MINIMAX CHARACTER OF HOTELLING'S T^2 TEST IN THE SIMPLEST CASE

By N. Giri,[1] J. Kiefer,[2] and C. Stein[3]

Cornell University and Stanford University

Summary. In the first nontrivial case, dimension $p = 2$ and sample size $N = 3$, it is proved that Hotelling's T^2 test of level α maximizes, among all level α tests, the minimum power on each of the usual contours where the T^2 test has constant power. A corollary is that the T^2 test is most stringent of level α in this case.

1. Introduction. Let $X_1, \cdots, X_N$ be independent normal p-vectors with common mean vector ξ and common 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 $\delta > 0$ (and finite) be specified. For testing the hypothesis $H_0: \xi = 0$ against $H_1: N\xi'\Sigma^{-1}\xi = \delta$ at significance level α, 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. Throughout this paper $0 < \alpha < 1$, so that $0 < C < 1$.

In this paper we are interested in a minimax question regarding the T^2 test, namely, whether or not that test maximizes, among all level α tests, the minimum power under H_1. We succeed in proving that the answer is affirmative in the first nontrivial case, $p = 2$, $N = 3$ (for each possible choice of δ and α), although there are strong indications, mentioned at the end of this section, that the answer is also affirmative for general p and N. However, analytical difficulties make it seem most unlikely that our method of proof can be generalized to handle more than a few of these cases. What is worse is that this proof yields no real understanding of why the result holds, nor of what it is which distinguishes this problem from others where Stein has shown that the best invariant procedure under the real linear group (which, among procedures based on the sufficient statistic $(\bar{X}, S)$, the T^2-test is, here) is not minimax. (See Stein (1955), Lehmann (1959), pp. 231 and 338, and James and Stein (1960), p. 376.) We nevertheless publish the present result in the hope that it may interest others to attack the problem.

The results previously proved for the T^2 test include the best invariant character under the real linear group, as mentioned above. For testing H_0 against $H_1': \xi'\Sigma^{-1}\xi > 0$, Simaika (1941) proved this test to be uniformly most powerful

Received July 5, 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) (NRO 47-005).

[3] Research supported by National Science Foundation Grant GP-40 at Stanford University.

of level α among tests whose power function depends only on $\xi'\Sigma^{-1}\xi$; this result also follows easily from the best invariant property. Stein (1956) showed that the test is admissible for testing H_0 against H_1', but the method used there yields nothing for the test of H_0 against H_1; the minimax proof of the present paper also yields no admissibility result for the problem of testing H_0 against H_1. Hsu (1945) showed that the T^2 test maximizes a certain integral over H_1 of the power, but, as he points out, this property is shared by many other tests, since the integral in question is infinite in value; thus, this result cannot be used to prove the desired minimax property. Of course, when $p = 1$ we have the usual properties of the symmetric Student's test, and when $N \leqq p$ it is easy to see that the infimum over H_1 of the power of every test equals the size of the test; hence, the case $p = 2$, $N = 3$ is the simplest one to be considered.

We now outline briefly our method of proof. We may restrict attention to the space of the minimal sufficient statistic $(\bar{X}, S)$. The examples of Stein mentioned above show that the Hunt-Stein theorem can not be applied for the real linear group G of $p \times p$ nonsingular matrices ($p \geqq 2$) which leave the present problem invariant, operating as $(\bar{X}, S; \xi, \Sigma) \rightarrow (g\bar{X}, gSg'; g\xi, g\Sigma g')$. However, the theorem does apply for the smaller group G_T of nonsingular lower-triangular matrices (zero above the main diagonal), which is solvable. (See Kiefer (1957), Lehmann (1959), p. 345.) Thus, there is a test of level α 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 over H_1. Whereas T^2 was a maximal invariant under G, with a single distribution under each of H_0 and H_1, the maximal invariant under G_T is a p-dimensional statistic $R = (R_1, \cdots, R_p)'$ with a single distribution under H_0 but with a distribution which depends continuously on a $(p - 1)$-dimensional parameter $\Delta = (\delta_1, \cdots, \delta_p)'$, $\delta_i \geqq 0$, $\sum_1^p \delta_i = \delta$ (fixed), under H_1. Thus, when $N > p > 1$ there is no UMP invariant test under G_T as there was under G. We compute the Lebesgue densities f_Δ^* and f_0^* of R, under H_1 and H_0. Because of the compactness of the reduced parameter spaces $\{0\}$ and $\Gamma = \{(\delta_1, \cdots, \delta_p): \delta_i \geqq 0, \sum_1^p \delta_i = \delta\}$ and the continuity of f_Δ^* in Δ, it follows (see Wald (1950)) that every minimax test for the reduced problem in terms of R, is Bayes. In particular, Hotelling's test $U = \sum_1^p R_i > C$, which is G_T-invariant, maximizes the minimum power over H_1 if and only if there is a probability measure λ on Γ such that, for some constant K,

$$\int_\Gamma \frac{f_\Delta^*(r_1, \cdots, r_p)}{f_0^*(r_1, \cdots, r_p)} \lambda(d\Delta) \begin{Bmatrix} \geqq \\ < \end{Bmatrix} K \quad \text{according to whether} \quad \sum_1^p r_i \begin{Bmatrix} \geqq \\ < \end{Bmatrix} C, \tag{1.1}$$

except possibly for a set of measure zero. (Here C depends on the specified α, and λ and K may depend only on C and the specified value $\delta > 0$.) An examination of the integrand in (1.1) will allow us to replace (1.1) by the equivalent

$$\int_\Gamma \frac{f_\Delta^*(r_1, \cdots, r_p)}{f_0(r_1, \cdots, r_p)} \lambda(d\Delta) = K \quad \text{if } \sum_1^p r_i = C. \tag{1.2}$$

