TESTS FOR THE EQUALITY OF TWO COVARIANCE MATRICES IN RELATION TO A BEST LINEAR DISCRIMINATOR ANALYSIS[1]

BY A. P. DEMPSTER

Harvard University

0. Summary. A pair of test statistics is proposed for the null hypothesis $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ when the data consists of a sample from each of the p-variate normal distributions $N(\boldsymbol{\mu}_1, \mathbf{\Sigma}_1)$ and $N(\boldsymbol{\mu}_2, \mathbf{\Sigma}_2)$. These tests are motivated in Section 1 and defined explicitly in Section 2. Section 3 proves a theorem which includes the null hypothesis distribution theory of the tests. Section 4 gives some details of the computation of the test statistics. An appendix describes the shadow property of concentration ellipsoids which facilitates the geometrical discussion earlier in the paper.

1. Introduction. Consider two p-variate normal populations with $p \times 1$ mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, and $p \times p$ covariance matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$. If $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}$, then all the information for discriminating between the populations is contained in a single linear combination of the p given variables, namely the *best linear discriminator* or b.l.d., which is determined, up to an arbitrary linear transformation, by $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\mathbf{\Sigma}$. Given a sample from each population, one might consider using a sample b.l.d. estimated on the assumption $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$, but first one might wish to test the null hypothesis $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$. This paper will propose new tests for this purpose, designed to be sensitive against particular kinds of differences between $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ which might render the use of the sample b.l.d. deceptive.

The tests will be motivated and described geometrically in terms of the p-dimensional affine space in which a $p \times 1$ p-variate observation vector $\mathbf{X}$ represents a general point. In this space, the first and second moment properties of a population with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$ are naturally represented by its *concentration ellipsoid* Λ [3] with equation

$$(\mathbf{X} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}) = 1. \tag{1.1}$$

Suppose the populations of this paper have concentration ellipsoids Λ_1 and Λ_2 with centers O_1 and O_2. Consider projecting both populations into the line joining O_1 and O_2 along a family Ω of parallel $(p - 1)$ dimensional planes, and suppose A_1 and A_2 are the points where the line segment O_1O_2 is cut by planes from Ω tangent to Λ_1 and Λ_2, respectively. Up to an arbitrary linear transformation, Ω determines a linear combination of the p given variables which is constant over a given member of Ω. This linear combination has a mean difference Δ

Received 8 October 1962; revised 25 June 1963.

[1] This research has been supported in part by the United States Navy through the office of Naval Research, under contract Nonr 1866 (37). Reproduction in whole or in part is permitted for any purpose of the United States Government.

between populations and standard deviations σ_1 and σ_2 in each population. The *shadow property* of concentration ellipsoids, which is described in Section 5, implies that the ratios $\Delta:\sigma_1:\sigma_2$ are the same as the affine line segment ratios $O_1O_2:O_1A_1:O_2A_2$.

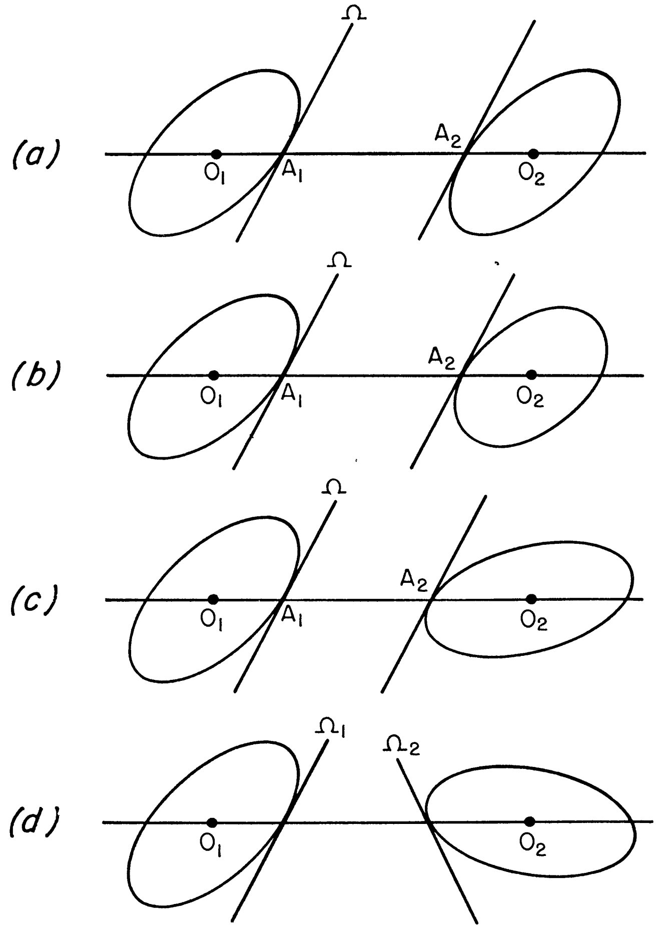

FIG. 1. Four possible relative positions of a pair of concentration ellipsoids.

