \Delta q_{jl}$ and $q_{ik} + \Delta q_{jl}$, respectively; the change in (3.3) is

$$\Delta q_{jl} \left(\frac{d_j}{\lambda_l} - \frac{d_j}{\lambda_k} - \frac{d_i}{\lambda_l} + \frac{d_i}{\lambda_k} \right) = \Delta q_{jl}(d_j - d_i) \left(\frac{1}{\lambda_l} - \frac{1}{\lambda_k} \right), \tag{3.4}$$

which is negative. This shows that (3.3) is minimized with respect to P by $\hat{p}_{ij} = 0$ for $i \varepsilon L_k$ and $j \varepsilon L_l$, $k \neq l$, and any orthogonal $\hat{P}_{kk} = (\hat{p}_{ij})$, $i, j \varepsilon L_k$. Then $\hat{\Gamma} = \hat{C}\hat{P}$. The resulting likelihood function

$$-\tfrac{1}{2}N \sum_{k=1}^{r} q_k \log \lambda_k - \tfrac{1}{2}n \sum_{k=1}^{r} \sum_{i \varepsilon L_k} (d_i/\lambda_k) \tag{3.5}$$

is maximized by (3.1). Q.E.D.

As an example, if Σ is assumed to be $\Psi + \sigma^2 I$, where Ψ is positive semi-definite of rank q, the maximum likelihood estimate of σ^2 is n/N times the mean of the $p - q$ smallest roots of S. (Note that if $\Sigma = \Psi + \sigma^2 I$, each root of Σ is the corresponding root of Ψ plus σ^2.)

COROLLARY 1. *The likelihood ratio criterion for testing the hypothesis*

$$\delta_{q_1 + \cdots + q_{k-1} + 1} = \delta_{q_1 + \cdots + q_k} \tag{3.6}$$

is

$$[\prod_{j \varepsilon L_k} d_j / (q_k^{-1} \sum_{j \varepsilon L_k} d_j)^{q_k}]^{\frac{1}{2}N}. \tag{3.7}$$

The criterion is a monotonic function of the geometric mean of the relevant roots divided by the arithmetic mean. When the hypothesis (3.6) is equality of all population roots, the criterion involves all sample roots

$$\prod_{j=1}^{p} d_j = |S| \text{ and } \sum_{i=1}^{p} d_i = \operatorname{tr} S.$$

The hypothesis then specifies that ellipsoids of constant densities are spheres, and it is called the hypothesis of sphericity [15]. In general the hypothesis (3.6) specifies that q_k of the principal axes are equal in length, $q_1 + \cdots + q_{k-1}$ are larger than the q_k equal ones, and $q_{k+1} + \cdots + q_r$ are smaller. The likelihood ratio criterion for a hypothesis consisting of several sets of equalities (3.6) is the product of the corresponding quantities (3.7).

The limiting distributions considered in the preceding sections yield the limiting distribution of the test criterion. The logarithm of (3.7) multiplied by -2 is asymptotically equivalent under the null hypothesis to

$$\begin{aligned}
&-n \log \prod_i d_i + nq_k \log \frac{\sum_i d_i}{q_k} \\
&= -n \sum_i \log (\lambda_k + n^{-\frac{1}{2}} h_i) + nq_k \log \frac{\sum_i (\lambda_k + n^{-\frac{1}{2}} h_i)}{q_k} \\
&= n \left\{ -\sum_i \log \left(1 + \frac{h_i}{\lambda_k n^{\frac{1}{2}}}\right) + q_k \log \left(1 + \frac{\sum_i h_i}{q_k \lambda_k n^{\frac{1}{2}}}\right) \right\} \qquad (3.8) \\
&= n \left\{ -\sum_i \left[\frac{h_i}{\lambda_k n^{\frac{1}{2}}} - \frac{h_i^2}{2\lambda_k^2 n} + \cdots \right] + q_k \left[\frac{\sum_i h_i}{q_k \lambda_k n^{\frac{1}{2}}} - \frac{(\sum_i h_i)^2}{2q_k^2 \lambda_k^2 n} + \cdots \right] \right\} \\
&= n \left\{ \sum_i \frac{h_i^2}{2\lambda_k^2 n} - \cdots - \frac{(\sum_i h_i)^2}{2q_k \lambda_k^2 n} + \cdots \right\},
\end{aligned}$$

where the sum is over the set L_k. By a familiar argument the asymptotic distribution of this quantity is the asymptotic distribution of

$$\begin{aligned}
[1/(2\lambda_k^2)] [\sum_i h_i^2 - (1/q_k) (\sum_i h_i)^2] &= [1/(2\lambda_k^2)] [\operatorname{tr} U_{kk}^2 - (1/q_k) (\operatorname{tr} U_{kk})^2] \\
&= [1/(2\lambda_k^2)] [\operatorname{tr} U_{kk} U_{kk}' - (1/q_k) (\operatorname{tr} U_{kk})^2] \qquad (3.9) \\
&= [1/(2\lambda_k^2)] [2 \sum_{i<j} u_{ij}^2 + \sum_i u_{ii}^2 - (1/q_k) (\sum_i u_{ii})^2].
\end{aligned}$$

In the limit u_{ij} $(i < j)$ are independently normal with means 0 and variances λ_k^2, and u_{ii} are independently normal with means 0 and variances $2\lambda_k^2$. Thus $\sum_{i<j} u_{ij}^2/\lambda_k^2$ is asymptotically χ^2 with $\frac{1}{2}q_k(q_k - 1)$ degrees of freedom;

$$\tfrac{1}{2}[\sum_i u_{ii}^2 - (\sum_i u_{ii})^2/q_k]/\lambda_k^2$$

is asymptotically χ^2 with $q_k - 1$ degrees of freedom; and (3.9) is asymptotically χ^2 with $\frac{1}{2}q_k(q_k + 1) - 1$ degrees of freedom. In the case of $q_k = p$, the criterion is the product of the criterion that the covariances are 0 (that is, that the variables are independent) and the criterion that the variances are equal given that the covariances are 0; asymptotically these two criteria correspond to the two χ^2's considered above. It should be noted that the hypothesis does not specify the size of the hypothesized equal roots.

Suppose that it is assumed that q_k roots are equal, let us say the last q_r roots. The estimate of λ_r is $\bar{d} = \sum_{p-q_r+1}^{p} d_i/q_r$. Then

$$n^{\frac{1}{2}}(\bar{d} - \lambda_r) = (\operatorname{tr} U_{rr})/q_r = (1/q_r) \sum_{p-q_r+1}^{p} u_{ii} \tag{3.10}$$

is asymptotically normally distributed with mean 0 and variance $2\lambda_r^2/q_r$. Thus asymptotically $(\frac{1}{2}nq_r)^{\frac{1}{2}}(\bar{d} - \lambda_r)/\lambda_r$ has the standardized normal distribution. Let t_α be such that

$$\int_{-t_\alpha}^{t_\alpha} [1/(2\pi^{\frac{1}{2}})]e^{-\frac{1}{2}u^2}\, du = 1 - \alpha. \tag{3.11}$$

Then asymptotically with probability $1 - \alpha$

$$[nq_r(\bar{d} - \lambda_r)^2]/(2\lambda_r^2) \leqq t_\alpha^2 . \tag{3.12}$$

This gives the asymptotic confidence interval (with asymptotic confidence $1 - \alpha$) of

$$\bar{d}/[1 + (2/\{nq_r\})^{\frac{1}{2}}t_\alpha] \leqq \lambda^* \leqq \bar{d}/[1 - (2/\{nq_r\})^{\frac{1}{2}}t_\alpha], \tag{3.13}$$

where n, q_r and α are such that the denominator on the right is positive.

