ADMISSIBLE BAYES CHARACTER OF T^2-, R^2-, AND OTHER FULLY INVARIANT TESTS FOR CLASSICAL MULTIVARIATE NORMAL PROBLEMS

By J. Kiefer[1] and R. Schwartz[2]

Cornell University and General Electric Company

0. Summary. In a variety of standard multivariate normal testing problems, it is shown that certain procedures, often fully invariant, similar, and/or likelihood ratio, are admissible Bayes procedures. The problems include the multivariate general linear hypothesis (where some of the procedures considered were previously shown to be admissible by other methods), the testing of independence of sets of variates (where the likelihood ratio test is shown, for the first time, to be admissible), tests about only some components of the means, classification procedures (for any number of populations), Behrens-Fisher problem, tests about values of or proportionality or equality of covariance matrices, etc. A general technique is developed for obtaining certain Bayes procedures for such problems from the corresponding Bayes procedures relative to *a priori* distributions of a certain type for problems where nuisance parameter means have been deleted.

1. Notation. Before discussing the contents of this paper, we list the notation which will be used throughout.

The letters k, m, n, p, q, r, N, with or without subscripts, will denote positive integers, usually the number of rows or columns of a matrix. S, T, U, V, W, X, Y, Z, with or without subscripts or superscripts, will denote random matrices (or vectors), which in the absence of subscripts always have p rows. S and T will be square. Other Roman capital letters will denote vectors and matrices. I_q denotes the $q \times q$ identity, and 0 denotes any matrix of zeros. V will denote the entire random matrix under observation in any problem, and will always have N columns (= vector observations), independently distributed, each p-variate normal. The decomposition of a $p \times q$ matrix B into blocks will be denoted, for example, by $B = \{B_{ij},\ 1 \leqq i \leqq k_1,\ 1 \leqq j \leqq k_2\}$, where B_{ij} is $p_i \times q_j$ with $\sum_i p_i = p$, $\sum_j q_j = q$. A decomposition into blocks of rows (resp., columns) alone will be denoted by $B = \{B_{(i)}\}$ (resp., $B = \{B_j\}$). Other decompositions will occasionally be denoted by superscripts. However, a notation like $V = (X, Y, U)$ will sometimes serve better than $V = (V_1, V_2, V_3)$ to distinguish the roles of different parts of V. Unprimed vectors will denote column vectors. B' denotes the transpose of B. The determinant of C is denoted by $|C|$, and its trace is denoted by tr C. If C is symmetric positive definite,

Received 27 August 1964.

[1] Research supported by the Office of Naval Research under contract No. Nonr-266(04) (NRO 47-005).

[2] Written, in part, while this author was a National Science Foundation Predoctoral Fellow.

$C^{\frac{1}{2}}$ will denote its unique symmetric positive definite square root. The average of the columns of any matrix X will be denoted by $\bar{X}$. The matrices S and T will be nonnegative definite symmetric, usually obtained in a problem as $S = (X - \bar{X})\ (X - \bar{X})'$ or $S = YY'$. Whenever such a matrix is positive definite on a set of probability one according to each θ in Ω, we shall invert it without further mention of the exceptional set.

Positive finite constants, depending on the problem but not on the parameter values, will be denoted by c or c_i. (A trivial exception to positivity and finiteness is the usage in (1.3).) The meaning of any c_i may change with the problem.

$\Omega = \{\theta\} = H_0 + H_1$ will denote the parameter space in any problem. Occasionally, to emphasize the symmetry of two hypotheses, we shall write $\Omega = H_1 + H_2$. The parameter θ will be decomposed into a collection of matrices (or vectors) ξ, ν, μ, Σ, etc., with or without subscripts; Σ without subscripts or superscripts will always be a covariance matrix which is $p \times p$ positive definite. Thus, Greek letters will be used to denote functions of θ; in addition, β, η, and γ will be reserved for other variables in terms of which it is convenient to write *a priori* densities, and of which components of θ may be functions. (See, for example, (4.1).) A Greek letter with subscript or superscript 0 (and perhaps other subscripts of superscripts) always denotes a specified value.

The use of $\sum$ to denote summation (e.g., as $\sum_i$ or $\sum_1^k$) will always be such that it cannot be mistaken for a parameter.

All probability laws of observable random variables, or functions thereof which we shall consider, will have Lebesgue densities on a Euclidean set. The Lebesgue density function of X when θ describes the underlying probability measure will be denoted by $f_X(x; \theta)$, or perhaps by $f_X(x; \beta(\theta))$ or $f_X(x; \beta)$ if this density depends on θ only through $\beta(\theta)$. We shall write $\exp a = e^a$, $\text{etr}\, A = \exp \text{tr}\, A$. Densities of particular interest are the multivariate normal density of a $p \times n$ matrix X of independent columns, each with nonsingular covariance matrix Σ, and with $EX = \xi$:

$$(1.1) \qquad \phi_{p,n}(x; \xi, \Sigma) = c_1|\Sigma|^{-n/2}\, \text{etr} - \tfrac{1}{2}\Sigma^{-1}(x - \xi)(x - \xi)';$$

and, if $n \geqq p$, the central Wishart density of $W = (X - \xi)(X - \xi)'$ in this setting:

$$(1.2) \qquad \psi_{p,n}(w; \Sigma) = c_2|\Sigma|^{-n/2}|w|^{(n-p-1)/2}\, \text{etr} - \tfrac{1}{2}\Sigma^{-1}w.$$

For (1.2), the domain is $\{w_{ij},\ i \geqq j: w \text{ positive definite}\}$.

A priori probability measures or positive constant multiples thereof will be denoted by Π. It is convenient to refrain from giving the explicit values of positive multiplicative constants and to require only $\Pi(\Omega) < \infty$ rather than $\Pi(\Omega) = 1$, and we shall do so. If $\Pi = \Pi_0 + \Pi_1$ with Π_i a finite measure on H_i, every Bayes critical region (for 0–1 loss function) is of the form

$$(1.3) \qquad \{v: \textstyle\int f_V(v; \theta)\Pi_1\,(d\theta)/\int f_V(v; \theta)\Pi_0\,(d\theta) > c\} \cup L_c$$

for some $c(0 \leqq c \leqq \infty)$, where L_c is a measurable subset of the set obtained from the set in braces in (1.3) by replacing $>$ by $=$. In all our applications every L_c will have probability 0 for all θ in Ω, so that our Bayes procedures will be essentially unique. (An exception occurs in Corollary 3.2, where a different argument is used.) *Hence, all our Bayes procedures are admissible.*

In each example we obtain a family of tests by varying c from 0 to ∞ in (1.3). When, for example, the tests are similar, this of course yields an admissible similar test of each possible significance level.

The Π_i which arise in our examples all have Lebesgue densities on Euclidean sets, or are measures assigning all mass to a single point θ_0, or are products of these. Sometimes it will be convenient to consider Π_i or one of its factors to be a density on a Euclidean set Γ which is mapped in a given way into Ω or one of its factors. For example, it will be simpler to compute with the Lebesgue density $c|I_p + \eta\eta'|^{-m}$ on Euclidean p-space $E^p = \Gamma = \{\eta\}$, than with the induced measure on the space of positive definite $\Sigma = (I_p + \eta\eta')^{-1}$. Such Lebesgue densities (which will be integrable but not necessarily of integral one) will be denoted by $d\Pi_i(\eta)/d\eta$. The integrating Lebesgue measure in such a case will be denoted by $d\eta$.

In each example it is possible to work either with the original V, or else with a sufficient statistic. Usually the computations are such that there is no particular gain in using the reduction to the latter form.

Throughout the paper densities will be continuous on the product of sample and parameter space, both of which will be Euclidean spaces or Borel subsets thereof. *A priori* densities will be of the same character. Thus, no measurability considerations will ever be required, and they will always be omitted.

The reader is referred to the books by Roy (1957) and Anderson (1958) for descriptions of various multivariate problems and procedures, and to the book by Lehmann (1959) for general hypothesis testing theory.

2. Introduction. Admissibility of various classical statistical tests has been proved using (1) Bayes procedures, (2) exponential or other special structure of Ω, (3) invariance, and (4) local properties. (Estimation problems do not concern us here, and the techniques of this paper yield little of interest in such problems.) Some examples of (1) can be found in Lehmann and Stein (1948), Karlin (1957), Lehmann (1959), and Ellison (1962). Method (2) has been used by Birnbaum (1955), Stein (1956b), Nandi (1963), Ghosh (1964), and, more recently, by Schwartz (1964b). (It is also indicated to be the approach of Roy and Mikhail (1960), but the method is inapplicable in at least one of the cases described in their abstract, that of testing independence; this will be discussed further in the next paragraph.) Aside from the trivial case of compact groups, only the one-dimensional translation parameter case of (3) has been studied, in Lehmann and Stein (1953). The most common occurrence of (4) is with unique uniformly or locally most powerful unbiased tests in cases where unbiasedness implies similarity on that part of the boundary of H_1 which is in H_0. (See Lehmann (1959).) Uniformly most powerful tests can be regarded as a special case of (1). A result like that of Wald (1942) on the analysis of variance test can be regarded, for example, as an application of (1) to the similar tests obtained from unbiasedness considerations (4).

The use of techniques (3), (4), and (2) in standard multivariate normal problems has been limited. Best invariant procedures under the full linear group need not be minimax, let alone admissible. Most powerful unbiased tests, or analogues of Wald's theorem, fail to exist. While the exponential structure can be used to prove admissibility of Hotelling's T^2-test (Stein (1956b)) and of a class of tests of the multivariate general linear hypotheses (Schwartz (1964b)), this technique (or its generalization to certain nonexponential families) cannot be used in the problem of testing independence of sets of variates when $p \geqq 3$. This is discussed by Stein (1956b); from a slightly different viewpoint, it can be seen, even in the bivariate case, that this method fails because $f_{R^2}(r^2; \rho^2)/f_{R^2}(r^2; 0)$ (where R and ρ are the sample and population correlation coefficients) is not unbounded as $\rho \to 1$. The admissibility of the usual test in the bivariate case $p = 2$ is proved in Lehmann (1959) using (4), but this approach also fails when $p > 2$.

Our main interest is in those procedures which are invariant under all linear-affine transformations which leave the problem invariant, and which we shall call *fully invariant*. We use the Bayes technique (1) to prove admissibility of certain fully invariant tests of the general multivariate linear hypothesis and of the hypothesis of independence of sets of variates, as well as in other testing problems. The T^2- and R^2-tests are special cases. Even in the case of a test such as Hotelling's, where admissibility was proved (Stein (1956b)) by method (2), our result yields additional information on the performance of the test; for the method of (2) only insures that no other test of the same size is superior to T^2 "far" from H_0, while our Bayes result, discussed further in Section 4, reflects the behavior of T^2 closer to H_0.

Nevertheless, the Bayes technique has severe limitations. As always, it may be hard to guess the Π with respect to which a given test is Bayes, or to carry out a very explicit integration for a given Π. Moreover, many natural admissible tests cannot be proved admissible by this approach for reasons other than that of lack of integrability for minimum sample sizes which is mentioned later for certain tests. For example, Birnbaum's treatment shows that, for the problem of testing that the mean of a bivariate normal vector, with known covariance matrix, is 0, any compact convex polygonal acceptance region is admissible; but analyticity considerations show it cannot be Bayes. Thus, in problems like that of Section 4, one cannot expect tests such as the familiar one of Roy which is based on the largest characteristic root of $(XX' + YY')^{-1}XX'$, to be Bayes; but such tests can be proved admissible by technique (2). (See the references given earlier.)