We are able to evaluate the unique value which K must take on in order that (1.2) can be satisfied, and are then faced with the question of whether or not there exists a probability measure λ satisfying the left half of (1.2). Writing $\lambda^*(A) = \lambda(\delta A)$, we show that λ^*, if it exists, depends on C and δ only through $C\delta$. The development thus far, which hold for general p and $N > p$, is carried out in Section 2. In Section 3 we then obtain a λ and carry out the proof that it satisfies the left half of (1.2) in the special case $p = 2$, $N = 3$.

The complexity of λ and of our proof that it satisfies the left half of (1.2) make it seem desirable to try other approaches for the general problem, but we have thus far succeeded with none of these. One attempt which must occur to most people who work on this problem is to consider instead the problem where, for fixed $\Sigma = H_\Sigma H_\Sigma'$ (say), the vector $\eta = H_\Sigma^{-1}\xi$ is uniformly distributed on the sphere $\eta'\eta = \delta$ under H_1; one can then use G_T on this modified problem, for which the presence of the minimax property for the T^2 test would imply its presence in the original problem; unfortunately, one obtains a test other than Hotelling's, and which is not minimax for the original problem.

As announced earlier in an abstract (Giri and Kiefer (1962)), it is easy to see that, for every α, N, and p, Hotelling's test has certain local and asymptotic minimax properties as $\delta \to 0$ and $\delta \to \infty$. This lends credence to our conjecture that the minimax result proved for T^2 in the present paper actually holds for all N and p.

The result for the test based on the multiple correlation coefficient R^2 when $p = 3$, $N = 4$, which is analogous to the result of the present paper, will be published elsewhere.

2. Reduction of the problem to (1.2). Since much of this development proceeds along standard lines, we shall omit some of the routine details. The reader may consult Lehmann (1959) for nomenclature and for a treatment of invariance and minimax theory in hypothesis testing.

We need only consider test functions which depend on the sufficient statistic $(\bar{X}, S)$, the Lebesgue density of which is

$$\begin{aligned} f_{\xi,\Sigma}(\bar{x}, s) = {} & c(\det \Sigma)^{-(N+p-1)/2} (\det s)^{(N-p-2)/2} \\ & \times \exp\{-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}(s + N(\bar{x} - \xi)(\bar{x} - \xi)')\} \end{aligned} \tag{2.1}$$

where $c = N^{p/2}/2^{Np/2}\pi^{p(p+1)/4} \prod_{i=1}^p \Gamma((N - i)/2)$.

We can compute a maximal invariant of $(\bar{X}, S)$ under the action of the group G_T of nonsingular lower triangular matrices which leave the problem invariant (as described in Section 1) in the usual fashion: If a function ϕ is invariant, then $\phi(\bar{x}, s) = \phi(g\bar{x}, gsg')$ for all g, s, $\bar{x}$. We may consider the domain of S to be the positive definite symmetric matrices, which have probability one for all ξ, Σ; then there is an F in G_T such that $S = FF'$. Putting $g = LF^{-1}$ where L is any diagonal matrix with values ± 1 in any order on the diagonal, we see that ϕ is a function only of the vector $LF^{-1}\bar{X}$ and hence, because of the freedom in the choice of L, of $|F^{-1}\bar{X}|$, or, equivalently, of the vector whose ith element is the square Z_i of the ith component of $F^{-1}\bar{X}$. Write $b_{[i]}$ for the i-vector consisting of the first i components of a p-vector b, and $C_{[i]}$ for the upper left-hand $i \times i$ submatrix of a $p \times p$ matrix C. Because of the way in which inverses are formed in G_T, $(F_{[i]})^{-1} = (F^{-1})_{[i]}$, so that

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

The vector $Z = (Z_1, \cdots, Z_p)'$ is thus a maximal invariant if it is invariant, and it easily seen to be the latter. Z_i is essentially Hotelling's statistic computed from the first i coordinates. We shall find it more convenient to work with the equivalent statistic $R = (R_1, \cdots, R_p)'$ where

$$\sum_1^i R_j = NZ_i/(1 + NZ_i)$$

or

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

It is easily verified that $R_i \geqq 0$, $\sum_i^p R_i \leqq 1$, and of course $\sum_1^p R_i = U = T^2/(N - 1 + T^2)$.

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 corresponding maximal invariant under H_0 takes on the single value $0 = (0, \cdots, 0)$. The Lebesgue density function f_Δ^* of $(R_1, \cdots, R_p)$ depends on Δ under H_1, and is a fixed f_0^* under H_0.