Suppose we wish to consider whether the smallest q_r roots of Σ are so small that we wish to ignore the last q_r principal components. If we assume that the last q_r roots of Σ are equal, say, λ_r, then we can use a one-sided confidence interval instead of (3.13). We then have

$$\lambda^* \leqq \bar{d}/[1 - (2/\{nq_r\})^{\frac{1}{2}}t_{2\alpha}]. \tag{3.14}$$

If $\bar{d}/[1 - (2/\{nq_r\})^{\frac{1}{2}}t_{2\alpha}]$ is sufficiently small, we may decide to study only the first $p - q_r$ principal components.

If the last q_r roots are all different

$$n^{\frac{1}{2}}\left(\left[\bar{d} - \left(\sum_{p-q_r+1}^{p} \delta_i\right)\Big/ q_r\right]\right) = (1/q_r) \sum_{p-q_r+1}^{p} h_i \tag{3.15}$$

is asymptotically normal with mean 0 and variance $2\sum_{p-q_r+1}^{p} \delta_i^2/q_r^2$, which is greater than $2q_r(\sum_{p-q_r+1}^{p} \delta_i/q_r)^2$. Thus the assumption of equality of roots will lead to underestimating the asymptotic variances of $\bar{d}$ when the roots are not equal and also to reducing the confidence interval. A more conservative procedure is to estimate the asymptotic covariance of $\bar{d}$ by the consistent estimate

$2\sum d_i^2/q_r^2$ and use as a confidence interval

$$\bar{d} - t_\alpha\left[\left(2\sum_{p-q_r+1}^{p} d_i^2\right)\Big/(nq_r^2)\right]^{\frac{1}{2}} < (1/q_r)\sum_{p-q_r+1}^{p}\delta_i < \bar{d} + t_\alpha\left[\left(2\sum_{p-q_r+1}^{p} d_i^2\right)\Big/(nq_r^2)\right]^{\frac{1}{2}}, \tag{3.16}$$

or

$$(1/q_r)\sum_{p-q_r+1}^{p}\delta_i < \bar{d} + t_{2\alpha}\left[\left(2\sum_{p-q_r+1}^{p} d_i^2\right)\Big/(nq_r^2)\right]^{\frac{1}{2}}. \tag{3.17}$$

The last inequality gives an upper asymptotic confidence bound on the average of the variances of the last q_r principal components. If the investigator finds this sufficiently small, he may want to neglect the last q_r components.

Another null hypothesis that might be considered is that some q_k δ_i's are equal to a specified number say λ. The likelihood ratio criterion for this hypothesis is

$$\left[\frac{\prod_i\left(\frac{n}{N}d_i\right)}{\lambda^{q_k}}\right]^{\frac{1}{2}N} e^{-\frac{1}{2}N\left[\frac{n}{N}\Sigma_i\frac{d_i}{\lambda}-q_k\right]}, \tag{3.18}$$

where the index i runs over the q_k relevant values. This criterion is the product of (3.7), the criterion for equality of roots, and

$$\left[\frac{\sum_i \frac{n}{N}d_i}{q_k\,\lambda}\right]^{\frac{1}{2}Nq_k} e^{-\frac{1}{2}N\left[\frac{n}{N}\Sigma_i\frac{d_i}{\lambda}-q_k\right]}, \tag{3.19}$$

the criterion that the common value of the roots is λ if the roots are assumed equal. This last criterion is a function only of $\sum_i d_i$. The asymptotic normality of this sum implies that the logarithm of (3.19) multiplied by (-2) has a limiting χ^2-distribution with 1 degree of freedom.

There are other test procedures that might be considered. If the last q_r roots are assumed equal, one might use the largest of the q_r smallest sample roots to evaluate the assumed equal population roots. The asymptotic theory would be based on the distribution of the largest of the q_r variables with density $f(d_{p-q_r+1}, \cdots, d_p; \lambda_r, q_r)$. A number of procedures have been suggested for testing $\Sigma = \sigma^2 I$ (see [16], for example); they depend on the p sample roots of S (and are hence invariant under orthogonal transformations); any such procedure based on the q_r smallest roots of S can be used to test equality of the q_r smallest roots of Σ.

If the smallest population roots are assumed different, one might want to determine whether the largest of these roots is small enough to be neglected; the asymptotic distribution of this root is normal. Another hypothesis is that the ratio of "unexplained variance" to the total is not greater than a fraction

f; that is, that

$$\sum_{i=q+1}^{p} \delta_i \leqq f \sum_{i=1}^{p} \delta_i \tag{3.20}$$

for some q. This hypothesis could be tested by use of

$$\sum_{q+1}^{p} d_i - f\sum_{1}^{p} d_i = -f\sum_{1}^{q} d_i + (1-f)\sum_{q+1}^{p} d_i , \tag{3.21}$$

which is asymptotically normal.

It might also be pointed out that the general asymptotic theory of the roots shows that the tests considered in this section are consistent.

4. Principal component analysis of a correlation matrix. Let $R = (r_{ij})$, where $r_{ij} = a_{ij}/(a_{ii}a_{jj})^{\frac{1}{2}}$. Let $A^* = (a_{ij}^*)$, where $a_{ij}^* = a_{ij}/(\sigma_{ii}\sigma_{jj})^{\frac{1}{2}}$; then A^* is distributed in a fashion similar to that of A, but with σ_{ij} replaced by ρ_{ij}. Let

$$n^{-\frac{1}{2}}(A^* - n\bar{R}) = X. \tag{4.1}$$

Then X is asymptotically normally distributed with mean 0 and covariances

$$\mathcal{E}x_{ij}x_{gh} = \rho_{ig}\rho_{jh} + \rho_{ih}\rho_{jg} , \tag{4.2}$$

and $n^{\frac{1}{2}}(R - \bar{R})$ is asymptotically distributed as

$$X - \tfrac{1}{2}(\bar{R}X_0 + X_0\bar{R}), \tag{4.3}$$

where X_0 is a diagonal matrix with x_{ii} as diagonal elements.

Let $B = (\beta_{ij})$ be an orthogonal matrix such that

$$\bar{R}B = B\Theta, \qquad B'\bar{R}B = \Theta, \tag{4.4}$$

where Θ is a diagonal matrix with diagonal elements $\theta_1 \geqq \cdots \geqq \theta_p$, the characteristic roots of $\bar{R}$. Then $n^{\frac{1}{2}}B'(R - \bar{R})B = n^{\frac{1}{2}}(B'RB - \Theta)$ is asymptotically distributed as

$$\begin{aligned} Z &= B'\,X\,B - \tfrac{1}{2}\left(\Theta\,B'\,X_0\,B + B'\,X_0\,B\,\Theta\right), \\ z_{kl} &= \sum_{i,j} \beta_{ik}x_{ij}\beta_{jl} - \tfrac{1}{2}(\theta_k + \theta_l)\sum_{i} \beta_{ik}x_{ii}\beta_{il} . \end{aligned} \tag{4.5}$$

Then

$$\begin{aligned} \mathcal{E}z_{kl}z_{mr} &= \theta_k\theta_l[\delta_{km}\delta_{lr} + \delta_{kr}\delta_{lm}] \\ &\quad - [(\theta_k + \theta_l)\theta_m\theta_r + \theta_k\theta_l(\theta_m + \theta_r)]\sum_i \beta_{ik}\beta_{il}\beta_{im}\beta_{ir} \\ &\quad + \tfrac{1}{2}(\theta_k + \theta_l)(\theta_m + \theta_r)\sum_{i,g} \beta_{ik}\beta_{il}\beta_{gm}\beta_{gr}\rho_{ig}^2 . \end{aligned} \tag{4.6}$$