Thus, we shall obtain admissibility of certain isolated but natural (and, often, well known) tests, rather than any general theorem characterizing Bayes tests. The tests obtained are often similar and unbiased, and are sometimes most powerful invariant and classical in origin (e.g., likelihood ratio tests). An admissible test of level α which is similar is of course also admissible with respect to its power function considered only on H_1, among level α tests.

There is no difficulty in constructing many noninvariant Bayes tests. For example (in a case of the setting of Section 4), if $V = (X, Y)$ with X $p \times 1$ and Y $p \times n (n \geqq p)$, with each column of V having covariance matrix Σ and with $EX = \xi$, $EY = 0$, for testing $H_0 : \xi = 0$ one easily shows that the following critical regions are Bayes: $X_{(1)} > c$ (where $X_{(i)}$ is 1×1), $X'X > c$, $X'X/\text{tr}\,(Y'Y) > c$, $\sum_1^p X_{(i)}^2/(Y'Y)_{ii} > c$, etc. The disadvantage of using any of these tests is of course that while each of them is similar for some subhypothesis of H_0, they have less satisfactory power characteristics under H_0 itself. This is why it is usually of greatest interest to find fully invariant tests. (In some examples, such as that of Section 6 (ii), it may be that the group of transformations which leaves the problem invariant is less relevant than a subgroup which leaves invariant some natural measure of distance from H_0; that is, it may be that $H_1 = \bigcup_\tau H_{1\tau}$ is invariant under a group G, but that each $H_{1\tau}$ is not.)

Such invariant tests often can be obtained, and in some cases have been obtained more than once in the literature since the first work of Jeffreys (1939) in this direction, as *formal* ("generalized") Bayes procedures with respect to invariant Π's of infinite mass. Of course, such a derivation cannot yield admissibility. In estimation problems as simple as that of estimating the mean of a standard univariate normal distribution with squared error loss, it is not hard to prove that the best invariant procedure cannot be Bayes. It was Lehmann and Stein (1948) who first showed that, in invariant testing problems, a different situation sometimes prevailed, and that best invariant procedures were sometimes genuinely Bayes for noninvariant reasons. An example of Section 7 (iii) gives a generalization of some of their univariate normal results. It will be seen that the rationale in choosing Π in other problems, such as the T^2 and R^2 generalizations of Sections 4 and 5, is somewhat different; this difference will be discussed further at the end of Section 3.

One consequence of the Bayesian character of certain invariant tests is that there is no possibility of proving an inadmissibility result for an essentially unique best invariant test where each hypothesis consists of a p-dimensional translation parameter family, analogous to the corresponding sweeping inadmissibility result of Stein (1956a) (see also Brown (1964)) in estimation problems. This is already evident in the example, covered by the results of Lehmann and Stein (1948), pp. 503–504, according to which, if $V = (V_1, \ldots, V_n)$ with $n > 1$, the V_i being independent normal p-vectors with common unknown mean and with covariance matrix $\sigma_i^2 I_p$ under H_i (with σ_i^2 specified), the essentially unique best invariant test, based on $\sum_1^n (V_i - \bar{V})'(V_i - \bar{V})$, is admissible.

It will be obvious (and will sometimes be illustrated explicitly) that in many examples there are many, and often infinitely many, linearly independent Π's relative to which a given procedure is Bayes. For example, Π_i will often assign all measure to a set where $\Sigma^{-1} = C + \eta\eta'$ where C can be taken to be an arbitrary fixed positive definite matrix (which we will usually take to be I_p) and η is a $p \times q$ random matrix; in some cases (e.g., (7.2)), even q may be varied for a fixed test. (A brief general discussion of these Π's will be found in the three paragraphs following (3.8).) The procedure is then Bayes relative to any finite (and, often, infinite) convex mixture of those Π's. This variability of the Π relative to which a given procedure is Bayes lends insight regarding the performance of the procedure. The richness of the family of Π's relative to which, for example, the T^2-test can be seen to be Bayes, may find a use in proving the minimax character of that test on a surface of constant power, and with computational ease compared with the calculation of Giri, Kiefer, and Stein (1963) in a special case. This approach has not yet succeeded.

Thus, no really novel minimax results are contained in this paper. An admissible test like (7.10), which (for appropriate c_i) has constant power on each H_i, is automatically minimax (the one-sided analogue being even simpler). Section 6(ii) gives an example where minimax properties follow from previously known results.

Regarding minimax properties, we remark that, at least locally, tests based on traces of appropriate matrices appear to be more satisfactory than those based on determinants (Section 4 and 6; see Schwartz (1964a)).

The reader will note that, in many respects, the results of Sections 4 and 5 are more satisfactory than those of some of the examples of Sections 6 and 7 (for example, 6(ii)) where the group involved is not merely a direct sum of full linear groups.

It would require too much space to list, in each setting considered herein, even a few of the tests which can be obtained by the methods of this paper. We shall therefore list a few such variants only in Sections 4 (multivariate general linear hypothesis) and 7(i); the applicability of the methods in other examples will be clear.

Moreover, there are many testing setups for normal and other exponential families which we shall omit entirely because of the space they would occupy, but in which our methods can be applied. A few of the problems we shall exclude are those concerned with hypothesized nonzero values of all or some of the canonical correlations, correlations, partial correlations, or eigenvalues of a covariance matrix; equality of such parameters of two covariance matrices; hypothesized values of certain parameters of both the mean and covariance matrix, or equality of such parameters of two mean vectors and covariance matrices; the hypotheses which arise in principal component and factor analysis.

In addition, there are multivariate analogues of many of the examples of Lehmann and Stein (1948), which we shall omit.

Our examples will be ones in which both means and covariance matrices are unknown. Where some of these are known (for example, if Σ and ν are known in Section 4), the problems are easier to solve and sometimes (as in the cases of Sections 4 and 6 if Σ is known) have well known solutions.

3. Preliminary results. We summarize here the integration results which will be used repeatedly.

From (1.1) we have, if η and z are $p \times m$ and t is $p \times p$ and positive definite,

$$\int_{E^{mp}} \operatorname{etr} \{ -\tfrac{1}{2}[t\eta\eta' - 2z\eta'] \} \, d\eta = c|t|^{-m/2} \operatorname{etr} \tfrac{1}{2}t^{-1}zz'. \tag{3.1}$$

If γ is a p-vector, an obvious diagonalization yields the well known relation

$$|I_p + \gamma\gamma'| = 1 + \gamma'\gamma. \tag{3.2}$$

Defining, for h real,

$$b_h = \int_{E^p} (1 + \gamma'\gamma)^{-h/2}\, d\gamma, \tag{3.3}$$

we clearly have

$$b_h < \infty \Leftrightarrow h > p. \tag{3.4}$$

With $\eta = (\eta_1, \eta_2, \ldots, \eta_m)$ $p \times m$, write $Q_j = I_p + \sum_{i=1}^{j} \eta_i\eta_i'$ and let $Q_j^{\frac{1}{2}}$ be the positive definite symmetric square root of Q_j. Using the change of variables $z_j = Q_{j-1}^{-\frac{1}{2}}\, \eta_j$ and (3.2)–(3.3), we have, for $j > 1$ and h real,

$$\begin{aligned} \int_{E^p} |Q_j|^{-h/2}\, d\eta_j &= \int |Q_{j-1} + \eta_j\eta_j'|^{-h/2}\, d\eta_j \\ &= |Q_{j-1}|^{(1-h)/2} \int |I_p + z_jz_j'|^{-h/2}\, dz_j \\ &= b_h |Q_{j-1}|^{(1-h)/2}. \end{aligned} \tag{3.5}$$

Hence,

$$\int_{E^{mp}} |I_p + \eta\eta'|^{-h/2}\, d\eta = \prod_{i=1}^{m} b_{h-i+1}, \tag{3.6}$$

which with (3.4) yields

$$\int_{E^{mp}} |I_p + \eta\eta'|^{-h/2}\, d\eta < \infty \Leftrightarrow h > m + p - 1. \tag{3.7}$$

We shall use the well known fact that, with η $p \times k$,

$$|I_k - \eta'(I_p + \eta\eta')^{-1}\eta| = |I_p + \eta\eta'|^{-1}. \tag{3.8}$$

This follows, for example, from a direct computation upon writing $\eta = A_pLA_k$ where A_j is orthogonal $j \times j$ and, according to whether $p \leqq k$ or $p \geqq k$, we have $L = (D_p, 0)$ or $L' = (D_k, 0)$ with D_j diagonal $j \times j$. This diagonalization also demonstrates the positive definiteness of the matrix whose determinant is on the left side in (3.8); we shall use this fact below.

A main idea in our construction of appropriate Π's is the frequent representation of covariance matrices $(I_p + \eta\eta')^{-1}$ where η is $p \times q$ for an appropriate q. Thus, we assign all measure to a set where $I_p - \Sigma$ is positive definite. The form $I_p + \eta\eta'$ and certain other forms used in the Π's were suggested by their appearance in the linear functionals used in approach (2) of Section 2 for the problem of testing the general linear hypothesis. (See Stein (1956) and Schwartz (1964a) and (1964b).)

The representation $(I_p + \eta\eta')^{-1}$ suggests a formal structure for the *a priori* densities $d\Pi_i(\eta)/d\eta$ which by (3.7) will in fact be integrable. It also yields integrals (and hence Bayes procedures) of simple functional form.

A second idea (which depends on the above-mentioned property of the set where Π is supported) is the elimination of means which are nuisance parameters by means of Lemma 3.1 below. Another, more obvious, idea, is the treatment of certain nuisance *components* of the p-vectors of means (for example, in Section 6) by letting Π assign all measure to a set where these components are independent of the others and have a single distribution under each H_i.

We shall now formalize a technique for proving that, in some cases, a procedure Δ which is Bayes for a problem $P = \{H_0, H_1, V\}$ remains Bayes when the problem is altered to P^* by the addition of certain nuisance parameters and corresponding observables. Suppose, for a problem P, that Δ is Bayes (i.e., satisfies (3.1)) for given Π_0, Π_1, and c. Suppose P is now altered to $P^* = \{H_0^*, H_1^*, V^*\}$ as follows: $V^* = (V, U_1, U_2, \ldots, U_m)$ where U_j is $q_j \times 1$ with m and q_j arbitrary positive integers; the U_j are independent of V and each other, U_j being normal with mean vector ν_j and nonsingular covariance matrix $\Sigma^{(j)}$. The $\Sigma^{(j)}$ might be related to some of the parameters of V or to each other (for example, several might be equal). However, we assume that, for each θ in a subset of H_i to which Π_i assigns all measure, there is a corresponding set in H_i^* for which the domain of $(\nu_1, \nu_2, \ldots, \nu_m)$ is $E^{\Sigma_1^m q_i}$ and for which $\Sigma^{(j)}$, which in this instance we shall write as $\Sigma^{(i,j)}$ for clarity, can be written as

$$\Sigma^{(j)} \equiv \Sigma^{(i,j)} = (C^{(j)} + D^{(i,j)})^{-1} \tag{3.9}$$

where $C^{(j)}$ is symmetric positive definite (and does not depend on i), and $D^{(i,j)}$ is symmetric nonnegative definite. (If $\Sigma^{(j)}$ is unrelated to θ, it can of course be treated trivially by letting each Π_i^* assign all measure to any specified value of $(\Sigma^{(j)}, \nu_j)$.) If $D^{(i,j)}$ is of rank $r_{i,j}$, we can then write

$$D^{(i,j)} = \Lambda^{(i,j)}\Lambda^{(i,j)\prime} \tag{3.10}$$

where Λ is $q_j \times r_{i,j}$. (Actually, certain restrictions on the ν_j can be imposed, as will become evident in the course of the proof of Lemma 3.1.) The possible distributions of V under H_i^* are the same as under H_i. As mentioned at the end of Section 1, the H_i and H_i^* are assumed to be Euclidean Borel sets.

