LOCAL AND ASYMPTOTIC MINIMAX PROPERTIES OF MULTIVARIATE TESTS

By N. Giri[1] and J. Kiefer[2]

Cornell and Stanford Universities

0. Summary. This paper contains details of the results announced in the abstract by the authors (1962). Techniques are developed for proving local minimax and "type D" properties and asymptotic (that is, far in distance from the null hypothesis) minimax properties in complex testing problems where exact minimax results seem difficult to obtain. The techniques are illustrated in the settings where Hotelling's T^2 test and the test based on the squared sample multiple correlation coefficient R^2 are customarily employed.

1. Introduction. In almost all of the standard hypothesis testing problems of multivariate analysis—in particular, in the normal ones—no meaningful non-asymptotic (in the sample size) optimum properties are known, either for the classical tests or for any other tests. The property of being a best invariant test under a group G of transformations which leave the problem invariant, which is possessed by some of these tests, is often unsatisfactory because the Hunt-Stein theorem is not valid; for example, this is the case if G is the real linear group of nonsingular $p \times p$ matrices where $p \geqq 2$. The only satisfactory properties known to us at this writing are the admissibility of Hotelling's T^2 test, proved by Stein (1956), and the minimax character in a few special cases of Hotelling's test, proved recently by the authors and Stein (1963), and of the test based on the multiple correlation coefficient, proved recently by the authors (1963).

The proof of local or asymptotic (far in distance from the null hypothesis) properties, for which we herein develop simple techniques, serves two purposes. Firstly, there is the obvious point of demonstrating such properties for their own sake. But well known and valid doubts have been raised as to the extent of meaningfulness of such properties. Secondly, then, and in our opinion more important, local or asymptotic properties can give an indication of what to look for in the way of genuine minimax or admissibility properties of certain procedures, even though the latter do not follow from the local or asymptotic properties. For example, if S_1 and S_2 are independent $p \times p$ central Wishart matrices ($p \geqq 2$) with expectations Σ and $\delta\Sigma$ per degree of freedom, and if it is desired to test $H_0:\delta = 1$ against $H_1:\delta = \lambda + 1$ (specified) > 1 or $H_1':\delta > 1$, then Stein showed (see Lehmann (1959), pp. 231, 338) that the best invariant test of level $\alpha(0 < \alpha < 1)$ under the real linear group G operating as $(S_1, S_2, \delta, \Sigma) \to (gS_1g', gS_2g', \delta, g\Sigma g')$, which is also the likelihood ratio test, is inadmissible and is not

Received 14 May 1963.

[1] Research supported by ONR Contract No. Nonr-401(03).

[2] Fellow of the John Simon Guggenheim Memorial Foundation; research supported in part by ONR Contract No. Nonr-266(04) (NR 047-005).

minimax for H_1. It was possible to obtain this result without much calculation because the best invariant procedure under the group G_T of nonsingular lower triangular matrices, which is transitive on $\{\Sigma\}$ and for which the Hunt-Stein theorem is valid, is not invariant under the real linear group. But in other examples such an expeditious demonstration may not be available; we note, therefore, that the result in the present example is already indicated in the local theory (as $\lambda \to 0$), which may thus be expected to indicate the direction of such results in other cases where (unlike the present case) the nonlocal theory is much more difficult. Such examples are in fact found in the T^2 and R^2 tests which are mentioned above and are treated in the paper: the local optimality of these tests, which will be seen in Section 2 not to be very difficult, was proved at a time when the T^2 and R^2 tests were not known to be minimax; such a simple local result, or its negation, lends credence to the genuine minimax property, or its negation, and thus indicates the direction of proof or disproof which seems most promising; although the R^2 and T^2 genuine minimax properties are known in only a few cases, one's belief is strengthened in the validity of these properties in general. When, as in the examples of Section 2, the local or asymptotic property is possessed by only one of the G_T-invariant procedures, it is of course more indicative than when, as in the example of Section 4, many tests share the property.

In the negative direction, the local or asymptotic theory not only indicates, but says something definite about the genuine minimax result: In the setting of Section 2 (resp., Section 4), if a level α test ϕ^* of $H_0:\delta = 0$ against $H_1:\delta = \lambda$ (specified) > 0 maximizes the minimum power under H_1 for *every* $\lambda > 0$, then it must clearly be locally minimax as $\lambda \to 0$ (resp., asymptotically minimax as $\lambda \to \infty$). Thus, the failure of ϕ^* to possess the local or asymptotic minimax property proves that it is not minimax (uniformly) for every λ.

One can refine the local and asymptotic notions we consider to versions obtained by including one or more error terms. For example, in Section 2 the local optimality property (2.4) can be refined to ask that a level α critical region which satisfies (2.4) with $\inf_\eta P_{\lambda,\eta}\{R\} = \alpha + C_1\lambda + o(\lambda)$ as $\lambda \to 0$, also satisfies

$$\lim_{\lambda\to 0} \frac{\inf_\eta P_{\lambda,\eta}\{R\} - \alpha - C_1\lambda}{\sup_{\phi\in Q_\alpha} \inf_\eta P_{\lambda,\eta}\{\phi \text{ rejects } H_0\} - \alpha - C_1\lambda} = 1.$$

This involves more calculations; typically, in the setting (2.3) of Section 2, two moments of the a priori distribution $\xi_{1,\lambda}$, rather than just one, become important. As further refinements are invoked, more moments are brought in. In the T^2 and R^2 cases mentioned earlier wherein genuine minimax results have been obtained, the limits of the corresponding a priori distributions as $\lambda \to 0$ are positive Lebesgue densities $f(\eta)$ whose moments are determined successively by further refinements. Thus, in those cases, the first moments of f coincide with those of $\xi_{1,\lambda}$ in Examples 1 and 2 of Section 2. A similar result for the first moment holds in Example 1 of Section 4 as $\lambda \to \infty$.

The local theory developed in Section 2 uses a slightly refined version of the well known result that Bayes procedures with constant risk are minimax. Such Bayesian techniques can be used also in the classical Neyman-Pearson local theory; for example, see Kiefer (1959), p. 280, for an application to regions of type C.

In Section 3 a variant of Isaacson's type D region is discussed. Somewhat surprisingly, the T^2 and R^2 tests are not of type D among G_T-invariant tests (except, of course in the lowest dimension). This fact can be interpreted, roughly, in terms of the classical tests having constant (maximin) power on a family of ellipsoids in a reduced parameter space, while certain other tests with other ellipsoids as local contours of the power function yield a greater Gaussian curvature for the power function. Unfortunately, the action of G_T in such problems destroys symmetry in the coordinates and makes it more difficult to achieve good intuition.

It is hoped that further, more difficult multivariate examples will be treated elsewhere.

The reader is referred to Lehmann (1959) and to Anderson (1958) for the standard nomenclature which we shall use.