In particular,

$$\mathcal{E}z_{kk}z_{ll} = 2\theta_k^2\delta_{kl} - 2\theta_k\theta_l(\theta_k + \theta_l)\sum_i \beta_{ik}^2\beta_{il}^2 + 2\theta_k\theta_l\sum_{i,g} \beta_{ik}^2\beta_{gl}^2\rho_{ig}^2 . \tag{4.7}$$

The characteristic roots of R, $d_1 \geqq \cdots \geqq d_p$, are the characteristic roots of $B'RB$. Let the characteristic vectors of R be $C = (c_1 \cdots c_p)$ and the characteristic vectors of $B'RB$ be E; then $BE = C$. We wish to find the asymptotic distribution of $d_1, \cdots, d_p$ and C. Because $\rho_{ii} = 1$, we have

$$\sum_i \theta_i = \operatorname{tr} \Theta = \operatorname{tr} \bar{R} = p, \tag{4.8}$$

$$\operatorname{diag} B\Theta B' = \operatorname{diag} I. \tag{4.9}$$

There are similar conditions on R, D and C. The theory in this case is much more complicated than for covariance matrices; we cannot give general results in as simple a form.

Consider first the case of all population roots equal. This condition implies the common root is 1 and $\bar{R} = \Theta = I$. Then we can take $B = I$, and $Z = X - X_0$. Thus $z_{ii} = 0$ and for $i \neq j$ $z_{ij} = x_{ij}$ have variance 1 and are uncorrelated. Then

$$R = CDC' = C(I + n^{-\frac{1}{2}}H)C' = I + n^{-\frac{1}{2}}CHC', \tag{4.10}$$

where $\operatorname{tr} H = 0$ and $\operatorname{diag} CHC' = \operatorname{diag} 0$. Then CHC' is asymptotically equivalent to Z. Since the limiting distribution of PZP' for orthogonal P is not the same as that of Z, C does not have the invariant distribution as a limiting distribution.

The second case of interest is that of two different roots. Some aspects of this case are considered in Section 5, and this case when the larger root is distinct and the smaller root is of multiplicity $p - 1$ is treated in Appendix A.

In the general case suppose the roots are $\lambda_1 > \lambda_2 > \cdots > \lambda_r$ of muliplicities $q_1, q_2, \cdots, q_r$, respectively. Then Θ, Z, D, H and E are partitioned as Δ, U, D, H, and E were previously. The limiting distribution of H_k, E_{kk}, and $F_{kl} = n^{\frac{1}{2}}E_{kl}$, $k \neq l$, is obtained from the limiting distribution of Z by use of

$$Z_{kk} = E_{kk}H_kE_{kk}', \qquad I = E_{kk}E_{kk}', \tag{4.11}$$

$$\begin{aligned} Z_{kl} &= \lambda_k E_{kk}F_{lk}' + \lambda_l F_{kl}E_{ll}', \qquad && k \neq l, \\ 0 &= E_{kk}F_{lk}' + F_{kl}E_{ll}', && k \neq l. \end{aligned} \tag{4.12}$$

5. Statistical inference about characteristic roots of a correlation matrix. We can use (3.7) to test the hypothesis that q_k of the θ's are equal (where d_j in (3.7) are roots of R). The limiting distribution of -2 times the logarithm of the criterion is the limiting distribution of $[\operatorname{tr} Z_{kk}^2 - (\operatorname{tr} Z_{kk})^2/q_k]/(2\lambda_k^2)$. The distribution is that of a quadratic form in normally distributed variables with 0 means. In general this distribution will not be χ^2. However, for $p = 3$, $r = 2$, and $q_2 = 2$ the criterion is asymptotically distributed as $(1 - \lambda_2^2/3)\chi_2^2$ (See Appendix A). For $p = 4$, $r = 2$ and $q_2 = 3$, the criterion is asymptotically distributed as $\chi_2^2 + (1 - \frac{1}{2}\lambda_2^2)\chi_3^2$. Lawley [13] has solved the general case when $q_2 = p - 1$.

We can consider an asymptotic confidence interval for $\lambda_r = \theta_{p-q_r+1} = \cdots =$

θ_p. We know that $\bar{d}$ converges stochastically to λ_r, but in general the variance of $n^{\frac{1}{2}}(\bar{d} - \lambda_r) = \operatorname{tr} Z_{rr}/q_r = \sum_{p-q_r+1}^{p} z_{ii}q_r$ will depend on the other roots and B as well. In one particular case, however we can get a simple evaluation of

$$\operatorname{Var}[(\operatorname{tr} Z_{rr})/q_r];$$

this is when $\theta_1 = \cdots = \theta_{q_1} = \lambda_1$, $\theta_{q_1+1} = \cdots = \theta_p = \lambda_2$. Then we have

$$\mathcal{E} z_{kk} z_{ll} = 2\lambda_2^2\Big(\delta_{kl} - 2\lambda_2 \sum_i \beta_{ik}^2 \beta_{il}^2 + \sum_{i,g} \beta_{ik}^2 \beta_{gl}^2 \rho_{ig}^2\Big), \qquad k, l \geqq q_1. \tag{5.1}$$

In this case $q_1\lambda_1 + q_2\lambda_2 = p$, and

$$1 = \sum_{j=1}^{p} \beta_{ij}^2 \delta_j = \lambda_1 \sum_{j=1}^{q_1} \beta_{ij}^2 + \lambda_2 \sum_{j=q_1+1}^{p} \beta_{ij}^2. \tag{5.2}$$

This implies $\sum_{j=q_1+1}^{p} \beta_{ij}^2 = q_2/p$. Thus

$$\begin{aligned} \mathcal{E}\Big(\sum_{q+1}^{p} z_{kk}\Big)^2 &= 2\lambda_2^2 \sum_{k,l=q_1+1}^{p} \Big(\delta_{kl} - 2\lambda_2 \sum_i \beta_{ik}^2\, \beta_{il}^2 + \sum_{i,g} \beta_{ik}^2\, \beta_{gl}^2\, \rho_{ig}^2\Big) \\ &= 2\lambda_2^2\Big(q_2 - 2\lambda_2 \sum_i (q_2/p)^2 + \sum_{i,g} (q_2/p)^2 \rho_{ig}^2\Big) \\ &= 2\lambda_2^2(q_2 - 2\lambda_2(q_2/p)^2 + (q_2/p)^2\{q_1\,\lambda_1^2 + q_2\,\lambda_2^2\}) = (2\lambda_2^2\, q_1\, q_2\, \lambda_1^2)/p \end{aligned} \tag{5.3}$$

and $\lambda_1 = [p - q_2\lambda_2]/q_1$. Then

$$[n^{\frac{1}{2}}(\bar{d} - \lambda_2)/\{\lambda_2(p - q_2\lambda_2)\}](pq_1q_2/2)^{\frac{1}{2}} \tag{5.4}$$

is asymptotically normally distributed with mean 0 and variance 1. From this result we can set up confidence intervals for λ_2.

6. Uses of principal component analysis. One of the uses of principal component analysis is in exploratory studies. The investigator has available many measurements—in fact, many more than he wants to subject to close scrutiny—and he wants to "reduce" the data. He is interested in studying what varies from individual to individual; hence variance is a measure of importance. He thus asks for a small number of linear combinations of the original variables that account for most of the variability. The sum of the variances of the full set of principal components is the sum of the variances of the original variables. The investigator will be satisfied with the principal components accounting for most of the total variance. This idea is made precise by the specification that he will be satisfied with q components if the sum of the variances of the $p - q$ other components is less than a certain amount, which is equivalent to the condition that the average of the $p - q$ smallest roots of Σ is less than a specified amount. He sees whether the (one-sided) confidence interval includes only values of the average less than the specified amount; if so, he is satisfied with the first q principal components.