LEMMA 3.1. *If Δ is Bayes relative to Π for problem P, then Δ^* is Bayes relative to some Π^* for problem P^*, where*

$$\Delta^*(v, u_1, u_2, \ldots, u_m) = \Delta(v). \tag{3.11}$$

PROOF. Write $\beta = (\Sigma^{(1)}, \Sigma^{(2)}, \ldots, \Sigma^{(m)})$ and $\nu = (\nu_1, \ldots, \nu_m)$. Let the conditional *a priori* distribution of the ν_j, given θ (the parameter of V) and β, be as follows: the ν_j are conditionally independent and, under H_i^* (that is, if $\theta \in H_i$), with *a priori* probability one,

$$(\Sigma^{(i,j)})^{-1}\nu_j = \Lambda^{(i,j)}\gamma^{(i,j)} \tag{3.12}$$

where $\gamma^{(i,j)}$ is $r_{i,j} \times 1$ and is normal with mean vector 0 and covariance matrix

$$(I_{r_{i,j}} - \Lambda^{(i,j)\prime}(C^{(j)} + D^{(i,j)})^{-1}\Lambda^{(i,j)})^{-1} = (B^{(i,j)})^{-1} \quad \text{(say)}.$$

Denote this conditional distribution of ν by $\Pi_{i,\theta,\beta}^*$. The (marginal) joint distribution of θ and β under H_i^* (that is, under $c_i\Pi_i^*$) is given as follows: θ has (marginal) distribution Π_i and, given θ, the conditional distribution of β is any probability measure $\Pi_{i,\theta}$ assigning measure one to a specified value (or set of values) of β which is possible for this value of θ (that is, which is consistent with any relation which may exist between θ and β). There is no measurability difficulty in this construction, since the H_i^* are Euclidean Borel sets. We must only verify that $B^{(i,j)}$ is positive definite. Letting $C^{\frac{1}{2}}$ denote the symmetric positive definite square root of $C^{(j)}$ and $\eta = C^{-\frac{1}{2}}\Lambda^{(i,j)}$, we can write $B^{(i,j)}$ as $I_{r_{i,j}} - \eta'(I_{q_j} + \eta\eta')^{-1}\eta$ and recall the second sentence following (3.8).

From (3.8) itself we then have $|B^{(i,j)}| = |I_{q_j} + \eta\eta'|^{-1} = |C^{(j)}||\Sigma^{(i,j)}|$. Hence, if ν_j satisfies (3.12) under H_i^*, we have, omitting most superscripts in the exponential after the second expression,

$$\begin{aligned}
&f_{U_j}(u; \nu_j, \Sigma^{(i,j)})\phi_{r_{i,j}}(\gamma^{(i,j)}; 0, (B^{(i,j)})^{-1}) \\
&\quad = c_3|\Sigma^{(i,j)}|^{-\frac{1}{2}}|B^{(i,j)}|^{\frac{1}{2}} \operatorname{etr} -\tfrac{1}{2}\{(\Sigma^{(i,j)})^{-1}(u - \nu_j)(u - \nu_j)' \\
(3.13)\qquad &\qquad + B^{(i,j)}\gamma^{(i,j)}\gamma^{(i,j)\prime}\} \\
&\quad = c_3|C^{(j)}|^{\frac{1}{2}} \operatorname{etr} -\tfrac{1}{2}\{C^{(j)}uu'\} \operatorname{etr} -\tfrac{1}{2}\{Duu' - 2\Lambda\gamma u' + \Lambda\gamma\gamma'\Lambda'\Sigma \\
&\qquad + \gamma\gamma' - \Lambda'(C + D)^{-1}\Lambda\gamma\gamma'\} \\
&\quad = c_3|C^{(j)}|^{\frac{1}{2}} \operatorname{etr} -\tfrac{1}{2}\{C^{(j)}uu'\} \operatorname{etr} -\tfrac{1}{2}(\gamma - \Lambda'u)(\gamma - \Lambda'u)'.
\end{aligned}$$

From this we conclude that if, under H_i^*, ν_j is the function of $\gamma^{(i,j)}$ given in (3.12), then

$$(3.14)\qquad \frac{\int_{E^{r_{1,j}}} f_{U_j}(u; \nu_j, \Sigma^{(1,j)})\phi_{r_{1,j}}(\gamma^{(1,j)}; 0, (B^{(1,j)})^{-1})\, d\gamma^{(1,j)}}{\int_{E^{r_{0,j}}} f_{U_j}(u; \nu_j, \Sigma^{(0,j)})\phi_{r_{0,j}}(\gamma^{(0,j)}; 0, (B^{(0,j)})^{-1})\, d\gamma^{(0,j)}} = c_4 .$$

Hence,

$$(3.15)\qquad \frac{\int f_{U_1,\cdots,U_m}(u_1, \cdots, u_m; \nu, \beta)\Pi_{1,\theta,\beta}^*(d\nu)}{\int f_{U_1,\cdots,U_m}(u_1, \cdots, u_n; \nu, \beta)\Pi_{0,\theta,\beta}^*(d\nu)} = c_5 .$$

Thus, writing $\theta^* = (\theta, \nu, \beta)$, we have

$$\begin{aligned}
(3.16)\quad &\int f_{V^*}(v, u_1, \ldots, u_m; \theta^*)\Pi_1^*(d\theta^*)/\int f_{V^*}(v, u_1, \ldots, u_m; \theta^*)\Pi_0^*(d\theta^*) \\
&\qquad = c_6[\int f_V(v; \theta)\Pi_1(d\theta)/\int f_V(v, \theta)\Pi_0(d\theta)],
\end{aligned}$$

so that a proper choice of c^* in $\Pi^* = c^*\Pi_1^* + \Pi_0^*$ yields the conclusion of the lemma.

A degenerate case of the above setup occurs when V and θ are absent; that is, when, on the basis of U_1 $U_2, \ldots, U_m$, it is desired to test some hypothesis concerning the $\Sigma^{(j)}$ (and/or a linear space of linear combinations of the ν_j which, under both H_i^*, has only 0 in common with the space spanned by $\Sigma^{(i,j)}\Lambda^{(i,j)}\gamma^{(i,j)}$). In that case let H_i^{**} be any subhypothesis of H_i^* which consists of the $(r_{i1} + \ldots + r_{im})$-dimensional Euclidean space of $(\gamma^{(i,1)}, \ldots, \gamma^{(i,m)})$, the values of the $\Sigma^{(j)}$ and other linear combinations of the ν_j being specified so that (ν, β) is completely determined by the $\gamma^{(i,j)}$. Then Π_i^* (formerly $\Pi_{i,\theta,\beta}^*$) has a continuous positive Lebesgue density on each H_i^{**}, and every procedure has continuous power function on each H_i^{**}. Hence, even though the Bayes procedure in this case is not essentially unique, we conclude:

COROLLARY 3.2. *For testing between the H_i^* based on $U_1, U_2, \ldots, U_m$, for each $\alpha(0 \leqq \alpha \leqq 1)$ the randomized test which accepts H_1^* with probability α for every sample value, is an admissible Bayes procedure.*

Lehmann and Stein (1948) also gave examples where the randomized test of Corollary 3.2 is Bayes.

We remark, incidentally, that the method used by Lehmann and Stein to handle means which are nuisance parameters (other than the use of a Π_i concentrated at a single point), and which differs from that of Lemma 3.1, is (roughly) to let the means have normal *a priori* densities under H_0 (say), with means equal to those under H_1, and with variances equal to the difference between hypothesized variances under H_1 and H_0. When one or more variances are nuisance parameters, as under H_0 in Student's problem when mean and variance are both specified under H_1 (that is, the problem of Section 6(vi) below, modified to specify Σ also, under H_1), their *a priori* density again reflects the difference between variances under H_0 and H_1. We are usually unable to make use of these techniques, but the test derived in (7.8) is an exception, which uses a direct multivariate analogue of the above technique for means. Lehmann and Stein generally consider simple alternatives and thereby often obtain uniformly most powerful tests against composite alternatives; these are stronger conclusions than ours, which are often obtained in settings where no such uniformly most powerful tests exist.

4. Multivariate general linear hypothesis. In the usual formulation, $V = (V_1, \ldots, V_N)$ with cov $V_i = \Sigma$ and $EV = \xi L$, H_0 being $\xi K = \xi^{(0)}K$ (specified, and which by a translation can be taken to be 0), with L and K known matrices. This can be transformed into the canonical form wherein $V = (X, Y, U)$ with $EX = \xi(p \times r)$, $EY = O(p \times n)$, $EU = \nu(p \times h)$, where, under Ω, ξ and ν have E^{pr} and E^{ph} as their domains, and all columns of V are again independent with common unknown covariance matrix Σ; H_0 is $\xi = 0$. We treat the problem in this canonical form.

According to Lemma 3.1, results for the general case follow from those for the case $h = 0$, which we hereafter treat. In part (v) below we shall discuss the case $n < p$; until that part, we suppose $n \geqq p$.