2. Locally minimax tests. Let X be a space with associated σ-field which, along with the other obvious measurability considerations, we will not mention in what follows. For each point (δ, η) in the parameter set Ω (where $\delta \geqq 0$), suppose that $p(\cdot; \delta, \eta)$ is a probability density function on X with respect to some σ-finite measure μ. (The range of η may depend on δ.) For fixed α, $0 < \alpha < 1$, we shall be interested in testing, at level α, the hypothesis $H_0: \delta = 0$ against the alternative $H_1: \delta = \lambda$, where λ is a specified positive value, and in giving a sufficient condition for a test to be approximately minimax in the sense of (2.4) below. This is a local theory, in the sense that $p(x; \lambda, \eta)$ is close to $p(x; 0, \eta)$ when λ is small. Thus, obviously, every test of level α would be locally minimax in the sense of the trivial criterion obtained by not subtracting α in the numerator and denominator of (2.4). As indicated in the introduction, our proof of (2.4) as it stands consists merely of considering local power behavior with sufficient accuracy to obtain an approximate version of the classical result that a Bayes procedure with constant risk is minimax. A result of the type obtained can be proved under various possible sets of conditions, of which we use a form convenient in many applications, listing possible generalizations and simplifications as remarks.

Throughout this section, such expressions as $o(1)$, $o(h(\lambda))$, etc., are to be interpreted as $\lambda \to 0$.

For each fixed α, $0 < \alpha < 1$, we shall consider critical regions of the form $R = \{x: U(x) > C_\alpha\}$ where U is bounded and positive and has a continuous d.f. for each (δ, η), equicontinuous in (δ, η) for $\delta <$ some δ_o, and where

$$(2.1) \qquad P_{0,\eta}\{R\} = \alpha, \qquad P_{\lambda,\eta}\{R\} = \alpha + h(\lambda) + q(\lambda, \eta),$$

where $q(\lambda, \eta) = o(h(\lambda))$ uniformly in η, with $h(\lambda) > 0$ for $\lambda > 0$ and $h(\lambda) = o(1)$.

We shall also be concerned with probability measures $\xi_{0,\lambda}$ and $\xi_{1,\lambda}$ on the sets $\delta = 0$ and $\delta = \lambda$, respectively, for which

$$\int p(x; \lambda, \eta)\xi_{1,\lambda}(d\eta) \Big/ \int p(x; 0, \eta)\xi_{0,\lambda}(d\eta) = 1 + h(\lambda)[g(\lambda) + r(\lambda)U(x)] + B(x, \lambda) \tag{2.2}$$

where $0 < c_1 < r(\lambda) < c_2 < \infty$ for λ sufficiently small, and where $g(\lambda) = O(1)$ and $B(x, \lambda) = o(h(\lambda))$ uniformly in x. It is clear from the form of (2.1) and (2.2) that, reparametrizing, there is no loss of generality in letting $h(\delta) = \delta$, but we retain the stated forms for use in applications.

REMARKS.

1. In many applications the set $\{\delta = 0\}$ is a single point. Also, the set $\{\delta = \lambda\}$ is often a convex finite-dimensional Euclidean set wherein each component η_i is $0(h(\lambda))$; in this case, if $p(x; \lambda, \eta)/p(x; 0, \eta)$ is of the form

$$1 + h(\lambda)U(x) + \sum_{i,j=1}^{k} S_i(x)a_{ij}(\lambda)\eta_j + B(x, \lambda, \eta) \tag{2.3}$$

with S_i and a_{ij} bounded and with $\sup_{x,\eta} B(x, \lambda, \eta) = o(h(\lambda))$, and if there exists any $\xi'_{1,\lambda}$ satisfying (2.2), then the degenerate $\xi''_{1,\lambda}$ which assigns measure 1 to the mean of $\xi'_{1,\lambda}$ also satisfies (2.2). Both of these simplifications occur in Examples 1 and 2 below.

2. Another simplification occurs if the $\xi_{i,\lambda}$ can be chosen to be independent of λ, as is the case in Examples 1 and 2. The assumptions on B and U can then be weakened. (See Remark 3.)

3. One can weaken the assumptions on U and B (and, similarly, on the S_i and a_{ij} of Remark 1), which are used only in order to verify (2.8) and (2.9) below. For example, the assumption on $B(x, \lambda)$ can be weakened to $P_{\lambda,\eta}\{|B(x,\lambda)| < \epsilon\, h(\lambda)\} \to 0$ as $\lambda \to 0$, uniformly in η for each $\epsilon > 0$. If the $\xi_{i,\lambda}$'s are independent of λ, the uniformity of this last condition is unnecessary. The boundedness of U and the equicontinuity of its distribution can be weakened similarly.

4. The following modifications are also trivial to introduce: consideration of critical regions of a more complicated form than $\{U > C_\alpha\}$; consideration of randomized tests rather than critical regions R; modification (in the absence of continuity of the power function) of the equality signs in (2.1) to $\leqq$ and $\geqq$ signs, respectively, with equality on the support of the $\xi_{i,\lambda}$. The conclusion of Lemma 1 clearly holds even if Q_α is modified to include every family $\{\phi_\lambda\}$ of tests of level $\alpha + o(h(\lambda))$. One can similarly consider optimality of a family $\{U_\lambda\}$ rather than of a single U, by replacing R by $R_\lambda = \{x: U_\lambda(x) > C_{\alpha,\lambda}\}$, where $P_{0,\eta}\{R_\lambda\} = \alpha - q_\lambda(\eta)$ with $q_\lambda(\eta) = o(h(\lambda))$.

LEMMA 1. *If U satisfies* (2.1) *and if for sufficiently small λ there exist $\xi_{0,\lambda}$ and $\xi_{1,\lambda}$ satisfying* (2.2), *then U is locally minimax of level α for testing $H_0: \delta = 0$ against $\delta = \lambda$ as $\lambda \to 0$; that is,*

$$\lim_{\lambda\to 0} \frac{\inf_\eta P_{\lambda,\eta}\{R\} - \alpha}{\sup_{\phi_\lambda \epsilon Q_\alpha} \inf_\eta P_{\lambda,\eta}\{\phi_\lambda \text{ rejects } H_0\} - \alpha} = 1, \tag{2.4}$$

where Q_α is the class of tests of level α.