For this use of principal component analysis in exploratory studies the variables have to be in the same units or at least in comparable units because the linear combinations are normalized by having the sum of coefficients squared to be unity and because the sum of variances is used. (When different units of measurement are involved the variables might be scaled according to inherent importance.) In these kinds of studies the principal components do not need to have any intrinsic meaning (that is, one cannot necessarily give them meaningful names).

In another situation, referred to in the introduction, all measurements are made in the same units and by the same kind of measuring device. Mathematically specified, the errors of measurement in all variables have the same variance, are mutually independent, and are independent of the true measurements (or systematic parts). Then the covariance matrix of the observed variables is $\Sigma = \Psi + \sigma^2 I$, where Ψ is the covariance matrix of the true measurements and σ^2 is the variance of error. The roots of Σ are σ^2 plus the roots of Ψ (positive semi-definite); if Ψ is of rank q, then the last $p - q$ roots are σ^2. In this case it is crucial that the units of the different variables are the same. Of course, the hypothesis holds if Σ is the sum of any positive semi-definite matrix of rank q and a nonnegative multiple of the identity matrix. The multiple might be larger than just the "error" variance. The hypothesis specifies nothing about the size of the common root (except that the other q are larger). The interpretation is that $\sigma^2 I$ represents randomness in the sense of independence and constant variance.

If one has in mind a specific value of the error variance, one can test that the last $p - q$ roots are equal and equal to the specified value. The criterion (3.18) can be used. Alternatively one can test equality of roots and then use a one-sided test that the common root has the specified value (against alternatives that it is larger). The two criteria are asymptotically independent.

Lawley [11], [12] has shown that if $(-\frac{1}{2}N)$ times the logarithm of the criterion (3.7) is multiplied by

$$n - q - \tfrac{1}{6}[2(p - q) + 1 + [2/(p - q)]] + \lambda^2 \sum_{i=1}^{q} [1/(\delta_i - \lambda)^2] \tag{6.1}$$

the resulting quantity has the moments of the proper χ^2-distribution to order $1/n^2$ when the last $p - q$ roots are equal to λ and (3.7) involves the smallest $p - q$ sample roots. The multiplying factor for (3.18) is

$$\begin{aligned} & n - q - \tfrac{1}{6}[2(p - q) + 1 - [2/(p - q + 1)]] \\ & \qquad - [1/(p - q + 1)] \left[\sum_{i=1}^{q} \delta_i/(\delta_i - \lambda) \right]^2 + \lambda^2 \sum_{i=1}^{q} [1/(\delta_i - \lambda)^2]. \end{aligned} \tag{6.2}$$

When measurements are made in different units and there is no substantive basis on which to scale them, resort has been made to the correlation matrix which is dimensionless. Let us see what the methods imply in such cases. Suppose

$$\Sigma = \Psi + K, \tag{6.3}$$

where K is diagonal. Here k_{ii} is interpreted as the variance of the error in the

ith variable. The correlations are

$$(6.4) \qquad \rho_{ii} = 1,\ \rho_{ij} = \psi_{ij}/[(\psi_{ii} + k_{ii})^{\frac{1}{2}}(\psi_{jj} + k_{jj})^{\frac{1}{2}}], \qquad i \neq j.$$

The correlation matrix is of the form $\Psi^* + \sigma^2 I$ if

$$(6.5) \qquad \sigma^2 = k_{ii}/(\psi_{ii} + k_{ii}) = 1/[(\psi_{ii}/k_{ii}) + 1];$$

that is, if the ratio ψ_{ii}/k_{ii} of "systematic" variance to error variance is the same for all variables. In the case of treating a covariance matrix the null hypothesis is true if the error variances are equal, but in the case of treating a correlation matrix the null hypothesis is true if the error variances are proportional to the systematic variances. It should be noted that the hypotheses are equivalent if all of the systematic variances are equal.

Another procedure with correlation matrices is to determine q linear combinations of the normalized variables with maximum variance; let $\Sigma = \Psi_1 + \Psi_2$, where Ψ_1 is made up from the q linear combinations of the variables not normalized and $\Psi_2 = (\psi_{ij}^{(2)})$ is the residual covariance matrix. The diagonal entries of the residual part of the correlation matrix are $\psi_{ii}^{(2)}/\sigma_{ii}$. Testing that the residual "variance" in the correlation matrix is sufficiently small is testing that $\sum_i \psi_{ii}^{(2)}/\sigma_{ii}$ is sufficiently small; that is, that the average fraction of residual variance in the variance of each variable is sufficiently small.

The procedures with correlation matrices seem a little incongruous, for first the individual variances are standardized to unity (in a sense eliminating variances) and then one goes back to maximizing variances of linear combinations. Besides the difficulty of interpretation there are the formal mathematical difficulties. The criterion for testing equality of the $p - q$ smallest roots of $\bar{R}$ is asymptotically distributed as a quadratic form, whose distribution is that of $\sum_i a_i \chi_{b_i}^2$. In most cases this is not χ^2, though possibly it can be approximated by $c\chi_d^2$ for suitable c and d. However, the coefficients a_i depend on unknown parameters. A conservative treatment is to use significance points from the χ^2 distribution with degrees of freedom appropriate to the covariance case.

We have considered procedures which split Σ (or alternatively $\bar{R}$) into a sum of two matrices. The lack of importance of the second matrix has been evaluated on two different bases; one is that it is proportional to the identity ($p - q$ smallest roots equal) and the other that the trace is small. Quite a different basis for considering the second matrix unimportant (that is, that the first matrix is the part of interest) is that the second matrix is diagonal with no specification of the size of the diagonal elements. This hypothesis does not involve the units of measurement and hence can be treated in terms of correlations as well as covariances. Factor analysis is primarily based on this approach; the first matrix (which the investigator hopes is of low rank) involves the common factors and the second matrix, which is diagonal, involves the specific factors and errors, which are uncorrelated from one variable to another. Statistical methods in factor analysis have been considered in another paper [3] and that discussion will not be repeated here. It might be emphasized, however, that factor analysis can be done without using variances of either the original variables or of components. Sometimes one of the procedures described in this paper is used when it is the hypothesis of independence that it is desired to test; it would seem preferable in such a case to use a test for independence (such as discussed in [3]) or at least look for an approximation to such a test, rather than test a hypothesis that is not called for.

Roy [16] has proposed statistical techniques which are useful in a different approach to the problem of reducing the number of variables in an exploratory study. He has indicated how to use S to set up simultaneous confidence bounds on σ_{ii} $(i = 1, \cdots, p)$ and as well as on all linear combination of the elements of Σ of the form $\gamma'\Sigma\gamma/(\gamma'\gamma)$. On this basis the investigator might eliminate variables that have sufficiently small variances and concentrate his attention on the other variables. The details of the screening of variables by use of such joint confidence bounds have not been worked out; presumably a variable would not be discarded unless its contribution to the variance of every linear combination were small. Since the subject of the present paper is principal component analysis, a detailed study of this procedure will not be made.

Selection of the principal components with largest variances may have the disadvantage that all of the original variables (or almost all of them) may enter into some of the selected principal components with nonzero weights. This feature suggests that selected principal components might be modified moderately to give some variables zero weights if the modification does not decrease the variance of the linear combinations much (or cause them to be correlated much).