Figure 1 illustrates ellipsoids Λ_1 and Λ_2. In Figure 1(a) $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$, so that Λ_1 is simply a translation of Λ_2, whereas in Figures 1(b), (c) and (d) increasingly general differences appear between $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$. The corresponding general situations will be called Cases (a), (b), (c) and (d).

Case (a) is defined by the condition $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$. Here the b.l.d. is defined as that linear combination of the given variables with maximum ratio Δ to $\sigma_1 = \sigma_2$. Geometrically, this corresponds to choosing Ω to maximize the ratio O_1O_2 to $O_1A_1 = O_2A_2$. It is clear that this occurs when A_i is the intersection of O_1O_2

with Λ_i for $i = 1, 2$, and Ω consists of planes parallel to the tangent planes to Λ_1 and Λ_2 at A_1 and A_2. When $\mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$, one can choose a family of planes Ω_1 which maximizes the ratio O_1O_2 to O_1A_1, and another family Ω_2 which maximizes the ratio O_1O_2 to O_2A_2, i.e., Ω_i consists of planes parallel to the tangent to Λ_i where Λ_i meets O_1O_2 for $i = 1, 2$. Case (b) refers to the situation where $\mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$ yet Ω_1 and Ω_2 are the same, and even $O_1A_1 = O_2A_2$ for the common Ω. Case (c) is the same as Case (b) except that $O_1A_1 \neq O_2A_2$ for the common Ω. Finally, under Case (d), Ω_1 and Ω_2 are different.

In Case (b) the common Ω may be used for discrimination just as in Case (a) with valid results; the only trouble here is that this discriminator wastes some information. In Case (c) the common Ω may be used, but care must be taken to allow for the different σ_1 and σ_2 of the discriminator, and again some information is wasted. In Case (d) no single linear discriminator is obviously indicated, but see the paper of Anderson and Bahadur [1].

A pair of test statistics will be suggested. These will be independent under the null hypothesis $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$. The first is intended to detect differences like Case (c) when Ω_1 and Ω_2 are the same, and the second is intended to detect different Ω_1 and Ω_2. These tests are based on sample analogues of the various line segments in Figure 1.

2. The proposed tests. Suppose a pair of independent samples from the p-variate normal distributions $N(\boldsymbol{\mu}_1, \mathbf{\Sigma}_1)$ and $N(\boldsymbol{\mu}_2, \mathbf{\Sigma}_2)$ consist of $p \times 1$ vectors $\mathbf{X}_{ij}$ for $j = 1, 2, \cdots, n_i$ and $i = 1, 2$. It will be assumed throughout that the sample sizes n_1 and n_2 satisfy $n_i \geqq p + 1$ for $i = 1, 2$.

In this situation $\boldsymbol{\mu}_i$ and $\mathbf{\Sigma}_i$ for $i = 1, 2$ are naturally estimated by

$$\bar{\mathbf{X}}_i = \sum_{j=1}^{n_i} \mathbf{X}_{ij}/n_i \quad \text{and} \quad \mathbf{S}_i = \mathbf{T}_i/(n_i - 1) \tag{2.1}$$

where

$$\mathbf{T}_i = \sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)(\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)'. \tag{2.2}$$

Or, if it may be assumed that $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}$, then the common $\mathbf{\Sigma}$ is naturally estimated by

$$\mathbf{S} = \mathbf{T}/(n_1 + n_2 - 2) \tag{2.3}$$

where

$$\mathbf{T} = \mathbf{T}_1 + \mathbf{T}_2. \tag{2.4}$$

These sample statistics may be used to define ellipsoids which estimate Λ_1 and Λ_2, i.e.,

$$(\mathbf{X} - \bar{\mathbf{X}}_i)'\mathbf{S}_i^{-1}(\mathbf{X} - \bar{\mathbf{X}}_i) = 1 \tag{2.5}$$

estimates Λ_i if $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are assumed different, or

$$(\mathbf{X} - \bar{\mathbf{X}}_i)'\mathbf{S}^{-1}(\mathbf{X} - \bar{\mathbf{X}}_i) = 1 \tag{2.6}$$

estimates Λ_i if $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are assumed the same.