PROOF. Write

$$\tau_\lambda = 1/\{2 + h(\lambda)[g(\lambda) + C_\alpha r(\lambda)]\}, \tag{2.5}$$

so that

$$(1 - \tau_\lambda)/\tau_\lambda = 1 + h(\lambda)[g(\lambda) + C_\alpha r(\lambda)]. \tag{2.6}$$

A Bayes critical region relative to the a priori distribution $\xi_\lambda = (1 - \tau_\lambda)\xi_{0,\lambda} + \tau_\lambda \xi_{1,\lambda}$ (for 0-1 losses) is, by (2.2) and (2.6),

$$B_\lambda = \{x : U(x) + B(x, \lambda)/r(\lambda)h(\lambda) > C_\alpha\}. \tag{2.7}$$

Write

$$P^*_{0,\lambda}\{A\} = \int P_{0,\eta}\{A\}\xi_{0,\lambda}(d\eta) \quad \text{and} \quad P^*_{1,\lambda}\{A\} = \int P_{\lambda,\eta}\{A\}\xi_{1,\lambda}(d\eta).$$

Let $V_\lambda = R - B_\lambda$ and $W_\lambda = B_\lambda - R$. Using the fact that $\sup_x |B(x, \lambda)/h(\lambda)| = o(1)$ and our continuity assumption on the d.f. of U, we have

$$P^*_{0,\lambda}\{V_\lambda + W_\lambda\} = o(1). \tag{2.8}$$

Also, for $U_\lambda = V_\lambda$ or W_λ,

$$P^*_{1,\lambda}\{U_\lambda\} = P^*_{0,\lambda}\{U_\lambda\}[1 + O(h(\lambda))]. \tag{2.9}$$

Write $r^*_\lambda(A) = (1 - \tau_\lambda)P^*_{0,\lambda}\{A\} + \tau_\lambda(1 - P^*_{1,\lambda}\{A\})$. From (2.5), (2.8), and (2.9), the integrated Bayes risk relative to ξ_λ is then

$$\begin{aligned} r^*_\lambda(B_\lambda) &= r^*_\lambda(R) + (1 - \tau_\lambda)(P^*_{0,\lambda}\{W_\lambda\} - P^*_{0,\lambda}\{V_\lambda\}) \\ &\qquad + \tau_\lambda(P^*_{1,\lambda}\{V_\lambda\} - P^*_{1,\lambda}\{W_\lambda\}) \\ &= r^*_\lambda(R) + (1 - 2\tau_\lambda)(P^*_{0,\lambda}\{W_\lambda\} - P^*_{0,\lambda}\{V_\lambda\}) \\ &\qquad + P^*_{0,\lambda}\{V_\lambda + W_\lambda\}O(h(\lambda)) \\ &= r^*_\lambda(R) + o(h(\lambda)). \end{aligned} \tag{2.10}$$

If (2.4) were false we could, by (2.1), find a family of tests $\{\phi_\lambda\}$ of level α such that ϕ_λ has power function $\alpha + g(\lambda, \eta)$ on the set $\delta = \lambda$, with

$$\limsup_{\lambda \to 0}\, [\inf_\eta g(\lambda, \eta) - h(\lambda)]/h(\lambda) > 0.$$

The integrated risk r'_λ of ϕ_λ with respect to ξ_λ would then satisfy

$$\limsup_{\lambda \to 0}\, (r^*_\lambda(R) - r'_\lambda)/h(\lambda) > 0,$$

contradicting (2.10).

EXAMPLE 1. (Hotelling's T^2 test). Let $X_1, \cdots, X_N$ be independently and identically distributed normal p-vectors, each with mean vector ξ and nonsingular covariance matrix Σ. Write $N\bar{X} = \sum_1^N X_i$ and $S = \sum_1^N (X_i - \bar{X})(X_i - \bar{X})'$. Let $\delta > 0$ be specified. For testing the hypothesis $H_0: \xi = 0$ against $H_1: N\xi'\Sigma^{-1}\xi = \delta$ at significance level α $(0 < \alpha < 1)$, a commonly employed procedure is Hotelling's T^2 test, which rejects H_0 when $T^2 = N(N-1)\bar{X}'S^{-1}\bar{X} > C'$ or, equivalently, when $U = T^2/(T^2 + N - 1) > C$, where C (or C') is chosen so as to yield a test of level α. We assume $N > p$, since it is easily shown that the denominator of (2.4) is zero in the degenerate case $N \leqq p$.

In our search for a locally minimax test as $\delta \to 0$, we may restrict attention to the space of the minimal sufficient statistic $(\bar{X}, S)$. The full linear group G of $p \times p$ nonsingular matrices leaves the problem invariant, operating as $(\bar{X}, S; \xi, \Sigma) \to (g\bar{X}, gSg'; g\xi, g\Sigma g')$. However, the Hunt-Stein theorem cannot be applied for this group if $p \geqq 2$, as Stein has demonstrated in several examples. (See Stein (1955), Lehmann (1959), pp. 231 and 338, and James and Stein (1960), p. 376.) However, the theorem does apply for the smaller group G_T of nonsingular lower triangular matrices (zero above the diagonal), which is solvable. (See Kiefer (1957), Lehmann (1959), p. 345.) Thus, for each δ there is a level α test which is almost invariant (hence, in the present problem, there is such a test which is invariant; see Lehmann (1957), p. 225) under G_T and which maximizes, among *all* level α tests, the minimum power under H_1. In terms of the local point of view, the denominator in (2.4) is unchanged by the restriction to G_T-invariant tests, and for any level α test ϕ there is a G_T-invariant level α test ϕ' for which the expression $\inf_\eta P_{\lambda,\eta}\{\phi'$ rejects $H_0\}$ is at least as large, so that a procedure which is locally minimax among G_T-invariant level α tests is locally minimax among all level α tests.

In place of the one-dimensional maximal invariant T^2 obtained under G, one now obtains a p-dimensional maximal invariant $Z = (Z_1, \cdots, Z_p)$ defined by

$$Z_i = \bar{X}'_{[i]}(S_{[i]})^{-1}\bar{X}_{[i]}$$

where we write $C_{[i]}$ for the upper left-hand $i \times i$ submatrix of a matrix C and $b_{[i]}$ for the i-vector consisting of the first i components of a vector b. Z_i is essentially Hotelling's statistic based on the first i coordinates. (This and the other straightforward computations which follow will be found in detail in Giri, Kiefer, and Stein (1963).) We shall find it more convenient to work with the equivalent statistic $Y = (Y_1, \cdots, Y_p)'$ where

$$Y_i = NZ_i/(1 + NZ_i) - NZ_{i-1}/(1 + NZ_{i-1}) \qquad (Z_{-1} = 0).$$

It is easily seen that $Y_i \geqq 0$, $\sum_1^p Y_i \leqq 1$, and $\sum_1^p Y_i = U = T^2/(N - 1 + T^2)$.

A corresponding maximal invariant $\Delta = (\delta_1, \cdots, \delta_p)$ in the parameter space of (μ, Σ) under G_T when H_1 is true is easily seen to be given by

$$\delta_i = N\xi'_{[i]}(\Sigma_{[i]})^{-1}\xi_{[i]} - N\xi'_{[i-1]}(\Sigma_{[i-1]})^{-1}\xi_{[i-1]} \qquad (\delta_1 = N\xi_1^2/\Sigma_{11}).$$

Here $\delta_i \geqq 0$ and $\sum_1^p \delta_i = \delta$. The nuisance parameter in this reduced setup is $\eta = (\eta_1, \cdots, \eta_p)$ where $\eta_i = \delta_i/\delta \geqq 0$, $\sum_1^p \eta_i = 1$. The corresponding maximal invariant under H_0 takes on the single value $0 = (0, \cdots, 0)$; we may for convenience also write $\eta = 0$ in that case.