We must now compute f_Δ^* and f_0^*. (Actually, we need only obtain f_Δ^*/f_0^* for use in (1.2), so we could proceed without keeping track of factors not depending on Δ in this derivation; however, it is not much extra work to keep track of these factors, so we shall do so.) We may put $\Sigma = I$ and $N^{\frac{1}{2}} \xi = (\delta_1^{\frac{1}{2}}, \delta_2^{\frac{1}{2}}, \cdots, \delta_p^{\frac{1}{2}})' = N^{\frac{1}{2}}\rho$ (say) in (2.1), since f_Δ^* depends only on ξ and Σ only through Δ. Let B be the unique lower triangular matrix with positive diagonal elements for which $BB' = S$, and let $V = B^{-1}\bar{X}$. One computes easily the Jacobians $\partial S/\partial B = 2^p \prod_1^p b_{ii}^{p+1-i}$ and $\partial \bar{X}/\partial V = \det B = \prod b_{ii}$, so that the joint density of V and B when $\Sigma = I$, $\xi = \rho$ is

$$h_{\rho,I}(v, b) = 2^p f_{\rho,I}(bv, bb') \prod b_{ii}^{p+2-i}.$$

Putting $W = (W_1, \cdots, W_p)'$ with $W_i = |V_i|$, and noting that the p-vector w with positive components can arise from any of the 2^p vectors $v = Mw$ where M is diagonal with diagonal entries ± 1, we can write $g = bM$ where g ranges over *all* matrices in G_T and obtain for the density of W

$$h^*_{\rho,I}(w) = 2^p \int f_{\rho,I}(gw, gg') \prod_i |g_{ii}|^{p+2-i} \prod_{i \geqq j} dg_{ij}$$

$$(2.2) \qquad = 2^p c \int \exp\{-\tfrac{1}{2} \operatorname{tr}[gg' + N(gw - \rho)(gw - \rho)']\} \cdot \prod_i |g_{ii}|^{N-i} \prod_{i \geqq j} dg_{ij},$$

the range of integration being from $-\infty$ to $+\infty$ in each variable. Let A be a lower triangular matrix for which $A(I + NWW')A' = I$. Then $A'A = (I + NWW')^{-1} = I - (1 + NW'W)^{-1}NWW'$, so that $NW'A'AW = NW'W/(1 + NW'W)$. Since $A_{[i]}(I_i + NW_{[i]}W'_{[i]})A'_{[i]} = I_i$, we obtain similarly

$$(2.3) \qquad NW'_{[i]}A'_{[i]}A_{[i]}W_{[i]} = NW'_{[i]}W_{[i]}/(1 + NW'_{[i]}W_{[i]}) = NZ_i/(1 + NZ_i) = \sum_{j=1}^{i} R_j,$$

so that $N^{\frac{1}{2}}AW$ is a vector whose ith element is $R_i^{\frac{1}{2}}$. Writing $gA^{-1} = q$, we have $\partial g/\partial q = \prod a_{ii}^{p-i+1}$. Also, $\operatorname{tr} NgW\rho' = (N^{\frac{1}{2}}\rho)'q(N^{\frac{1}{2}}AW) = \sum_{i \geqq j} (\delta_i R_j)^{\frac{1}{2}} q_{ij}$. From the equality of the second and fourth expressions of (2.3) we see that $W_i^2 = N^{-1}R_i/(1 - \sum_1^i R_j)(1 - \sum_1^{i-1} R_j)$, so that

$$\partial W/\partial R = \prod_{i=1}^{p} [(\partial W_i^2/\partial R_i)/(\partial W_i^2/\partial W_i)]$$
$$= N^{-p/2}2^{-p} \prod_{i=1}^{p} \left[R_i^{-\frac{1}{2}} \left(1 - \sum_1^{i-1} R_j\right)^{\frac{1}{2}} \left(1 - \sum_1^{i} R_j\right)^{-\frac{3}{2}} \right]$$
$$= N^{-p/2}2^{-p} \left(1 - \sum_1^{p} R_j\right)^{-\frac{3}{2}} \prod_{i=1}^{p} R_i^{-\frac{1}{2}} \prod_{i=1}^{p-1} \left(1 - \sum_1^{i} R_j\right)^{-1}.$$

Since $\prod_1^i a_{jj}^2 = \det(A'_{[i]}A_{[i]}) = 1/\det(I_i + NW_{[i]}W'_{[i]}) = 1/(1 + NW'_{[i]}W_{[i]}) = 1 - \sum_1^i R_j$ and $g_{ii} = a_{ii}q_{ii}$, (2.2) yields

$$f^*_\Delta(r) = h^*_{\rho,I}(w(r))\partial w/\partial r = \left[cN^{-p/2} \left(1 - \sum_1^p r_j\right)^{(N-p-2)/2} e^{-\delta/2} \Big/ \prod_1^p r_i^{\frac{1}{2}} \right]$$
$$\times \int \exp\{-\tfrac{1}{2} \sum_{i \geqq j} [q_{ij}^2 - 2(\delta_i r_j)^{\frac{1}{2}} q_{ij}]\} \prod_i |q_{ii}|^{N-i} \prod_{i \geqq j} dq_{ij},$$

the integration again being from $-\infty$ to $+\infty$ in each variable. For $i > j$, integration with respect to q_{ij} yields a factor $(2\pi)^{\frac{1}{2}} \exp\{\delta_i r_j/2\}$. For $j = i$, we obtain a factor

$$(2\pi)^{\frac{1}{2}} e^{r_i\delta_i/2} E[\chi_1^2(r_i\delta_i)]^{(N-i)/2} = 2^{(N-i+1)/2}\Gamma((N - i + 1)/2)\phi((N - i + 1)/2, \tfrac{1}{2}; r_i\delta_i/2),$$

where $\chi_1^2(\beta)$ is a noncentral chi-square variable with one degree of freedom and noncentrality parameter $\beta(= E\chi_1^2(\beta) - 1)$, and where ϕ is the confluent hypergeometric function (sometimes denoted as ${}_1F_1$),

$$\phi(a, b; x) = \sum_{j=0}^{\infty} [\Gamma(a + j)\Gamma(b)/\Gamma(a)\Gamma(b + j)j!]x^j. \tag{2.4}$$

