The Annals of Statistics
1984, Vol. 12, No. 2, 451–469

AN ALTERNATIVE TO STUDENT'S t-TEST FOR PROBLEMS WITH INDIFFERENCE ZONES

BY LAWRENCE D. BROWN[1] AND HAROLD SACKROWITZ

Cornell University and Rutgers University

Consider a sample from a normal population with mean, μ, and variance unknown. Suppose it is desired to test $H_0: \mu \le \mu_0$ versus $H_1: \mu \ge \mu_1$, with the region $H_1^I: \mu_0 < \mu < \mu_1$ being a (nonempty) indifference zone. It is shown that the usual Student's t-test is inadmissible for this problem. An alternative test is proposed.

The two sided problem with indifference region is also discussed. By contrast with the above result, the usual Student's t-test is admissible here. However the two sided version of the alternative test mentioned above does offer some practical advantages relative to the two sided t-test.

A 3-decision version of the two sided problem is also discussed. Here the t-test is inadmissible, and is dominated by the appropriate version of the alternative test.

The results concerning tests are also reformulated as results about confidence procedures.

1. Introduction. Consider the classical problem of testing the mean, μ, of a normal sample having variance, σ^2. In many situations it is possible to identify an indifference region of μ values in the alternative hypothesis adjacent to the null hypothesis. These are values for which the decision to accept is not a significant error.

The idea for such a formulation can be found in the pioneering work of Neyman and Pearson (1933). They wrote

> "There is an essential difference in character between errors of type I and II ... if H_0 is accepted when some alternative [$\mu \in H_1$] is true the consequences that follow will depend upon the nature of [μ] and its difference from H_0. ... Problems may occur in which all type II errors can be ... divided into two classes—those which do not matter and those which do—all the latter being treated as of equal consequence."

Introduction of an indifference zone can be motivated as a crude approximation to a more precise formulation which would give type II errors continuously greater significance as the "distance" increases between the true parameter and the null hypothesis. The latter situation can of course be mathematically modeled and treated by assigning a loss function to type II errors; the loss increasing as the distance increases from the null hypothesis. Procedures would then be compared by comparing their risk functions. In the one-sided problem of testing

Received September 1983; revised December 1983.

[1] Research supported in part by N. S. F. research grant MCS 8200031.

AMS 1980 *subject classifications.* Primary 62F03, secondary 62A99.

Key words and phrases. t-test, indifference zone, confidence intervals, admissibility.

$H_0{:}\mu \le \mu_0$ versus $H_1{:}\mu > \mu_0$ if the loss from type II error were small for $\mu_0 < \mu < \mu_1$ and rose steeply as μ increased past μ_1 then this comparison would lead to much the same choice of procedure as would the easier analysis based on the cruder assumption of an indifference zone $H_1^I{:}\mu_0 < \mu < \mu_1$. An explicit discussion along these lines can be found, for example, in Arrow (1960).

For the one sided problem we propose a test which dominates the usual t-test in the presence of such an indifference zone. Figures for the power presented and discussed here indicate that this test offers a noticeable advantage over the usual t-test only when n is small to moderate and when σ is rather small relative to the length of the indifference region.

For the two sided problem with indifference zone we propose using the two sided version of the above test even though in the classical formulation we show it does not dominate the t-test. Furthermore, as we argue with much precedent, an appropriate solution to the familiar two sided problem involves a three decision formulation; and for such a formulation the appropriate procedure is indeed a combination of two one sided procedures.

Part I: one sided problems.

2. Notation and definitions. In a statistical problem with sample space $\mathscr{X}$ and parameter space Θ it may be desired to test a null hypothesis H_0 versus an alternative $H_1 = \Theta - H_0$. However, there may be a set of values $H_1^I \subset H_1$ which represent deviations from the null hypothesis whose detection would be virtually inconsequential. This is the indifference region.

A test corresponds to a critical function $\varphi{:}\mathscr{X} \to [0, 1]$ describing the probability given $x \in \mathscr{X}$ of rejecting. The power of the test is denoted by

$$\pi_\varphi(\theta) = E_\theta(\varphi(X)).$$

It is primarily desired that the power $\pi_\varphi(\theta)$ be small on H_0. Thus a test is called level α if

$$\sup_{\theta \in H_0} \pi_\varphi(\theta) = \alpha. \tag{2.1}$$

Subject to this restriction, it is desirable for the power to be large on the effective alternative, $H_1 - H_1^I$. Thus, if φ_1 and φ_2 are two level α tests we say φ_2 *dominates* φ_1 if

$$\pi_{\varphi_2}(\theta) \ge \pi_{\varphi_1}(\theta) \quad \text{for} \quad \theta \in H_1 - H_1^I, \tag{2.2}$$

with strict inequality for some $\theta \in H_1 - H_1^I$. We say that φ_2 *completely dominates* φ_1 if (2.2) holds and also

$$\pi_{\varphi_2}(\theta) \le \pi_{\varphi_1}(\theta) \quad \text{for} \quad \theta \in H_0. \tag{2.3}$$

We will say that a level α test φ is *admissible* if there is no other level α test which dominates it.

Note that values of $\pi_\varphi(\theta)$ for $\theta \in H_1^I$ are irrelevant for domination and admissibility considerations. In fact, admissibility as defined by (2.2) is isomorphic to admissibility in a statistical testing problem with null hypothesis H_0

and alternative $H_1 - H_1^I$. A third formulation which leads to a closely related definition of admissibility involves the action space $\{a_0 = \text{accept}, a_1 = \text{reject}\}$ and the loss function

$$L(\theta, a_i) = \begin{cases} 1 & \theta \in H_0, \quad i = 1, \quad \text{or} \\ & \theta \in H_1 - H_1^I, \quad i = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2.4}$$

Admissibility of φ with respect to this loss function means there does not exist another procedure which completely dominates it (apart from the minor technical point that now the strict inequality may occur in (2.3) instead of (2.2)).

Throughout this paper let $X_1, \cdots, X_n$ be a sample from a normal population with mean μ and variance σ^2. Let $\Theta = \{(\mu, \sigma^2): \mu \in \mathbb{R}, \sigma^2 > 0\}$. The customary sufficient statistics are

$$\bar{X} = (1/n) \textstyle\sum_{i=1}^n X_i, \quad S^2 = (1/n) \sum_{i=1}^n (X_i - \bar{X})^2. \tag{2.5}$$

Fix μ_0, μ_1. The null, alternative, and indifference regions for the one sided problem are

$$H_0: \mu \le \mu_0, \quad H_1: \mu > \mu_0 \quad \text{and} \quad H_1^I: \mu_0 < \mu < \mu_1. \tag{2.6}$$

Student's t-test (one-sided) is given by

$$\varphi_t(\bar{x}, s) = \begin{cases} 1 & (\bar{x} - \mu_0)/s > C_\alpha \\ 0 & \text{otherwise.} \end{cases} \tag{2.7}$$

C_α is chosen as $t_\alpha/\sqrt{n-1}$ where t_α can be found in the usual tables of Student's t-distribution for $(n - 1)$ degrees of freedom. Of course, $\pi_{\varphi_t}((\mu_0, \sigma^2)) = \alpha$ for all $\sigma > 0$.

In the classical formulation, for which $H_1^I = \phi$, Student's t-test is Uniformly Most Powerful Unbiased and admissible. However, when $H_1^I \neq \phi$, Student's t-test is not admissible. A test which completely dominates it (when $\alpha < ½$) is described in Section 3 and is easily implemented with the aid of tables contained in Chow (1982).

