The Annals of Statistics
1997, Vol. 25, No. 6, 2345–2367

AN UNBIASED TEST FOR THE BIOEQUIVALENCE PROBLEM

BY LAWRENCE D. BROWN,[1] J. T. GENE HWANG[2] AND AXEL MUNK[3]

University of Pennsylvania, Cornell University and Ruhr-Universität Bochum

It is shown that the standard two one-sided tests procedure for bioequivalence is a biased test. Better tests exist. In this paper, an unbiased α-level test and other tests which are uniformly more powerful than the two one-sided tests procedure are constructed. Its power can be noticeably larger than that of the α-level two one-sided tests procedure.

1. Introduction. Recently there has been great interest in the problem of demonstrating bioequivalence of treatments, especially in the pharmaceutical industry. One of the reasons is that developing a marketable drug which is bioequivalent (i.e., equivalent in efficacy) to a well-established drug typically costs 1 percent or less relative to the cost of developing a new one.

The statistical formulation of the problem is by now agreed upon to be the following hypothesis testing problem. Let m_1 and m_2 be the parameters corresponding to the two treatments. For example, the parameter of interest may be one of the pharmacokinetic parameters (AUC, $C_{\max}$ or $T_{\max}$) of the concentration–time curves. Here AUC, $C_{\max}$ and $T_{\max}$ stand for the area under the curve, the maximum concentration level and the time until the maximum concentration level is achieved.

Bioequivalence of the two treatments is defined as $\Delta_1 \le \rho \le \Delta_2$, where $\rho = m_1/m_2$ and Δ_i, $i = 1, 2$, is the tolerance limit prespecified by a regulatory agency. The statistical problem is then to test

$$H_0\colon \rho \ge \Delta_2 \text{ or } \rho \le \Delta_1 \quad \text{versus } H_1\colon \Delta_1 < \rho < \Delta_2.$$

If one can reject H_0, then one can declare bioequivalence. The regulatory agencies in both the United States and the European community use $\Delta_1 = 0.8$ and $\Delta_2 = 1.25$.

It is also recommended that a logarithmic transformation be applied and hence the ratio problem is turned into a difference problem, that is,

$$\theta = \ln(\rho) = \mu_1 - \mu_2$$

where $\mu_i = \ln m_i$. For the aforementioned recommended choice of Δ_1 and Δ_2 of the regulatory agencies, the hypotheses become

$$H_0\colon |\theta| \ge \Delta \quad \text{versus } H_1\colon |\theta| < \Delta, \tag{1.1}$$

Received July 1995; revised May 1997.

[1]Supported in part by NSF Grant DMS-96-26118.

[2]Formerly Jiunn T. Hwang. Supported in part by NSF Grant DMS-91-07842 and by Taiwan National Science Council Grant NSC-82-0208-M-007-136.

[3]Supported in part by the Deutsche Forschungsgemeinschaft Grant Br655/4-2.

AMS 1991 *subject classifications*. Primary 62F04; secondary 62P10.

Key words and phrases. Two one-sided tests procedure, uniformly more powerful, pharmacokinetic parameters.

where $\Delta = \ln 1.25$. Note that in this special case the transformation leads to a symmetric interval of θ. In general, even if an asymmetric interval is obtained, one can make a simple location transformation to turn it into a symmetric interval. Therefore symmetry in (1.1) can be assumed without loss of generality.

The standard test recommended in the United States [FDA (1992)] and the European Communities [EC-GCP (1993)] is the two one-sided tests procedure. To describe this test, let D be an estimate of θ with

$$D \sim N(\theta, \sigma^2) \tag{1.2}$$

and S^2 be an independent observation such that

$$S^2/\sigma^2 \sim \text{chi squared distribution with } v \text{ degrees of freedom.} \tag{1.3}$$

The above canonical form applies to many different models such as the general linear model, ANOVA and crossover designs. One of the simplest special cases involves a 2×2 crossover design. Let X_i, $1 \le i \le n$, i.i.d. observations, be the response of the ith subject after receiving the treatment. Similarly let Y_i, $1 \le i \le n$, i.i.d., be the response of the ith subject after receiving the reference treatment.

