THE ADMISSIBILITY OF HOTELLING'S T^2-TEST

By Charles Stein

Stanford University

1. Summary. In Section 3 we shall prove a theorem based on a method of A. Birnbaum [1] and E. Lehmann concerning the admissibility of certain tests of simple hypotheses in multivariate exponential families. In Section 4 we compute the supporting hyperplanes of the convex acceptance region in some of the most common applications of Hotelling's T^2-test and show that the theorem of Section 3 implies the admissibility of this test. In Section 5 we point out some of the limitations of the method of this paper.

2. Introduction. We recall the definition of admissibility of a statistical test. Let $\mathfrak{X}$ be a set, $\mathfrak{B}$ a σ-algebra of subsets of $\mathfrak{X}$, Θ a set, Θ_0 a nonempty proper subset of Θ, and for each $\theta \varepsilon \Theta$, let P_θ be a probability measure on $\mathfrak{B}$. We observe a random element X distributed in $\mathfrak{X}$ according to P_θ, with θ an unknown element of Θ, and we want to test the hypothesis $H_0 : \theta \varepsilon \Theta_0$. A test is a $\mathfrak{B}$-measurable function φ on $\mathfrak{X}$ to the closed interval [0, 1], with the interpretation that if we observe X, we reject H_0 with probability $\varphi(X)$. The test φ_0 is said to be admissible if there does not exist a test φ such that

$$\int \varphi \, dP_{\theta_0} \leqq \int \varphi_0 \, dP_{\theta_0} \qquad \text{for all } \theta_0 \varepsilon \Theta_0, \tag{1}$$

$$\int \varphi \, dP_{\theta} \geqq \int \varphi_0 \, dP_{\theta} \qquad \text{for all } \theta \varepsilon \Theta - \Theta_0, \tag{2}$$

with strict inequality for some $\theta_0 \varepsilon \Theta_0$ or some $\theta \varepsilon \Theta - \Theta_0$; i.e., a φ which has a smaller probability of error for some parameter point and does not have a larger probability of error for any parameter point.

An exponential family consists of a finite-dimensional real linear space $\mathfrak{X}$, a measure μ on the σ-algebra $\mathfrak{B}$ of all ordinary Borel subsets of $\mathfrak{X}$, a subset Θ of the adjoint space $\mathfrak{X}'$ (the linear space of all real-valued linear functions on $\mathfrak{X}$) such that for all $\xi \varepsilon \Theta$

$$\psi(\xi) = \int e^{\xi x} \, d\mu(x) < \infty, \tag{3}$$

and P, the function on Θ to the set of probability measures on $\mathfrak{B}$ given by

$$P_\xi(A) = \frac{1}{\psi(\xi)} \int_A e^{\xi x} \, d\mu(x), \tag{4}$$

for all $A \varepsilon \mathfrak{B}$.

It is well known that any set of nonsingular multivariate normal distributions in the same space can be expressed as an exponential family. If $Y =$

Received December 8, 1954.

$(Y_1 \cdots Y_p)$ is a random p-dimensional column vector (written horizontally to facilitate printing) normally distributed with mean y_0 and nonsingular covariance matrix s_0, its density (with respect to ordinary Lebesgue measure in p-space) is

$$\begin{aligned}(5)\quad &\frac{1}{(2\pi)^{p/2}(\det s_0)^{1/2}} \exp\left[-\tfrac{1}{2}(y - y_0)' s_0^{-1}(y - y_0)\right] \\ &\qquad = \frac{\exp(-\frac{1}{2}y_0' s_0^{-1} y_0)}{(2\pi)^{p/2}(\det s_0)^{1/2}} \exp\left\{(y_0' s_0^{-1})y - \tfrac{1}{2} \operatorname{tr}\left[s_0^{-1}(yy')\right]\right\}.\end{aligned}$$

We take $\mathfrak{X}$ to be the $(p + p(p + 1)/2)$-dimensional real linear space of pairs (y, s) with s a symmetric $p \times p$ matrix and denote by (η, Γ) with η a p-dimensional vector and Γ a symmetric $p \times p$ matrix the element of $\mathfrak{X}'$ defined by

$$(6)\qquad (\eta, \Gamma)(y, s) = \eta' y - \tfrac{1}{2} \operatorname{tr} \Gamma s.$$