Note that the indifference region has been carved only out of the alternative hypothesis, and not also out of the null hypothesis. This is a reflection of the asymmetry present in the conventional theory of hypothesis tests which grants primary importance to the significance level of a test and a secondary status to the type II error.

There are situations for which an asymmetric formulation is not appropriate. For example, in testing the mean yields μ_1 and μ_2 of two equally new processes it may be desired to decide either $\mu_1 \ge \mu_2$ or $\mu_2 \ge \mu_1$. (There may also be an indifference region in a problem of this type.) In such a situation the two decisions—whatever they are labeled—possess a symmetric logical status. The asymmetric formulation used in our paper is then not appropriate. (Some of the technical results of our paper do however remain useful after an appropriate reinterpretation.)

It should also be noted that there are existing approaches to the question of how to "test hypotheses," or of what to do instead. The mathematical results of this paper would be of interest in some of these approaches and not in others.

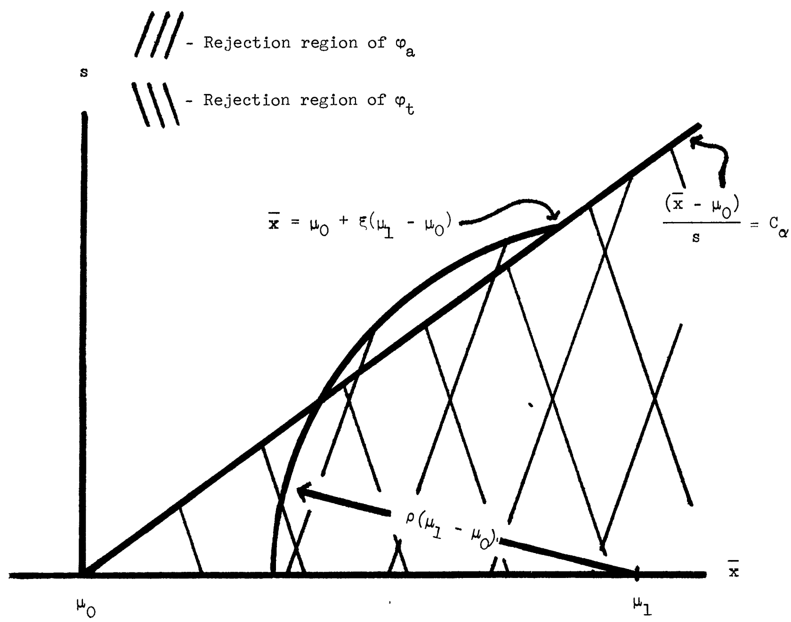

FIG. 3.1. *Rejection regions of φ_t and φ_a.*

3. The new test; description and power. It is convenient to describe and picture the new test in the $(\bar{x}, s)$ coordinate system. The reader should refer to Figure 3.1, which shows both the rejection region of the new test and of the classical one-sided t-test as defined in (2.7).

The new test, with critical function φ_a, is defined by

$$(3.1)\qquad \varphi_a(\bar{x}, s) = \begin{cases} 1 & \text{if} \quad \bar{x} - \mu_0 < \xi(\mu_1 - \mu_0) \quad \text{and} \\ & \qquad s^2 + (\bar{x} - \mu_1)^2 < \rho^2(\mu_1 - \mu_0)^2 \\ & \text{or} \quad \bar{x} - \mu_0 \geq \xi(\mu_1 - \mu_0) \quad \text{and} \quad (\bar{x} - \mu_0)/s > C_\alpha \\ 0 & \text{otherwise.} \end{cases}$$

Chow (1982) contains unique values of ξ, ρ, and C_α for tests at various levels and for various values of n. Precise analytic definitions of ξ and ρ are given in Section 4. C_α is of course the critical value for the usual Student's t-test if based on $(\bar{x} - \mu_0)/s$. Thus, C_α is $t_\alpha/\sqrt{n - 1}$ where t_α is available from ordinary tables for Student's one-sided t-test. (It can be seen from (3.1) or Figure 3.1 that ξ is actually an algebraic function of ρ and C_α.)

The procedure φ_a is not scale invariant in the usual sense. This is due to the fact that a fixed indifference region is not invariant with respect to scale changes of the variable to be observed. However, it is important to note that the procedure φ_a is invariant relative to (what we call) the overall formulation of the problem. That is, making the transformations $\bar{x} \to a\bar{x} + b$, $s \to as$ $(a > 0)$ and $\mu_i \to a\mu_i + b$, $i = 1, 2$ does not change the value of φ_a. Although the classical definition does not include the transformation on μ_i it seems reasonable here and has the following effect. If we make a change of units (for example) from feet to inches,

TABLE 3.1 *Power of level α test—φ_a—when $\mu = \mu_0$ for various values of σ, n, α.*

n	$\alpha = .05$			$\alpha = .025$		
	$\sigma/(\mu_1 - \mu_0) =$			$\sigma/(\mu_1 - \mu_0) =$		
	.5	1	2	.5	1	2
5	.027	.049	.05	.009	.024	.025
7	.036	.05	.05	.014	.025	.025
10	.043	.05	.05	.020	.025	.025
15	.05	.05	.05	.024	.025	.025
20	.05	.05	.05	.025	.025	.025

our decision will remain the same provided only that we change from feet to inches in the description of the indifference zone. From a practical point of view, this would seem to be quite satisfactory.

It is of interest to compare the power of φ_a with that of φ_t. Lengthy tables of the power are available in Chow (1982), from which the following is taken. Table 3.1 shows the power of the alternative level α test (φ_a) when $\mu = \mu_0$ for a few values of σ^2, n and $\alpha = .025$ or .05.

All entries in Table 3.1 are no more than α. (To a sufficient number of decimals, all entries would be strictly less than α.) This is guaranteed by the construction of φ_a and Theorem 5.1. What is of more interest to note is that the power is near α, the power of φ_t, unless both n and $\sigma/(\mu_1 - \mu_0)$ are rather small. However if both are small, then the power may be much less than α. Thus for such values of n and $\sigma/(\mu_1 - \mu_0)$ the probability of type I error (= power) of φ_a is markedly superior to that of φ_t.

On the alternative of interest, $\mu \geq \mu_1$, the power of φ_a is very slightly greater than that of φ_t. However the difference in power is practically negligible (less than .001 when $n \geq 4$) in all the cases examined in Chow (1982). Thus the performance of φ_a on the alternative of interest is technically superior to that of φ_t, but it appears that for all practical purposes the tests perform equally well on this region.

Two other qualitative features of the power function of φ_a may be worth noting. (i) On most of the indifference zone, $\mu_0 < \mu < \mu_1$, the power of φ_a is less than that of φ_t. This would of course be undesirable if points of the indifference zone were instead to be considered part of the true alternative. The difference in power here is negligible unless both n and $\sigma/(\mu_1 - \mu_0)$ are rather small (as in Table 3.1) and μ is not too close to μ_1. (ii) For values of $\mu < 0$ the pattern of Table 3.1 persists only when μ is moderately close to μ_0—for example when $(\mu - \mu_0)/\sigma = -\frac{1}{2}$, all the figures in Chow (1982) show a practically negligible difference in power (less than .001).

Section 5 contains further comment on the practical implications of these results on the power of φ_a as compared to φ_t.

4. The new test; theoretical results. This section contains a formal description of the new test and a statement and proof of the fact that it dominates Student's t-test in the indifference zone formulation.