More specifically, for a 2×2 crossover design with subject effects (and without period effect), we may define

$$D_i = X_i - Y_i, \qquad 1 \le i \le n,$$

which are i.i.d. Under the customary assumption that D_i is normally distributed, letting

$$D = \Sigma D_i / n$$

and

$$S^2 = \Sigma (D_i - D)^2 / n$$

implies (1.2) and (1.3) with $v = n - 1$. In the simple case where X_i and Y_i have the common standard deviation σ_0, then σ in (1.2) and (1.3) equals

$$\sigma = (n/2)^{1/2} \sigma_0.$$

The α-level two one-sided tests procedure proposed in Schuirmann (1987) corresponds to the rejection region C_2, which consists of D and S such that

$$\Delta \ge |D| + t_\alpha \frac{S}{\sqrt{v}}, \tag{1.4}$$

where t_α is the upper α quantile of a t distribution with v degrees of freedom. [A related but more general type of test has been proposed in Berger (1982).] The two one-sided tests procedure is, however, biased. In fact even for $\theta = 0$, it can be seen that the power function satisfies

$$\lim_{\sigma \to \infty} P(C_2) = 0.$$

Several attempts have been made to improve upon the two one-sided tests procedure. See, for example, Anderson and Hauck (1983) and Rocke (1984, 1985). These alternative tests, however, are only approximately α-level. Existence of a better test with exact level α was established in Hwang and Liu (1992). Munk (1993) constructed a test which is shown numerically to be α-level ($\alpha = 0.1$) for degrees of freedom larger than 10 and which is uniformly more powerful than the two one-sided tests procedure. For a recent related confidence interval approach, see Hsu, Hwang, Liu and Ruberg (1994).

In this paper we construct an unbiased α-level test. In the construction, it is seen that the rejection region of the unbiased test always contains properly the rejection region (1.4). Therefore the unbiased test is uniformly more powerful than the α-level two one-sided tests procedure. See Theorem 4.1. The improvement in power may be quite noticeable.

2. Description of the test. A precise construction of the unbiased test is given in Section 4. As described in Section 5, it is feasible to numerically implement this construction and picture the critical region, that is, the rejection region, C_U, of the unbiased test. This is the region on which one can declare bioequivalence.

Figure 1 shows a typical picture of C_U and some other regions, as described below. The coordinates of this plot are (D, S), as defined previously. The triangle ABC and its inside is the rejection region C_2 of the two one-sided tests

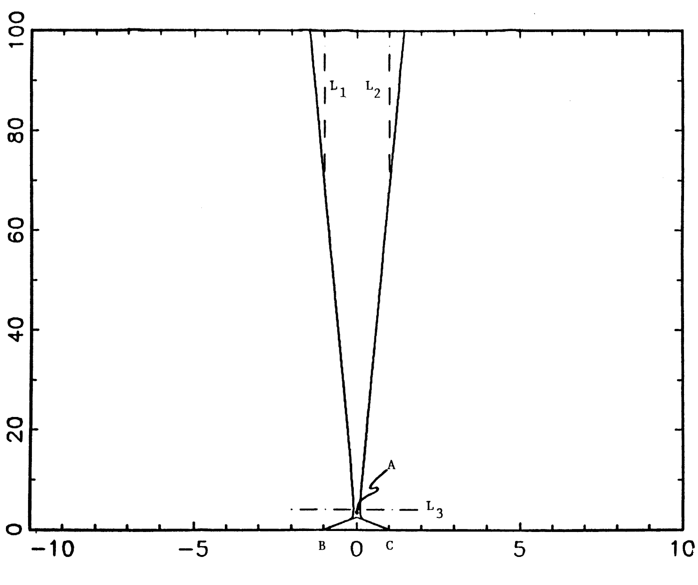

FIG. 1. *Rejection regions with* $\alpha = 0.05$, $v = 14$ *and* $\Delta = 1$. C_2 = *the triangle* ABC *and its inside*; C_U = *the union of* C_2 *and the region on top of* C_2 *bounded by the two solid curves*; $C_M = C_U$ *after removing the two regions outside the vertical lines* L_1 *and* L_2. $C_T = C_U$ *after removing the region above* L_3.

procedure. The region bounded by the solid curves forms C_U. Note that the boundaries C_2 and C_U coincide for small S, and C_U strictly contains C_2. Therefore C_U is uniformly more powerful than C_2 for any θ. This is also true for any level α.

Anderson and Hauck (1983) and Hauck and Anderson (1984) have proposed a test whose boundaries are qualitatively rather similar to those we display in Figure 1. Their test is motivated by a partly ad hoc argument. Numerical results for the few examples which are reported in their papers indicate that the level of their nominally $\alpha = 0.05$ test may actually be in the range 0.057 to 0.061, which is of course close, but not equal, to the nominal value. See Frick (1987) for some numerical studies about their levels. Their test is not unbiased (unlike ours), but the qualitative similarity does reinforce the value of their proposal.

When one of the authors, Hwang, presented a preliminary version of this paper in an invited session about bioequivalence in the August 1993 ASA Annual Meeting in San Francisco, two criticisms were raised about the unbiased test. It was argued that the rejection region of C_U is unbounded, and also that for any D it is possible to establish bioequivalence when S is large enough. Similar criticisms have been made about the test proposed by Anderson and Hauck (1983) which, as we have noted, has boundaries with the same qualitative features as ours. See Schuirmann (1987), pages 673–676 for a careful presentation of these arguments. See also Schuirmann (1996), point 1 and Berger and Hsu's counterargument (1996), page 317 in the Rejoinder, by comparing C_{L1} to the usual t-interval.