Let f be the mapping of $\mathcal{Y}$ (the space of p-dimensional vectors) into $\mathfrak{X}$ given by

$$(7)\qquad f(y) = (y, yy'),$$

let ν be ordinary Lebesgue measure in $\mathcal{Y}$, and let μ, defined by

$$(8)\qquad \mu(A) = \nu(f^{-1}A),$$

be the induced measure in $\mathfrak{X}$. Let Θ be the set of $(\eta, \Gamma) \varepsilon \mathfrak{X}'$ with Γ positive definite, and define $P_{(\eta,\Gamma)}$ by (4). Then $(\mathfrak{X}, \mu, \Theta, P)$ is an exponential family and $P_{(\eta,\Gamma)}$ is the distribution of (Y, YY') if we put

$$(9)\qquad \Gamma = s_0^{-1}, \qquad \eta = s_0^{-1} y_0 .$$

Since the function f is $1 - 1$ and preserves measurability (and is therefore sufficient), this shows that the set of all nonsingular normal distributions in p-space (and therefore any subset) is an exponential family.

3. A theorem on admissibility for tests in exponential families.

THEOREM: *Let $(\mathfrak{X}, \mu, \Theta, P)$ be an exponential family and Θ_0 a nonempty proper subset of Θ. Let A be a closed convex subset of $\mathfrak{X}$ such that for every $\xi \varepsilon \mathfrak{X}'$ and real c for which*

$$(10)\qquad \{x : \xi x > c\} \cap A = \phi \ (\textit{the empty set})$$

there exists $\theta_1 \varepsilon \tilde{\Theta} = \{\xi : \int e^{\xi x} d\mu(x) < \infty\}$ such that there exist arbitrarily large λ for which $\theta_1 + \lambda\xi \varepsilon \Theta - \Theta_0$. Then the test φ_0, defined by

$$(11)\qquad \varphi_0(x) = \begin{cases} 0 & \textit{if } x \varepsilon A, \\ 1 & \textit{if } x \not\varepsilon A, \end{cases}$$

i.e.,

$$(12)\qquad \varphi_0(x) = 1 - \chi_A(x),$$

is admissible for testing the hypothesis that a random element X of $\mathfrak{X}$ is distributed according to some P_θ with $\theta \varepsilon \Theta_0$ against the alternatives $\theta \varepsilon \Theta - \Theta_0$.

The reader will observe that the theorem essentially gives conditions under which a test is admissible for testing any simple hypothesis in the given exponential family. However, the above statement is more convenient for the application we are going to make. This theorem is an extension of a result and method of proof which appeared in a first draft of Birnbaum [1], suggested, I believe, by E. Lehmann. It is related to Theorem 3 in the final version of Birnbaum's paper.

PROOF. We shall suppose a test φ strictly better than φ_0 and obtain a contradiction. Thus, suppose

$$\int \varphi(x)\, dP_{\theta_0}(x) \leqq \int [1 - \chi_A(x)]\, dP_{\theta_0}(x), \tag{13}$$

$$\int \varphi(x)\, dP_\theta(x) \geqq \int [1 - \chi_A(x)]\, dP_\theta(x) \tag{14}$$

for all $\theta_0 \varepsilon \Theta_0$ and $\theta \varepsilon \Theta - \Theta_0$, with strict inequality for at least one $\theta_0 \varepsilon \Theta_0$ or one $\theta \varepsilon \Theta - \Theta_0$. By (13),

$$\int [1 - \chi_A(x) - \varphi(x)]\, dP_{\theta_0}(x) \geqq 0, \tag{15}$$

so that, since P_θ and μ are mutually absolutely continuous for any $\theta \varepsilon \tilde{\Theta}$ either

$$\mu\{x{:}1 - \chi_A(x) - \varphi(x) \neq 0\} = 0, \tag{16}$$

or

$$\mu\{x{:}1 - \chi_A(x) - \varphi(x) > 0\} > 0. \tag{17}$$

Since (16) would imply equality everywhere in (13) and (14), it is impossible. But