It is more natural to state and prove the results of this section in terms of the usual canonical parameters and statistics of the exponential family corresponding to $(\bar{X}, S^2)$. It is also convenient to *assume that* $\mu_0 = 0$, and we shall do so throughout this section without loss of generality. Thus, let

$$V = \sum X_i = n\bar{X}, \quad W = \sum X_i^2 = n(S^2 + \bar{X}^2),$$

and let

$$\eta = \mu/\sigma^2 \quad \text{and} \quad \theta = -\tfrac{1}{2}\sigma^2.$$

The joint density function of (V, W) in terms of the natural parameters $(\eta, \theta) \in \mathscr{N} = \{(\eta, \theta): \theta < 0\}$ is

$$(4.1) \quad f_{\eta,\theta}(v, w) = \begin{cases} K(\eta, \theta)(w - v^2/n)^{(n-3)/2}\exp(\eta v + \theta w) & \text{if} \quad w \geq v^2/n \\ 0 \quad \text{otherwise} \end{cases}$$

where

$$K(\eta, \theta) = (\sqrt{n\pi}\,\Gamma((n-1)/2))^{-1}(-2\theta)^{n/2}\exp(n\eta^2/4\theta).$$

For simplicity write

$$f_{\eta,\theta}(v, w)\, dv\, dw = K(\eta, \theta)\exp(\eta v + \theta w)\ \omega(dv, dw)$$

where ω is defined by

$$\omega(E) = \iint_{E\cap\{(v,w): w \geq v^2/n\}} \left(\frac{w - v^2}{n}\right)^{(n-3)/2} dv\, dw.$$

In terms of (V, W) the critical function of the usual t test when $\alpha < \tfrac{1}{2}$ is given by

$$(4.2) \qquad \tilde{\varphi}_t(v, w) = \begin{cases} 1 & \text{if} \quad v > 0 \quad \text{and} \quad w \leq k_\alpha v^2 \\ 0 & \text{otherwise} \end{cases}$$

where $k_\alpha = (1 + C_\alpha^2)/nC_\alpha^2$. The rejection region for this test is depicted in Figure 4.1, along with the rejection region for the new test. In terms of the parameters (η, θ) the basic hypotheses are $\tilde{H}_0 = \{(\eta, \theta) \in \mathscr{N}: \eta \leq 0\}$ and $\tilde{H}_1 = \{(\eta, \theta) \in \mathscr{N}: \eta > 0\}$ and the indifference zone $H_1^I: 0 < \mu < \mu_1$ is

$$\tilde{H}_1^I = \{(\eta, \theta): 0 < \eta < -2\mu_1\theta\}.$$

This region is also shown in Figure 4.1.

In order to define the new test, introduce the sets

$$E_0(c) = \{(v, w): \max(2\mu_1 v + c, v^2/n) \leq w < k_\alpha v^2\}$$

$$E_1(c) = \{(v, w): k_\alpha v^2 \leq w \leq 2\mu_1 v + c\}.$$

Again see Figure 4.1. It is clear from the geometry and the absolute continuity of ω that there is a unique value, call it c^*, for which

$$\omega(E_0(c^*)) = \omega(E_1(c^*)).$$

Define the point (v^*, w^*) by $\{(v^*, w^*)\} = \overline{E_0(c^*)} \cap E_1(c^*)$.

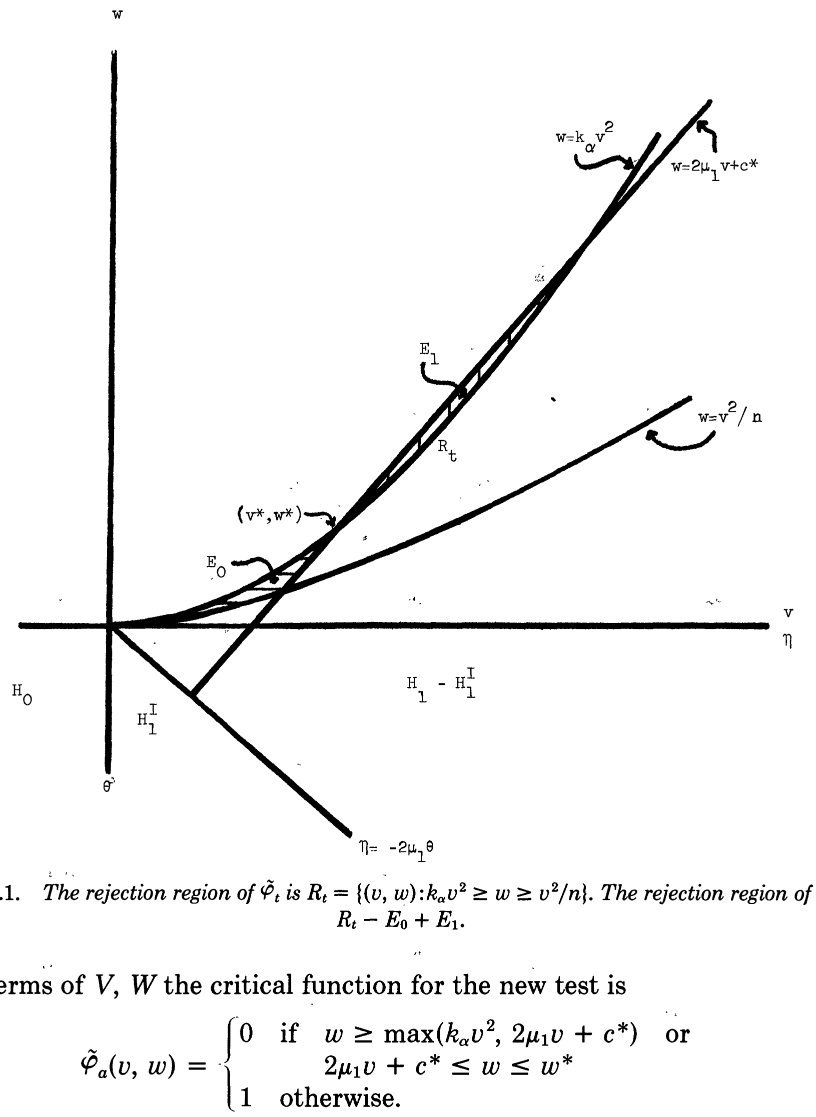

FIG. 4.1. *The rejection region of* $\tilde{\varphi}_t$ *is* $R_t = \{(v, w): k_\alpha v^2 \geq w \geq v^2/n\}$. *The rejection region of* $\tilde{\varphi}_a$ *is* $R_t - E_0 + E_1$.

In terms of V, W the critical function for the new test is

$$\tilde{\varphi}_a(v, w) = \begin{cases} 0 & \text{if} \quad w \geq \max(k_\alpha v^2, 2\mu_1 v + c^*) \quad \text{or} \\ & \quad 2\mu_1 v + c^* \leq w \leq w^* \\ 1 & \text{otherwise.} \end{cases} \tag{4.3}$$

In words, the rejection region for $\tilde{\varphi}_a$ is that of φ_t except for $E_0(c^*)$, and with the addition of $E_1(c^*)$. ($E_i(c^*)$ thus denotes the set where $\tilde{\varphi}_t$ and $\tilde{\varphi}_a$ differ and $\tilde{\varphi}_a$ decides on H_i.) It can easily be seen that in terms of $\bar{x}$, s the new test is described as in Section 3.