7. Justification of the asymptotic distribution of roots and vectors in the general case. The proof of the theorem of Section 2 is based on the following special case of a theorem due to Rubin [17] and proved in Appendix D:

THEOREM ON LIMITING DISTRIBUTIONS. *Let $F_n(u)$ be the cumulative distribution function of a random matrix U_n. Let V_n be a (matrix-valued) function of U_n, $V_n = f_n(U_n)$, and let $G_n(v)$ be the (induced) distribution of V_n. Suppose $\lim_{n\to\infty} F_n(u) = F(u)$ [in every continuity point of $F(u)$] and suppose for every continuity point u of $f(u)$, $\lim_{n\to\infty} f_n(u_n) = f(u)$, when $\lim_{n\to\infty} u_n = u$. Let $G(v)$ be the distribution of the random matrix $V = f(U)$, where U has the distribution $F(u)$. If the probability of the set of discontinuities of $f(u)$ according to $F(u)$ is* 0, *then*

$$\lim_{n\to\infty} G_n(v) = G(v) \tag{7.1}$$

[in every continuity point of $G(v)$].

In our case H, E_{kk}, and F_{kl} are functions of U, and depend on n. To apply the theorem we need to show that if a sequence of symmetric (nonrandom) matrices U_n converges to U then the corresponding sequences of $[H_k(n), E_{kk}(n), F_{kl}(n)]$ converges to the solution of

$$U_{kk} = E_{kk}H_kE'_{kk}, \qquad I = E_{kk}E'_{kk}, \tag{7.2}$$

$$\begin{aligned} U_{kl} &= \lambda_k E_{kk} F'_{lk} + \lambda_l F_{kl} E'_{ll}, & k \neq l. \\ 0 &= E_{kk} F'_{lk} + F_{kl} E'_{ll}, & k \neq l, \end{aligned} \tag{7.3}$$

for almost every symmetric U. The distribution of (random) H_k, E_{kk} and F_{kl} can be defined either in the $p(p+1)/2$ dimensional space of h_i, e_{ij}, $i < j$, and f_{ij}, $i < j$, or in the $p^2 + p$ dimensional space of h_i, e_{ij}, and f_{ij}. (In the latter case the probability is concentrated on a surface.)

First let us show that as $U_n \to U$, $h_i(n) \to h_i$. Consider $d_1(n) > \cdots > d_{q_1}(n)$, the q_1 largest characteristic roots of $\Delta + n^{-\frac{1}{2}} U_n$. Then

$$h_1(n) = n^{\frac{1}{2}}(d_1(n) - \lambda_1) > \cdots > h_{q_1}(n) = n^{\frac{1}{2}}(d_{q_1}(n) - \lambda_1)$$

are the q_1 largest roots of

$$\begin{aligned} 0 &= |\Delta + n^{-\frac{1}{2}} U_n - (\lambda_1 + n^{-\frac{1}{2}} h) I| \\ &= \begin{vmatrix} n^{-\frac{1}{2}}[U_{11}(n) - hI] & n^{-\frac{1}{2}} U^*(n) \\ n^{-\frac{1}{2}} U^*(n) & \Delta^* - \lambda_1 I + n^{-\frac{1}{2}}[U^{**}(n) - hI] \end{vmatrix}, \end{aligned} \tag{7.4}$$

where

$$\Delta = \begin{pmatrix} \lambda_1 I & 0 \\ 0 & \Delta^* \end{pmatrix}, \qquad U_n = \begin{bmatrix} U_{11}(n) & U^*(n)' \\ U^*(n) & U^{**}(n) \end{bmatrix}. \tag{7.5}$$

When we factor $n^{-\frac{1}{2}}$ out of the first q_1 columns of the matrix we obtain the determinant

$$\begin{vmatrix} U_{11}(n) - hI & n^{-\frac{1}{2}} U^*(n)' \\ U^*(n) & \Delta^* - \lambda_1 I + n^{-\frac{1}{2}}[U^{**}(n) - hI] \end{vmatrix}, \tag{7.6}$$

which is a polynomial of degree p in ϕ. As $n \to \infty$ and $U_n \to U$, the coefficients of the polynomial approach the coefficients of the polynomial (of degree q_1)

$$|U_{11} - hI| \cdot |\Delta^* - \lambda_1 I|, \tag{7.7}$$

and q_1 roots of (7.6) set equal to 0 approach the q_1 roots of $|U_{11} - hI| = 0$ (because the roots are continuous functions of the coefficients). These q_1 roots of (7.7) must be the largest ones because for each $i > q_1$, $\lambda_1 + n^{-\frac{1}{2}} h_i(n) \to \lambda_j$ for some $j > 1$. This proves the result for the first q_1 roots; the other roots are treated similarly.

Now consider the vector associated with the first root $h_1(n)$. Let

$$U_{11}(n) = \begin{pmatrix} u_{11}(n) & u_1^*(n)' \\ u_1^*(n) & \bar{U}(n) \end{pmatrix}, \qquad U^*(n)' = \begin{pmatrix} U_1^*(n)' \\ \bar{U}^*(n)' \end{pmatrix}. \tag{7.8}$$

Then the components of the first vector are proportional to the cofactors of the first column in

$$\begin{pmatrix} u_{11}(n) - h_1(n) & u_1^*(n)' & n^{-\frac{1}{2}} U_1^*(n)' \\ u_1^*(n) & \bar{U}(n) - h_1(n) I & n^{-\frac{1}{2}} \bar{U}^*(n)' \\ U_1^*(n) & \bar{U}^*(n) & \Delta^* - \lambda_1 I + n^{-\frac{1}{2}}[U^{**}(n) - h_1(n) I] \end{pmatrix}. \tag{7.9}$$

The first cofactor is

$$\text{(7.10)}\qquad \begin{vmatrix} \bar{U}(n) - h_1(n)I & n^{-\frac{1}{2}}\bar{U}^*(n)' \\ \bar{U}^*(n) & \Delta^* - \lambda_1 I + n^{-\frac{1}{2}}[U^{**}(n) - h_1(n)I] \end{vmatrix} = a_{1n} + n^{-\frac{1}{2}} b_{1n},$$

where $a_{1n} \to a_1$ and $b_{1n} \to b_1$; this can be seen by expanding (7.10) by minors of the first $q_1 - 1$ columns. Expansion of other cofactors by minors of the first $q_1 - 1$ columns shows that for $i \leqq q_1$, the element is $a_{in} + n^{-\frac{1}{2}} b_{in}$ and for $i > q_1$ it is $n^{-\frac{1}{2}} b_{in}$. Then the normalized elements of the first characteristic vector are

$$\text{(7.11)}\qquad \begin{aligned} &(a_{in} + n^{-\frac{1}{2}} b_{in}) \Big/ \left[\sum_1^{q_1} a_{in}^2 + 2n^{-\frac{1}{2}} \sum_1^{q_1} a_{in}\, b_{in} + (1/n) \sum b_{in}^2\right]^{\frac{1}{2}}, \qquad i = 1, \cdots, q_1, \\ &(n^{-\frac{1}{2}} b_{in}) \Big/ \left[\sum_1^{q_1} a_{in}^2 + 2n^{-\frac{1}{2}} \sum_1^{q_1} a_{in}\, b_{in} + (1/n) \sum b_{in}^2\right]^{\frac{1}{2}}, \qquad i = q_1 + 1, \cdots, p. \end{aligned}$$

This shows that the first q_1 elements of the first characteristic vector converge to finite limits and that $n^{\frac{1}{2}}$ times each other element converges. A similar argument shows that the elements of $E_{kk}(n)$ converge and the elements of $F_{kl}(n) = n^{\frac{1}{2}} E_{kl}(n)$ converge.