Thus, finally, for $r \varepsilon H = \{r: r_i > 0, 1 \leqq i \leqq p; \Sigma r_i < 1\}$, we have

$$\begin{aligned} f_\Delta^*(r) &= \left[\pi^{-p/2}\Gamma(N/2)(1 - \textstyle\sum_1^p r_j)^{(N-p-2)/2} \Big/ \Gamma((N - p)/2) \prod_1^p r_i^{\frac{1}{2}}\right] \\ &\quad \times \exp\left\{-\delta/2 + \sum_{j=1}^p r_j \sum_{i>j} \delta_i/2\right\} \prod_{i=1}^p \phi((N - i + 1)/2, \tfrac{1}{2}; r_i\, \delta_i/2). \end{aligned} \tag{2.5}$$

Of course, $f_0^*(r)$ is just the expression preceding the exponential in (2.5), while $f_\Delta^*(r)/f_0^*(r)$ is the exponential and the product following it.

The continuity in Δ over its compact domain Γ is evident, so we can conclude that the minimax character of the critical region $U \geqq C$ is equivalent to the existence of a probability measure λ satisfying (1.1). Clearly (1.1) implies (1.2). On the other hand, if there are a λ and a K for which (1.2) is satisfied and if $r^* = (r_1^*, \cdots, r_p^*)'$ is such that $\sum r_i^* = C' > C$, writing $f = f_\Delta^*/f_0^*$ and $r^{**} = Cr^*/C'$, we see at once that $f(r^*) = f((C'/C)r^{**}) > f(r^{**}) = K$ because of the form of f_Δ^*/f_0^* and the fact that $C'/C > 1$ and $\sum r_i^{**} = C$. This and a similar argument for the case $C' < C$ show that (1.2) implies (1.1). (Of course, we do not assert that the left side of (1.2) still depends only on $\sum r_i$ if $\sum r_i \neq C$.)

The computation of the next section is somewhat simplified by the fact that, for fixed C and δ, we can at this point compute the unique value of K for which (1.2) can possibly be satisfied. Let $\hat{R} = (R_1, \cdots, R_{p-1})$, and write $f_\Delta^*(\hat{r} \mid u)$ for the version of the conditional Lebesgue density of $\hat{R}$ given that $\sum_1^p R_i = u$ which is continuous in $\hat{r}$ and u for $r_i > 0$, $\sum_1^{p-1} r_i < u < 1$, and is 0 elsewhere; write $f_\delta^{**}(u)$ for the denstiy of $U = \sum_1^p R_i$ which is continuous for $0 < u < 1$ and vanishes elsewhere (and which depends on Δ only through δ). Then (1.2) can be written as

$$\int f_\Delta^*(\hat{r} \mid C)\, d\lambda(\Delta) = \left[K \frac{f_0^{**}(C)}{f_\delta^{**}(C)}\right] f_0^*(\hat{r} \mid C) \quad \text{for} \quad r_i > 0, \sum_1^{p-1} r_i < C. \tag{2.6}$$

The integral of (2.6), being a probability mixture of probability densities, is itself a probability density in $\hat{r}$, as is $f_0^*(\hat{r} \mid C)$. Hence, the expression in square brackets equals one. It is well known that, for $0 < C < 1$,

$$f_\delta^{**}(C) = \frac{\Gamma(N/2)\, e^{-\delta/2}}{\Gamma(p/2)\Gamma((N - p)/2)} C^{(p-2)/2}(1 - C)^{(N-p-2)/2}\phi(N/2, p/2; C\delta/2). \tag{2.7}$$

(See Anderson (1958) or use (2.5).) Hence, from (2.5), (1.2) becomes

$$\int_\Gamma \exp\left\{\sum_{j=1}^p r_j \sum_{i>j} \frac{\delta_i}{2}\right\} \prod_{i=1}^p \phi((N - i + 1)/2, \tfrac{1}{2}; r_i\, \delta_i/2)\, d\lambda(\Delta) = \phi(N/2, p/2; C\delta/2) \tag{2.8}$$

for all r with $r_i > 0$, $\sum r_i = C$. Write Γ_1 for the unit $(p-1)$-simplex $\{(\beta_1, \cdots, \beta_p): \beta_i \geqq 0, \sum_i^p \beta_i = 1\}$. Writing $\gamma = C\delta$ and making the change of variables $\beta_i = \delta_i/\delta$, $t_i = \gamma r_i/C$, and writing λ^* for the measure on Γ_1 associated with λ on Γ $(\lambda^*(A) = \lambda(\delta A))$, (2.8) becomes

$$\int_{\Gamma_1} \exp\left\{\sum_{j=1}^{p} t_j \sum_{i>j} \frac{\beta_i}{2}\right\} \prod_{i=1}^{p} \phi((N-i+1)/2, 1/2; \beta_i t_i/2)\, d\lambda^*(\beta_1, \cdots, \beta_p) = \phi(N/2, p/2; \gamma/2) \tag{2.9}$$

for all $(t_1, \cdots, t_p)$ with $\sum t_i = \gamma$ and $t_i > 0$ (hence, by analyticity, for all $(t_1, \cdots, t_p)$ with $\sum t_i = \gamma$). Thus, λ^*, if it exists, depends on C and δ only through their product γ. (When $p = 1$, Γ_1 is a single point, but the dependence on γ is genuine in other cases.)