THEOREM 4.1. *If* $\alpha < 1/2$ *the usual one-sided t-test is inadmissible in the indifference formulation* (2.6) *of Section* 2. *The test* $\tilde{\varphi}_a$ *completely dominates* $\tilde{\varphi}_t$ *and is admissible.*

PROOF. Note that, since $(v^*, w^*) = \bar{E}_0 \cap E_1$,

$$\inf\{e^{\eta v + \theta w}: (v, w) \in E_0\} = e^{\eta v^* + \theta w^*} = \sup\{e^{\eta v + \theta w}: (v, w) \in E_1\} \tag{4.4}$$

for all $(\eta, \theta) \in \tilde{H}_0$, and

$$\sup\{e^{\eta v+\theta w}:(v, w) \in E_0\} = e^{\eta v^*+\theta w^*} = \inf\{e^{\eta v+\theta w}:(v, w) \in E_1\} \tag{4.5}$$

for all $(\eta, \theta) \in \tilde{H}_1 - \tilde{H}_1^I$. To verify (4.5) it is important to observe that the line $2\mu_1 v + c^*$ (which bounds E_1 and E_0 on opposite sides) contains (v^*, w^*) and is perpendicular to the line $\eta = -2\mu_1\theta$ which forms the lower boundary of $H_1 - H_1^I$.

Hence, for $(\eta, \theta) \in \tilde{H}_0$ we have for the difference in power

$$\begin{aligned}\tilde{\pi}_{\tilde{\varphi}_t}(\eta, \theta) - \tilde{\pi}_{\tilde{\varphi}_a}(\eta, \theta) &= E_{\eta,\theta}(\tilde{\varphi}_t(V, W) - \tilde{\varphi}_a(V, W)) \\ &= K(\eta, \theta)\left[\int_{E_0} e^{\eta v+\theta w}\omega(dv\, dw) - \int_{E_1} e^{\eta v+\theta w}\omega(dv\, dw)\right] \\ &\geq K(\eta, \theta)e^{\eta v^*+\theta w^*}(\omega(E_0) - \omega(E_1)) = 0.\end{aligned}$$

And, for $(\eta, \theta) \in \tilde{H}_1 - \tilde{H}_1^I$, we have similarly

$$\begin{aligned}\tilde{\pi}_{\tilde{\varphi}_t}(\eta, \theta) - \tilde{\pi}_{\tilde{\varphi}_a}(\eta, \theta) &= K(\eta, \theta)\left[\int_{E_0} e^{\eta v+\theta w}\omega(dv\, dw) - \int_{E_1} e^{\eta v+\theta w}\omega(dv\, dw)\right] \\ &\leq K(\eta, \theta)e^{\eta v^*+\theta w^*}(\omega(E_0) - \omega(E_1)) = 0.\end{aligned}$$

This proves that $\tilde{\varphi}_a$ completely dominates $\tilde{\varphi}_t$.

The admissibility of $\tilde{\varphi}_a$ follows easily from the Theorem of Stein (1956). (See also Birnbaum, 1955.) The key observation here is that the acceptance region for $\tilde{\varphi}_a$ is the intersection of the domain of (V, W) with half spaces whose outward normals (eventually) lie in the effective alternative. □

REMARK. Lehmann and Stein (1948) show that for $\alpha \geq 1/2$ the t-test is uniformly most powerful. Hence the assumption in the theorem that $\alpha < 1/2$ is necessary as well as sufficient for validity of the conclusion of the theorem.

It should be clear from the above proof how to construct a variety of other tests which dominate the t-test in the indifference formulation. Here is a formal statement.

COROLLARY 4.2. *Let F_0, F_1 be two measurable sets such that*

(i) *F_0 and F_1 are separated by some line with slope $2\mu_1$ and by some line with slope 0, with F_0 below and to the left of F_1,*
(ii) *F_0 is below $w = k_\alpha v^2$ and F_1 is above $w = k_\alpha v^2$,*
(iii) *$\omega(F_0) = \omega(F_1) \neq 0$.*

Let

$$\varphi(v, w) = \begin{cases} 1 & \text{if } \tilde{\varphi}_t(v, w) = 1 \text{ and } (v, w) \notin F_0, \\ 1 & \text{if } (v, w) \in F_1 \\ 0 & \text{otherwise.}\end{cases}$$

Then φ completely dominates $\tilde{\varphi}_t$ in the indifference formulation.

PROOF. The proof follows the pattern of the proof of Theorem 4.1. □

It is of considerable interest to discover whether there may exist some alternative test dominating φ_t which noticeably improves on the power of φ_t (and φ_a) over some substantial part of $H_1 - H_1^I$. We have not found such a test, and examination of Figure 4.1 and the power figures for φ_a has made us skeptical concerning the existence of such a test.

5. Discussion. The tables of power summarized in Section 3 withdraw much of the potential direct practical importance of the theoretical result contained in Theorem 4.1. It is clear from the figures in Section 3 that the new procedure, φ_a, is virtually equivalent to the usual t-test unless n is moderately small or $\sigma/(\mu_1 - \mu_0)$ is quite small. (This is also to be expected on theoretical grounds since as n grows the t-test begins to mimic closely the normal test for known σ, which is U.M.P.) Still, the t-test is frequently used in situations where n is indeed small, and occasionally $\sigma/(\mu_1 - \mu_0)$ may be quite small. Hence, it is worthwhile to further discuss some aspects of the figures in Section 3.

In the case of φ_a relative to φ_t, there is no appreciable gain in power over $H_1 - H_1^I$. The only real difference in performance is the reduction in power over H_0. Is such a reduction of sufficient merit to justify the postulation of an indifference zone and the surrender of similarity?

It seems to us that this question does not have an unequivocal answer. Two-decision procedures have a variety of applications. One class of applications involve what Fisher (1955) called "acceptance procedures" as a "technological and commercial apparatus." This is to be contrasted with Fisher's original conception of a significance test as a tool in "the work of scientific discovery by physical or biological experimentation." (Again, see Fisher, 1955.) Of course, both of these categories are fairly general and each involve a variety of possible criteria and goals for the statistical procedures. There are also a variety of other current uses of two decision procedures.

For a situation fitting Fisher's original conception of a test as a tool in scientific discovery, it may well be that the only important feature of the type I error is its supremum which is of course the level of the test. Thus, here, reduced power only over a (small) part of H_0 seems a questionable benefit. (This would be so even if it were possible to identify a valid indifference region in H_1, which would rarely be the case in applications of this sort.)

In the case of acceptance procedures, any rejection may be followed by a costly reevaluation or realignment. If so, it would be an economic advantage to control the rate of type I error over any part of H_0 which might obtain. (In such cases the role of the significance level is often reduced to a mere description of one facet of the procedure being used, rather than a fundamental feature of it. Also, these are often situations where one should contemplate the use of procedures at various levels, not merely the conventional .05, .01, or etc.) This, then, is a type of situation where it could be worthwhile identifying a valid indifference zone in H_1 and then adopting the procedure, φ_a, which reduces probability of type I error over part of H_0 (in return for reduced power over part of H_1^I).

In Section 9 we consider the construction of two-sided confidence intervals. Here some of the hesitation disappears. Suppose it is not a serious error to cover parameter values μ' with $|\mu' - \mu| < \delta$ where μ represents the true value. Our proposed two-sided confidence intervals have greater probability of covering the true value and not significantly greater probability of covering values of μ' with $|\mu' - \mu| > \delta$. Again, the probability of correct coverage is noticeably greater for smallish n and σ/δ. Thus for the construction of confidence intervals the introduction of an indifference zone leads to a more clearcut advantage.