Now let us return to (2.21) to (2.24), where these are now considered to refer to (nonrandom) $U_{kl}(n)$, $H_k(n)$, $E_{kk}(n)$ and $F_{kl}(n)$. Then as $n \to \infty$ the terms $n^{-\frac{1}{2}} M_{kl}$ and $n^{-\frac{1}{2}} W_{kl}$ converge to 0 because their elements are polynomials in $n^{-\frac{1}{2}}$ and elements of $H_k(n)$, $E_{kk}(n)$ and $F_{kl}(n)$ which converge to finite limits. Then the limits satisfy (7.2) and (7.3) as was to be shown.

The discontinuities only occur because of indeterminacies in E_{kk} due to multiple roots of U_{kk} but such matrices are of (Lesbesgue) measure 0.

APPENDIX A. *The case of the correlation matrix when the $p - 1$ smallest roots are equal.* Let $\delta_2 = \cdots = \delta_p = \lambda$ and $B = (\beta_1 B_2)$. Since

$$p = \operatorname{tr} \bar{R} = \delta_1 + (p - 1)\lambda, \qquad \delta_1 = 1 + (p - 1)(1 - \lambda).$$

Then

$$\text{(A1)}\qquad \bar{R} = B\Theta B' = \delta_1 \beta_1 \beta_1' + \lambda B_2 B_2' = \delta_1 \beta_1 \beta_1' + \lambda(I - \beta_1 \beta_1').$$

In particular

$$\text{(A2)}\qquad \rho_{ii} = 1 = \delta_1 \beta_{i1}^2 + \lambda(1 - \beta_{i1}^2),$$

and hence $\beta_{i1} = \pm p^{-\frac{1}{2}}$. Let us assume that the signs of x_i are taken so $\beta_{i1} = p^{-\frac{1}{2}}$. Let $\rho = 1 - \lambda$; then $\lambda = 1 - \rho$ and $\delta_1 = 1 + (p - 1)\rho$. Then

$$\bar{R} = \lambda I + (\delta_1 - \lambda)\beta_1\beta_1' = (1-\rho)I + p\rho\begin{pmatrix} 1/p & \cdots & 1/p \\ \vdots & & \vdots \\ 1/p & \cdots & 1/p \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}. \tag{A3}$$

This is the case of all pairs of variables having the same correlation.

If $Y = X - \frac{1}{2}(\bar{R}X_0 + X_0\bar{R})$, then

$$Z = B'YB = \begin{pmatrix} \beta_1' \\ B_2' \end{pmatrix} Y(\beta_1 B_2) = \begin{pmatrix} \beta_1' Y\beta_1 & \beta_1' YB_2 \\ B_2' Y\beta_1 & B_2' YB_2 \end{pmatrix}. \tag{A4}$$

Then

$$\begin{aligned} \operatorname{tr} Z_{22} &= \operatorname{tr} B_2'YB_2 = \operatorname{tr} YB_2B_2' = \operatorname{tr} Y(I - \beta_1\beta_1') \\ &= \operatorname{tr} Y - \beta_1'Y\beta_1 = -(1/p)\sum_{i,j} y_{ij} = -(2/p)\sum_{i<j} y_{ij}\,. \end{aligned} \tag{A5}$$

$$\begin{aligned} \operatorname{tr} (Z_{22})^2 &= \operatorname{tr} B_2'YB_2B_2'YB_2 = \operatorname{tr} YB_2B_2'YB_2B_2' \\ &= \operatorname{tr} Y(I - \beta_1\beta_1')Y(I - \beta_1\beta_1') \\ &= \operatorname{tr} Y^2 - \operatorname{tr} Y\beta_1\beta_1'Y - \operatorname{tr} Y^2\beta_1\beta_1' + \operatorname{tr} Y\beta_1\beta_1'Y\beta_1\beta_1' \\ &= \operatorname{tr} YY' - 2\,\beta_1'Y^2\beta_1 + (\beta_1'Y\beta_1)^2 \\ &= \sum_{i,j} y_{ij}^2 - (2/p)\sum_{i,j,k} y_{ik}y_{jk} + (1/p^2)(\sum_{i,j} y_{ij})^2. \end{aligned} \tag{A6}$$

Then we find $Q_p(Z_{22}) = \operatorname{tr} Z_{22}^2 - (\operatorname{tr} Z_{22})^2/(p-1)$ in terms of Y. For example,

$$Q_3(Z_{22}) = 8/9[y_{12}^2 + y_{13}^2 + y_{23}^2 - y_{12}y_{13} - y_{12}y_{23} - y_{13}y_{23}]. \tag{A7}$$

To find the limiting distribution of Q we use the fact that Y is asymptotically normally distributed with variances and covariances

$$\mathcal{E}y_{ij}^2 = \lambda^2(1+\rho)^2, \qquad i \neq j, \tag{A8}$$

$$\mathcal{E}y_{ij}y_{ik} = \tfrac{1}{2}\lambda^2\rho(2+3\rho), \qquad i \neq j \neq k \neq i, \tag{A9}$$

$$\mathcal{E}y_{ij}y_{kh} = 2\lambda^2\rho^2, \qquad \text{no subscripts equal.} \tag{A10}$$

As an example of the asymptotic distribution of the criterion consider the case of $p = 3$. Then the criterion is

$$Q^* = Q_3(Z_{22})/(2\lambda^2) = 4/9(u_{12}^2 + u_{13}^2 + u_{23}^2 - u_{12}u_{13} - u_{12}u_{23} - u_{13}u_{23}), \tag{A11}$$

where $y_{ij} = \lambda u_{ij}$ and $\mathcal{E}u_{ij} = (1+\rho)^2$, $i \neq j$, and

$$\mathcal{E}u_{ij}u_{ik} = \tfrac{1}{2}\rho(2+3\rho), \qquad i \neq j \neq k \neq i. \tag{A12}$$

Then the distribution of Q^* is $b_1\chi_1^2 + b_2\chi_1^2$, where b_1 and b_2 are the nonzero roots of

$$\text{(A13)}\quad \left|\begin{pmatrix} 4/9 & -2/9 & -2/9 \\ -2/9 & 4/9 & -2/9 \\ -2/9 & -2/9 & 4/9 \end{pmatrix} \cdot \begin{pmatrix} (1+a)^2 & (a/2)(2+3a) & (a/2)(2+3a) \\ (a/2)(2+3a) & (1+a)^2 & (a/2)(2+3a) \\ (a/2)(2+3a) & (a/2)(2+3a) & (1+a)^2 \end{pmatrix} - uI\right| = 0.$$

These are both $1 - \lambda^2/3$. Thus Q^* is asymptotically distributed as $(1 - \lambda^2/3)\chi_2^2$ for $p = 3$.