3. The case $p = 2$, $N = 3$. Representing the integration in (2.9) in terms of $\beta_2 (0 \leqq \beta_2 \leqq 1)$ and noting that $\phi(\frac{3}{2}, \frac{1}{2}; x/2) = (1 + x)e^{x/2}$, we obtain from (2.9), on writing $t_1 = \gamma - t_2$, $\beta_1 = 1 - \beta_2$,

$$\int_0^1 [1 + (\gamma - t_2)(1 - \beta_2)]\, \phi(1, \tfrac{1}{2}; \beta_2 t_2/2)\, d\lambda^*(\beta_2) = e^{(t_2-\gamma)/2}\phi(3/2, 1; \gamma/2). \tag{3.1}$$

One could presumably try to solve (3.1) for λ^* by using the theory of the Meijer transform (with kernel $\phi(1, \frac{1}{2}; x/2)$). We proceed instead by expanding both sides of (3.1) as power series in t_2. Writing $\mu_i = \int_0^1 \beta^i\, d\lambda^*(\beta)$, $0 \leqq i < \infty$ for the ith moment of λ^* and

$$B = e^{-\gamma/2}\phi(\tfrac{3}{2}, 1; \gamma/2), \tag{3.2}$$

we obtain the equations

$$\begin{aligned} &\text{(a) } 1 + \gamma - \gamma\mu_1 = B \\ &\text{(b) } -(2r-1)\mu_{r-1} + (2r + \gamma)\mu_r - \gamma\mu_{r+1} = B[\Gamma(r + \tfrac{1}{2})/r!\Gamma(\tfrac{1}{2})] \quad r \geqq 1 \end{aligned} \tag{3.3}$$

as equivalent to (3.1). (Of course, $\mu_0 = 1$ for λ^* to be a probability measure.) One could now try to show that the sequence $\{\mu_i\}$ defined by $\mu_0 = 1$ and (3.3) satisfies the classical necessary and sufficient condition for it to be the moment sequence of a probability measure on [0, 1] or, equivalently, that the Laplace transform $\sum_0^\infty \mu_j(-t)^j/j!$ is completely monotone on $[0, \infty)$, but we have been unable to proceed successfully in this way. Instead, we shall obtain, in the next paragraph, a function $m_\gamma(x)$ which we then prove, in the succeeding paragraphs below, to be the Lebesgue density $d\lambda^*(x)/dx$ of an absolutely continuous probability measure λ^* satisfying (3.3) (and, hence, (3.1)). That proof does not rely on the somewhat heuristic development of the next paragraph, but we nevertheless sketch that development to give the reader an idea of where the m_γ of (3.8) came from, rather than merely to pull it out of thin air.

The generating function $\psi(t) = \sum_{j=0}^\infty \mu_j t^j$ of the sequence $\{\mu_i\}$ satisfies a differential equation which is obtained in the usual fashion by multiplying (3.3)

(b) by t^{r-1} and summing from 1 to ∞:

$$
\begin{aligned}
(3.4)\qquad & 2t^2(1-t)\psi'(t) - (t^2 - \gamma t + \gamma)\psi(t) \\
&= Bt[(1-t)^{-\frac{1}{2}} - 1] + \gamma[t(1-\mu_1) - 1] \\
&= Bt(1-t)^{-\frac{1}{2}} - t - \gamma.
\end{aligned}
$$

(A corresponding use, instead, of the Laplace transform to obtain (3.8) below, is more involved.) This is solved by treatment of the corresponding homogeneous equation and by variation of parameter, to yield

$$(3.5)\qquad \psi(t) = \frac{e^{-\gamma/2t}}{(1-t)^{\frac{1}{2}}}\int_0^t e^{\gamma/2T}\left[\frac{-1}{2T(1-T)^{\frac{1}{2}}} - \frac{\gamma}{2T^2(1-T)^{\frac{1}{2}}} + \frac{B}{2T(1-T)}\right]dT,$$

the integration being understood to start from the origin along the negative real axis of the complex plane. The constant of integration has been chosen to make ψ continuous at 0 with $\psi(0) = 1$, and (3.5) defines a single-valued function on the complex plane minus a cut along the real axis from 1 to ∞. In fact, the analyticity of ψ on this region can easily be demonstrated by considering the integral of ψ on a closed curve about 0 avoiding 0 and the cut, making the inversion $w = 1/t$, shrinking the path down to the cut $0 \leqq w \leqq 1$, and using (3.30) below. Now, if there did exist an absolutely continuous λ^* whose suitably regular derivative m_γ satisfied

$$(3.6)\qquad \int_0^1 m_\lambda(x)/(1-tx)\,dx = \psi(t),$$

we could obtain m_γ by using the simple inversion formula

$$(3.7)\qquad m_\gamma(x) = (2\pi i x)^{-1}\lim_{\epsilon\downarrow 0}[\psi(x^{-1} + i\epsilon) - \psi(x^{-1} - i\epsilon)].$$