In summary, Theorem 4.1 shows that it is in principle possible to take advantage of the identification of an indifference zone. This may also be true in situations other than the ones discussed in this paper. In the Student's t problem under consideration the advantage is somewhat limited in practice, as shown by the power function of φ_a described in Section 3. Whether it is really desirable to substitute φ_a for φ_t must depend on many factors. There are relatively concrete factors such as the sample size and a (crude) a priori determination of the likely range of values of $\sigma/(\mu_1 - \mu_0)$ (and possibly of $(\mu - \mu_0)/(\mu_1 - \mu_0)$). But it is equally important to consider the purpose and philosophy of the particular statistical test in question, and the true practical meaning in this context of an indifference zone and of the decisions "Accept" and "Reject."

Part II: two-sided tests.

6. The customary formulation: definitions and theoretical results. It is possible to routinely extend the one-sided formulation of Section 2 to the two-sided case. In this routine extension $\mu_1 \le \mu_0 \le \mu_2$ are given constants, $H_0 = \{(\mu, \sigma^2): \mu = \mu_0\}$, $H_1 = \{(\mu, \sigma^2): \mu \ne \mu_0\}$ and $H_1^I = \{(\mu, \sigma^2): \mu_1 \le \mu < \mu_0 \text{ or } \mu_0 < \mu \le \mu_2\}$. To simplify the notation set $\mu_0 = 0$, without loss of generality. Let ψ_t denote the critical function of the usual symmetric two-sided t-test for this problem:

$$\psi_t(\bar{x}, s) = \begin{cases} 1 & \text{if} \quad |\bar{x}|/s > C_{\alpha/2} \\ 0 & \text{otherwise.} \end{cases}$$

Then, another test, with critical function ψ, *dominates* ψ_t if

$$\Pi_\psi((0, \sigma^2)) \le \Pi_{\psi_t}((0, \sigma^2)) \equiv \alpha \quad \text{for} \quad \sigma^2 > 0$$

and

$$\Pi_\psi((\mu, \sigma^2)) \ge \Pi_{\psi_t}((\mu, \sigma^2)) \quad \text{for} \quad (\mu, \sigma^2) \in H_1 - H_1^I$$

with strict inequality for at least one value of $(\mu, \sigma^2) \in H_0 \cup (H_1 - H_1^I)$. If no test dominates ψ_t then it is admissible.

Here is the major theoretical result in this setting.

THEOREM 6.1. *ψ_t is admissible in the above setting.*

(The proof is fairly technical though of some interest for its relation to other general results on testing in exponential families, such as those in Marden and Perlman (1980) and Marden (1981). It is deferred to the appendix.)

7. The customary formulation—an alternative to the t-test. The admissibility of ψ_t does not necessarily compel its use. The proof of Theorem 6.1 shows only that for any level α procedure $\psi \neq \psi_t$, one will have $\Pi_{\psi}((\mu, \sigma^2)) < \Pi_{\psi_t}((\mu, \sigma^2))$ for some values of $(\mu, \sigma^2) \in H_1 - H_1^I$ *with* σ^2 *sufficiently large*. But when σ^2 is very large, both ψ_t and ψ will have power nearly α.

In searching for an alternative to ψ_t the obvious choice is the two-sided version of φ_a. Suppose for simplicity $\mu_1 = -\mu_2$. Then this obvious choice is the level α two-sided procedure, ψ_a related to the level $\alpha/2$ one-sided procedure, φ_a, by

$$\psi_a(\bar{x}, s) = \varphi_a(|\bar{x}|, s).$$

(When $\mu_1 \neq -\mu_2$ the same principle can be followed of combining two level $\alpha/2$ one-sided tests. The resulting procedure will not, of course, be symmetric.) It can be easily shown that ψ_a is admissible.

The power of ψ_a at $(0, \sigma^2) \in H_0$ is exactly twice that of φ_a at the same point. Hence powers of ψ_a on H_0 at levels .05 and .1 may be found by doubling the figures in Table 3.1. Obviously ψ_a is level α as a consequence of Theorem 4.1.

Tables of the difference in power, $\Pi_{\psi_a}((\mu, \sigma^2)) - \Pi_{\psi_t}((\mu, \sigma^2))$ for $\mu \neq 0$ may be found in Chow (1982). These tables show that this difference in power is indeed negative for μ sufficiently near μ_1. However *all differences in power for* $(\mu, \sigma^2) \in H_1 - H_1^I$ *appear to be negligible* (less than .001).

From a practical perspective, within this customary formulation, ψ_a may be substituted for ψ_t. As in the one-sided situation, one then sees a reduction in type I error rate for small to moderate values of n and smallish values of $\sigma/(\mu_1 - \mu_0)$ and the type II error rate of over $H_1 - H_1^I$ is virtually unaltered. In this sense the two-sided case is a duplicate of the one-sided case, and the discussion of Section 5 applies to this two-sided case as well.

8. Alternate formulations: discussion and results. Concerning two-sided testing problems like the two-sided t-problem, Lehmann (1950) wrote,

> "It seems that in nearly any problem of this kind that would arise in practice one would want to decide when rejecting the hypothesis, whether the true parameter value lies above or below the hypothetical one.... It would therefore seem most natural to formulate such problems as 3-decision problems."

To emphasize this point of view (with which we agree) we present another quotation; this one is from Cox and Hinkley (1974, page 106).

> "In the overwhelming majority of applications it is essential to consider the direction of departure in interpreting the result of the significance test; that is, to conclude that there is evidence that $\theta > \theta_0$ when in fact $\theta < \theta_0$ would normally be very misleading. Yet the composite power function treats such conclusions in the same way as the correct conclusion.
>
> "This argument shows that we are really involved with two tests, one to examine the possibility $\theta > \theta_0$ and the other for $\theta < \theta_0$. [The recommended procedure is essentially] the commonly used procedure of

forming a critical region of size α by combining two regions each of size $\sigma/2$, one in each direction. It is different in many cases from the unbiased region."

These authors were addressing only the two-sided problem without indifference region, but their remarks have equal force also when there is such a region. In that case there are three possible decisions: d_0 = accept H_0, d_- = decide $\mu < \mu_0 = 0$, and d_+ = decide $\mu > \mu_0 = 0$. There are 5 regions of the parameter space—$(-\infty, \mu_1)$, $[\mu_1, 0)$, $\{0\}$, $(0, \mu_2]$, (μ_2, ∞).

One may introduce suitable admissibility criteria involving error probabilities as was done in Section 2. However it is slightly more convenient here to use a loss function formulation. Thus, dominance and admissibility considerations may reasonably be based on one of the loss function L_1 or L_2 summarized in Table 8.1. With either L_1 or L_2 one could define the *level* of a procedure ρ as $\sup R_i((\mu, \sigma^2), \rho)$ where, as usual, $R_i((\mu, \sigma^2), \rho) = E_{\mu,\sigma^2}(L_i((\mu, \sigma^2), \rho))$.

This table can be compared with the loss structure corresponding to the formulation in Section 6, which is summarized in Table 8.2.

Both loss functions L_1 and L_2 seem to us quite reasonable for this three decision indifference formulation. The loss function L_1 is given in Sobel and Wald (1949), who presented two-sided, three decision formulation with indifference zones for a sequential testing problem. The loss function L_2 appears to us to be the correct indifference zone analog of the loss functions presented in Lehmann (1957). Fortunately the theoretical results and the qualitative numerical results agree here for both these loss functions, so at present we need not choose between them.