A straightforward but slightly tedious computation yields for the Lebesgue density of Y on $H = \{y: y_i > 0, 1 \leqq i \leqq p; \sum_1^p y_i < 1\}$ the function

$$(2.11) \quad \begin{aligned} f_\Delta(y) &= \pi^{-p/2}\Gamma(N/2)\left(1 - \sum_1^p y_j\right)^{(N-p-2)/2} \Big/ \Gamma[(N-p)/2] \prod_1^p y_i^{\frac{1}{2}} \\ &\times \exp\left\{-\frac{\delta}{2} + \sum_{j=1}^p y_j \sum_{i>j} \frac{\delta_i}{2}\right\} \prod_{i=1}^p \phi((N-i+1)/2, 1/2; y_i\delta_i/2), \end{aligned}$$

where ϕ is the confluent hypergeometric function (sometimes denoted by ${}_1F_1$),

$$(2.12) \quad \phi(a, b; x) = \sum_{j=0}^\infty [\Gamma(a+j)\Gamma(b)/\Gamma(a)\Gamma(b+j)j!]x^j.$$

We now verify the assumptions of Lemma 1 for $U = \sum_1^p Y_i$. Those just preceding (2.1) are obvious. In (2.1) we can take $h(\lambda) = b\lambda$ with b a positive constant. Of course, $P_{\lambda,\eta}\{R\}$ does not depend on η. From (2.11) and (2.12), we have

$$(2.13) \quad \frac{f_{\lambda,\eta}(y)}{f_{0,0}(y)} = 1 + \frac{\lambda}{2}\left\{-1 + \sum_{j=1}^p y_i\left[\sum_{i>j} \eta_i + (N-j+1)\eta_j\right]\right\} + B(y, \eta, \lambda),$$

where $B(y, \eta, \lambda) = o(\lambda)$ uniformly in y and η. We have the setup of Remark 1 above, and (2.2) is satisfied by letting $\xi_{0,\lambda}$ give measure one to the single point $\eta = 0$, while $\xi_{1,\lambda}$ gives measure one to the single point η^* (say) whose jth coordinate is $(N-j)^{-1}(N-j+1)^{-1}p^{-1}N(N-p)$, so that $\sum_{i>j} \eta_i^* + (N-j+1)\eta_j^* = N/p$ for all j. Applying Lemma 1, we have

THEOREM 1. *For every p, N, and α, Hotelling's T^2 test is locally minimax for testing $\delta = 0$ against $\delta = \lambda$ as $\lambda \to 0$.*

EXAMPLE 2. (The R^2 test) With $X_1, \cdots, X_N$ as in Example 1, partition Σ as

$$\Sigma = \begin{Vmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{Vmatrix}$$

where Σ_{22} is $(p-1) \times (p-1)$. Write $\rho^2 = \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}/\Sigma_{11}$. It is desired to test the hypothesis $H_0: \rho^2 = 0$ that the first component is independent of the others, against the alternative $H_1: \rho^2 = \delta$, where δ is specified, $0 < \delta < 1$. It is clear that the transformations $(\xi, \Sigma, \bar{X}, S) \to (\xi + b, \Sigma, \bar{X} + b, S)$ leave the problem invariant and, along with the group G_T' considered below, generate a group which satisfies the Hunt-Stein conditions and in which these transformations form a normal subgroup; the action of these transformations is to reduce the problem to that where $\xi = 0$ (known) and $S = \sum_1^N X_iX_i'$ is sufficient, where N has been reduced by one from what it was originally. We therefore treat this latter formulation, considering $X_1, \cdots, X_N$ to have zero mean. We assume $N \geqq p \geqq 2$, the case $N < p$ now being degenerate.

We now consider the group G_T' of nonsingular lower triangular matrices whose first column contains only zeros except for the first element. It is easily seen that this group, operating as $(S; \Sigma) \to (gSg', g\Sigma g')$, leaves the problem invariant.

The development now parallels that of Example 1. A maximal invariant $R = (R_2, \cdots, R_p)$ is defined by

$$\textstyle\sum_2^i R_j = S_{12[i]}(S_{22[i]})^{-1}S_{21[i]}/S_{11}, \qquad 2 \leqq i \leqq p.$$

Thus, $R_j \geqq 0$, $\sum_2^p R_j \leqq 1$, and $\sum_2^p R_j = U$ is the squared sample multiple correlation coefficient between first and other components (usually denoted by R^2). The corresponding maximal invariant in the parameter space, $\Delta = (\delta_2, \cdots, \delta_p)$, is given by

$$\textstyle\sum_2^i \delta_j = \Sigma_{12[i]}(\Sigma_{22[i]})^{-1}\Sigma_{21[i]}/\Sigma_{11}, \qquad 2 \leqq i \leqq p.$$

Thus, $\delta_i \geqq 0$, $\sum_2^p \delta_i = \rho^2$, the squared population multiple correlation coefficient. We write $\eta = (\eta_2, \cdots, \eta_p)$ with $\eta_i = \delta_i/\delta$ as before, and $(\delta, \eta) = (0, 0)$ under H_0. The Lebesgue density of R when $\rho^2 = \lambda$ can be computed (see Giri and Kiefer (1963) for details) to be

$$(2.14)\ f_{\lambda,\eta}(r)$$

$$= \frac{(1-\lambda)^{\frac{1}{2}(p-1)}\left(1-\sum_2^p r_i\right)^{\frac{1}{2}(N-p-1)}}{\left[1+\sum_2^p r_i((1-\lambda)/\gamma_i - 1)\right]^{\frac{1}{2}N}\Gamma[\frac{1}{2}(N-p+1)]\pi^{\frac{1}{2}(p-1)}}$$

$$\times \prod_2^{p-1}\{r_i^{\frac{1}{2}}\gamma_i^{\frac{1}{2}}(\pi_i+1)^{\frac{1}{2}(N-i+2)}\Gamma[\tfrac{1}{2}(N-i+2)]\}^{-1}$$

$$\times \sum_{\beta_2=0}^{\infty}\cdots\sum_{\beta_j=0}^{\infty}\Gamma\left(\sum_2^p \beta_j + \tfrac{1}{2}N\right)$$

$$\times \prod_2^p\left\{\frac{\Gamma[\frac{1}{2}(N-i+2)+\beta_i]}{(2\beta_i)!}\left[\frac{4r_i(1-\lambda)/\gamma_i(1+\pi_i^{-1})}{1+\sum_2^p r_j[(1-\lambda)/\gamma_j - 1]}\right]^{\beta_i}\right\}$$

where $\gamma_i = 1 - \sum_2^i \delta_j = 1 - \lambda\sum_2^i \eta_j$, $\pi_i = \delta_i/\gamma_i = \lambda\eta_i(1 - \lambda\sum_2^i \eta_j)$. (The expression $1/(1+\pi_i^{-1})$ means 0 if $\delta_i = 0$.) From this we obtain