(i) Let both Π_1 and Π_0 assign all their measure to θ's for which $\Sigma^{-1} = I_p + \eta\eta'$ for some $(p \times r)\eta$. Also, under H_1 all measure is assigned to ξ's of the form $\xi = \Sigma\eta$ (where $\Sigma^{-1} = I_p + \eta\eta'$). The Π_i's can be considered as absoutely continuous measures on the space E^{pr} of η's, and are given by

$$(4.1)\qquad \begin{aligned} d\Pi_1(\eta)/d\eta &= |I_p + \eta\eta'|^{-(r+n)/2} \operatorname{etr} \tfrac{1}{2}\{\eta'(I_p + \eta\eta')^{-1}\eta\}, \\ d\Pi_0(\eta)/d\eta &= |I_p + \eta\eta'|^{-(r+n)/2}. \end{aligned}$$

The integrability of these densities follows from (3.7) (since $n \geqq p$) and the boundedness of the nonnegative definite matrix in braces (which, according to the comment two sentences below (3.8), yields a positive definite matrix when subtracted from I_p). We then have

$$
\begin{aligned}
&\int f_{X,Y}(x, y; \theta)\Pi_1\,(d\theta)/\int f_{X,Y}(x, y; \theta)\Pi_0\,(d\theta)\\
&= \frac{\int |I_p + \eta\eta'|^{(r+n)/2}\,\mathrm{etr}\,\{-\tfrac{1}{2}(I_p + \eta\eta')(xx' + yy') + \eta x' - \tfrac{1}{2}(I_p + \eta\eta')^{-1}\eta\eta'\}\Pi_1\,(d\eta)}{\int |I_p + \eta\eta'|^{(r+n)/2}\,\mathrm{etr}\,\{-\tfrac{1}{2}(I_p + \eta\eta')(xx' + yy')\}\Pi_0\,(d\eta)} \qquad (4.2)\\
&= \frac{\mathrm{etr}\,\{\tfrac{1}{2}(xx' + yy')^{-1}xx'\}\int \mathrm{etr}\,\{-\tfrac{1}{2}(xx' + yy')\cdot(\eta - (xx' + yy')^{-1}x)(\eta - (xx' + yy')^{-1}x)'\}\,d\eta}{\int \mathrm{etr}\,\{-\tfrac{1}{2}(xx' + yy')\eta\eta'\}\,d\eta}\\
&= \mathrm{etr}\,\{\tfrac{1}{2}(xx' + yy')^{-1}xx'\}.
\end{aligned}
$$

(As stated in Section 1, we shall not require, here and in other examples, a discussion of the exceptional set where $XX' + YY'$ is not invertible.) Since $\mathrm{tr}\,(XX' + YY')^{-1}XX' = c$ with probability zero for each θ, we conclude that, for each $c \geqq 0$, the critical region

$$\mathrm{tr}\,(XX' + YY')^{-1}XX' \geqq c \qquad (4.3)$$

is an admissible Bayes procedure. It is fully invariant, similar, and (as a consequence of the results of Das Gupta, Anderson, and Mudholkar (1964)) unbiased.

We now give an indication of some of the many modifications in the Π_i which, as described in Section 2, still yield (4.3), and also list a few modifications which yield other tests. It will be obvious that some of these modifications can be combined. Again, Lemma 3.1 applies in all cases.

(ii) In place of I_p in (4.1) we can put any positive definite symmetric $p \times p$ matrix B and write $\Sigma^{-1} = B + \eta\eta'$, and in place of $\xi = \Sigma\eta$ we can put $\xi = b\Sigma\eta$ for any nonzero scalar b, multiplying the exponent in (4.1) by b^2. We still obtain (4.3). This means that, for each such B and b, there is a Π relative to which (4.3) is Bayes, and such that Π assigns all probability to a set of (Σ, ξ) for which Σ is smaller than B^{-1} in the sense that $B^{-1} - \Sigma$ is nonnegative definite, and for which $\mathrm{tr}\,\Sigma^{-1}\xi\xi' < pb^2$. Thus, the test (4.3) has good performance "near" H_0, in agreement with the local minimax character and in contrast with the "distant" goodness (obtained by method (2) of Section 2), both in Schwartz (1964a). We note that, for fixed b, the Π_1's corresponding to different B's assign all measure to disjoint sets of (ξ, Σ).

(iii) Letting $\Sigma^{-1} = I_p + \eta\eta'$ as before, suppose we now let $\xi = \Sigma\eta\beta$ under H_1, where β is $r \times r$. Let k be a fixed number satisfying $0 < k < 1$. Under H_1 the conditional density of β, given η, is $\phi_{r,r}(\beta; 0, [k^{-1}I_r - \eta'(I_p + \eta\eta')^{-1}\eta]^{-1})$. The marginal density of η under H_1 is

$$c_7|I_p + \eta\eta'|^{-(r+n)/2}|k^{-1}I_r - \eta'(I_p + \eta\eta')^{-1}\eta|^{-r/2}, \qquad (4.4)$$

which is again integrable, since the second determinant of (4.4) is bounded away from zero. Π_0 is again given by the second line of (4.1). The product of $f_{X,Y}$ with the density of β and η under H_1 is then

$$(4.5)\quad c_8 \operatorname{etr}\{-\tfrac{1}{2}(xx' + yy')\} \operatorname{etr}\{-\tfrac{1}{2}(yy' + (1-k)xx')\eta\eta'\} \cdot \operatorname{etr}\{-\tfrac{1}{2}k^{-1}(\beta - kx'\eta)(\beta - kx'\eta)'\}.$$

The integration of the last factor of (4.5) with respect to β yields a constant, and comparing the integration with respect to η of the middle factor with the corresponding integration (with $k = 0$) under Π_0, we obtain the critical region

$$(4.6)\qquad |YY' + XX'|/|YY' + (1-k)XX'| \geqq c$$

as an admissible Bayes test for $0 < k < 1$. For $k = 1$, one obtains the likelihood ratio test

$$(4.7)\qquad |YY' + XX'|/|YY'| \geqq c,$$

but one change is needed in the previous derivation: the integrability of (4.4), which by (3.8) equals $c_7|I_p + \eta\eta'|^{-n/2}$ when $k = 1$, is now assured, according to (3.7), if and only if $n > p + r - 1$. Thus, only under this restriction does our method show that the test (4.7) is admissible Bayes, although the admissibility without this restriction can be proved by method (2) of Section 2. (See Schwartz (1964b).) It is not known whether or not (4.7) is Bayes when $n < p + r$, except when $r = 1$, when the treatment of part (i) applies.

(iv) Without giving any details, we list a few of the many other examples of *a priori* distributions with respect to which the Bayes procedure can be computed and is fully invariant:

(a) For $0 < k_1 < 1$ and $k_2 > 0$, modify (iii) by letting $\xi = \Sigma\eta(k_2I_r + \beta)$ under H_1, the conditional density of the $r \times r$ matrix β, given η, being

$$g(\eta) \operatorname{etr}\{-\tfrac{1}{2}[k_1^{-1}\beta\beta' - \eta'(I_p + \eta\eta')^{-1}\eta(k_2I_r + \beta)(k_2I_r + \beta')]\},$$

where

$$g(\eta) = |k_1^{-1}I_r - \eta'(I_p + \eta\eta')^{-1}\eta|^{r/2} \operatorname{etr}\{\tfrac{1}{2}k_2^2\eta'(I_p + \eta\eta')^{-1}\eta \cdot [-I_r + (k_1^{-1}I_r - \eta'(I_p + \eta\eta')^{-1}\eta)^{-1}\eta'(I_p + \eta\eta')^{-1}\eta]\};$$

the marginal density of η under H_1 is now $c_8|I_p + \eta\eta'|^{-(r+n)/2}/g(\eta)$. The Bayes critical region is

$$(4.8)\quad \operatorname{etr}\{\tfrac{1}{2}k_2^2(YY' + (1-k_1)XX')^{-1}XX'\} \cdot |YY' + XX'|^{r/2}/|YY' + (1-k_1)XX'|^{r/2} \geqq c.$$

For $k_1 = 1$ we have the same modification as in (iii), so that in that case the approach only proves that (4.8) is Bayes if $n > p + r - 1$.

(b) If $r \geqq p$, alter $d\Pi_i(\eta)/d\eta (i = 0, 1)$ in (4.1) by multiplying it by $|\eta\eta'|^{t/2}$ where $p - r - 1 < t < n - p + 1$; these inequalities, needed for integrability of the altered $d\Pi_i(\eta)/d\eta$, come from an obvious modification of (3.7) and corresponding considerations near $|\eta\eta'| = 0$. The integrand in both the numerator and denominator of the third expression of (4.2) is multiplied by $|\eta\eta'|^{t/2}$. Using a result of Constantine (1963) (p. 1279), the resulting fully invariant critical region can be written as

$$(4.9)\qquad {}_1F_1((r+t)/2, r/2, \tfrac{1}{2}(XX'+YY')^{-1}XX') \geqq c,$$

where ${}_1F_1$ is the hypergeometric function of matrix argument (which is a polynomial multiplied by an exponential if $t/2$ is an integer; see Herz (1955)).

(c) If (i) is modified only by putting $\xi = \Sigma\eta B$ with a corresponding change in $d\Pi_1(\eta)/d\eta$ in (4.1), where B is a fixed $r \times r$ matrix, we obtain

$$(4.10)\qquad \mathrm{tr}\,(XX'+YY')^{-1}XB'BX' \geqq c,$$

which is not fully invariant (unless B is orthogonal). However, if we instead put $\xi = \Sigma\eta B\epsilon$ where B is again fixed and (4.1) is altered under H_1 by letting the $r \times r$ orthogonal matrix ϵ be uniformly distributed over the orthogonal group and independent of η, we obtain, according to James (1964) (formulas (25) and (30)), integrating first over η and then over ϵ, the fully invariant test

$$(4.11)\qquad {}_0F_0(\tfrac{1}{2}B'B, X'(XX'+YY')^{-1}X) \geqq c,$$

where ${}_0F_0$ is the hypergeometric function of two matrix arguments.

(v) We now consider the case $n < p$. If $n < p < n + r$, the test (4.3) (for example) is a nontrivial fully invariant test, but the Bayes approach of (i) fails; the admissibility in this case is still obtained by the method of Schwartz (1964b). When $n + r \leqq p$, the only fully invariant tests are the trivial fully randomized ones (which will be seen below to be inadmissible). However, there are reasonable admissible tests which are not fully invariant. For example, let $W = AV$ for any fixed nonsingular $p \times p$ matrix A, and let $W' = (W'_{(1)}, W'_{(2)})$ be a decomposition of W with $W_{(1)}$ having n rows. We now let both Π_i assign all their measure to those θ for which $W_{(1)}$ and $W_{(2)}$ are independent and the columns of $W_{(2)}$ have any specified distribution (for example, $W_{(2)}$ can have density $\phi_{p-n,r+n}(\,\cdot\,; 0, I_{p-n})$ under both H_i). The parameters of the distribution of $W_{(1)}$ have *a priori* densities on E^{nr} given by (4.1) with p replaced by n. The derivation proceeds as before to show that the test (4.3), with (X, Y) replaced by $W^{(1)}$, is admissible Bayes. Lemma 3.1 again extends this result to the case where $h > 0$.

We remark that, when, $n < p < n + r$ (which is only possible when $r > 1$), the test (4.3) can be shown from the results of Das Gupta, Anderson, and Mudholkar (1964) to have nontrivial mimimum power on the set $H_{1c'} = \{\mathrm{tr}\,\Sigma^{-1}\xi\xi' \geqq c'\}$, so that the trivial randomized test of the same size cannot be maximin on any fully invariant set contained in $H_{1c'}$ for some $c' > 0$. However, when $n + r \leqq p$ the trivial randomized test is maximin on such a set (although inadmissible if $n > 0$, as is shown by comparison with the test based on $W_{(1)}$); this follows from the fact that the maximal invariant under the group of lower triangular matrices does not depend on the last row of X (see Giri, Kiefer, and Stein (1963) for this type of computation).