Appendix B. *An asymptotic test for a given principal component.* Consider testing the null hypothesis that γ_1, the vector of coefficients of the first principal component, is a specified vector $\gamma_1^0(\gamma_1^{0\prime}\gamma_1^0 = 1)$ under the assumption that this characteristic vector of Σ corresponds to a characteristic root of multiplicity 1. We use the fact that $n^{\frac{1}{2}}(c_1 - \gamma_1) = y$, say, has a limiting normal distribution with mean 0 and covariance matrix

$$\text{(B1)}\qquad \sum_{j=2}^{p} [\delta_1\delta_j/(\delta_1 - \delta_j)^2]\gamma_j\gamma_j' = \Gamma^*\Lambda^2\Gamma^{*\prime},$$

where

$$\text{(B2)}\qquad \Gamma^* = (\gamma_2 \cdots \gamma_p),$$

$$\Lambda = \begin{pmatrix} (\delta_1\delta_2)^{\frac{1}{2}}/(\delta_1 - \delta_2) & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & (\delta_1\delta_p)^{\frac{1}{2}}/(\delta_1 - \delta_p) \end{pmatrix}.$$

Then $z = \Lambda^{-1}\Gamma^{*\prime}y$ has a limiting normal distribution with mean 0 and covariance matrix

$$\text{(B3)}\qquad \Lambda^{-1}\Gamma^{*\prime}(\Gamma^*\Lambda^2\Gamma^{*\prime})\Gamma^*\Lambda^{-1} = I.$$

Thus

$$\text{(B4)}\qquad z'z = y'\Gamma^*\Lambda^{-2}\Gamma^{*\prime}y$$

has a limiting χ^2-distribution with $p - 1$ degrees of freedom. The matrix of the quadratic form in y is

$$\text{(B5)}\qquad \begin{aligned} \Gamma^*\Lambda^{-2}\Gamma^{*\prime} &= \Gamma^*\begin{pmatrix} (\delta_1/\delta_2) - 2 + \delta_2/\delta_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & (\delta_1/\delta_p) - 2 + (\delta_p/\delta_1) \end{pmatrix}\Gamma^{*\prime} \\ &= \delta_1(\Sigma^{-1} - (1/\delta_1)\gamma_1\gamma_1') - 2(I - \gamma_1\gamma_1') + (1/\delta_1)(\Sigma - \delta_1\gamma_1\gamma_1') \\ &= \delta_1\Sigma^{-1} - 2I + (1/\delta_1)\Sigma, \end{aligned}$$

since $\Gamma\Delta^{-1}\Gamma' = \Sigma^{-1}$, $\Gamma\Gamma' = I$ and $\Gamma\Delta\Gamma' = \Sigma$. Thus

(B6) $$n(c_1 - \gamma_1)'(\delta_1\Sigma^{-1} - 2I + (1/\delta_1)\Sigma)(c_1 - \gamma_1)$$

has a limiting χ^2-distribution with $p - 1$ degrees of freedom.

For γ_1, the characteristic vector of Σ with root δ_1 and of Σ^{-1} with root δ_1^{-1},

(B7) $$(\delta_1\Sigma^{-1} - 2I + (1/\delta_1)\Sigma)\gamma_1 = 0.$$

Then (B6) is

(B8) $$nc_1'(\delta_1\Sigma^{-1} - 2I + (1/\delta_1)\Sigma)c_1 = n(\delta_1 c_1'\Sigma^{-1}c_1 + (1/\delta_1)c_1'\Sigma c_1 - 2).$$

If we want to test the null hypothesis that the population covariance matrix is a given Σ of which a given vector γ_1^0 is the first characteristic vector against the alternative that the covariance matrix has a different vector as first characteristic vector, we can use (B8) and refer this criterion to the χ^2-distribution with $p - 1$ degrees of freedom.

If the entire matrix Σ is not specified, we can estimate Σ, Σ^{-1} and δ_1 consistently by S, S^{-1} and d_1, respectively. Then the asymptotic equivalent of (B6),

(B9) $$\begin{aligned} n(c_1 - \gamma_1)'[d_1S^{-1} - 2I + (1/d_1)S](c_1 - \gamma_1) \\ = n[d_1\gamma_1'S^{-1}\gamma_1 + (1/d_1)\gamma_1'S\gamma_1 - 2] \end{aligned}$$

has a limiting χ^2-distribution with $p - 1$ degrees of freedom when γ_1 is the first characteristic vector of Σ. Thus for a given vector γ_1, (B9) can be used to test the hypothesis that it is the first characteristic vector of Σ.

An asymptotic confidence region for the first characteristic vector of Σ consists of all vectors $\gamma_1(\gamma_1'\gamma_1 = 1)$ such that (B9) is not greater than the significance point of the χ^2-distribution at significance level the complement of the desired confidence coefficient.

For convenience the above discussion has been carried on in terms of the first characteristic vector. Similar results are true for any other characteristic vector corresponding to a root of multiplicity 1.

Mallows [14] has considered confidence regions for characteristic vectors without specifying that the vector is associated with any particular ordered root.

Appendix C. *Some inequalities.* Since d_1 is the largest characteristic root of S,

(C1) $$d_1 = \max_{c'c=1} c'Sc \geqq \gamma_1'S\gamma_1 .$$

As noted in Section 1, $\gamma_1'x$ has variance δ_1 and from Section 2 we see that $\gamma_1'A\gamma_1$ is distributed as

(C2) $$\gamma_1'\sum_{\alpha=1}^{n} x_\alpha x_\alpha'\gamma_1 = \sum_{\alpha=1}^{n} (\gamma_1'x_\alpha)^2,$$

where $\gamma_1'x_\alpha$ is distributed accoring to $N(0, \delta_1)$. Hence, $\gamma_1'A\gamma_1 = n\gamma_1'S\gamma_1$ is distributed as $\delta_1\chi_n^2$ (χ^2 with n degrees of freedom). Hence

(C3) $$\Pr\{d_1 \leqq d\delta_1\} \leqq \Pr\{\gamma_1'S\gamma_1 \leqq d\delta_1\} = \Pr\{\chi_n^2 \leqq n\, d\}.$$

The expectations of the left and right hand sides of (C1) satisfy

(C4) $$\mathcal{E}\, d_1 \geqq \delta_1 .$$

Since the inequality (C1) depends only on d_1 being the largest characteristic root of a symmetric random matrix S, the inequality (C4) holds if δ_1 is the largest characteristic root of $\mathcal{E}S$. The inequality (C4) has been proved by van der Vaart [18].

Suppose one takes a second sample; let its covariance matrix be S^* with largest characteristic root d_1^*. Consider the variance in the second sample of the first principal component $c_1'x$ as determined by the first sample. This variance is $c_1'S^*c_1$. Then

$$c_1'S^*c_1 \leqq \max_{c'c=1} c'S^*c = d_1^*. \tag{C5}$$

Given c_1, the conditional expectation of this variance satisfies

$$\mathcal{E}(c_1'S^*c_1 \mid c_1) = c_1'\Sigma c_1 \leqq \delta_1 . \tag{C6}$$

Thus

$$\mathcal{E}c_1'S^*c_1 \leqq \delta_1 . \tag{C7}$$

We see that the variance in the second sample of the first principal component determined by the first sample is always no greater than the variance in the second sample of the first principal component determined by that sample and on the average underestimates the population variance of the first principal component determined by the population. One might say the variance of the first principal component "shrinks." This corresponds to the well known shrinkage of the multiple correlation coefficient.

The corresponding inequalities for the smallest roots are

$$d_p \leqq \gamma_p'S\gamma_p , \tag{C8}$$

$$\Pr\{d_p \leqq d\delta_p\} \geqq \Pr\{\chi_n^2 \leqq n\, d\}, \tag{C9}$$

$$\mathcal{E}d_p \leqq \delta_p , \tag{C10}$$

$$c_p'S^*s_p \geqq d_p^*, \tag{C11}$$

$$\mathcal{E}c_p'S^*c_p \geqq \delta_p . \tag{C12}$$

APPENDIX D. *Proof of the Theorem on Limiting Distributions.* The theorem of Section 7, which was used to justify the derivation of asymptotic distributions, has not been proved in publication even though it has been used and quoted for more than 10 years (for example, in [1]). A relatively simple proof is given here. The theorem was stated in more general form by Rubin in an unpublished paper [17] (and was used in a paper published by Rubin and the present author in 1950). The proof given by Rubin is correspondingly more sophisticated.[3]

[3] Another special case of Rubin's theorem is contained in Herman Chernoff's lectures, On Order Relations and Convergence in Distribution (dittoed), reported by John W. Pratt. These notes and discussions with Chernoff and Pratt have been helpful. The criterion of convergence used in these notes and in this paper is also given in P. Billingsley (1956). Invariance principle for dependent random variables. *Trans. Amer. Math. Soc.* **83** 252–268. It is to be hoped that Rubin will publish his more general theorem.