$$(2.15)\quad \frac{f_{\lambda,\eta}(r)}{f_{0,0}(r)} = 1 + \frac{N\lambda}{2}\left\{-1+\sum_{j=2}^p r_j\left[\sum_{i>j}\eta_i + (N-j+2)\eta_j\right]\right\} + B(r,\eta,\lambda)$$

where $B(r, \eta, \lambda) = o(\lambda)$ uniformly in r and η. We see that the assumptions of Lemma 1 are again satisfied for $U = \sum_2^p y_i$ (="R^2"), with $h(\lambda) = b\lambda$ again. In fact, (2.15) becomes (2.13) if we replace $N\lambda$ in (2.15) by λ, p by $p-1$, and the index range $2 \leqq j \leqq p$ by $1 \leqq j' = j - 1 \leqq p - 1$; thus, $\xi_{1,\lambda}$ now gives measure one to the point whose jth coordinate $(2 \leqq j \leqq p)$ is $(N-j+1)^{-1}\cdot(N-j+2)^{-1}(p-1)^{-1}N(N-p+1)$. (Of course, it is no coincidence that (2.13) and (2.15) correspond: (2.11) involves ratios of noncentral to central chi-square variables, while (2.14) involves similar ratios with random noncentrality parameters; the first order terms in the expansions, which involve only expectations of these quantities, correspond to each other.) We conclude

THEOREM 2. *For every p, N, and α, the critical region which consists of large values of the squared sample multiple correlation coefficient R^2 is locally minimax for testing $\rho^2 = 0$ against $\rho^2 = \lambda$ as $\lambda \to 0$.*

3. Type D and E regions. The notion of a type D or E region is due to Isaacson (1951). Kiefer (1958) showed that the usual F-test of the univariate linear hypothesis has this property. Lehmann (1959a) showed that, in finding regions which are of type D, invariance could be invoked in the manner of the Hunt-Stein theorem; and that this could also be done for type E regions (if they exist) provided that one works with a group which operates as the identity on the nuisance parameter set (H of the next paragraph).

Suppose, for a parameter set $\Omega' = \{(\theta, \eta) : \theta \varepsilon \Theta, \eta \varepsilon H\}$ with associated distributions, with Θ a Euclidean set, that every test function ϕ has a power function $\beta_\phi(\theta, \eta)$ which, for each η, is twice continuously differentiable in the components of θ at $\theta = 0$, an interior point of Θ. Let Q_α be the class of locally strictly unbiased level α tests of $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$; our assumption on β_ϕ implies that all tests in Q_α are similar and that $\partial\beta_\phi/\partial\theta_i \mid_{\theta=0} = 0$ for ϕ in Q_α. Let $\Delta_\phi(\eta)$ be the determinant of the matrix $B_\phi(\eta)$ of second derivatives of $\beta_\phi(\theta, \eta)$ with respect to the components of θ (that is, the Gaussian curvature) at $\theta = 0$. We assume the parametrization to be such that $\Delta_{\phi'}(\eta) > 0$ for all η for at least one ϕ' in Q_α. A test ϕ^* is said to be of *type E* if $\phi^* \varepsilon Q_\alpha$ and $\Delta_{\phi^*}(\eta) = \max_{\phi \varepsilon Q_\alpha} \Delta_\phi(\eta)$ for all η. If H is a single point, ϕ^* is said to be of *type D*.

In the examples which interest us, such as those treated in the previous section, it seems doubtful that type E regions exist. (In terms of Lehmann's development, H is not left fixed by many transformations.) Without pursuing the question of when such regions exist, we introduce two possible optimality criteria, in the same spirit as the type D and E criteria, which will always be fulfilled by some test under minimum regularity assumptions: Write $\bar{\Delta}(\eta) = \max_{\phi \varepsilon Q_\alpha} \Delta_\phi(\eta)$. A test ϕ^* will be said to be of *type D_A* if $\phi \varepsilon Q_\alpha$ and

$$\max_\eta [\bar{\Delta}(\eta) - \Delta_{\phi^*}(\eta)] = \min_{\phi \varepsilon Q_\alpha} \max_\eta [\bar{\Delta}(\eta) - \Delta_\phi(\eta)]$$

and of *type D_M* if

$$\max_\eta [\bar{\Delta}(\eta)/\Delta_{\phi^*}(\eta)] = \min_{\phi \varepsilon Q_\alpha} \max_\eta [\bar{\Delta}(\eta)/\Delta_\phi(\eta)].$$

These criteria resemble stringency and regret criteria employed elsewhere in statistics; the subscripts "A" and "M" stand for "additive" and "multiplicative" regret principles. The possession of these properties is invariant under the product of any transformation on Θ (acting trivially on H) of the same general type as those for which type D regions retain their property, and an arbitrary 1-1 transformation on H (acting trivially on Θ), but, of course, not under more general transformations on Ω'. Obviously, a type E test automatically satisfies these weaker criteria.

Suppose now that a problem is invariant under a group of transformations G for which the Hunt-Stein theorem holds and which acts trivially on Θ; that is, such that $g(\theta, \eta) = (\theta, g\eta)$ for g in G, in a usual abuse of notation. If ϕg is, as usual, the test function defined by $\phi g(x) = \phi(gx)$, a trivial computation then shows that $\Delta_{\phi g}(\eta) = \Delta_\phi(g\eta)$ and hence that $\bar{\Delta}(\eta) = \bar{\Delta}(g\eta)$. Also, if ϕ is better than ϕ' in the sense of either of the above criteria, then ϕg is clearly better than $\phi' g$. All of the requirements of Lehmann's development are easily seen to be satisfied, so that we can conclude that there is an almost invariant (hence, in our examples, an invariant) test which is of type D_A or D_M. (This differs from the way in which invariance is used in the application of (b), p. 883, of Lehmann (1959a), where it is used to reduce Θ rather than H, as here.)

If, furthermore, G is transitive on H, then $\bar{\Delta}(\eta)$ is constant, as is $\Delta_\phi(\eta)$ for an invariant ϕ (which we therefore write simply as Δ_ϕ). In this case we conclude that *if ϕ^* is invariant and if ϕ^* is of type D among invariant ϕ (that is, if Δ_{ϕ^*} maximizes Δ_ϕ over all invariant ϕ), then ϕ^* is of type D_A and D_M among all ϕ.*

Our main tool for verifying optimality in these senses is a trivial one:

LEMMA 0. *Let L be a class of non-negative definite symmetric $m \times m$ matrices, and suppose J is a fixed nonsingular member of L. If* tr $J^{-1}B$ *is maximized (over B in L) by $B = J$, then* det B *is also maximized by J. Conversely, if L is convex and J maximizes* det B, *then* tr $J^{-1}B$ *is maximized by $B = J$.*

PROOF. Write $J^{-1}B = A$. If $A = I$ maximizes $\operatorname{tr} A$, we have $(\det A)^{1/m} \leqq \operatorname{tr} A/m \leqq 1 = \det I$. Conversely, if I maximizes $\det A$, it also maximizes $\operatorname{tr} A$, since $\operatorname{tr} B > \operatorname{tr} I$ implies $\det(\alpha B + (1-\alpha)I) > 1$ for α small and positive.