However, there is nothing in the theory of the Stieltjes transform which tells us that an m_γ satisfying (3.7) does satisfy (3.6) (and, hence, (3.1)), so we use (3.7) only as a formal device to obtain an m_γ which we shall then prove, in the remaining paragraphs, satisfies (3.1). From (3.5) and (3.7) we obtain, for $0 < x < 1$,

$$(3.8)\qquad m_\gamma(x) = \frac{e^{-\gamma x/2}}{2\pi x^{\frac{1}{2}}(1-x)^{\frac{1}{2}}}\left\{\int_0^\infty e^{-\gamma u/2}\left[\frac{B}{1+u} - \frac{u^{\frac{1}{2}}}{(1+u)^{3/2}}\right] + B\int_0^x \frac{e^{\gamma u/2}}{1-u}\,du\right\}.$$

In order to prove that $d\lambda^*(x) = m_\gamma(x)\,dx$ (with m_γ defined by (3.8)) satisfies (3.1) with λ^* a probability measure, we must show that

(3.9)

(a) $m_\gamma(x) \geqq 0$ for almost all x, $0 \leqq x \leqq 1$,

(b) $\int_0^1 m_\gamma(x)\,dx = 1$,

(c) $\mu_1 = \int_0^1 x m_\gamma(x)\,dx$ satisfies (3.3) (a)

(d) $\mu_r = \int_0^1 x^r m_\gamma(x)\,dx$ satisfies (3.3) (b) for $r \geqq 1$.

Condition (3.9) (a) follows at once from (3.8) and the fact that $B > 1$ and $u^{\frac{1}{2}}(1 + u) < (1 + u)^{3/2}$ for $u > 0$. To prove (3.9) (d), we note that m_γ as defined by (3.8) satisfies the differential equation

$$(3.10)\quad m_\gamma'(x) + m_\gamma(x)[\gamma/2 + (1 - 2x)/2x(1 - x)] = B/2\pi x^{\frac{1}{2}}(1 - x)^{3/2},$$

so that an integration by parts yields, for $r \geqq 1$,

$$(3.11)\quad \begin{aligned}(r + 1)\,\mu_r - r\mu_{r-1} &= \int_0^1 [(r + 1)\,x^r - rx^{r-1}]\,m_\gamma(x)\,dx \\ &= \int_0^1 (x^r - x^{r+1})\,m_\gamma'(x)\,dx = \mu_r(1 - \gamma/2) + \gamma\mu_{r+1}/2 - \mu_{r-1}/2 \\ &\quad + B\Gamma(r + \tfrac{1}{2})/2\pi^{\frac{1}{2}}\,r!\end{aligned}$$

which is (3.3) (b). The proof of (3.9) (b) and (c) relies on certain identities involving hypergeometric functions. In the next paragraph we list some of the readily available properties of hypergeometric functions which will be used in the proof.

The material summarized in the present paragraph can be found, for example, in Erdélyi (1953), Chapter 6. The confluent hypergeometric function (2.4) has an integral representation when $c > a > 0$ given by

$$(3.12)\quad \phi(a, c; x) = \frac{\Gamma(c)}{\Gamma(a)\,\Gamma(c - a)} \int_0^1 e^{xt}\,t^{a-1}\,(1 - t)^{c-a-1}\,dt.$$

The associated solution ψ to the hypergeometric equation has the representation

$$(3.13)\quad \psi(a, c; x) = \frac{1}{\Gamma(a)} \int_0^\infty e^{-xt}\,t^{a-1}\,(1 + t)^{c-a-1}\,dt$$

if $a > 0$. We shall use the fact that the general definition of ψ, as used in what follows when $a = 0$, satisfies

$$(3.14)\quad \psi(0, c; x) = 1.$$

We shall also use the differential properties

$$(3.15)\quad \frac{d}{dx}\,\phi(a, c; x) = \left(\frac{a}{c} - 1\right)\phi(a, c + 1; x) + \phi(a, c; x)$$

$$(3.16)\quad \frac{d}{dx}\,\psi(a, c; x) = \psi(a, c; x) - \psi(a, c + 1; x),$$

$$(3.17)\quad \frac{d}{dx}\,\psi(a, c; x) = x^{-1}\,a\,[(a - c + 1)\,\psi\,(a + 1, c; x) - \psi\,(a, c; x)],$$

and the identities

$$(a - c + 1)\phi(a, c; x) - a\phi(a + 1, c; x) + (c - 1)\phi(a, c - 1; x) = 0, \tag{3.18}$$

$$c\phi(a, c; x) - c\phi(a - 1, c; x) - x\phi(a, c + 1; x) = 0, \tag{3.19}$$

$$\psi(a, c; x) - a\psi(a + 1, c; x) - \psi(a, c - 1; x) = 0, \tag{3.20}$$

$$(c - a)\psi(a, c; x) - x\psi(a, c + 1; x) + \psi(a - 1, c; x) = 0. \tag{3.21}$$