Rather than consider the geometrical picture provided by the four ellipsoids (2.5) and (2.6) for $i = 1, 2$, it is convenient to consider the equivalent picture as in Figure 2 with ellipsoids L_1, L_2 and L centered at the origin O defined by

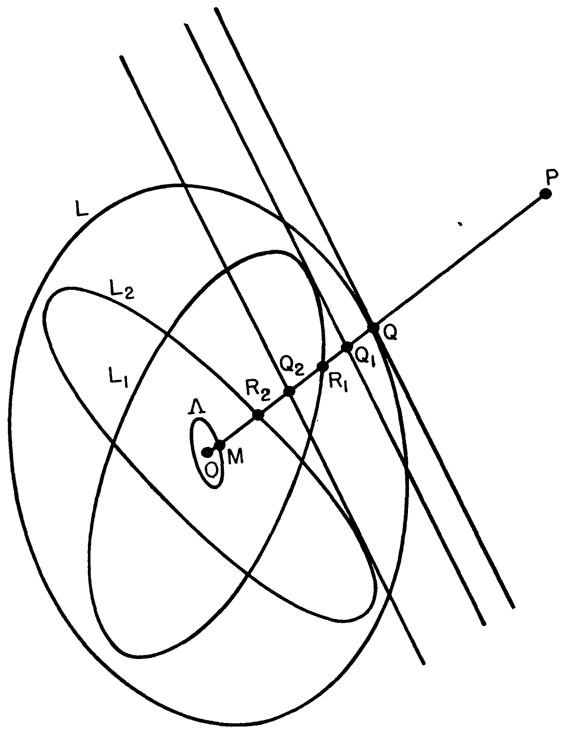

FIG. 2. Basic geometrical figures for defining the statistics of interest.

equations

$$\mathbf{X}'\mathbf{T}_1^{-1}\mathbf{X} = 1, \quad \mathbf{X}'\mathbf{T}_2^{-1}\mathbf{X} = 1, \quad \text{and} \quad \mathbf{X}'\mathbf{T}^{-1}\mathbf{X} = 1, \tag{2.7}$$

and with vector OP where P has coordinates

$$\mathbf{D} = [n_1n_2/(n_1 + n_2)]^{\frac{1}{2}}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2). \tag{2.8}$$

Figure 2 also shows an ellipsoid Λ with equation

$$\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X} = 1, \tag{2.9}$$

where $\mathbf{\Sigma} = \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ under the null hypothesis of interest. Suppose the line segment OP meets Λ, L, L_1 and L_2 in M, Q, R_1 and R_2, respectively. Suppose Ω is the family of planes parallel to the tangent to L at Q, and define Q_1 and Q_2

to be the points of intersection with OP of the tangent planes to L_1 and L_2 in the family Ω.

Under the assumption $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$, Ω defines the natural estimator of the sample b.l.d. based on (2.8) and the pooled within sample dispersion matrix (2.4). Now the sample b.l.d. has a sum of squares about the mean of sample 1, a sum of squares about the mean of sample 2, and the pool of these sums of squares, and, from the shadow property, these three quantities are proportional to OQ_1^2, OQ_2^2 and $OQ^2 = OQ_1^2 + OQ_2^2$. It is therefore plausible to use the ratio

$$C_3 = OQ_1^2/OQ_2^2 \tag{2.10}$$

to test the null hypothesis $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ against an alternative like Case (c) in Figure 1.

To further test whether Case (d) occurs it is plausible to use some weighted combination of the ratios OR_1^2/OQ_1^2 and OR_2^2/OQ_2^2. A particular weighted combination which has a nice distribution property is

$$C_4 = H(OQ_1^2/OR_1^2 - 1, OQ_2^2/OR_2^2 - 1; OQ_2^2/OQ^2, OQ_1^2/OQ^2) \tag{2.11}$$

where $H(x, y; p, q)$ denotes the harmonic mean of x and y with weights p and q, i.e.,

$$H(x, y; p, q) = (p + q)/(p \cdot x^{-1} + q \cdot y^{-1}). \tag{2.12}$$

The null hypothesis distribution theory follows from the theorem of Section 3, and is very simple. In fact C_3 and C_4 are independent and are both distributed like ratios of independent χ^2 random variables.