Of course, the usefulness of this tool lies in the fact that the generalized Neyman-Pearson lemma allows us to maximize tr QB_ϕ (for fixed Q) more easily than to maximize $\Delta_\phi = \det B_\phi$, among similar level α tests. We can find, for each Q, a ϕ_Q which maximizes tr QB_ϕ; a ϕ^* which maximizes Δ_ϕ is then obtained by finding a ϕ_Q for which $B_{\phi_Q} = Q^{-1}$.

In examples of the type which interest us, the reduction by invariance under a group G which is transitive on H often results in a reduced problem wherein the maximal invariant is a vector $Y = (Y_1, \cdots, Y_m)$ whose distribution depends only on $\gamma = (\gamma_1, \cdots, \gamma_m)$ where $\gamma_i = \theta_i^2$ (where $\theta = (\theta_1, \cdots, \theta_m)$), and such that the density f_γ of Y with respect to a σ-finite measure μ is of the form

$$f_\gamma(y) = f_0(y)\{1 + \sum_1^m \gamma_i(h_i + \sum_j a_{ij}y_j)\} + Q(y, \gamma) \tag{3.1}$$

where the h_i and a_{ij} are constants and $Q(y, \gamma) = o(\sum \gamma_i)$ as $\gamma \to 0$, and where we can differentiate under the integral sign to obtain, for invariant test functions ϕ of level α,

$$\begin{aligned} \beta_\phi(\theta, \eta) = \alpha\left(1 + \sum_1^m \gamma_i h_i\right) \\ + \sum_i \gamma_i \sum_j a_{ij} \int y_j \phi(y) f_0(y)\mu(dy) + o(\sum \gamma_i) \end{aligned} \tag{3.2}$$

as θ (or γ, where $\gamma_i = \theta_i^2$) $\to 0$. We shall call this the *symmetric reduced regular* (SRR) case.

In the SRR case, every invariant ϕ has a diagonal B_ϕ whose ith diagonal entry, by (3.2), is $2[h_i\alpha + E_0\{\sum_j a_{ij}Y_j\phi(Y)\}]$. By the Neyman-Pearson lemma, $\operatorname{tr} QB_\phi$ is maximized over such ϕ by a ϕ^* of the form

$$\phi^*(y) = \begin{Bmatrix}1\\0\end{Bmatrix} \text{ if } \sum_{i,j} a_{ij}q_iy_j \begin{Bmatrix}>\\<\end{Bmatrix} C \tag{3.3}$$

where C is a constant and q_i is the ith diagonal element of Q (we need only consider diagonal Q's at this point). From this and the previous remarks, we conclude

LEMMA 2. *In the* SRR *case, an invariant test* ϕ^* *of level* α *is of type* D *among invariant* ϕ *(and, hence, of type* D_A *and* D_M *among all* ϕ*) if and only if* ϕ^* *is of the form* (3.3) *with* $q_i^{-1} = \text{const } b_{\phi^*i}$, *where* b_{ϕ^*i} *is the* ith *diagonal element of* B_{ϕ^*} *(that is,* $q_i^{-1} = \text{const } [h_i\alpha + E_0\{\sum_j a_{ij}Y_j\phi^*(Y)\}]$*).*

EXAMPLE 1. In the setting of Example 1 of Section 2, we let $\theta = N^{\frac{1}{2}}F^{-1}\xi$ where F is the (unique) member of G_T with positive diagonal elements and such that $FF' = \Sigma$. Then Θ is Euclidean p-space, G_T operates transitively on $H = \{$positive definite symmetric $\Sigma\}$ but trivially on Θ, and we have the SRR case with Y_i as in (2.11) and $\gamma_i = \delta_i$ of (2.11). We thus have (3.2) with $h_i = -\frac{1}{2}$, $a_{ij} = 1$ (resp., 0, $N - j + 1$) if $i > j$ (resp., $i < j$, $i = j$). Hotelling's T^2 test has a power function which depends only on $\sum \gamma_i$, so that, with the above parametrization for θ, we have B_{T^2} a multiple of the identity. Also, Hotelling's critical region is of the form $\sum y_i > C$. But, when all q_i are equal, the critical region corresponding to (3.3) is of the form $\sum_j (N + p + 1 - 2j)y_j > C$, which is not Hotelling's region if $p > 1$. We conclude, perhaps somewhat surprisingly in view of Theorem 1,

THEOREM 3. *For* $0 < \alpha < 1 < p < N$, *Hotelling's* T^2*-test is not of type* D *among* G_T*-invariant tests, and hence is not of type* D_A *or* D_M *(nor of type* E*) among all tests.*

The actual computation of a ϕ^* of type D among G_T-invariant tests appears difficult in view of the fact that we must (by (2.11)) compute an integral of the form

$$E_0\{Y_r^h\phi(Y)\} = \int_{\{\Sigma c_jy_j>C\}} \frac{\pi^{-\frac{1}{2}p}\Gamma(N/2)y_r^{h-\frac{1}{2}}(1 - \sum y_j)^{\frac{1}{2}(N-p-2)}}{\Gamma[\frac{1}{2}(N - p)]\prod_{i\neq r} y_i^{\frac{1}{2}}} \prod dy_i \tag{3.4}$$

for $h = 0$ or 1 for various choices of the c_j's and C. When α is close to 0 or 1, one can carry out approximate computations, as illustrated by the discussion of the next paragraph.

As $\alpha \to 1$, one can see that the complement $\bar{R}$ of the critical region becomes a simplex with one corner at 0. When $p = 2$, if we write $\rho = 1 - \alpha$ and consider critical regions of level α of the form $by_1 + y_2 > C$ where $0 < L^{-1} \leqq b \leqq L$, L being fixed but large (this keeps $\bar{R}$ close to the origin), we obtain easily from (3.4) that $\rho = E_0(1 - \phi(Y)) = (N - 2)C/2b^{\frac{1}{2}} + o(C)$ as $C \to 0$. Similarly, $E_0\{(1 - \phi(Y))Y_i\} = (N - 2)C^2b^{i-\frac{3}{2}}/8 + o(C) = \rho^2b^{i-\frac{1}{2}}/2(N - 2) + o(\rho)$ as

$\rho \to 0$, while $E_0 Y_i = 1/N$. We therefore obtain, from (2.13), for the power near H_0,

$$\alpha + \frac{\rho}{2}\left\{\gamma_1\left[1 - \frac{N\rho + o(\rho)}{2(N-2)b^{\frac{1}{2}}}\right] + \gamma_2\left[1 - \frac{\rho + o(\rho)}{2(N-2)b^{\frac{1}{2}}} - \frac{(N-1)\rho b^{\frac{1}{2}} + o(\rho)}{2(N-2)}\right]\right\} + o\left(\sum \gamma_i\right), \tag{3.5}$$

where the $o(\rho)$ and $o(\sum \gamma_i)$ terms are uniform in γ and ρ, respectively. The product Δ_ϕ of the coefficients of γ_1 and γ_2 is easily seen to be maximized when $b = (N+1)/(N-1) + o(1)$, as $\rho \to 0$; with more care, one can obtain further terms in an expansion in ρ for the type D choice of b. The argument is completed by showing that $b < L^{-1}$ implies that $\bar{R}$ lies in a strip so close to the y_1-axis as to make $E_0\{(1 - \phi(Y))Y_1\}$ too large and $E_0\{(1 - \phi(Y))Y_2\}$ too small to yield a ϕ as good as that with $b = (N+1)/(N-1)$, with a similar argument if $b > L$.