The obvious extensions of ψ_t and ψ_a to this setting are ρ_t and ρ_a with ρ_t given at level α by

$$\rho_t((\bar{x}, s)) = \begin{cases} d_+ & \text{if} \quad \bar{x} > 0 \quad \text{and} \quad \varphi_t(\bar{x}, s) = 1 \\ d_0 & \text{if} \quad \varphi_t(\bar{x}, s) = 0 \\ d_- & \text{if} \quad \varphi_t^{(-)}(\bar{x}, s) = \varphi_t(-\bar{x}, s) = 1 \end{cases}$$

TABLE 8.1 *Values of $L_1(L_2)$.*

Decision	True Parameter Value: $\mu\varepsilon$				
	$(-\infty, \mu_1)$	$[\mu_1, 0)$	$\{0\}$	$(0, \mu_2]$	(μ_2, ∞)
d_-	0 (0)	0 (0)	1 (1)	1 (1)	1 (2)
d_0	1 (1)	0 (0)	0 (0)	0 (0)	1 (1)
d_+	1 (2)	1 (1)	1 (1)	0 (0)	0 (0)

TABLE 8.2 *Values of the loss function for "the customary formulation."*

Decision	True Parameter Value: $\mu\varepsilon$				
	$(-\infty, \mu_1)$	$[\mu_1, 0)$	$\{0\}$	$(0, \mu_2]$	(μ_2, ∞)
Accept	1	0	0	0	1
Reject	0	0	1	0	0

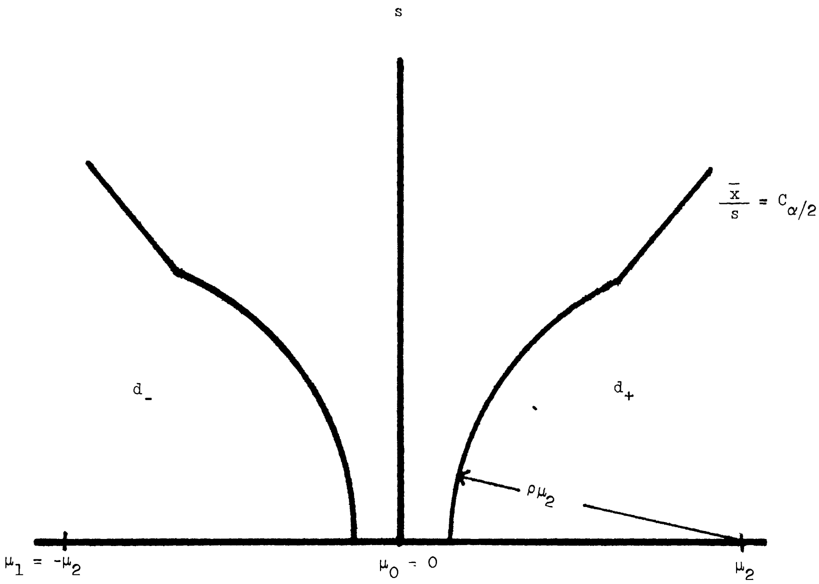

FIG. 8.1. *Decision regions for ρ_a.*

where φ_t is the level $\alpha/2$ t-test. ρ_a is defined by

$$\rho_a((\bar{x}, s)) = \begin{cases} d_+ & \text{if } \varphi_a^{(+)}(\bar{x}, s) = 1 \\ d_0 & \text{if } \varphi_a^{(+)}(\bar{x}, s) = 0 = \varphi_a^{(-)}(\bar{x}, s) \\ d_- & \text{if } \varphi_a^{(-)}(\bar{x}, s) = 1 \end{cases}$$

where $\varphi_a^{(+)}$ ($\varphi_a^{(-)}$) is the level $\alpha/2$ alternate procedure for testing $\mu < 0$ versus $\mu > 0$ ($\mu > 0$ versus $\mu < 0$) with indifference zone $0 < \mu \le \mu_2$ ($\mu_1 \le \mu < 0$). See Figure 8.1.

Here is the main theoretical result for this setting.

THEOREM 8.1. *With respect to either L_1 or L_2, ρ_a is admissible and dominates ρ_t. (Hence, of course, ρ_t is inadmissible.)*

PROOF. The admissibility of ρ_a can be proved from Stein (1956) as was the admissibility of φ_a. For loss function L_2 the dominance of ρ_a over ρ_t follows from that of φ_a over φ_t upon application of the methods of Lehmann (1957). Thus, let $\varphi_a^{(-)} = \varphi_a(-\bar{x}, s)$, etc. Then if $\mu < \mu_1$,

$$E_{(\mu,\sigma^2)}(L_2((\mu, \sigma^2), \rho_a)) = \Pi\varphi_a^{(+)}(\mu, \sigma^2) + (1 - \Pi\varphi_a^{(-)}(\mu, \sigma^2))$$
$$\le \Pi_{\varphi_t}(\mu, \sigma^2) + (1 - \Pi\varphi_t^{(-)}(\mu, \sigma^2)) = E_{(\mu,\sigma^2)}(L_2((\mu, \sigma^2), \rho_t))$$

by Theorem 4.1. The desired comparisons of values of the risk function on other portions of the parameter space can be obtained similarly. The proof for L_1 is also similar to the above since

$$E_{(\mu,\sigma^2)}(L_1((\mu, \sigma^2), \rho_a)) = (1 - \Pi\varphi_a^{(-)}(\mu, \sigma^2)), \quad \text{etc.} \quad \Box$$

(Note that ρ_a actually dominates ρ_t also in the sense of decreasing $\Pr_{\mu,\sigma^2}\{d_+\}$ for $\mu \in (-\infty, \mu_1)$, etc.)

The preceding theoretical result can leave one slightly more reassured that ρ_a is to be preferred to ρ_t than was the case in comparing ψ_a to ψ_t—there is no need here to rely on the partial results for power in the numerical tables. However, it remains true here, as in previous cases, that the only appreciable differences in risk between ρ_a and ρ_t seem to occur when $\mu = 0$, n is small to moderate and $\sigma/(\mu_0 - \mu_1)$ or $\sigma/(\mu_2 - \mu_0)$ is rather small. These differences in risk may again be obtained from Table 3.1. If we restrict attention to the symmetric case which has $\mu_1 = -\mu_2$ then the difference in power is 2((entry in Table 3.1) $- \alpha$). Consequently, the discussion of Section 5 still applies here with obvious alterations to fit the different formulation.

9. Confidence intervals. Conditionally on S the acceptance regions for the one-sided test φ_a are translation invariant intervals so long as $\xi < 1$. (This occurs for $\alpha = .025$ and $\alpha = .05$ except for $\alpha = .025$ and $n = 2, 3$.) This is also true regarding ψ_a in the two-sided case. Hence in either case these procedures may routinely be used to form confidence intervals. As the two-sided case is of more interest, we will concentrate on that. We also discuss the relation of Theorem 8.1 to these confidence intervals.

Let ψ_a denote the level α procedure of Section 7 with $\mu_0 = 0$, as before, and $-\mu_1 = \mu_2 = \delta$. Define the confidence set

$$I_{1-\alpha}(\bar{x}, s) = \{\mu : \psi_a(\bar{x} - \mu, s) = 0\}.$$

Then $I_{1-\alpha}(\bar{x}, s)$ forms a system of location invariant $(1 - \alpha)$ confidence intervals for μ. Although in the usual sense these procedures are not scale invariant, they are invariant relative to the overall formulation (as discussed in Section 3 of the confidence interval model). These confidence intervals have the following properties:

(1) $$\inf_{\sigma^2} \Pr_{\mu,\sigma^2}\{\mu \in I_{1-\alpha}\} = 1 - \alpha;$$

so the intervals truly have confidence coefficient $(1 - \alpha)$. However,

(2) as can be seen in Table 3.1, if n is small to moderate and σ/δ is small then $\Pr_{\mu,\sigma^2}\{\mu \in I_{1-\alpha}\}$ exceeds $(1 - \alpha)$ by a noticeable amount. Hence the confidence procedure may have a markedly better probability than $(1 - \alpha)$ of containing the true value. (The usual t intervals have exactly probability $(1 - \alpha)$ no matter what (μ, σ^2) are.)