3. Null hypothesis distribution theory. In this section the notation $\chi_r^2(\tau^2)$ will denote the non-central χ^2 distribution on r d.f. and non-centrality parameter τ^2, i.e., the distribution of $(U_1 + \tau)^2 + U_2^2 + \cdots + U_r^2$ where $U_1, U_2, \cdots, U_r$ are independent $N(0, 1)$. The notation $G_{r,s}$ will denote the distribution of the ratio of independent χ_r^2 and χ_s^2 random variables. The symbol $\sim$ should be read "is distributed like."

The following theorem is a little more general than is required for the tests of Section 2. The ratio C_1/C_2 is, apart from a constant factor, the two-sample T^2 or D^2 statistic. No special use is suggested for the statistic C_5.

THEOREM. *Suppose the* $p \times 1$ *random vectors* $\mathbf{X}_{ij}$ *for* $j = 1, 2, \cdots, n_i$ *and* $i = 1, 2$ *are distributed like independent samples from* $N(\boldsymbol{\mu}_i, \mathbf{\Sigma})$ *for* $i = 1, 2$. *Suppose* $n_i \geqq p + 1$ *for* $i = 1, 2$ *and suppose* $\mathbf{\Sigma}$ *has rank* p. *Suppose the line segments* OP, OM, OQ, OQ_1, OQ_2, OR_1 *and* OR_2 *are defined from the* $\mathbf{X}_{ij}$ *as in Section* 2. *Then the random variables*

$$C_1 = OP^2/OM^2, \tag{3.1}$$

$$C_2 = OQ^2/OM^2, \tag{3.2}$$

$$C_3 = OQ_1^2/OQ_2^2, \tag{3.3}$$

$$C_4 = H(OQ_1^2/OR_1^2 - 1, OQ_2^2/OR_2^2 - 1; OQ_2^2/OQ^2, OQ_1^2/OQ^2) \tag{3.4}$$

$$C_5 = [OQ_2^2/OQ_1^2][OR_1^2/OR_2^2][(OQ_2^2 - OR_2^2)/(OQ_1^2 - OR_1^2)] \tag{3.5}$$

are independent with the $\chi_p^2(\tau^2)$, $\chi^2_{n_1+n_2-p-1}$, G_{n_1-1,n_2-1}, G_{p-1,n_1+n_2-2p} *and* G_{n_1-p,n_2-p} *distributions, respectively, where*

$$\tau^2 = [n_1 n_2/(n_1 + n_2)](\mathbf{\mu}_1 - \mathbf{\mu}_2)'\mathbf{\Sigma}^{-1}(\mathbf{\mu}_1 - \mathbf{\mu}_2). \tag{3.6}$$

The proof of this theorem will be shortened by not deriving the distributions of C_1 and C_2 which have already been given by Bowker [2].

The remaining distributions are most simply derived in terms of a geometrical approach different from that used in Sections 1 and 2. In the new approach the rows of the $p \times (n_1 + n_2)$ data matrix

$$(\mathbf{X}_{11}, \mathbf{X}_{12}, \cdots, \mathbf{X}_{1n_1}, \mathbf{X}_{21}, \mathbf{X}_{22}, \cdots, \mathbf{X}_{2n_2}) \tag{3.7}$$

are regarded as p vectors in an $(n_1 + n_2)$-dimensional Euclidean space E. In the manner usual in analysis of variance, E may be written as the direct sum of four mutually orthogonal subspaces

$$E = E_G \oplus E_B \oplus E_I \oplus E_{II} \tag{3.8}$$

of dimensions 1, 1, $n_1 - 1$ and $n_2 - 1$, respectively. Here, E_G is the subspace spanned by the $1 \times (n_1 + n_2)$ vector,

$$(1, 1, \cdots, 1, 1, 1, \cdots, 1), \tag{3.9}$$

and corresponds to the single d.f. for the grand mean. E_B is the subspace spanned by the $1 \times (n_1 + n_2)$ vector

$$(n_1^{-1}, n_1^{-1}, \cdots, n_1^{-1}, -n_2^{-1}, -n_2^{-1}, \cdots, -n_2^{-1}) \tag{3.10}$$

and corresponds to the single d.f. for the difference between sample means. E_I and E_{II} refer to the remaining subspaces describing differences within the first and second samples, respectively.