When ρ is very close to 0, we see that all choices of $b > 0$ give substantially the same power, $\alpha + \rho(\gamma_1 + \gamma_2)/2 + O(\rho^2)$, so that the relative departure from being of type D, of the T^2 test or any other critical region of the form $bY_1 + Y_2 > C_\alpha$ (b fixed and positive), approaches 0 as $\alpha \to 1$. We do not know how great the departure of Δ_{T^2} from $\bar{\Delta}$ can be for arbitrary α.

One can treat similarly the case $p > 2$ and also the case $\alpha \to 0$.

EXAMPLE 2. We have already noted the correspondence between (2.13) and (2.15). Thus, for the setting of Example 2 of Section 2, we obtain by an argument like that used for Theorem 3,

THEOREM 4. *For $0 < \alpha < 1$, $p > 2$, and $N \geqq p$ or $N > p$ depending on whether or not the mean ξ is known, the critical region consisting of large values of the squared sample multiple correlation coefficient R^2 is not of type D among G_T-invariant tests, and hence is not of type D_A or D_M (nor of type E) among all tests.*

Approximate computations of G_T-invariant type D tests can be carried out when α is close to 0 or 1, as in Example 1.

4. Asymptotically minimax tests. In this section we treat the setting of Section 2 when $\lambda \to \infty$, and expressions such as $o(1)$, $o(H(\lambda))$, etc., are to be interpreted in this light. We are now interested in minimaxing a probability of error which is going to zero. Readers who are familiar with asymptotically large sample size theory (referred to in the remark below) will recall that, in that setting, it is difficult directly to compare approximations to such small probabilities for different families of tests, and one instead compares their logarithms. While our considerations are asymptotic in a sense not involving sample sizes (although some examples, such as that of testing whether the mean of a normal variate with unit variance is 0 or λ, fall equivalently into either framework), we encounter the same difficulty, which accounts for the form of (4.4).

As in Section 2, various possible sets of assumptions could be used. We choose one which differs slightly from the form used in Section 2. Modifications are remarked on, below.

Suppose then that the region $R = \{x: U(x) \geqq C_\alpha\}$ satisfies (in place of (2.1))

$$(4.1) \qquad P_{0,\eta}\{R\} = \alpha, \qquad P_{\lambda,\eta}\{R\} = 1 - \exp\{-H(\lambda)[1 + o(1)]\},$$

where $H(\lambda) \to +\infty$ with λ and the $o(1)$ term is uniform in η. Suppose, replacing (2.2), that

$$(4.2) \qquad \int p(x; \lambda, \eta)\xi_{1,\lambda}(d\eta) \Big/ \int p(x; 0, \eta)\xi_{0,\lambda}(d\eta) = \exp\{H(\lambda)[G(\lambda) + R(\lambda)U(x)] + B(x, \lambda)\}$$

where $\sup_x |B(x, \lambda)| = o(H(\lambda))$ and $0 < c_1 < R(\lambda) < c_2 < \infty$. Our only other regularity assumption is that C_α is a point of increase from the left of the d.f. of U, when $\delta = 0$, uniformly in η; that is,

$$(4.3) \qquad \inf_\eta P_{0,\eta}\{U \geqq C_\alpha - \epsilon\} > \alpha$$

for every $\epsilon > 0$.

REMARK. The reader will find no difficulty in giving analogues here of the specializations and generalizations remarked upon in Section 2. One further variation, which is more relevant here (where $1 - P_{\lambda,\eta}\{R\} \to 0$ as $\lambda \to \infty$) than it would be in the setting of Section 2, is to let $\alpha \to 0$ as $\lambda \to \infty$ in such a way that both $P_{0,\eta}\{R\}$ and $1 - P_{\lambda,\eta}\{R\}$ go to zero, perhaps at different rates. One obtains asymptotic minimax results which have some formal resemblance to familiar results which are asymptotic (in the sense that the sample size $n \to \infty$) for testing between simple hypotheses, as considered by Chernoff, Bahadur, Hodges and Lehmann, and others.

LEMMA 3. *If U satisfies* (4.1) *and* (4.3), *and if for sufficiently large λ there exist $\xi_{0,\lambda}$ and $\xi_{1,\lambda}$ satisfying* (4.2), *then U is asymptotically logarithmically minimax of level α for testing $H_0: \delta = 0$ against $\delta = \lambda$ so $\lambda \to \infty$; that is,*

$$(4.4) \qquad \lim_{\lambda\to\infty} \frac{\inf_\eta \{-\log[1 - P_{\lambda,\eta}\{R\}]\}}{\sup_{\phi_\lambda \varepsilon Q_\alpha} \inf_\eta \{-\log[1 - P_{\lambda,\eta}\{\phi_\lambda \text{ rejects } H_0\}]\}} = 1.$$

PROOF. Suppose, contrary to (4.4), that there is an $\epsilon > 0$ and an unbounded sequence Γ of values λ with corresponding tests ϕ_λ in Q_α for which

$$(4.5) \qquad P_{\lambda,\eta}\{R\} > 1 - \exp\{-H(\lambda)(1 + 5\epsilon)\}$$

for all η.