$$\{x{:}1 - \chi_A(x) - \varphi(x) > 0\} = A' \cap B, \tag{18}$$

where A' is the complement of A and

$$B = \{x{:}\varphi(x) < 1\}. \tag{19}$$

Cover A' with a denumerable collection $\mathcal{S}$ of open half-spaces disjoint from A. Then, by (17) and (18), for some half-space $S = \{x{:}\xi x > c\} \varepsilon \mathcal{S}$,

$$\mu(A' \cap B \cap S) > 0. \tag{20}$$

By hypothesis there exist $\theta_1 \varepsilon \tilde{\Theta}$ and arbitrarily large $\lambda > 0$ such that

$$\theta_\lambda = \theta_1 + \lambda\xi \varepsilon \Theta - \Theta_0. \tag{21}$$

Then

$$\begin{aligned}
&\int [1 - \chi_A(x) - \varphi(x)]\, dP_{\theta_\lambda}(x) \\
&\quad = \frac{1}{\psi(\theta_\lambda)} \int [1 - \chi_A(x) - \varphi(x)] e^{\theta_\lambda x}\, d\mu(x) \\
&\quad = \frac{\psi(\theta_1)}{\psi(\theta_\lambda)} \int [1 - \chi_A(x) - \varphi(x)] e^{\lambda \xi x}\, dP_{\theta_1}(x) \\
(22)\quad &\quad = \frac{\psi(\theta_1)}{\psi(\theta_\lambda)} e^{\lambda c} \int [1 - \chi_A(x) - \varphi(x)] e^{\lambda(\xi x - c)}\, dP_{\theta_1}(x) \\
&\quad = \frac{\psi(\theta_1)}{\psi(\theta_\lambda)} e^{\lambda c} \left\{ \int_{\{x:\xi x > c\}} [1 - \chi_A(x) - \varphi(x)]\, e^{\lambda(\xi x - c)}\, dP_{\theta_1}(x) \right. \\
&\qquad \left. + \int_{\{x:\xi x \leqq c\}} [1 - \chi_A(x) - \varphi(x)] e^{\lambda(\xi x - c)}\, dP_{\theta_1}(x) \right\}.
\end{aligned}$$

Since the first integral in the final expression approaches $+\infty$ as $\lambda \to +\infty$ and the second is bounded, this is >0 for sufficiently large λ, contradicting (14).

4. Admissibility of Hotelling's T^2-test. Let $Y, Z_1, \cdots, Z_m, U_1 \cdots, U_n$ be independently normally distributed random p-dimensional vectors (with $p \leqq n$) with means given by

$$(23)\qquad \mathcal{E}Y = y_0, \qquad \mathcal{E}Z_j = z_{j0}, \qquad \mathcal{E}U_i = 0$$

and common unknown, nonsingular covariance matrix s_0. Suppose y_0 and the z_{j0} are also unknown. Hotelling's T^2-test for the hypothesis $H_0 : y_0 = 0$ is to accept H_0 if and only if

$$(24)\qquad Y'(YY' + \sum U_i U_i')^{-1} Y \leqq c,$$

where the positive constant c is chosen so as to give the desired significance level. We shall show that this test is admissible as a test against unrestricted alternatives.

The joint probability density function of $Y, Z_1, \cdots, Z_m, U_1, \cdots, U_n$ (with respect to ordinary Lebesgue measure ν in the $(m + n + 1)p$-dimensional coordinate space) is

$$(25)\qquad \frac{\exp\{-\frac{1}{2}[y_0' s_0^{-1} y_0 + \sum z_{i0}' s_0^{-1} z_{i0}]\}}{(2\pi)^{(m+n+1)p/2} (\det s_0)^{(m+n+1)/2}} \times \exp\{-\tfrac{1}{2} \operatorname{tr} s_0^{-1}[\sum u_i u_i' + \sum z_j z_j' + yy'] + y_0' s_0^{-1} y + \sum z_{i0}' s_0^{-1} z_i\}.$$