The first task of this proof is to relate the line segment ratios of Figure 2 to various angles in the new geometrical space. The information in Figure 2 is based only on the components of the data vectors in $E_B \oplus E_I \oplus E_{II}$ and so the components in E_G will not be considered further. Denote by E_D the p-dimensional subspace of $E_B \oplus E_I \oplus E_{II}$ spanned by the components of the data vectors in $E_B \oplus E_I \oplus E_{II}$.

Any vector $\mathbf{U}$ in E_D corresponds to a linear combination of the p given variables, and the components of $\mathbf{U}$ along E_B and $E_I \oplus E_{II}$ have, for squared lengths, the between sample mean square and the pooled within sample sum of squares for this linear combination of variables. The ratio of this mean square and sum of squares is therefore given by $\cot^2 \theta$ where θ is the angle between $\mathbf{U}$ and E_B. Henceforth, suppose $\mathbf{U}$ corresponds to the sample b.l.d., i.e., $\mathbf{U}$ is a vector in E_D with maximum ratio $\cot^2 \theta$ or, equivalently, minimum θ. Then, from the shadow property in Figure 2,

$$\cot^2 \theta = OP^2/OQ^2. \tag{3.11}$$

Suppose the component of $\mathbf{U}$ in $E_I \oplus E_{II}$ is denoted by $\mathbf{U}_1$ and the components of $\mathbf{U}_1$ in E_I and E_{II} are denoted by $\mathbf{U}_I$ and $\mathbf{U}_{II}$. Suppose the angles from $\mathbf{U}_1$ to $\mathbf{U}_I$ and $\mathbf{U}_{II}$ are denoted by θ_1 and $\theta_2 = \frac{1}{2}\pi - \theta_1$. Then

$$\cos^2 \theta_1 = OQ_1^2/OQ^2 \quad \text{and} \quad \cos^2 \theta_2 = OQ_2^2/OQ^2, \tag{3.12}$$

i.e., $\cos^2 \theta_1$ and $\cos^2 \theta_2$ represent the fractions of the pooled sum of squares OQ^2 contributed by the first and second samples. The final step, which is to bring in OR_1^2 and OR_2^2, is somewhat more complicated and requires the following build-up of theory.

Since only the direction of $\mathbf{U}$ is determined, it may be assumed for simplicity that $\mathbf{U}$ has unit length, so that $\mathbf{U}_1$, $\mathbf{U}_I$ and $\mathbf{U}_{II}$ have lengths $\sin \theta$, $\sin \theta \cos \theta_1$ and $\sin \theta \cos \theta_2$, respectively.

Suppose $\mathbf{U}_2$ is a unit vector orthogonal to $\mathbf{U}_1$ in the space spanned by $\mathbf{U}_I$ and $\mathbf{U}_{II}$, and suppose E_1 and E_2 denote the subspaces spanned by $\mathbf{U}_1$ and $\mathbf{U}_2$. Suppose $E_{I'}$ and $E_{II'}$ denote the subspaces of E_I and E_{II} orthogonal to $\mathbf{U}_I$ and $\mathbf{U}_{II}$, respectively. Then

$$E_I \oplus E_{II} = E_1 \oplus E_2 \oplus E_{I'} \oplus E_{II'} \tag{3.13}$$

is a decomposition of $E_I \oplus E_{II}$ into four mutually orthogonal subspaces of dimensions 1, 1, $n_1 - 2$ and $n_2 - 2$ respectively.

Suppose $E_{D'}$ denotes the $(p - 1)$-dimensional subspace of E_D orthogonal to $\mathbf{U}$. $E_{D'}$ is orthogonal to $\mathbf{U}$, E_B and $\mathbf{U}_1$, and so lies entirely in $E_2 \oplus E_{I'} \oplus E_{II'}$. Suppose $\mathbf{F}_I$ denotes the $(p - 1)$-dimensional orthogonal projection of $E_{D'}$ into $E_2 \oplus E_{I'}$, and define ϕ_1 to be the angle between F_I and E_2. Then ϕ_1 is the angle between E_2 and a vector $\mathbf{W}_I$ in F_I, where, for simplicity, $\mathbf{W}_I$ may be taken to have unit length. Let $F_{I'}$ denote the $(p - 2)$-dimensional subspace of F_I orthogonal to $\mathbf{W}_I$. In a similar manner, F_{II} may be defined as the orthogonal projection of $E_{D'}$ into $E_2 \oplus E_{II'}$, and ϕ_2, $\mathbf{W}_{II}$ and $F_{II'}$ may be defined in a manner analogous to ϕ_1, $\mathbf{W}_I$ and $F_{I'}$. Note that $F_{I'}$ and $F_{II'}$ are subspaces of $E_{I'}$ and $E_{II'}$ respectively.