The fully invariant tests for the problem of this section are well known to depend only on the nonzero latent roots t_i (say) of $(XX'+YY')^{-1}XX'$. The test (4.3), based on $\sum_i t_i$, has received much less attention than the likelihood ratio test (4.7) (based on $\prod_i (1-t_i)^{-1}$), Roy's test (based on $\max_i t_i$), or

Hotelling's T_0^2-test (based on $\sum_i t_i/(1 - t_i)$). All of these tests of course reduce to Hotelling's T^2-test when $r = 1$, to the (univariate) analysis of variance F-test when $p = 1$, and to Student's two-tailed t-test when $p = r = 1$. The test (4.3) was suggested by Pillai (1955), who has studied the distribution of the statistic under H_0, and it has also been studied by Schwartz (1964a), who proved its admissibility and certain other optimum properties.

5. Independence of sets of variates. Here $V = (Y, U)$ where under Ω the columns of V are independent with common unknown nonsingular covariance matrix Σ, Y is $p \times n$ with $EY = 0$, and U is $p \times h$ with $EU = \nu$ (unknown). Let $V' = (V'_{(1)}, V'_{(2)}, \dots, V'_{(k)})$ where $V_{(i)}$ has p_i rows and $\sum_1^k p_i = p$. Under H_0 the $V_{(i)}$ are independent, so that

$$\Sigma = \begin{pmatrix} \Sigma_{11} & 0 & 0 & \cdots & 0 \\ 0 & \Sigma_{22} & 0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 0 \\ 0 & 0 & \cdots & 0 & \Sigma_{kk} \end{pmatrix}$$

where Σ_{ii} is $p_i \times p_i$. The problem in this form usually arises (by means of an orthogonal transformation on the right) from that of observing $V = (V_1, V_2, \dots, V_m)$ where V_i is $p \times h$, $n = (m - 1)h$, $EV_i = \nu$ for all i, and V has independent columns, each with covariance matrix Σ. In any event, we can consider the case $h = 0$ and then obtain the general result from Lemma 3.1. We shall also assume $n \geqq p$; the results when $n < p$ are parallel to those discussed for the case $n + r \leqq p$ in the previous section, in the existence of admissible tests which are better than the trivial randomized test, which are based (for example) on only some of the rows of Y, and which (like all tests) have trivial minimax properties.

We let Π assign all measure to Σ^{-1}'s of the form $I_p + \eta\eta'$ under H_1, where η is $p \times 1$, and to Σ^{-1}'s of the form

$$\begin{pmatrix} I_{p_1} + \eta_{(1)}\eta'_{(1)} & 0 & \cdots & 0 \\ 0 & I_{p_2} + \eta_{(2)}\eta'_{(2)} & \cdots & \vdots \\ \vdots & 0 & \vdots & 0 \\ 0 & \cdots & 0 & I_{p_k} + \eta_{(k)}\eta'_{(k)} \end{pmatrix}$$

under H_0, where $\eta_{(i)}$ is $p_i \times 1$. We set

$$(5.1)\qquad \begin{aligned} d\Pi_1(\eta)/d\eta &= |I_p + \eta\eta'|^{-n/2}, \\ d\Pi_0(\eta)/d\eta &= \textstyle\prod_{i=0}^k |I_{p_i} + \eta_{(i)}\eta'_{(i)}|^{-n/2}. \end{aligned}$$

According to (3.7), the densities (5.1) are integrable on E^p provided $n > p$. We obtain, under H_0,

$$(5.2)\quad f_Y(y, \Sigma)\, d\Pi_0(\eta)/d\eta = c_1 \operatorname{etr}\{-\tfrac{1}{2}yy'\} \exp\{-\tfrac{1}{2}\textstyle\sum_i \eta'_{(i)}y_{(i)}y'_{(i)}\eta_{(i)}\},$$

and, under H_1,

$$(5.3)\qquad f_Y(y; \Sigma)\, d\Pi_1(\eta)/d\eta = c_2 \operatorname{etr}\{-\tfrac{1}{2}yy'\} \exp\{-\tfrac{1}{2}\eta'yy'\eta\}.$$

Hence, from (3.1), we obtain that, for $c \geqq 1$ and $n > p$,

$$\prod_{i=1}^{k} |Y_{(i)}Y'_{(i)}|/|YY'| \geqq c \tag{5.4}$$

is an admissible Bayes critical region.

The derivation of the test (5.4) required $n > p$, and thus that approach does not handle the "minimum sample size". In the special case $k = 2$, $p_1 = 1$, a slightly different trick, used by Lehmann and Stein (1948), will work even when $n = p$. Let Π assign all measure to Σ^{-1}'s of the form

$$I_p + \begin{pmatrix} 1 & \eta' \\ \eta & \eta\eta' \end{pmatrix} \quad \text{under} \quad H_1,$$

where η is $(p - 1) \times 1$, and to Σ^{-1}'s of the form

$$I_p + \begin{pmatrix} 1 - b & 0 \\ 0 & \eta\eta' \end{pmatrix} \quad \text{under} \quad H_0,$$

where η again is $(p - 1) \times 1$ and $0 \leqq b \leqq 1$.

We set

$$\frac{d\,\Pi_1(\eta)}{d\eta} = \left| I_p + \begin{pmatrix} 1 & \eta' \\ \eta & \eta\eta' \end{pmatrix} \right|^{-p/2},$$

$$\frac{d\,\Pi_0(\eta)}{d\eta} = \left| I_p + \begin{pmatrix} 1 - b & 0 \\ 0 & \eta\eta' \end{pmatrix} \right|^{-p/2},$$

which are integrable on E^{p-1}. Consider the particular Bayes test which rejects if

$$\int f_Y(y; \Sigma)\Pi_1(d\eta)/\int f_Y(y; \Sigma)\Pi_0(d\eta) \geqq 1. \tag{5.5}$$

Carrying out the integrations according to (3.1) yields the rejection region

$$\exp\left(\tfrac{1}{2}Y_{(1)}Y'_{(2)}(Y_{(2)}Y'_{(2)})^{-1}Y_{(2)}Y'_{(1)}\right)/\exp\left(\tfrac{1}{2}bY_{(1)}Y'_{(1)}\right) \geqq 1. \tag{5.6}$$

Taking logarithms of both sides, we finally get the rejection region

$$Y_{(1)}Y'_{(2)}(Y_{(2)}Y'_{(2)})^{-1}Y_{(2)}Y'_{(1)}/Y_{(1)}Y'_{(1)} \geqq b, \tag{5.7}$$

which in this special case is equivalent to (5.4).

As in Section 4, an infinite-dimensional set of Π's will yield (5.4), and other fully invariant tests can be obtained. (See also the second paragraph of Section 7(i).) The test (5.4) is the likelihood ratio test. For $k = 2$, $p_1 = 1$, it is the R^2-(multiple correlation coefficient-) test, which when $p = 2$ reduces to the classical two-sided test based on the sample correlation coefficient.

The technique of Section 6(ii) below can be applied to yield tests concerning the independence of subsets of the components which involve a total of fewer than p of the components.

6. Other problems of testing means.

(i) *Generalized Behrens-Fisher problem.* Let $V = (V^{(1)}, V^{(2)}, \cdots, V^{(k)})$ where

$V^{(i)}$ is $p \times (mn_i)$. Under Ω the n_i submatrices $V_1^{(i)}, V_2^{(i)}, \cdots, V_{n_i}^{(i)}$ of $V^{(i)}$, each of size $p \times m$, are identically distributed, with $EV_t^{(i)} = \xi^{(i)}$ and with each column of $V^{(i)}$ having covariance matrix $\Sigma^{(i)}$. The problem is to test H_0: $\xi^{(1)} = \xi^{(2)} = \cdots = \xi^{(q)}$. One of the ways of treating this problem is to reduce it to that of Section 4 by considering Π's which assign all measure to a set where $a_1\Sigma^{(1)} = a_2\Sigma^{(2)} = \cdots = a_q\Sigma^{(q)}$, where the positive numbers a_i (or, equivalently, their ratios) are specified, and where also $\Sigma^{(q+1)} = \cdots = \Sigma^{(k)} = I_p$, $\xi^{(q+1)} = \cdots = \xi^{(k)} = 0$. Writing

$$
\begin{aligned}
S &= \textstyle\sum_{i=1}^{q} \sum_{t=1}^{n_i} a_i(V_t^{(i)} - \bar{V}^{(i)})(V_t^{(i)} - \bar{V}^{(i)})', \\
\tilde{V} &= (\textstyle\sum_1^q n_i a_i)^{-1} \sum_1^q n_i a_i \bar{V}^{(i)}, \\
T &= \textstyle\sum_{i=1}^{q} n_i a_i(\bar{V}^{(i)} - \tilde{V})(\bar{V}^{(i)} - \tilde{V})', \\
U &= (\textstyle\sum_1^q n_i a_i)^{\frac{1}{2}}\tilde{V},
\end{aligned}
\tag{6.1}
$$

and reducing the problem to the canonical form of Section 4, we can then use all of the results of Section 4 with XX', YY', r, and n replaced, respectively, by T, S, $(q-1)m$, and $m\sum_1^q (n_i - 1)$. Thus, for example, if $m\sum_1^q (n_i - 1) \geqq p$, the critical region

$$\operatorname{tr}(S+T)^{-1}T \geqq c \tag{6.2}$$

is an admissible Bayes procedure. In the case $k = q = 2$, $m = p = 1$, we obtain Student's test if we set all a_i equal, and Welch's (simplest) test if a_i is proportional to $1/n_i(n_i - 1)$. These choices thus yield corresponding generalizations for general m and p. (However, these admissibility results have limited interest because of the lack of similarity.)

Admissible tests of equality of a subset of the components of the $\xi^{(i)}$ can be obtained, in the manner of (ii) below, by basing a test such as (6.2) only on these components.

(ii) *Tests concerning a subset of components of* ξ. If in the first paragraph of Section 4 we began with the more general form $EV = L_1\xi L_2$, with $K_1\xi K_2 = 0$ under H_0, we would have in the canonical form, in place of (X, Y), a random matrix with an expectation matrix some of whose elements (or linear combinations thereof) are 0 under Ω, and some additional ones of which are 0 under H_0. Bayes procedures for this problem can be found using modifications of the methods of Section 4.

Since the tests obtained in the most general case are less simple than those obtained in other cases, we shall for the sake of brevity mention here only the following special case: $V = (X, Y, U)$ with the assumptions of Section 4 except that $\xi' = (\xi'_{(1)}, \xi'_{(2)}, 0)$, where $\xi_{(i)}$ is $p_i \times r$ and $p_1 + p_2 \leqq p$ (here p_1 may be 0). Write $V' = (V'_{(1)}, V'_{(2)}, V'_{(3)})$, in the same way. Under H_0, $\xi_{(2)} = 0$. Here we shall give an example of a reasonable class of admissible tests which are similar but not fully invariant except when $p_3 = 0$ (i.e., under

$$X \to \begin{pmatrix} ABC \\ 0DE \\ 00F \end{pmatrix} X + \begin{pmatrix} G \\ 0 \\ 0 \end{pmatrix}$$

where $|A||D||F| \neq 0$), but which are intuitively appealing and, as described in the next paragraph, have further justification. Let Π assign all measure to the set where the submatrices $V_{(1)}$, $V_{(2)}$, and $V_{(3)}$ are independent, and where the columns of $V_{(1)}$ and $V_{(3)}$ have zero means and identity covariance matrices. All results of Section 4 then apply, with p_2, $X_{(2)}$, $Y_{(2)}$, $\xi_{(2)}$, $\nu_{(2)}$, Σ_{22} replacing p, X, Y, ξ, ν, Σ.