Thus the given family of distributions is equivalent to an exponential family $(\mathfrak{X}, \mu, \Theta, P)$. $\mathfrak{X}$ is the $[p(p + 1)/2 + (m + 1)p]$-dimensional space of all $(s, y, z_1, \cdots, z_m)$, where s is a symmetric $p \times p$ matrix and $y, z_1, \cdots, z_m$ are p-dimensional vectors. The measure μ is given by

$$(26)\qquad \mu(A) = \nu(f^{-1}A),$$

where f is the function on the original $(m + n + 1)p$-dimensional space to $\mathfrak{X}$, defined by

$$f(y, z_1, \cdots, z_m, u_1, \cdots, u_n) = (\sum u_i u_i' + \sum z_j z_j' + yy', y, z_1, \cdots, z_m). \tag{27}$$

It is convenient to designate by $(\Gamma, \eta, \zeta_1, \cdots, \zeta_m)$, with Γ a $p \times p$ symmetric matrix and $\eta, \zeta_1, \cdots, \zeta_m$ p-dimensional vectors, the element of the adjoint space $\mathfrak{X}'$, defined by

$$(\Gamma, \eta, \zeta_1, \cdots, \zeta_m)(s, y, z_1, \cdots, z_m) = -\tfrac{1}{2} \operatorname{tr} \Gamma s + \eta' y + \sum \zeta_j' z_j . \tag{28}$$

The parameter space Θ consists of the $(\Gamma, \eta, \zeta_1, \cdots, \zeta_m)$, with Γ positive definite. The correspondence between this designation of the parameter point and that in terms of $(s_0, y_0, z_{10}, \cdots, z_{m0})$ is given by

$$\Gamma = s_0^{-1}, \qquad \eta = s_0^{-1} y_0, \qquad \zeta_j = s_0^{-1} z_{j0} . \tag{29}$$

P is the function given by

$$P_{(\Gamma,\eta,\zeta_1,\cdots,\zeta_m)}(A) = \frac{1}{\psi(\Gamma,\eta,\zeta_1,\cdots,\zeta_m)} \cdot \int_A \exp\,[(\Gamma, \eta, \zeta_1, \cdots, \zeta_m)(s, y, z_1, \cdots, z_m)]\, d\mu(s, y, z_1, \cdots, z_m). \tag{30}$$

In terms of the sample point in $\mathfrak{X}$, the acceptance region for Hotelling's T^2-test is

$$\{(s, y, z_1, \cdots, z_m) : s - \sum z_j z_j' \text{ is positive definite and } y'(s - \sum z_j z_j')^{-1} y \leqq c\} . \tag{31}$$

LEMMA. *The set* (31) *is contained in the intersection of all half-spaces of the form*

$$\left\{(s, y, z_1, \cdots, z_m) : \eta' y + \sum k_j \eta' z_j - \tfrac{1}{2} \operatorname{tr} \eta\eta' s \leqq \frac{c + \sum k_j^2}{2}\right\} \tag{32}$$

and those of the form

$$\left\{(s, y, z_1, \cdots, z_m) : \sum_j k_j \eta' z_j - \tfrac{1}{2} \operatorname{tr} \eta\eta' s \leqq \frac{\sum k_j^2}{2}\right\}, \tag{33}$$

where η ranges over the set of all p-dimensional vectors different from 0 *and $k_1, \cdots, k_m$ range over the real line. Also* (31) *differs from this intersection by a set of probability* 0.

We first show that any point $(s, y, z_1, \cdots, z_m)$ in the set (31) lies in each of these half-spaces. Since $s - \sum z_j z_j'$ is positive definite, we have, by Schwarz' inequality

$$\begin{aligned} \sum k_j \eta' z_j &\leqq \sqrt{\sum k_j^2} \sqrt{\sum (\eta' z_j)^2} \\ &\leqq \tfrac{1}{2} \sum k_2^j + \tfrac{1}{2} \operatorname{tr} \eta\eta' \sum z_j z_j' \leqq \tfrac{1}{2} \sum k_j^2 + \tfrac{1}{2} \operatorname{tr} \eta\eta' s, \end{aligned} \tag{34}$$

so that (33) holds.