To bring OR_1^2 into the discussion, suppose that $\mathbf{V}$ is a vector in the orthogonal projection of E_D into $E_B \oplus E_I$ making the minimum angle with E_B. The components of $\mathbf{V}$ along E_B and E_I have lengths proportional to OP and OR_1. (This is the same kind of result as (3.11).) By comparison, the orthogonal projection $\mathbf{U}^*$ of $\mathbf{U}$ into $E_B \oplus E_I$ has components along E_B and $\mathbf{E}_I$ which have lengths proportional to OP and OQ_1. Now $\mathbf{V}$ can be expressed as $\mathbf{U}^* + \mathbf{U}^{**}$ where $\mathbf{U}^{**}$ is the contribution from the orthogonal projection of $E_{D'}$ into $E_B \oplus E_I$. Since $\mathbf{U}^{**}$ has no component along E_B, the ratio OR_1/OQ_1 is simply the ratio of the lengths of the components of $\mathbf{V}$ and $\mathbf{U}^*$ along E_I.

The projection of $E_{D'}$ into $E_B \oplus E_I$ may be carried out in two stages. Firstly, components along $E_{II'}$ may be removed, leaving F_I which may be decomposed into $\mathbf{W}_I$ and $F_{I'}$. Secondly, components along $\mathbf{U}_{II}$ may be removed. The second stage does not alter $F_{I'}$ which already lay in $E_{I'}$, but $\mathbf{W}_I$, which has components of length $\cos \phi_1$ and $\sin \phi_1$ along E_2 and $E_{I'}$, is reduced to a vector

$\mathbf{W}_I^*$ with components of lengths $\cos \phi_1 \sin \phi_1$ and $\sin \phi_1$ along $\mathbf{U}_I$ and $E_{I'}$. Now $\mathbf{U}^{**}$ must be chosen to maximize the component of $\mathbf{V}$ along E_I. No useful contribution to $\mathbf{U}^{**}$ can come from $F_{I'}$ because $F_{I'}$ is orthogonal to both $\mathbf{U}_I$ and $\mathbf{W}_1^*$, and therefore $\mathbf{U}^{**}$ must consist of a multiple of $\mathbf{W}_1^*$. But W_1^* makes angle η_1 with $\mathbf{U}_I$ where

$$\cot \eta_1 = \cos \phi_1 \sin \theta_1 / \sin \phi_1 \tag{3.14}$$

and so a multiple of $\mathbf{W}_1^*$ may be used to reduce the length of $\mathbf{U}_I$ to a fraction $\sin \eta_1$ of its original length, i.e.,

$$OR_1^2/OQ_1^2 = \sin^2 \eta_1 = \tan^2 \phi_1/(\tan^2 \phi_1 + \sin^2 \theta_1). \tag{3.15}$$

A similar formula holds for OR_2^2/OQ_2^2, or, in slightly altered form,

$$OQ_i^2/OR_i^2 - 1 = \sin^2 \theta_i \cot^2 \phi_i \qquad \text{for } i = 1, 2. \tag{3.16}$$

The completion of the proof of the theorem now requires little more than giving the null hypothesis distributions of $\cos^2 \theta_1$, $\cot^2 \phi_1$ and $\cot^2 \phi_2$. Under the null hypothesis $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$, the projection of E_D into $E_I \oplus E_{II}$ is spherically distributed in $E_I \oplus E_{II}$ independently of the components of E_D along E_B. It follows that, conditional on C_1 and C_2 fiixed, $\mathbf{U}_1$ has a spherical distribution in $E_I \oplus E_{II}$ and so