A useful integration formula (Erdélyi, op. cit., p. 285, equation 16) is, for $y > 0$,

$$\int_0^\infty e^{-x}(x + y)^{-1} \phi(\tfrac{1}{2}, 1; x)\, dx = \psi(\tfrac{1}{2}, 1; y) \tag{3.22}$$

The formula obtained by the (obviously permissible) differentiation under the integral sign of (3.22) with respect to y and use of (3.17) is

$$\int_0^\infty e^{-x}(x + y)^{-2} \phi(\tfrac{1}{2}, 1; x)\, dx = -[\Gamma(\tfrac{1}{2})/2y][\psi(\tfrac{3}{2}, 1; y)/2 - \psi(\tfrac{1}{2}, 1; y)]. \tag{3.23}$$

The function m_γ defined by (3.8) can be written, using (3.13), in terms of hypergeometric functions, for $0 < x < 1$, as

$$\begin{aligned} m_\gamma(x) &= \frac{e^{-\gamma x/2}}{2\pi\,[x(1 - x)]^{\frac{1}{2}}} \left\{ \int_0^\infty e^{-\gamma u/2}\, [e^{-\gamma/2}\varphi(\tfrac{3}{2}, 1; \gamma/2)(1 + u)^{-1} \right. \\ &\qquad \left. - u^{\frac{1}{2}}(1 + u)^{-3/2}]\, du + e^{-\gamma/2}\, \phi(\tfrac{3}{2}, 1; \gamma/2) \int_0^x e^{\gamma u/2}(1 - u)^{-1}\, du \right\} \\ &= \frac{e^{-\gamma x/2}}{2\pi\,[x(1 - x)]^{\frac{1}{2}}} \left\{ e^{-\gamma/2}\, \phi(\tfrac{3}{2}, 1; \gamma/2)\psi(1, 1; \gamma/2) - \Gamma(\tfrac{3}{2})\psi(\tfrac{3}{2}, 1; \gamma/2) \right. \\ &\qquad \left. + \phi(\tfrac{3}{2}, 1; \gamma/2) \left[\int_{1-x}^\infty v^{-1}\, e^{-\gamma v/2}\, dv - \int_1^\infty v^{-1}\, e^{-\gamma v/2}\, dv \right] \right\} \\ &= \frac{1}{2\pi\,[x(1 - x)]^{\frac{1}{2}}} \{e^{-\gamma/2}\, \phi(\tfrac{3}{2}, 1; \gamma/2)\psi(1, 1; \gamma(1 - x)/2) \\ &\qquad - e^{-\gamma x/2}\Gamma(\tfrac{3}{2})\psi(\tfrac{3}{2}, 1; \gamma/2)\}. \end{aligned} \tag{3.24}$$

We now prove (3.9) (b). From (3.13), (3.12), and (3.22), we have

$$\begin{aligned} &\int_0^1 \frac{1}{2\pi\,[x(1 - x)]^{\frac{1}{2}}}\, \psi(1, 1; \gamma(1 - x)/2)\, dx \\ &= \int_0^1 \frac{1}{2\pi\,[x(1 - x)]^{\frac{1}{2}}}\, \psi(1, 1; \gamma x/2)\, dx \\ &= \int_0^1 \frac{dx}{2\pi\,[x(1 - x)]^{\frac{1}{2}}} \int_0^\infty (1 + t)^{-1}\, e^{-\gamma x t/2}\, dt \\ &= \frac{1}{2}\int_0^\infty (1 + t)^{-1}\, \phi(\tfrac{1}{2}, 1; \gamma t/2)\, e^{-\gamma t/2}\, dt = \tfrac{1}{2}\, \Gamma(\tfrac{1}{2})\psi(\tfrac{1}{2}, 1; \gamma/2). \end{aligned} \tag{3.25}$$

From this, (3.24), and (3.12), we have, putting $\gamma/2 = z$,

$$H(z) \equiv \frac{4e^z}{\Gamma(\frac{1}{2})} \int_0^1 m_{2z}(x)\,dx = 2\phi(\tfrac{3}{2},1;z)\psi(\tfrac{1}{2},1;z) - \phi(\tfrac{1}{2},1;z)\psi(\tfrac{3}{2},1;z). \tag{3.26}$$

We shall show that

$$H'(z) - H(z) = 0, \tag{3.27}$$

from which $H(z) = Ce^z$. (This identity is probably known, but we were unable to find it in the literature.) By direct evaluation in terms of elementary integrals when $\gamma = 0$ (or by using (3.26) and the expansion of ψ and ϕ near $z = 0$), we have $\int_0^1 m_0(x)\,dx = 1$; hence, (3.9) (b) follows from (3.27). To prove (3.27), we use (3.15) and (3.16), which yield (omitting everywhere the argument z)

$$\begin{aligned} H' - H = {} & \phi(\tfrac{3}{2},2)\psi(\tfrac{1}{2},1) + 2\phi(\tfrac{3}{2},1)\psi(\tfrac{1}{2},1) - 2\phi(\tfrac{3}{2},1)\psi(\tfrac{1}{2},2) \\ & + \tfrac{1}{2}\phi(\tfrac{1}{2},2)\psi(\tfrac{3}{2},1) - \phi(\tfrac{1}{2},1)\psi(\tfrac{3}{2},1) + \phi(\tfrac{1}{2},1)\psi(\tfrac{3}{2},2). \end{aligned} \tag{3.28}$$