(3) On the other hand, the numerical results indicate that if $|\mu' - \mu| > \delta$ then

$$|\Pr_\mu\{\mu' \in I_{(1-\alpha)}\} - \Pr_\mu\{\mu' \in T_{(1-\alpha)}\}| < .001$$

where $T_{1-\alpha}$ denotes the usual level α two-sided Student's t confidence interval. Hence the $I_{1-\alpha}$ intervals have virtually the same probability of covering false values of μ' not too close to μ as do the T intervals.

(4) The I intervals are noticeably deficient in classical terms with respect to the T intervals only in the respect that they have a higher probability of covering false values of μ' near μ (i.e. $|\mu' - \mu| < \delta$).

In many circumstances it is a minor or even negligible disadvantage to cover false values of μ' which are near μ. On the other hand it is a definite advantage to increase the probability of including the true value (other important factors being equal).

Thus, in forming confidence intervals it is a clear advantage to use the intervals $I_{1-\alpha}$ in place of $T_{1-\alpha}$ if one can identify a zone of the form $|\mu' - \mu| < \delta$ for which one is indifferent to false coverage. The advantage increases as n and σ/δ decrease.

The preceding discussion relates to the classical formulation aiming to increase the probability of covering the true value and decrease the probability of covering false values. The inadmissibility result of Theorem 8.1 is relevant to a modification of this formulation. We now present this modification as a logical complement to Theorem 8.1. We are not at present convinced as to the utility, or disutility, of this formulation in statistical practice.

Let (μ, μ') be a pair of possible mean values. μ represents the true value, and μ' represents some other value which may be covered. If $\mu' \neq \mu$ then μ' is a false value. Because of the indifference formulation we will be primarily concerned only with values (μ, μ') for which $\mu' = \mu$ or $|\mu' - \mu| > \delta$.

Let V be a system of $(1 - \alpha)$ confidence intervals—that is, $V(\bar{x}, s) = [v_1(\bar{x}, s), v_2(\bar{x}, s)]$. The objective is now to make large

$$\Pr_\mu\{\mu \in V\}; \tag{9.1}$$

and for $\mu' \geq \mu + \delta$ ($\mu' \leq \mu - \delta$) to make small

$$\Pr_\mu\{\mu' \leq v_2(\bar{x}, s)\} \quad (\Pr\{\mu' \geq v_1(\bar{x}, s)\}). \tag{9.2}$$

This latter criterion should be interpreted as follows: If $\mu + \delta < \mu'$ then it is an equally serious mistake when μ is true to (i) state a confidence interval containing μ' (i.e. $v_1 \leq \mu' \leq v_2$) or to (ii) state a confidence interval falling to the right of μ' (i.e. $\mu' < v_1$. This of course implies, in particular, that $\mu \notin V$.) A symmetric criterion holds when $\mu - \delta > \mu'$. Condition (9.2) above, is what distinguishes this formulation from the classical formulation. It must be based on the premise that for any μ' significantly to the right of the true value, μ, it is as serious a deficiency to state an interval clear to the right of μ' as it is to state an interval which contains μ'.

THEOREM 9.1 *In the preceding formulation the usual Student's t intervals, $T_{(1-\alpha)} = (t_1, t_2)$ are inadmissible, and are dominated by the admissible intervals $I_{(1-\alpha)} = (i_1, i_2)$. Thus, in particular*

$$1 - \alpha = \Pr_\mu\{\mu \in T_{(1-\alpha)}\} \leq \Pr_\mu\{\mu \in I_{(1-\alpha)}\} \tag{9.3}$$

and, for $\mu + \delta < \mu'$

$$\Pr_\mu\{\mu' \leq t_2\} \geq \Pr_\mu\{\mu' \leq i_2\}, \tag{9.4}$$

and for $\mu - \delta > \mu'$

$$\Pr_\mu\{\mu' \geq t_1\} \geq \Pr_\mu\{\mu' \geq i_1\}. \tag{9.5}$$

In addition, it is also true that for $\mu < \mu' \leq \mu + \delta$

$$\Pr_\mu\{\mu' < t_1\} \geq \Pr_\mu\{\mu' < i_1\} \tag{9.6}$$

with a similar statement being valid for $\mu - \delta \leq \mu' < \mu$.

(Property (9.6) is also a desirable feature of the intervals I. At least, it is not undesirable—one could reasonably argue that it is irrelevant what happens when $|\mu - \mu'| \leq \delta$ due to the indifference formulation.)

PROOF. The proof relies on Theorem 8.1 and a standard trick. For example, for the situation $\mu + \delta < \mu'$ one has

$$\begin{aligned}\Pr_\mu\{\mu' \leq i_2(\bar{X}, S)\} &= \Pr_\mu\{0 \leq i_2(\bar{X}, S) - \mu'\} = \Pr_\mu\{0 \leq i_2(\bar{X} - \mu', S)\} \\ &= \Pr_{\mu-\mu'}\{0 \leq i_2(\bar{X}, S)\} = \Pr_{\mu-\mu'}\{\rho_a(\bar{X}, S) = d_0 \text{ or } d_+\}.\end{aligned}$$

Similar statements are, naturally, true for T and ρ_t, and (9.4) then follows from Theorem 8.1. The rest of the theorem is proved similarly. □

APPENDIX

This section contains the material necessary to establish Theorem 6.1. The following simple lemma is used several times in the proof of Theorem 6.1. The variables V, W and other quantities which appear here are defined in Section 4.

LEMMA A1. *Let $m: R \times (-\infty, 0) \times (0, \infty) \to R$ and $\rho: R \times (0, \infty) \to [-1, 1]$ (both measurable). Let $S \subset R \times (0, \infty)$ (again, measurable). Suppose*

(i) $\forall (v, w) \in S$, *$m(v, w, \cdot)$ is nonnegative and monotone nondecreasing,*
(ii) $\sup\{|m(v, w, \theta)| : w \geq v^2/n, (v, w) \notin S, -1 \leq \theta < 0\} < \infty$,
(iii) $\rho(v, w)m(v, w, \theta) \leq 0 \ \forall v, w \in S$,
(iv) $\lim_{\theta \nearrow 0} m(v, w, \theta) = m(v, w, 0)$ *exists and is finite for all* $(v, w) \in R \times (-\infty, 0)$,
(v) $\omega(S^c) < \infty$.

Then

$$\begin{aligned}\lim_{\theta \nearrow 0} &\int \rho(v, w)m(v, w, \theta)\omega(dv\,dw) \\ &= \int \rho(v, w)m(v, w, 0)\omega(dv\,dw) \geq -\infty\end{aligned} \tag{1}$$

(thus the limit in (1) *exists as an extended real number.)*

A symmetric result holds if the inequality in (iii) is reversed.