We remark that any minimax or local minimax property (on a set described in terms of ξ and Σ) of a test of Section 4 also holds (in terms of $\xi_{(2)}$ and Σ_{22}) for the corresponding test based on $V_{(2)}$ in our present setting; this follows from the validity of such a minimax property on the subset of Ω where $V_{(3)}$ and $V_{(1)}$ have zero means and identity covariance matrices and are independent of $V_{(2)}$, together with the fact that the power of the test on Ω depends only on $\xi_{(2)}$ and Σ_{22}. We note that a property described in terms of $\xi_{(2)}$ and Σ_{22} (unlike one described in terms of the eigenvalues of $(\Sigma_{22} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32})^{-1}\xi_{(2)}\xi_{(2)}'$ is not fully invariant except when $p_3 = 0$, which is why corresponding minimax procedures which are fully invariant need not exist even in the case $p_1 = 0$, $p_2 = p_3 = 1$. The rationale behind consideration of a smaller group (with $C = 0$, $E = 0$, with respect to which appropriate tests based on $V_{(2)}$ are invariant) was discussed in Section 2. Cochran and Bliss, Stein, and Olkin and Shrikhande have considered some of the problems of this subsection.

(iii) *Testing equality of components; Scheffé tests.* As another example of the general problem outlined in the first paragraph of (ii) above, suppose $\xi' = (\xi_{(1)}', \xi_{(2)}', \xi_{(3)}', \cdots, \xi_{(k)}', \xi_{(k+1)}', 0)$ where $\xi_{(i)}$ is $p_0 \times m$ for $1 \leq i \leq k$. Under H_0, $\xi_{(1)} = \xi_{(2)} = \cdots = \xi_{(k)}$. By letting Π assign all measure to the set where the $k + 2$ submatrices of V (corresponding to the above subdivision of ξ) are independent, with the last two having 0 means and identity covariance matrices, and with $\Sigma_{11} = \Sigma_{22} = \cdots = \Sigma_{kk}$ (unknown), the treatment of this problem can be reduced to that of the Behrens-Fisher problem of (i) above. The resulting tests are of course not invariant under such transformations as $X_{(i)} \to X_{(i)} + \sum_{j=1}^{k+2} A^{(j)}X_{(j)} (1 \leq i \leq k)$, where $A^{(j)}$ is $p_0 \times p_0$ and the transformation is nonsingular; these leave the problem invariant.

Somewhat different tests can be obtained in this last problem by the following device: For simplicity, suppose $k = 2$, and dispose of the last two submatrices of V as above. (The test will again not be fully invariant.) The first two submatrices $V_{(1)}$ and $V_{(2)}$ are, however, independent with probability zero under Π; and Π assigns all measure to the set where $\Sigma_{11} = \Sigma_{22} = I_{p_0} - \Sigma_{12}$ and $\xi_{(1)} = -\xi_{(2)}$ (which last includes $\xi_{(1)} = \xi_{(2)} = 0$ under Π_0). Write $V_{(1)}^* = V_{(1)} - V_{(2)}$ and $V_{(2)}^* = V_{(1)} + V_{(2)}$. Then Π assigns all measure to the set where the parameters of $V^{*\prime} = (V_{(1)}^{*\prime}, V_{(2)}^{*\prime})$ are as follows: $\xi_{(1)}^*$ arbitrary under H_1 and 0 under H_0; $\xi_{(2)}^* = 0$, $\Sigma_{22}^* = 2I_{p_0}$, $\Sigma_{12}^* = 0$, Σ_{11}^* arbitrary positive definite under both H_i. We can now use the techniques of Section 4, just as in part (ii) above. We conclude, for example, that if $n \geq p_0$ the critical region

$$\text{tr}\,[(Y_{(1)} - Y_{(2)})(Y_{(1)} - Y_{(2)})' + (X_{(1)} - X_{(2)})(X_{(1)} - X_{(2)})']^{-1} \cdot (X_{(1)} - X_{(2)})(X_{(1)} - X_{(2)})' \geq c \tag{6.3}$$

is an admissible similar Bayes procedure. When $p_0 = r = 1$, this is Scheffé's test for the Behrens-Fisher problem (in canonical form); however, in that context it is usually assumed that $\Sigma_{12} = 0$, which is not (and cannot be) assumed in the above development.

(iv) *Tests concerning a subset of components of* $\Sigma^{-1}\xi$. This problem, most recently considered by Giri, (1964), (1965), arises from discriminant analysis. The example of (ii) above is modified by setting $(\Sigma^{-1}\xi)' = \Gamma' = (\Gamma'_{(1)}, \Gamma'_{(2)}, 0)$, with $\Gamma_{(2)} = 0$ under H_0. This problem can be treated by modifications of the methods used above. For example, as in (ii) we can obtain tests by letting Π assign all measure to the set where $V_{(2)}$ is independent of both $V_{(1)}$ and $V_{(3)}$. As in the case of (ii), one can obtain many reasonable tests which are not fully invariant (i.e., under

$$X \rightarrow \begin{pmatrix} A00 \\ BC0 \\ DEF \end{pmatrix} X + \begin{pmatrix} G \\ 0 \\ 0 \end{pmatrix}$$

where $|A||C||F| \neq 0$).

(v) *Classification.* Suppose $V = (V^{(1)}, V^{(2)}, V^{(3)})$ where $V^{(j)} = (V_1^{(j)}, \cdots, V_{m_j}^{(j)})$, each $V_t^{(j)}$ being $p \times r$, the columns of V being independent with common unknown covariance matrix Σ, and $EV_t^{(j)} = \xi^{(j)}(p \times r)$. It is desired to test $H_1 : \xi^{(3)} = \xi^{(1)}$ against $H_2 : \xi^{(3)} = \xi^{(2)}$; that is, a sample of size m_j from population $j (j = 1, 2, 3)$ is to be used to classify "population 3" as either "population 1" or "population 2". Let $m = m_1 + m_2 + m_3$. As an example, we shall prove the admissibility of the likelihood ratio criterion, analogous to (4.7). Write $(m - 2)^{-1}r^{-1}S^{(j)}$ for the usual best unbiased estimator of Σ under H_j, and write

$$\begin{aligned} Y^{(j)} &= (m_j + m_3)^{-\frac{1}{2}}(m_j\bar{V}^{(j)} + m_3\bar{V}^{(3)}), \\ Z^{(j)} &= m_{3-j}^{\frac{1}{2}}\bar{V}^{(3-j)}, \\ U^{(j)} &= (Y^{(j)}, Z^{(j)}), EU^{(j)} = \nu^{(j)}. \end{aligned}$$

Let $Q^{(j)}$ be any orthogonal $mr \times mr$ matrix such that $VQ^{(j)} = (W^{(j)}, U^{(j)})$ where $W^{(j)}$ is $p \times (m - 2)r$. We now consider our problem in terms of $W^{(j)}, U^{(j)}$. Under H_j the $U^{(j)}$ corresponds to the nuisance parameter $\nu^{(j)}$, which we dispose of in the manner of (3.13) with $C^{(j)} = I_p$ (again writing $\Sigma^{-1} = I_p + \eta\eta'$ with $\eta\; p \times 1$); although Lemma 3.1 does not apply directly (because the nuisance variable $U^{(j)}$ differs under the two hypotheses), we conclude from (3.13) that an admissible Bayes procedure for our problem can be obtained in the form: select H_1 or H_2 according to whether

$$(6.5) \qquad \frac{\operatorname{etr}\{-\frac{1}{2}u^{(1)}u^{(1)\prime}\} \int_{E^p} f_{W^{(1)}}(w^{(1)}; (I_p + \eta\eta')^{-1})\, d\Pi_1(\eta)}{\operatorname{etr}\{-\frac{1}{2}u^{(2)}u^{(2)\prime}\} \int_{E^p} f_{W^{(2)}}(w^{(2)}; (I_p + \eta\eta')^{-1})\, d\Pi_2(\eta)} > \text{ or } < c.$$

Under both H_1 and H_2, we let $d\Pi_i(\eta)/d\eta = |I_p + \eta\eta'|^{-(m-2)r/2}$, which according to (3.7) is integrable provided $(m - 2)r > p$. Using (3.1) and the fact that

$W^{(j)}W^{(j)\prime} = S^{(j)}$ and $U^{(1)}U^{(1)\prime} + W^{(1)}W^{(1)\prime} = U^{(2)}U^{(2)\prime} + W^{(2)}W^{(2)\prime}$, we obtain

$$(6.6) \qquad |S^{(2)}|/|S^{(1)}| > \text{ or } < c$$

as an admissible Bayes rule for classifying population 3 as being the same as 1 or 2 (respectively), provides $(m - 2)r > p$. This procedure is fully invariant, under all transformations of the form $V_i^{(j)} \to AV_i^{(j)} + B$ with A (nonsingular) and B independent of i and j. It is the likelihood ratio criterion (Anderson (1958), pp. 140–141).

This test enjoys a kind of similarity, in that it has constant power on the set where $\Sigma^{-1}(\xi^{(1)} - \xi^{(2)})(\xi^{(1)} - \xi^{(2)})'$ has specified eigenvalues; in particular, the power of the test is constant over the set where $\xi^{(1)} = \xi^{(2)}$.

Other admissible procedures can be obtained by modifications of the type considered earlier. In addition, admissible procedures can be obtained for classification into one of $k(>2)$ populations: since the left side of (6.6) is proportional to the ratio of two *a posteriori* probabilities, the analogue of (6.6) is, in an obvious notation, to choose the classification j which minimizes $c_j|S^{(j)}|$.

The test (6.6) is of course also admissible for the problem where H_1 and H_2 are enlarged so as not to assume $\Sigma^{(1)} = \Sigma^{(2)}$, H_i being $\xi^{(3)} = \xi^{(i)}$, $\Sigma^{(3)} = \Sigma^{(i)}$. However, certain additional tests which may seem more appropriate can be obtained similarly in this case. For example, a fully invariant test can be obtained by putting $\Sigma^{(i)} = (I_p + \eta_i\eta_i')^{-1}$ under H_i with η_i $p \times 1$, and with

$$d\Pi_i(\eta)/d\eta = |I_p + \eta_i\eta_i'|^{(m_i+m_3-1)r/2}|I_p + \eta_{3-i}\eta_{3-i}'|^{(m_3-i-1)r/2};$$

there are integrable if $(m_i - 1)r > p$ for $i = 1, 2$. Writing $T^{(i)} = V^{(i)}V^{(i)\prime} - m_i\bar{V}^{(i)}\bar{V}^{(i)\prime}$, an analysis similar to that used to obtain (6.6) yield the procedure

$$(6.7) \qquad |S^{(2)} - T^{(1)}||T^{(1)}|/|S^{(1)} - T^{(2)}||T^{(2)}| > \text{ or } < c.$$

A classification problem with known covariance matrices was considered by Ellison (1962). See also Das Gupta (1965).