To this expression add the following four left hand side expressions, each of which equals zero (where a and c are the arguments as they appear in (3.12) and (3.13) and where, as in (3.28), we again omit display of the common argument z of ϕ and ψ):

$$\begin{aligned} &\psi(\tfrac{3}{2},1) \text{ times (3.18) with } a = \tfrac{1}{2}, c = 2; \\ &\psi(\tfrac{3}{2},2) \text{ times (3.19) with } a = \tfrac{3}{2}, c = 1; \\ &2\phi(\tfrac{3}{2},1) \text{ times (3.20) with } a = \tfrac{1}{2}, c = 2; \\ &-\phi(\tfrac{3}{2},2) \text{ times (3.21) with } a = \tfrac{3}{2}, c = 1; \end{aligned} \tag{3.29}$$

one obtains $H' - H = 0$, as desired.

We now verify (3.9) (c). We first note, from (3.12) and from (3.19) with $a = \frac{3}{2}, c = 1$, that

$$\int_0^1 \frac{(1 + \gamma y)\,e^{\gamma y/2}}{2\pi\,[y(1-y)]^{\frac{1}{2}}}\,dy = \phi(\tfrac{1}{2},1;\gamma/2)/2 + \gamma\phi(\tfrac{3}{2},2;\gamma/2)/4 = \phi(\tfrac{3}{2},1;\gamma/2)/2 \tag{3.30}$$

An alternative way of writing (3.9) (c) is, by (3.9) (b) (which we have just proved),

$$1 = [\phi(\tfrac{3}{2},1;\gamma/2)]^{-1}\,e^{\gamma/2} \int_0^1 [1 + \gamma(1-x)]\,m_\gamma(x)\,dx. \tag{3.31}$$

The right side of (3.31) may be expressed, using (3.25) and (3.30), as

$$\frac{1}{\phi(\frac{3}{2}, 1; \gamma/2)} \int_0^1 \frac{(1 + \gamma y)}{2\pi[y(1 - y)]^{\frac{1}{2}}} \{\phi(\tfrac{3}{2}, 1; \gamma/2)\psi(1, 1; \gamma y/2) - e^{\gamma y/2}\Gamma(\tfrac{3}{2})\psi(\tfrac{3}{2}, 1; \gamma/2)\} \, dy$$

$$\begin{aligned} (3.32) \qquad &= \int_0^1 \frac{\gamma y \psi(1, 1; \gamma y/2)}{2\pi[y(1 - y)]^{\frac{1}{2}}} \, dy + \int_0^1 \frac{\psi(1, 1; \gamma y/2)}{2\pi[y(1 - y)]^{\frac{1}{2}}} \, dy \\ &\quad - \frac{\Gamma(\frac{3}{2})\psi(\frac{3}{2}, 1; \gamma/2)}{\phi(\frac{3}{2}, 1; \gamma/2)} \int_0^1 \frac{(1 + \gamma y)e^{\gamma y/2}}{2\pi[y(1 - y)]^{\frac{1}{2}}} \, dy \\ &= \int_0^1 \frac{\gamma y \psi(1, 1; \gamma y/2)}{2\pi[y(1 - y)]^{\frac{1}{2}}} \, dy + \frac{\Gamma(\frac{1}{2})\psi(\frac{1}{2}, 1; \gamma/2)}{2} - \frac{\Gamma(\frac{3}{2})\psi(\frac{3}{2}, 1; \gamma/2)}{2}. \end{aligned}$$

To evaluate the integral on the last line of (3.32), we use (3.21) with $a = 1$, $c = 0$ and (3.14), (3.13), (3.12), and (3.23), to write

$$\begin{aligned} \int_0^1 \frac{\gamma y \psi(1, 1; \gamma y/2)}{2\pi[y(1 - y)]^{\frac{1}{2}}} \, dy &= \int_0^1 \frac{[\psi(0, 0; \gamma y/2) - \psi(1, 0; \gamma y/2)]}{\pi[y(1 - y)]^{\frac{1}{2}}} \, dy \\ (3.33) \qquad &= 1 - \int_0^1 \frac{dy}{\pi[y(1 - y)]^{\frac{1}{2}}} \int_0^\infty (1 + t)^{-2} e^{-(\gamma y t/2)} \, dt \\ &= 1 - \int_0^\infty (1 + t)^{-2} \phi(\tfrac{1}{2}, 1; \gamma t/2) e^{-\gamma t/2} \, dt \\ &= 1 + \Gamma(\tfrac{1}{2})[\psi(\tfrac{3}{2}, 1; \gamma/2)/2 - \psi(\tfrac{1}{2}, 1; \gamma/2)]/2. \end{aligned}$$

Thus, (3.32) and (3.33) imply (3.13) and, hence, (3.9) (c).

REFERENCES

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

ERDÉLYI, A. (1953). (Editor) *Higher Transcendental Functions*, **1** McGraw-Hill, 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.

HSU, P. L. (1945). On the power function of the E^2-test and the T^2-test. *Ann. Math. Statist.*, **16** 278–286.

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

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

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

SIMAIKA, J. B. (1941). An optimum property of two statistical tests. *Biometrika* **32** 70–80.

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

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

WALD, A. (1950). *Statistical Decision Functions*. Wiley, New York.