$$C_3 = \cot^2 \theta_1 \sim G_{n_1-1,n_2-1}. \tag{3.17}$$

Next, if $\mathbf{U}$ and $\mathbf{U}_1$ are regarded as fixed, which also fixes C_1, C_2 and C_3, the conditional distribution of the subspace $E_{D'}$ is spherically symmetrical in $E_2 \oplus E_{I'} \oplus E_{II'}$. The relationship of $E_{D'}$ to the subspaces E_2, $E_{I'}$ and $E_{II'}$ determines $\cot^2 \phi_1$ and $\cot^2 \phi_2$ whose joint distribution is therefore independent of C_1, C_2 and C_3 and determined by the spherical distribution of $E_{D'}$. In fact, from the following lemma,

$$(\cot^2 \phi_1, \cot^2 \phi_2) \sim (Z/Z_1, Z/Z_2) \tag{3.18}$$

where Z, Z_1 and Z_2 are independently distributed like χ^2_{p-1}, $\chi^2_{n_1-p}$ and $\chi^2_{n_2-p}$.

Lemma. *Suppose H_D is an r-dimensional spherically distributed subspace of a Euclidean space $H \oplus H_1 \oplus H_2$ where H, H_1 and H_2 are mutually orthogonal subspaces of dimensions 1, m_1 and m_2, respectively. Suppose the orthogonal projection of H_D into $H \oplus H_i$ makes angle Ψ_i with H, $i = 1, 2$. Then*

$$(\cot^2 \Psi_1, \cot^2 \Psi_2) \sim (Y/Y_1, Y/Y_2) \tag{3.19}$$

where Y, Y_1 and Y_2 are independently distributed like χ^2_r, $\chi^2_{m_1-r+1}$ and χ_{m_2-r+1}.

Proof. Suppose $\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_{m_1+m_2+1}$ are $r \times 1$ vectors whose $r \cdot (m_1 + m_2 + 1)$ entries are all independent $N(0, 1)$ random variables. Suppose $\mathbf{Q}_1 = \sum_2^{m_1+1} \mathbf{Z}_i \mathbf{Z}_i'$ and $\mathbf{Q}_2 = \sum_{m_1+2}^{m_1+m_2+1} \mathbf{Z}_i \mathbf{Z}_i'$. Suppose $Y = \mathbf{Z}_1'\mathbf{Z}_1$, $Y_1 = \mathbf{Z}_1'\mathbf{Z}_1/\mathbf{Z}_1'\mathbf{Q}_1^{-1}\mathbf{Z}_1$, and $Y_2 = \mathbf{Z}_1'\mathbf{Z}_1/\mathbf{Z}_1'\mathbf{Q}_2^{-1}\mathbf{Z}_1$. Then from the distribution of C_1 and C_2 in the theorem, i.e., from the theory proved in [2], the random variables Y, Y_1 and Y_2 are distributed as independent χ^2_r, $\chi^2_{m_1-r+1}$ and $\chi^2_{m_2-r+1}$ random variables.

Next suppose that the r rows of

$$(3.20)\qquad (\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_{m_1+1}, \mathbf{Z}_{m_1+2}, \cdots, \mathbf{Z}_{m_1+m_2+2})$$

span an r-dimensional subspace H_D of a Euclidean space $H \oplus H_1 \oplus H_2$ where H, H_1 and H_2 are subspaces spanned by variation in coordinates 1, 2 to $m_1 + 1$ and $m_1 + 2$ to $m_1 + m_2 + 1$, respectively. Then H_D, H, H_1 and H_2 satisfy the hypothesis of the lemma, and, analogous to (3.11),

$$(3.21)\qquad \cot^2 \Psi_i = \mathbf{Z}_1'\mathbf{Q}_i^{-1}\mathbf{Z}_1 = Y/Y_i,$$

for $i = 1, 2$, as required.

To complete the proof of the theorem, one need only note that from (3.4), (3.5), (3.12) and (3.16)

$$(3.22)\qquad C_4 = (\tan^2 \phi_1 + \tan^2 \phi_2)^{-1} \quad \text{and}$$

$$(3.23)\qquad C_5 = \tan^2 \phi_1 / \tan^2 \phi_2,$$

and from (3.18) these have the required distributions.