Since $y'(s - \sum z_j z_j')^{-1} y \leqq c$, we have, again by Schwarz' inequality,

$$
\begin{aligned}
\eta' y + \sum k_j \eta' z_j &\leqq \sqrt{\eta'(s - \sum z_j z_j')\eta}\sqrt{y'(s - \sum z_j z_j')^{-1} y} \\
&\quad + \sqrt{\sum k_j^2}\sqrt{\sum (\eta' z_j)^2} \\
&\leqq \tfrac{1}{2}\eta'(s - \sum z_j z_j')\eta + \tfrac{1}{2} y'(s - \sum z_j z_j')^{-1} y \qquad (35) \\
&\quad + \tfrac{1}{2}\sum k_j^2 + \tfrac{1}{2}\sum (\eta' z_j)^2 \\
&\leqq \tfrac{1}{2}[\mathrm{tr}\ \eta\eta' s + c + \sum k_j^2],
\end{aligned}
$$

so that (32) holds. Thus, (31) is contained in the intersection of all sets of the form (32) and (33).

Next, consider a point $(s, y, z_1, \cdots, z_m)$ for which (32) and (33) hold for all η and $k_1, \cdots, k_m$. Putting $k_j = \eta' z_j$ in (33), we obtain

$$\sum (\eta' z_j)^2 - \tfrac{1}{2}\ \mathrm{tr}\ \eta\eta' s \leqq \frac{\sum (\eta' z_j)^2}{2},$$

i.e.,

$$\mathrm{tr}\ \eta\eta'(s - \sum z_j z_j') \geqq 0. \qquad (36)$$

Since (36) holds for all η, it follows that $s - \sum z_j z_j'$ is positive semidefinite, and thus, except for a set of probability 0, positive definite. Then by (32), with $k_j = \eta' z_j$,

$$\eta' y + \sum (\eta' z_j)^2 - \tfrac{1}{2}\mathrm{tr}\ \eta\eta' s \leqq \tfrac{1}{2}[c + \sum (\eta' z_j)^2],$$

so that

$$\eta' y \leqq \tfrac{1}{2}[c + \tfrac{1}{2}\mathrm{tr}\ \eta\eta'(s - \sum z_j z_j')].$$

With

$$\eta = (s - \sum z_j z_j')^{-1} y,$$

this becomes

$$y'(s - \sum z_j z_j')^{-1} y \leqq \frac{c}{2} + \tfrac{1}{2} y'(s - \sum z_j z_j')^{-1} y, \qquad (37)$$

which shows that the given point is in the set (31).

We return to the proof of the admissibility of Hotelling's T^2-test. The set (31) is essentially the intersection of all sets of the form (32) and (33) whose defining relations can be rewritten in the notation of (28)

$$(\eta\eta', \eta, k_1\eta, \cdots, k_m\eta)(s, y, z_1, \cdots, z_m) \leqq \tfrac{1}{2}[c + \sum k_j^2] \qquad (38)$$

$$(\eta\eta', 0, k_1\eta, \cdots, k_m\eta)(s, y, z_1, \cdots, z_m) \leqq \tfrac{1}{2}\sum k_j^2. \qquad (39)$$

Thus, if $\xi = (\Gamma, \eta, \zeta_1, \cdots, \zeta_m)$ is any point in $\mathfrak{X}'$ for which

$$\{x : \xi x > c\} \cap A = \phi, \qquad (40)$$

where A is the intersection of the half-spaces mentioned in the lemma, then ξ must be a limit of positive linear combinations of elements of the form (38) or (39). In particular, Γ must be positive semidefinite. Consequently, for any parameter point,

$$\theta_1 = (\Gamma^{(1)}, \eta^{(1)}, \zeta_1^{(1)}, \cdots, \zeta_m^{(1)})$$

and any $\lambda > 0$, the first component $\Gamma^{(1)} + \lambda\Gamma$ of $\theta_1 + \lambda\xi$ must be positive definite and, for sufficiently large λ, the second component $\eta^{(1)} + \lambda\eta$ must be different from 0. The theorem of Section 3 enables us to conclude that A is an admissible acceptance region.