PROOF. Split the region of integration in (1) as $\int = \int_S + \int_{S^c}$. Then, $\lim_{\theta \nearrow 0} \int_{S^c}$ exists and has the desired value by condition (ii), (iv) and the boundness of ρ. And, $\lim_{\theta \nearrow 0} \int_S$ exists and has the desired value by (i), (iii), and (iv). □

PROOF OF THEOREM 6.1. Let φ be the critical function of a level α test at least as good as the t-test with critical function ψ_t. Let

$$\rho(v, w) = \psi_t(v, w) - \varphi(v, w)$$

and

$$\Delta(\eta, \theta) = \int \rho(v, w) e^{\eta v + \theta w} \omega(dv\, dw) = K^{-1}(\eta, \theta)(\Pi_{\psi_t}(\eta, \theta) - \Pi_{\varphi}(\eta, \theta)). \tag{2}$$

Let $a \geq \mu_2 V(-\mu_1)$ so that $H_1^I \supset \{(\mu, \sigma^2) : \mu = \pm a\}$. Recall that $\mu = -\eta/2\theta$. Since φ is at least as good as ψ_t, (2) yields

$$\Delta(0, \theta) \geq 0, \quad \Delta(-2a\theta, \theta) \leq 0, \quad \Delta(2a\theta, \theta) \leq 0.$$

To simplify notation, combine terms and rewrite this as

$$\Delta_0(\theta) = \Delta(0, \theta) \geq 0, \quad \Delta_1(\theta) = \tfrac{1}{2}(\Delta(-2a\theta, \theta) + \Delta(2a\theta, \theta)) \leq 0. \tag{3}$$

Suppose we can show

$$\int w \rho(v, w) \omega(dv\, dw) = 0 \tag{4}$$

and

$$\int v^2 \rho(v, w) \omega(dv\, dw) \leq 0. \tag{5}$$

Then $\int (w - kv^2)\rho(v, w)\omega(dv\, dw) \geq 0$ where $k = k_{\alpha/2}$ as defined in (4.2). But, $(w - kv^2)\rho(v, w) \leq 0$ for all v, w by the definition of ψ_t. Thus (4) and (5) imply $\rho(v, w) = 0$ a.e. (ω); and this means that ψ_t is admissible. The proof will therefore be complete when (4) and (5) have been established.

Examining ψ_t, we note that its acceptance region is bounded by the parabola $w = kv^2$. The slope of this curve is $dw/dv = 2kv$. Hence for $|v| \geq a/k$ the perpendicular ray to the tangent at $(\pm |v|, kv^2)$ goes out away from the acceptance region to $\pm\infty$ inside the effective alternative, $\tilde{H}_1 - \tilde{H}_1^I = \{(\eta, \theta) : |\eta| \geq -2a\theta\}$. It follows from the Stein-Birnbaum argument (Stein, 1956) that $\rho(v, w) \leq 0$ a.e. (ω) on $\{(v, w) : w \geq kv^2 \text{ or } w \leq 2a|v| - a^2/k \text{ or } w \geq a^2/k\} \supset A \cup B$ where

$$A = \{(v, w) : w \geq a^2/k\}, \quad B = \{(v, w) : w \geq 2a|v|\}.$$

By Lemma A1 with $S = A$ and $m(v, w, \theta) = e^{\theta w}$

$$0 \leq \lim_{\theta \nearrow 0} \Delta_0(\theta) = \int \rho(v, w)\omega(dv\, dw).$$

Similarly with $S = B$ and $m(v, w, \theta) = \frac{1}{2}(e^{\theta(w-2av)} + e^{\theta(w+2av)})$.

$$0 \geq \lim_{\theta \nearrow 0} \Delta_1(\theta) = \int \rho(v, w)\omega(dv\, dw).$$

Hence

$$\lim_{\theta \nearrow 0} \Delta_0(\theta) = 0 = \lim_{\theta \nearrow 0} \Delta_1(\theta). \tag{6}$$

Standard theory of exponential families yields that Δ_0 and Δ_1 are infinitely differentiable and that derivatives can be computed under the integral sign.

Again apply Lemma A1. This time let $S = A$ and $m(v, w, \theta) = we^{\theta w}$ to get

$$\lim_{\theta \nearrow 0} \frac{d}{d\theta} \Delta_0(\theta) = \int w\rho(v, w)\omega(dv\, dw). \tag{7}$$

Similarly, letting $S = B$ and $m = \frac{1}{2}[(w - 2av)e^{\theta(w-2av)} + (w + 2av)e^{\theta(w+2av)}]$ yields

$$\lim_{\theta \nearrow 0} \frac{d}{d\theta} \Delta_1(\theta) = \int w\rho(v, w)\omega(dv\, dw). \tag{8}$$

In particular, the limits in (7) and (8) exist and so by (3) and (6) must satisfy

$$\lim_{\theta \nearrow 0} \frac{d}{d\theta} \Delta_0(\theta) \leq 0 \leq \lim_{\theta \nearrow 0} \frac{d}{d\theta} \Delta_1(\theta).$$

This verifies (4) since in fact

$$\lim_{\theta \nearrow 0} \frac{d}{d\theta} \Delta_0(\theta) = \lim_{\theta \nearrow 0} \frac{d}{d\theta} \Delta_1(\theta) = \int w\rho(v, w)\omega(dv\, dw).$$

Again using Lemma A1 and (3) and (6) yields

$$0 \leq \lim_{\theta \nearrow 0} \frac{d^2}{d\theta^2} \Delta_0(\theta) = \int w^2\rho(v, w)\omega(dv\, dw)$$

and

$$\begin{aligned} 0 &\leq \lim_{\theta \nearrow 0} \frac{d^2}{d\theta^2} \Delta_1(\theta) \\ &= \lim_{\theta \nearrow 0} \int \rho(v, w)\left(\frac{1}{2}\right)[(w - 2av)^2 e^{\theta(w-2av)} + (w + 2av)^2 e^{\theta(w+2av)}]\omega(dv\, dw) \\ &= \int (w^2 + 4av^2)\rho(v, w)\omega(dv\, dw). \end{aligned}$$

This verifies (5), and completes the proof. □

REFERENCES

ARROW, K. J. (1960). Decision theory and the choice of a level of significance for the t-test. In Olkin, I. et al (eds.) *Contributions to Probability and Statistics* 70–78. Stanford University Press, Stanford.

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

CHOW, M. S. (1982). Tables for an alternative to the t-test. Cornell University, Statistics Center preprint series. Ithaca, N.Y.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics.* Chapman, London.

FISHER, R. A. (1955). Statistical methods and scientific deduction. *J. Roy. Statist. Soc. B* **17** 69–78.

HODGES, J. L. and LEHMANN, E. L. (1954). Testing the approximate validity of statistical hypotheses. *J. Roy. Statist. Soc. B* **16** 261–268.

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

LEHMANN, E. L. (1950). Some principles of the theory of testing hypotheses. *Ann. Math. Statist.* **21** 1–26.

LEHMANN, E. L. (1957). A theory of some multiple decision problems, I. *Ann. Math. Statist.* **28** 1–25.

MARDEN, J. I. (1981). Invariant tests on covariance matrices. *Ann. Statist.* **9** 1258–1266.

MARDEN, J. I. and PERLMAN, M. D. (1980). Invariant tests for means with covariates. *Ann. Statist.* **8** 25–63.

NEYMAN, J. and PEARSON, E. S. (1933). The testing of statistical hypotheses in relation to probabilities a priori. *Proc. Camb. Phil. Soc.* **29** 492–510.

SOBEL, M. and WALD, A. (1949). A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of a normal distribution. *Ann. Math. Statist.* **20** 502–522.

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

DEPARTMENT OF MATHEMATICS
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903