There are two cases, (4.6) and (4.9). If $\lambda \varepsilon \Gamma$ and

$$(4.6) \qquad -1 - G(\lambda) \leqq R(\lambda)C_\alpha + 2\epsilon,$$

consider the a priori distribution given by the $\xi_{i,\lambda}$ and by τ_λ satisfying

$$(4.7) \qquad \tau_\lambda/(1 - \tau_\lambda) = \exp\{H(\lambda)(1 + 4\epsilon)\}.$$

The integrated risk of any Bayes procedure B_λ must satisfy

$$(4.8) \qquad \begin{aligned} r_\lambda^*(B_\lambda) \leqq r_\lambda^*(\phi_\lambda) &\leqq (1 - \tau_\lambda)\alpha + \tau_\lambda \exp\{-H(\lambda)(1 + 5\epsilon)\} \\ &= (1 - \tau_\lambda)[\alpha + \exp\{-\epsilon H(\lambda)\}], \end{aligned}$$

by (4.5) and (4.7). But, according to (4.2), a Bayes critical region is

$$B_\lambda = \{x: U(x) + B(x, \lambda)/R(\lambda)H(\lambda) \geqq [-(1 + 4\epsilon) - G(\lambda)]/R(\lambda)\}.$$

Hence, if λ is so large that $\sup_x |B(x, \lambda)/H(\lambda)R(\lambda)| < \epsilon/c_2$, we have, from (4.6),

$$B_\lambda \supset \{x: U(x) > C_\alpha - \epsilon/c_2\} = B'_\lambda \quad \text{(say)}.$$

The assumption (4.3) implies that $P_{0,\eta}\{B'_\lambda\} > \alpha + \epsilon'$ with $\epsilon' > 0$, contradicting (4.8) for sufficiently large λ.

On the other hand, if $\lambda \varepsilon \Gamma$ and

$$-1 - G(\lambda) > R(\lambda)C_\alpha + 2\epsilon, \tag{4.9}$$

let

$$\tau_\lambda/(1 - \tau_\lambda) = \exp\{H(\lambda)(1 + \epsilon)\}. \tag{4.10}$$

Then, by (4.2),

$$B_\lambda = \{x: U(x) + B(x, \lambda)/R(\lambda)H(\lambda) \geqq [-(1 + \epsilon) - G(\lambda)]/R(\lambda)\}.$$

Hence, if $\sup_x |B(x, \lambda)/H(\lambda)R(\lambda)| < \epsilon/2c_2$, we conclude from (4.9) that $B_\lambda \subset R$, so that, by (4.1) and (4.10),

$$r^*(B_\lambda) > \tau_\lambda \exp\{-H(\lambda)[1 + o(1)]\} = (1 - \tau_\lambda)\exp\{H(\lambda)(\epsilon - o(1))\}. \tag{4.11}$$

But

$$r^*(B_\lambda) \leqq r^*(\phi_\lambda) \leqq (1 - \tau_\lambda)\alpha + \tau_\lambda \exp\{-H(\lambda)(1 + 5\epsilon)\}$$
$$= (1 - \tau_\lambda)[\alpha + \exp\{-4\epsilon H(\lambda)\}],$$

which contradicts (4.11) for sufficiently large λ.

Example 1. In the setting of Example 1 of Section 2, with

$$U = T^2/(N - 1 + T^2) = \sum_1^p Y_i$$

again, (2.11) and (2.12) yield (since $\phi(a, b; x) = \exp\{x(1 + o(1))\}$ as $x \to \infty$)

$$\frac{f_{\lambda,\eta}(y)}{f_{0,0}(y)} = \exp\left\{\frac{\lambda}{2}\left[-1 + \sum_{j=1}^p y_j \sum_{i \geqq j} \eta_i\right][1 + B(y, \eta, \lambda)]\right\} \tag{4.12}$$

with $\sup_{y,\eta} |B(y, \eta, \lambda)| = o(1)$ as $\lambda \to \infty$. From this and the smoothness of $f_{0,0}$ (or from the well known form of the density of T^2) we see (for example, putting $\eta_p = 1$, the density of U being independent of η) that

$$P_{\lambda,\eta}\{U < C_\alpha\} = \exp\{\tfrac{1}{2}\lambda(C_\alpha - 1)[1 + o(1)]\} \tag{4.13}$$

as $\lambda \to \infty$; thus, (4.1) is satisfied with $H(\lambda) = (1 - C_\alpha)/2$. Next, letting $\xi_{1,\lambda}$ assign measure one to the point $\eta_1 = \cdots = \eta_{p-1} = 0$, $\eta_p = 1$, and $\xi_{0,\lambda}$ assign measure one to $(0, 0)$, we obtain (4.2). Finally, (4.3) is trivial. We conclude, from Lemma 3, THEOREM 5. *For every α, p, N, Hotelling's test is asymptotically minimax for testing $\delta = 0$ against $\delta = \lambda$ as $\lambda \to \infty$.*

Although no critical region of the form $\sum_1^p a_i Y_i > C$ other than Hotelling's would have been locally minimax in the considerations of Section 2, many regions of this form are asymptotically minimax (which, of course, makes Theorem 5 less of an argument in support of the use of the T^2 test):

THEOREM 6. *If $C < 1$ and $1 = b_1 \leqq b_2 \leqq \cdots \leqq b_p$, then the critical region $\{\sum_1^p b_j Y_j > C\}$ is asymptotically minimax (among tests of the same size) as $\lambda \to \infty$.*

PROOF. Since the maximum of $\sum_1^p y_j \sum_{i \geqq j} \eta_j$ subject to $\sum_1^p b_j y_j \leqq C$ is clearly achieved at $y_1 = C$, $y_2 = \cdots = y_p = 0$, integration of $f_{\lambda,\eta}(y)$ over a small region near that point yields (4.13) with C_α replaced by C. Since the b_j's are nondecreasing in j, it is obvious from (4.12) that $\xi_{1,\lambda}$ can be chosen to yield (4.2) with $U = \sum b_j Y_j$. Again, (4.3) is trivial.

The reader may find it interesting, in the case $p = 2$, to note geometrically what happens to (4.13) if $C > 1$, and to note the dependence of the power on η if $b_2 < b_1$.

The result of Theorem 5 is obviously related, in the underlying structure which yields it, to Stein's (1956) admissibility result, although neither implies the other. It is interesting to note also that the same departure from this structure (in the behavior as $\rho^2 \to 1$) which prevents Stein's method from proving the admissibility of the R^2 test, also prevents us from applying Lemma 3 to Example 2 of Section 2 as $\rho^2 \to 1$.

REFERENCES

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

GIRI, N. and KIEFER, J. (1962). Minimax properties of Hotelling's and certain other multivariate tests (abstract). *Ann. Math. Statist.* **33** 1490–1491.

GIRI, N. and KIEFER, J. (1963). Minimax character of the R^2 test in the simplest case. To be published.

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.

ISAACSON, S. L. (1951). On the theory of unbiased tests of simple statistical hypotheses specifying the values of two or more parameters. *Ann. Math. Statist.* **22** 217–234.

JAMES, W. and STEIN, C. (1960). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 361–379.

KIEFER, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. *Ann. Math. Statist.* **28** 573–601.

KIEFER, J. (1958). On the nonrandomized optimality and randomized nonoptimality of symmetrical designs. *Ann. Math. Statist.* **29** 675–699.

KIEFER, J. (1959). Optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* **21** 272–319.

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

LEHMANN, E. L. (1959a). Optimum invariant tests. *Ann. Math. Statist.* **30** 881–884.

STEIN, C. (1955). On tests of certain hypotheses invariant under the full linear group (abstract). *Ann. Math. Statist.* **26** 769.

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