5. Limitations of the method. Examining the proof of the theorem of Section 3 and the end of the proof of Section 4, we see that, in showing that for any test essentially different from a T^2-test but of the same size, there exists an alternative for which the other test is worse, we have looked at parameter points which are arbitrarily far out, in particular, points for which $\eta^1\Gamma^{-1}\eta = y_0's_0^{-1}y_0$ is arbitrarily large. How large this has to be taken depends on the test with which we are comparing the T^2-test. This is unsatisfactory, since it leaves open the possibility that there is a test which is appreciably better than Hotelling's T^2-test for all values of $\eta'\Gamma^{-1}\eta$ which are of practical importance and worse only where both tests have power very close to 1.

A question which comes closer to answering our concerns in practice is whether Hotelling's T^2-test is admissible for testing H_0 against the class of alternatives $\eta'\Gamma^{-1}\eta = \lambda$, with λ a given positive constant. The methods of this paper are completely inadequate for this purpose. In the case $p = 1$, $m = 0$ (Student's t-test with no unknown means as nuisance parameters), the affirmative answer is given in Lehmann and Stein [2]. For $p \geqq 2$ or $m > 0$ the answer is unknown. For $p \geqq 2$ it is not even known whether the appropriate T^2-test is minimax for the problem of testing against a given $\eta'\Gamma^{-1}\eta$, with constant losses a, $b > 0$ for errors of the first and second kinds. An example by the author [4] shows that this does not follow from the invariance of the problem under the full linear group and the minimax property of Hotelling's T^2-test among all invariant tests. The strongest known optimum property of Hotelling's T^2-test seems to be that of Simaika [3] that of all tests whose power depends only on $\eta'\Gamma^{-1}\eta$, it is uniformly most powerful.

Nevertheless, it is clear that, in most applications, Hotelling's T^2-test cannot be substantially improved upon for all η, Γ with $\eta'\Gamma^{-1}\eta$ fixed. For if n/p is large, the test is nearly equivalent to the χ^2-test which results if one knows Γ. This χ^2-test is of course admissible against such alternatives. The proof is essentially given by Wald [5].

For some common multivariate tests of composite hypotheses, the methods of the present paper give no information. For example, let $Y_1, \cdots, Y_n$, with $n \geqq p$, be independently normally distributed random p-dimensional vectors with mean 0 and unknown nonsingular covariance matrix s_0. Suppose we want to test the hypothesis that the first coordinate of the Y_i is independent of the last $(p - 1)$ coordinates, i.e., that $s_{012} = 0$, where

$$s_0 = \begin{pmatrix} s_{011} & s_{012}' \\ s_{012} & s_{022} \end{pmatrix} \tag{41}$$

with s_{011} a 1×1 matrix. The usual test is to accept H_0 if the sample multiple correlation coefficient is small, i.e., if

$$S_{12}' S_{22}^{-1} S_{12} \leqq c S_{11}, \tag{42}$$

where $S = \sum_i Y_i Y_i'$. By a calculation analogous to that of the lemma in Section 4, one can show that the convex cone determined by (42) differs by a set of probability 0 from the intersection of all half-spaces of the form

$$\{s: -c s_{11} + 2\xi' s_{12} - \operatorname{tr} \xi\xi' s_{22} \leqq 0\}, \tag{43}$$

where ξ ranges over the set of all non-zero $(p - 1)$-dimensional vectors. However, the matrix

$$\begin{pmatrix} c & -\xi' \\ -\xi & \xi\xi' \end{pmatrix}$$

is not positive semidefinite for $c < 1$, the only case of interest. A similar argument shows that the methods of this paper can never prove the admissibility of an acceptance region which is a cone when the alternative hypothesis is a subset of the class of normal distributions with mean 0.

REFERENCES

[1] A. BIRNBAUM, "Characterizations of complete classes of tests of some multiparametric hypotheses, with applications to likelihood ratio tests," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 21–36.

[2] E. LEHMANN AND C. STEIN, "The admissibility of certain invariant statistical tests involving a translation parameter," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 473–9.

[3] J. SIMAIKA, "On an optimum property of two important statistical tests," *Biometrika*, Vol. 32 (1941), p. 62.

[4] C. STEIN, "On tests of certain hypotheses invariant under the full linear group," (abstract), *Ann. Math. Stat.*, Vol. 26 (1955), p. 769.

[5] A. WALD, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 426–482.