(vi) *Testing between two possible values of the mean.* In Section 4, suppose (in the canonical form) that H_0 is $\xi = \xi^{(00)}$, while H_1 is $\xi = \xi^{(01)}$. Letting $V^{(i)} = (X - \xi^{(i)}, Y, Z)$, we can use any of the Π_0's of Section 4 on $V^{(i)}$ under the present H_i. For example, comparing the denominator of the second expression of (4.2) for $i = 0, 1$ when x is replaced by $x - \xi^{(i)}$, we obtain

$$(6.8) \quad \operatorname{etr}\{\tfrac{1}{2}(\xi^{(01)} - \xi^{(00)})X'\}|(X - \xi^{(00)})(X - \xi^{(00)})' + YY'|^{r/2}/ |(X - \xi^{(01)})(X - \xi^{(01)})' + YY'|^{r/2} \geqq c$$

as an admissible Bayes critical region. Other forms can be obtained similarly.

Modifications in these H_i can be made along the lines of problem formulations considered earlier in this section; for example, the hypotheses can specify values of only a subset of the elements of ξ, and the subset can differ under the two H_i.

(vii) *Lehmann-Stein examples.* Many of the considerations of Lehmann and Stein (1948) have obvious multivariate analogues, but of course one no longer obtains uniformly most powerful one-sided tests (as in equation (6.9) of Lehmann and Stein). For example, in the setting of Section 4 with $h = 0$, to test $H_0 : \xi = 0$ against $H_1 : \xi = \xi^{(0)}, \Sigma = \Sigma^{(0)}$ (simple alternative), write $V^* = \Sigma^{(0)-\frac{1}{2}}V$ and $\xi^* = \Sigma^{(0)-\frac{1}{2}}\xi^{(0)}$, to reduce H_1 to $H_1^* : \xi = \xi^*, \Sigma = I_p$. In this form the problem is considered on pages 509–510 of Lehmann and Stein.

7. Tests concerning the covariance matrix.

(i) *Equality of covariance matrices.* This usually arises from $V^* = (V^{*(1)}, V^{*(2)}, \cdots, V^{*(k)})$ where the columns of each $V^{*(j)}$ are identically distributed. (The reduced form obtained below also applies if several $V^{*(j)}$ are assumed to have equal $\Sigma^{(j)}$'s or, equivalently, if $V^{*(j)}$ has identically distributed $p \times d_j$ submatrices with d_j no longer necessarily 1.) After an orthogonal transformation on the right of each $V^{*(j)}$, the problem is reduced to one where, under Ω, $V = (V^{(1)}, V^{(2)}, \cdots, V^{(k)}, W)$ with $EV^{(i)} = 0, EV_1^{(i)}V_1^{(i)\prime} = \Sigma^{(i)}$ (unknown), $EW = \nu$ with all elements unknown and unrelated. Here $V^{(i)}$ is $p \times n_i$. H_0 is $\Sigma^{(1)} = \Sigma^{(2)} = \cdots = \Sigma^{(k)}$ (the common value being unspecified). As in previous sections, we may suppose W absent, since Lemma 3.1 then handles the case where W is present; in the application of Lemma 3.1 in parts (i) and (ii) of this section, a given column of W will sometimes have a covariance matrix of different form under H_0 and H_1, but the $C^{(j)}$ of (3.9) can always be taken to be a scalar multiple of I_p. Let $q_i(0 \leqq i \leqq k)$ be positive integers. Let Π_0 assign all measure to the set where $\Sigma^{(1)} = \cdots = \Sigma^{(k)} = (I_p + \eta\eta')^{-1}$ where η is $p \times q_0$, and let Π_1 assign all measure where $\Sigma^{(i)} = (I_p + \eta_i\eta_i')^{-1}$ where η_i is $p \times q_i$. Furthermore, put $N = \sum_1^k n_i$ and

$$d\Pi_0(\eta)/d\eta = |I_p + \eta\eta'|^{-N/2} \tag{7.1}$$

$$d\Pi_1(\eta)/d\eta = \textstyle\prod_{i=1}^k |I_p + \eta_i\eta_i'|^{-n_i/2}.$$

According to (3.7), if $q_0 \leqq N - p$ and $q_i \leqq n_i - p$ for $1 \leqq i \leqq k$, the densities of (7.1) are integrable. In this case an integration of the type we have performed repeatedly yields, writing $S^{(i)} = V^{(i)}V^{(i)\prime}$,

$$|\textstyle\sum_{i=1}^k S^{(i)}|^{q_0}/\prod_{i=1}^k |S^{(i)}|^{q_i} \geqq c \tag{7.2}$$

as an admissible Bayes procedure, which is similar and fully invariant if $\sum_1^k q_i = q_0$. Such a test can be obtained for the simplest choice $q_i = 1$ (with $q_0 = k$) provided $n_i > p$ for $1 \leqq i \leqq k$. The likelihood ratio test (resp., Barlett's modification thereof) can be obtained in this way for some sets of values n_i, by setting $q_i = c_1(n_i + 1)$ (resp., $q_i = c_1 n_i$) and $q_0 = \sum_i q_i$, where $c_1 < 1$; the obvious choice $c_1 = 1$ does not make (7.1) integrable. This dependence on divisibility properties of the n_i can be overcome for sufficiently large n_i by using the analogue of modification (iv) (b) of Section 4, which we shall now describe.

Take all $q_i \geqq p$, and multiply $d\Pi_1(\eta)$ (resp., $d\Pi_0(\eta)$) in (7.1) by $\prod_1^k |\eta_i\eta_i'|^{t_i/2}$ (resp., by $|\eta\eta'|^{t_0/2}$) with $p - 1 < t_i + q_i < n_i - p + 1$ (resp., $p - 1 < t_0 + q_0 < N - p + 1$) for integrability. There exist such t_i provided $\min_i n_i > 2(p - 1)$, so that slightly larger sample sizes are needed than for small q_i in (7.1). We obtain the test of (7.2) with q_i replaced by $q_i + t_i$ for $0 \leqq i \leqq k$.

Setting $q_i + t_i = c_1(n_i + 1)$ (resp., $= c_1 n_i$) for $1 \leqq i \leqq k$ and $q_0 + t_0 = \sum_1^k (q_i + t_i)$, where c_1 is slightly larger than $(p - 1)/\min_i n_i$ (resp., $(p - 1)/\min (n_i + 1)$), we obtain the likelihood ratio test (resp., Bartlett's modification), provided $\min_i n_i > 2(p - 1)$. The use of this modification was pointed out to us by Professor Olkin (to whom we are also indebted for other helpful comments) in the equivalent form of replacing $|\eta\eta'|^{t/2}/|I + \eta\eta'|^{n/2}$ by the density $|\Lambda|^{(q-p+t-1)/2}/|I + \Lambda|^{n/2}$ on the positive definite matrices $\Lambda = \Sigma^{-1} - I$. This technique can also be applied in Section 5 where, however, it only exhibits a wider variety of *a priori* densities relative to which (5.4) is Bayes, and requires somewhat larger sample sizes. As we have seen in Section 4(iv)(b), the use of such densities there leads to a different test from that obtained for $t = 0$. Modifications of (7.2) and other forms of tests can be obtained by using other forms of Π, as in previous sections; in particular, EW can be treated in the manner of (iii) below instead of by means of Lemma 3.1.

As a one-sided variant of the above, suppose H_1 is altered to state that $\Sigma^{(i)} - \Sigma^{(i+1)}$ is nonnegative definite for $1 \leqq i \leqq k$, and not zero for all i. In that case $(\Sigma^{(i+1)})^{-1} - (\Sigma^{(i)})^{-1}$ is nonnegative definite, so that a possible choice of $d\Pi_1(\eta)/d\eta$ is $\prod_1^k |I_p + \sum_1^i \eta_t\eta_t'|^{-n_i/2}$ where η_t is $p \times q_t$ and $(\Sigma^{(i)})^{-1} = I_p + \sum_1^i \eta_t\eta_t'$ under Π_1. Using (3.5), a modification of the argument which led to (3.7) shows that this is integrable if $\sum_i^k n_t \geqq p + q_i$ for $1 \leqq i \leqq k$, and, with the Π_0 of (7.1), produces the admissible Bayes critical region

$$(7.3) \qquad |\textstyle\sum_{t=1}^k S^{(t)}|^{q_0} / \prod_{i=1}^k |\sum_{t=i}^k S^{(t)}|^{q_i} \geqq c,$$

which is fully invariant if $\sum_1^k q_i = q_0$. The technique of the previous paragraph can also be used to obtain (7.3) with nonintegral q_i. Among possible alternative forms is one obtained from (7.4) below.

The technique of Section 6(ii) can be applied to yield corresponding tests about a subset of the components.

(ii) *Proportionality of covariance matrices.* Suppose the setup of (i) is changed for $i = 0, 1$ to $H_i : a_{i1}\Sigma^{(1)} = a_{i2}\Sigma^{(2)} = \cdots = a_{ik}\Sigma^{(k)}$, where only the positive values a_{ij} are known. In this problem each H_i is acted upon transitively by the full linear group of nonsingular $p \times p$ matrices, as well as by the group of nonsingular lower triangular matrices. Hence, every procedure invariant under either group has constant power under each H_i, so that for each group there is a best invariant procedure of any specified size. Now, it can be checked directly, using the Neyman-Pearson lemma, that the essentially unique best triangular invariant test is not invariant under the full linear group (Lehmann (1959), p. 338 treats special cases of this fact), so that no fully invariant procedure can even be minimax, let alone admissible. Thus, we cannot hope to find any fully invariant admissible procedures for this problem. Whether the best triangular invariant procedure is admissible is unknown; we have been unsuccessful in showing that it is Bayes. Instead one can use our previous methods to find admissible procedures of simple structure which, however, are not triangular invariant.

As an example of the latter approach, we can, under H_i, let $a_{ij}\Sigma^{(j)} = (I_p + \eta\eta')^{-1}$ and use the Π_0 of (7.1) with q_{i0} replacing q_0. This yields

$$\text{(7.4)} \quad \operatorname{etr}\{\textstyle\sum_1^k a_{0i}S^{(i)}\}|\sum_1^k a_{0i}S^{(i)}|^{q_{00}}/\operatorname{etr}\{\sum_1^k a_{1i}S^{(i)}\}|\sum_1^k a_{1i}S^{(i)}|^{q_{10}} \geqq c$$

as an admissible Bayes test provided $q_{j0} \leqq N - p$ for $j = 0, 1$.

For a two-sided version of the above, suppose H_1 is altered to read: $b_{h1}\Sigma^{(1)} = b_{h2}\Sigma^{(2)} = \cdots = b_{hk}\Sigma^{(k)}$ for either $h = 1$ or 2, where the b_{hj} are specified positive constants. Letting $\Pi_1 = c_3\Pi_{11} + c_4\Pi_{12}$ where Π_{1h} is obtained from the one-sided Π_1 by replacing a_{1j} by b_{hj}, we obtain

$$\text{(7.5)} \quad \operatorname{etr}\{\textstyle\sum_{i=1}^k a_{0i}S^{(i)}\}|\sum_{i=1}^k a_{0i}S^{(i)}|^{q_{00}}\Big/ \sum_{h=1}^2 c_h \operatorname{etr}\{\sum_{i=1}^k b_{hi}S^{(i)}\}|\sum_{i=1}^k b_{hi}S^{(i)}|^{q_{h0}} \geqq 1$$

as an admissible Bayes test if $q_{h0} \leqq N - p$ for $h = 0, 1, 2$.