The convergence of $F_n(u)$ to $F(u)$ is equivalent to

$$\text{(D1)} \qquad \limsup_{n\to\infty} \Pr\{U_n \varepsilon S\} \leqq \Pr\{U \varepsilon S\}$$

for every closed (Borel) set S in the Euclidean space of matrices u. We wish to show the convergence of $G_n(v)$ to $G(v)$ which is equivalent to

$$\text{(D2)} \qquad \limsup_{n\to\infty} \Pr\{V_n \varepsilon T\} \leqq \Pr\{V \varepsilon T\}$$

for every closed (Borel) set T in the Euclidean space of matrices v, where $V_n = f_n(U_n)$ and $V = f(U)$. Let

$$\text{(D3)} \qquad S_n = \{u \mid f_n(u) \varepsilon T\},$$

$$\text{(D4)} \qquad S = \{u \mid f(u) \varepsilon T\},$$

$$\text{(D5)} \qquad R_n = S_n \cup S_{n+1} \cup \cdots,$$

$$\text{(D6)} \qquad R^* = \bar{R}_1 \cap \bar{R}_2 \cap \cdots = \lim_{n\to\infty} \bar{R}_n,$$

where $\bar{R}_n$ is the closure of R_n and $\bar{R}_1 \supset \bar{R}_2 \supset \cdots$ since $R_1 \supset R_2 \supset \cdots$. Then

$$\text{(D7)} \qquad \begin{aligned} \limsup_{n\to\infty} \Pr\{V_n \varepsilon T\} &= \limsup_{n\to\infty} \Pr\{U_n \varepsilon S_n\} \\ &\leqq \limsup_{n\to\infty} \Pr\{U_n \varepsilon R_m\}, \qquad m = 1, 2, \cdots, \end{aligned}$$

since $S_n \subset R_m$ for $n \geqq m$. Next (D7) implies

$$\text{(D8)} \qquad \begin{aligned} \limsup_{n\to\infty} \Pr\{V_n \varepsilon T\} &\leqq \limsup_{n\to\infty} \Pr\{U_n \varepsilon \bar{R}_m\} \\ &\leqq \Pr\{U \varepsilon \bar{R}_m\}, \qquad m = 1, 2, \cdots, \end{aligned}$$

because of the convergence of the distributions of U_n. Since the left hand side of (D8) does not depend on m

$$\text{(D9)} \qquad \limsup_{n\to\infty} \Pr\{V_n \varepsilon T\} \leqq \lim_{m\to\infty} \Pr\{U \varepsilon \bar{R}_m\} = \Pr\{U \varepsilon R^*\},$$

the limit existing because $\bar{R}_m$ is a monotone sequence. The assumption

$$\text{(D10)} \qquad \Pr\{U \varepsilon D\} = 0,$$

where D is the set of discontinuities of $f(u)$, implies

$$\text{(D11)} \qquad \Pr\{U \varepsilon S \cup D\} \leqq \Pr\{U \varepsilon S\} + \Pr\{U \varepsilon D\} = \Pr\{U \varepsilon S\}.$$

The theorem follows from (D9), (D11) and the following lemma:

LEMMA D1.

$$\text{(D12)} \qquad R^* \subset S \cup D.$$

PROOF. If $u \varepsilon R^*$, then $u \varepsilon \bar{R}_n$ ($n = 1, 2, \cdots$); that is, in R_n there is a sequence $\{u_k^{(n)}\}$ such that $\lim_{k\to\infty} u_k^{(n)} = u$. We can pick an element from each sequence, say $u_{k(n)}^{(n)} = u_n$, so $u_n \to u$. (For instance, let $u_{k(n)}^{(n)}$ be the $u_k^{(n)}$ with smallest index k so $|u_k^{(n)} - u| < 2^{-n}$.) Since $u_1 \varepsilon R_1$, $u_1 \varepsilon S_m$ for some $m \geqq 1$. Let

$m(1)$ be the smallest m for which this is true, and let $u_1 = u^*_{m(1)}$. Then

$$u_{m(1)+1} \varepsilon S_m$$

for some $m \geqq m(1) + 1$. Let $m(2)$ be the smallest such m, and define $u_{m(1)+1} = u^*_{m(2)}$, etc. Then $u^*_{m(k)} \to u$. Since

$$\lim_{n\to\infty} f_{m(n)}(u^*_{m(n)}) = f(u) \tag{D13}$$

for $u \varepsilon D$, $f_{m(n)}(u^*_{m(n)}) \varepsilon T$ and T is closed, $f(u) \varepsilon T$ and the lemma is proved.

It might be noted that the proof is valid for metric spaces and thus justifies a more general theorem than the one given in Section 7 and used in this paper.

REFERENCES

[1] Anderson, T. W. (1951). The asymptotic distribution of certain characteristic roots and vectors. *Proc. Second Berkeley Symp. on Math. Statist. and Prob.* 103–130. Univ. of California Press.

[2] Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

[3] Anderson, T. W. and Rubin, Herman. (1956). Statistical inference in factor analysis. *Proc. Third Berkeley Symp. on Math. Statist. and Prob.* **5** 111–150. Univ. of California Press.

[4] Bartlett, M. S. (1950). Tests of significance in factor analysis. *British J. Psych. (Statist. Sect.)*, **3** 77–85.

[5] Bartlett, M. S. (1951). The effect of standardization on a χ^2 approximation in factor analysis. *Biometrika* **38** 337–344.

[6] Bartlett, M. S. (1951). A further note on tests of significance in factor analysis. *British J. Psych. (Statist. Sect.)*, **4** 1–2.

[7] Bartlett, M. S. (1954). A note on the multiplying factors for various χ^2 approximations. *J. Roy. Statist. Soc. Ser. B.* **16** 296–298.

[8] Girshick, M. A. (1939). On the sampling theory of roots of determinantal equations. *Ann. Math. Statist.* **10** 203–224.

[9] Hotelling, Harold (1933). Analysis of a complex of statistical variables into principal components. *J. Educ. Psych.* **24** 417–441, 498–520.

[10] Lawley, D. N. (1953). A modified method of estimation in factor analysis and some large sample results. *Uppsala Symposium on Psychological Factor Analysis, 17–19 March* 1953, 35–42, Almqvist and Wicksell, Uppsala.

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

[12] Lawley, D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, **43** 295–303.

[13] Lawley, D. N. (1963). On testing a set of correlation coefficients for equality. *Ann. Math. Statist.* **34** 149–151.

[14] Mallows, C. L. (1960). Latent vectors of random symmetric matrices. *Biometrika*, **48** 133–149.

[15] Mauchly, J. W. (1940). Significance test for sphericity of a normal n-variate distribution. *Ann. Math. Statist.* **11** 204–209.

[16] Roy, S. N. (1957). *Some Aspects of Multivariate Analysis*. Wiley, New York.

[17] Rubin, Herman The topology of probability measures on a topological space. Unpublished.

[18] van der Vaart, H. Robert (1961). On certain characteristics of the distribution of the latent roots of a symmetric random matrix under general conditions. *Ann. Math. Statist.* **32** 864–873.