4. Computational details. Since the statistics C_1, C_2, C_3, C_4 and C_5 introduced above are defined in terms of line segment ratios, it may not be obvious how to compute them from the basic statistics $\mathbf{D}$, $\mathbf{T}_1$, $\mathbf{T}_2$ and $\mathbf{T}$ defined in Section 2. Referring to Figure 2, it may be easily checked that

$$(4.1)\qquad OP^2/OQ^2 = \mathbf{D}'\mathbf{T}^{-1}\mathbf{D}$$

and

$$(4.2)\qquad OP^2/OR_i' = \mathbf{D}'\mathbf{T}_i^{-1}\mathbf{D},$$

for $i = 1, 2$.

To bring in OQ_1 and OQ_2, it may be noted that the expression in (4.1) may be written $(\mathbf{T}^{-1}\mathbf{D})'\mathbf{T}(\mathbf{T}^{-1}\mathbf{D})$ which is the norm of the vector $\mathbf{T}^{-1}\mathbf{D}$ under the inner product $\mathbf{T}$, or, in statistical terms, (4.1) is the pooled within sample sum of squares for the linear combination $\mathbf{T}^{-1}\mathbf{D}$ of the original p variates. (This linear combination defines the usual sample b.l.d.) Now, OQ_1^2 and OQ_2^2 are proportional to the contributions to this pooled sum of squares from sample 1 and sample 2, respectively. Consequently,

$$(4.3)\qquad OQ_i^2/OQ^2 = \mathbf{D}'\mathbf{T}^{-1}\mathbf{T}_i\mathbf{T}^{-1}\mathbf{D}/\mathbf{D}'\mathbf{T}^{-1}\mathbf{D},$$

for $i = 1, 2$.

Formulas (4.1), (4.2) and (4.3) determine all of the required line segment ratios.

5. Appendix: The shadow property of concentration ellipsoids. The concentration ellipsoid Λ (1.1) of a p-variate population is well-known, or easily shown, to be affinely invariant, i.e., its definition is unchanged under any linear transformation of the coordinates $\mathbf{X}$. Consequently its properties can be described in coordinate-free geometrical language. One such property is the *shadow property* which was used earlier in the paper.

Suppose a p-variate population with concentration ellipsoid Λ is projected along a family Ω of parallel $(p - s)$-dimensional planes into an s-dimensional plane π. Then the shadow property states that *the concentration ellipsoid* Λ' *in* π *of the projected population is given by the boundary of the shadow of* Λ *cast in* π *by the projection.* The property is illustrated in Figure 3 where Λ may be regarded as a 3-dimensional ellipsoid, π a 2-dimensional plane, and Ω a family of parallel lines.

The proof of this property is straightforward in terms of a suitably chosen coordinate system, and so is omitted. It may be noted that one can begin with the case of a finite population and then deduce the result for an infinite population or distribution as a limiting case. It may also be noted that the result continues to hold when the population lies in an r-dimensional plane with $r < p$. In this

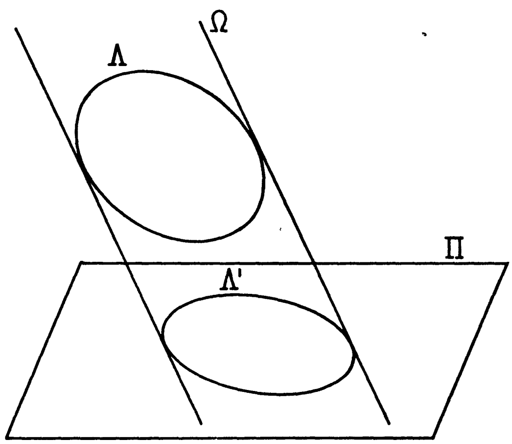

FIG. 3. Shadow property.

case the covariance matrix of the population has rank $r < p$, and so (1.1) cannot be used to define Λ directly. However, Λ can be defined as an r-dimensional ellipsoid using (1.1) for any set of r coordinates in the plane of the population, and for this more general definition the shadow property still holds.

In this paper the shadow property is used in the special case $s = 1$, i.e., where projection is into a line. Note that a 1-dimensional concentration ellipsoid consists of two points whose mid-point and radius are the mean and standard deviation of the population.

REFERENCES

[1] ANDERSON, T. W. and BAHADUR, R. A. (1962). Classification into two multivariate normal distributions with different covariance matrices. *Ann. Math. Statist.* **33** 420–431.

[2] BOWKER, ALBERT H. (1960). A representation of Hotelling's T^2 and Anderson's classification statistic W in terms of simple statistics. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling.* 142–149. Stanford Univ. Press.

[3] CRAMÉR, HARALD (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press.