Nonintegral q's can be obtained in (7.4) and (7.5) as in the paragraph following (7.2).

The technique of Section 6(ii) can be applied to yield corresponding tests concerning only a subset of the components.

(iii) *Testing that covariance matrices have specified values.* With the V^* of (i), suppose $\Sigma_0^{(1)}, \Sigma_0^{(2)}, \cdots, \Sigma_0^{(k)}$ are specified positive definite matrices. We now consider $H_0 : \Sigma^{(i)} = \Sigma_0^{(i)} (1 \leqq i \leqq k)$. For the sake of brevity we shall detail only the case $k = 1$, the modifications required when $k > 1$ being straightforward. Thus, we consider $V = (V_1, V_2, \cdots, V_n, W)$, the vectors V_i and W having common covariance matrix Σ, with $EV_i = 0$ and $EW = \nu$. H_0 is $\Sigma = \Sigma_0$ (specified), and we shall consider various possible alternatives.

First suppose we also have specified two positive definite matrices Σ_{0L} and Σ_{0U} such that $\Sigma_0 - \Sigma_{0L}$ and $\Sigma_{0U} - \Sigma_0$ are both positive definite, and suppose that H_1 is that Σ is one of the pair Σ_{0L}, Σ_{0U} (or else that H_1 is: either $\Sigma_{0L} - \Sigma$ or $\Sigma - \Sigma_{0U}$ is nonnegative definite). In this case we can imitate Lehmann (1959), p. 332, as follows: Set $0 < c_1 < 1$, let ν_0 be a fixed vector, and let $\Pi_0, \Pi_{1a}, \Pi_{1d}$, and $\Pi_1 = c_1\Pi_{1a} + (1 - c_1)\Pi_{1d}$ be probability measures such that

$$\text{(7.6)} \qquad \Pi_0\{\Sigma = \Sigma_0\} = 1, \qquad d\Pi_0(\nu)/d\nu = \phi_{p,1}(\nu; \nu_0, \Sigma_{0U} - \Sigma_0);$$

$$\text{(7.7)} \qquad \begin{aligned} &\Pi_{1a}\{\Sigma = \Sigma_{0L}\} = 1, \qquad d\Pi_{1a}(\nu)/d\nu = \phi_{p,1}(\nu; \nu_0, \Sigma_{0U} - \Sigma_{0L});\\ &\Pi_{1d}\{\Sigma = \Sigma_{0U}, \nu = \nu_0\} = 1.\end{aligned}$$

Then, writing $S = \sum_1^n V_iV_i'$ and $G = \Sigma_{0U}^{-1}(W - \nu_0)(W - \nu_0)'$, we obtain the admissible Bayes critical region

$$\text{(7.8)} \qquad \frac{c_1 |\Sigma_{0L}|^{-n/2} |\Sigma_{0U}|^{-\frac{1}{2}} \operatorname{etr}\{-\frac{1}{2}\Sigma_{0L}^{-1}S - \frac{1}{2}G\} + (1 - c_1)|\Sigma_{0L}|^{-(n+1)/2} \cdot \operatorname{etr}\{-\frac{1}{2}\Sigma_{0U}^{-1}S - \frac{1}{2}G\}}{|\Sigma_0|^{-n/2}|\Sigma_{0U}|^{-\frac{1}{2}} \operatorname{etr}\{-\frac{1}{2}\Sigma_0^{-1}S - \frac{1}{2}G\}} \geqq c,$$

or, for arbitrary nonnegative c_3 and c_4 (not both zero),

$$\text{(7.9)} \qquad c_3 \operatorname{etr}\{(\Sigma_0^{-1} - \Sigma_{0L}^{-1})S\} + c_4 \operatorname{etr}\{(\Sigma_0^{-1} - \Sigma_{0U}^{-1})S\} \geqq 1.$$

If either $c_3 = 0$ or $c_4 = 0$ (i.e., $c_1 = 1$ or 0), we obtain a one-sided test. If there are positive numbers $a_L < 1 < a_U$ such that $\Sigma_{0L} = a_L\Sigma_0$ and $\Sigma_{0U} = a_U\Sigma_0$, the test (7.9) simplifies to

$$\text{(7.10)} \qquad \operatorname{tr} \Sigma_0^{-1}S \geqq c_5 \quad \text{or} \quad \leqq c_6,$$

where c_5 and c_6 are arbitrary ($c_5 = \infty$ or $c_6 = 0$ yielding one-sided tests); this is in the form which is familiar when $p = 1$. The test (7.10) is thus also admissible for $H_0 : \Sigma = \Sigma_0$ against $H_1 : \Sigma = a\Sigma_0$ for some $a \neq 1$. It is thus also admissible for $H_0 : \Sigma = \Sigma_0$ against $H_1 : \Sigma \neq \Sigma_0$. Recalling Corollary 3.2, we note that we have made only the minimal assumption $n \geqq 1$ here.

The techniques used earlier may also be applied to obtain other tests for the present problem. For example, to test $H_0 : \Sigma = \Sigma_0$ against $H_1 : \Sigma_0 - \Sigma$ is nonnegative definite, we can treat ν by Lemma 3.1 and, in terms of $V_1, \cdots, V_n$, let $\Sigma^{-1} = \Sigma_0^{-1} + \eta\eta'$ under H_1, with η $p \times 1$ and $d\Pi_1(\eta)/d\eta = |\Sigma_0^{-1} + \eta\eta'|^{-n/2}$. This is integrable if $n > p$, and in that case yields the admissible Bayes critical region

$$\text{(7.11)} \qquad |S| \leqq c.$$

The technique of Section 6(ii) can be applied to yield tests concerning the values of covariance matrices of subsets of the components.

(iv) *Testing symmetry or sphericity of the covariance matrix.* Suppose $V = (Y, W)$ where under Ω the independent columns of V have common covariance matrix Σ, where $EY = 0$ ($p \times n$), and where $EW = \nu$ (unknown). Here $p \geqq 3$. H_0 is the hypothesis of "symmetry", that Σ is of the form $\lambda I_p + \beta J_p$ for some (unknown) $\lambda > 0$ and β, where J_p is a $p \times p$ matrix of 1's. It will be seen that Π assigns all measure where $\beta \geqq 0$ and hence that W can be handled by using Lemma 3.1, so we treat the case where W is absent. Let $\Sigma^{-1} = I_p + \eta\eta'$ under H_1, with η $p \times 1$, and let $\Sigma^{-1} = (1 + a\eta^2)I_p + b\eta^2 J_p$ under H_0, where η is 1×1 and $a \geqq 0$, $b \geqq 0$, with not both of a, b equal to zero. Moreover, for $q > 0$ let

$$\text{(7.12)} \qquad \begin{aligned} d\Pi_0(\eta)/d\eta &= |(1 + a\eta^2)I_p + b\eta^2 J_p|^{-n/2}|\eta|^q, \\ d\Pi_1(\eta)/d\eta &= |I_p + \eta\eta'|^{-n/2}. \end{aligned}$$

These are integrable if $np - q > 1$ and $n > p$. It is easy to verify that Π_1 gives measure 0 to H_0. Using (3.1) and writing $S = YY'$ as before, we obtain

$$\text{(7.13)} \qquad |S|/[a \operatorname{tr} S + b\textstyle\sum_{i,j=1}^{p} S_{i,j}]^{(q+1)/2} \geqq c$$

as an admissible Bayes test. If $q = 2p - 1$, we obtain a scale-invariant test. This is of course really a test of the hypothesis that H_0 holds with correlation $-b/[a + (p - 1)b]$ between any two different components of V_1.

When β and $b = 0$ in the above development, (7.13) becomes the likelihood ratio test of "sphericity". The ratio of the denominator of (7.13) with $b = 0$ to the same expression with $b > 0$ gives an admissible test of sphericity assuming symmetry.

REFERENCES

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

BIRNBAUM, A. (1955). Characterization of complete classes of tests of some multiparametric hypotheses, with applications to likelihood ratio tests. *Ann. Math. Statist.* **26** 21–36.

BROWN, L. (1964). Admissibility of translation-invariant estimators. Submitted to *Ann. Math. Stat.*

CONSTANTINE, A. G. (1963). Some non-central distribution problems of multivariate linear hypothesis. *Ann. Math. Statist.* **34** 1270–1285.

DAS GUPTA, S. (1965). Optimum classification rules. To be published in *Ann. Math. Statist.*

DAS GUPTA, S., ANDERSON, T. W., and MUDHOLKAR, G. S. (1964). Monotonicity of the power functions of some tests of the multivariate linear hypothesis. *Ann. Math. Statist.* **35** 200–205.

ELLISON, B. E. (1962). A classification problem in which information about alternative distributions is based on samples. *Ann. Math. Statist.* **35** 213–223.

GIRI, N., KIEFER, J., and STEIN, C. (1963). Minimax character of Hotelling's T^2-test in the simplest case. *Ann. Math. Statist.* **34** 1524–1535.

GIRI, N. (1964). On the likelihood ratio test of a normal multivariate testing problem. *Ann. Math. Statist.* **35** 181–189.

GIRI, N. (1965). On the likelihood ratio test of a normal multivariate testing problem, II. *Ann. Math. Statist.* **36** 1061–1065.

GHOSH, M. N. (1964). On the admissibility of some tests of Manova. *Ann. Math. Statist.* **35** 789–794.

HERZ, C. S. (1955). Bessel functions of matrix argument. *Ann. Math.* **61** 474–523.

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

JEFFREYS, H. (1939). *Theory of Probability.* Oxford Univ. Press.

KARLIN, S. (1957). Polya type distributions, II. *Ann. Math. Statist.* **28** 281–308.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

LEHMANN, E. L. and STEIN, C. (1948). Most powerful tests of composite hypotheses, I. Normal distributions. *Ann. Math. Statist.* **19** 495–516.

LEHMANN, E. L. and STEIN, C. (1953). The admissibility of certain invariant statistical tests involving a translation parameter. *Ann. Math. Statist.* **24** 473–479.

NANDI, H. K. (1963). On the admissibility of a class of tests. *Calcutta Statist. Assoc. Bull.* **15** 13–18.

PILLAI, K. C. S. (1955). Some new test criteria in multivariate analysis. *Ann. Math. Statist.* **26** 117–121.

ROY, S. N. (1957). *Some Aspects of Multivariate Analysis.* Wiley, New York.

ROY, S. N. and MIKHAIL, W. F. (1960). On the admissibility of a class of tests in normal multivariate analysis (abstract). *Ann. Math. Statist.* **31** 536.

SCHWARTZ, R. (1964a). Properties of a test in Manova (abstract). *Ann. Math. Statist.* **35** 939.

SCHWARTZ, R. (1964b). Admissible invariant tests in Manova (abstract). *Ann. Math. Statist.* **35** 1398.

STEIN, C. (1956a). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 197–206. Univ. of California Press.

STEIN, C. (1956b). The admissibility of Hotelling's T^2-test. *Ann. Math. Statist.* **27** 616–623.

WALD, A. (1942). On the power function of the analysis of variance test. *Ann. Math. Statist.* **13** 434–439.