We tried to find a fundamental argument for the assertion that a reasonable rejection region should not be unbounded by using a likelihood approach, a Bayesian approach, and so on. However, we did not succeed. Therefore we are not convinced it should not be unbounded. It can even be shown [Munk (1992)] that every test with a bounded critical region has a vanishing power at any given θ when the variance is large. Nevertheless, for those who find an unbounded region unappealing, we recommend using the truncated procedure C_T illustrated in Figure 1. The rejection region of C_T is C_U after removing points above L_3, which is drawn in Figure 1. Formally, L_3 passes through the two points on the boundaries of C_U with smallest $|D|$. This region has the added monotonicity property (which some may find appealing) that if (D, S_1) is in the rejection region and $S_2 < S_1$ then (D, S_2) will also be in that region.

An alternate modification for which we do see a practical argument is also shown in Figure 1, as C_M, which is C_U except that points with $|D| > \Delta$ are removed. That is, the points outside L_1 and L_2 are removed. The practical argument for this modification is that often a bioequivalence decision is followed by a point estimate of θ (and/or an associated confidence interval). The usual point estimate (and/or center of a confidence interval) is $\widehat{\theta} = D$. If S is quite large, the procedure given by C_U can thus leave the statistician in the embarrassing position of rejecting the null hypothesis that $|\theta| \geq \Delta$ while at the same time estimating a value $\widehat{\theta}$ for which $|\widehat{\theta}| > \Delta$. For this reason, the modification C_M may be appealing, since the rejection region is cut off at L_1 and L_2. (Formally, C_M is formed by truncating C_U at $D = \pm\Delta$.)

3. Numerical results. By using the approach in Section 5, we construct the rejection region C_U (see also Figure 7). Since the transformation $(D, S) \to (D/\Delta, S/\Delta)$ can be used to reduce the problem of testing H_0: $|\theta| \geq \Delta$ down to the case where $\Delta = 1$ as in Section 3, we shall focus on $\Delta = 1$. (Therefore if C_U is the critical region for testing $|\theta| \geqq 1$, then for a general Δ, the test that rejects H_0 if and only if $(D/\Delta, S/\Delta) \in C_U$ is the α-level unbiased test.)

As shown in Figure 1 as well as in Theorem 4.1, the rejection region of the unbiased test contains that of the two one-sided tests procedure. Therefore the unbiased test is uniformly more powerful than the two one-sided test.

The power functions are plotted in Figures 2–5. Note that for a given v and $\Delta = 1$, the only unspecified parameters are σ and θ. For $v = 19$, $\alpha = 0.05$, the powers are plotted against θ while σ varies from picture to picture. The dashed line corresponds to the two one-sided tests whereas the solid line corresponds to the unbiased test. As expected, the power of the unbiased test is always higher. The dash and dot lines (more visible in the figures for $\sigma \geq 0.7$) corresponds to the power of the truncated test C_T. These powers are calculated by numerically integrating the normal cumulative distribution functions with respect to S, using the 20-point quadrature integration routine in the com-

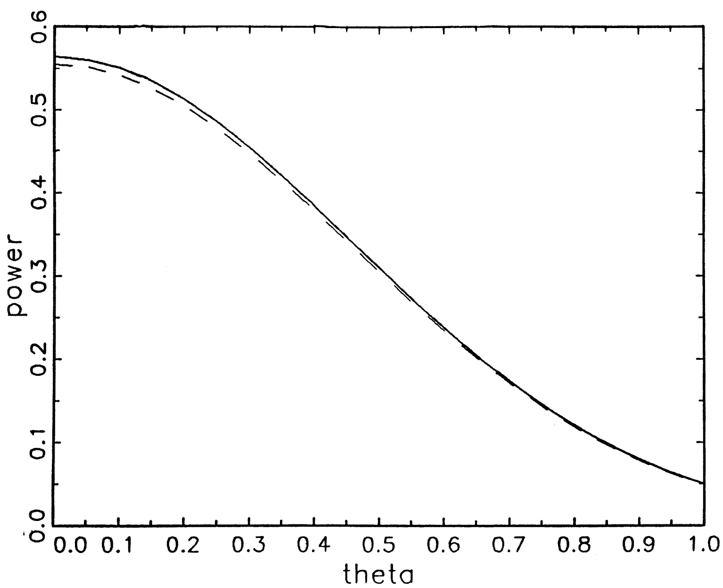

FIG. 2. *Power functions $\sigma = 0.4$. The power functions of C_U, C_2 and C_T are plotted against θ using a solid line, a dashed line, and a dash-and-dot line, respectively. In this figure and Figure 3, the dash and dot line is invisible since it coincides with the solid line.*

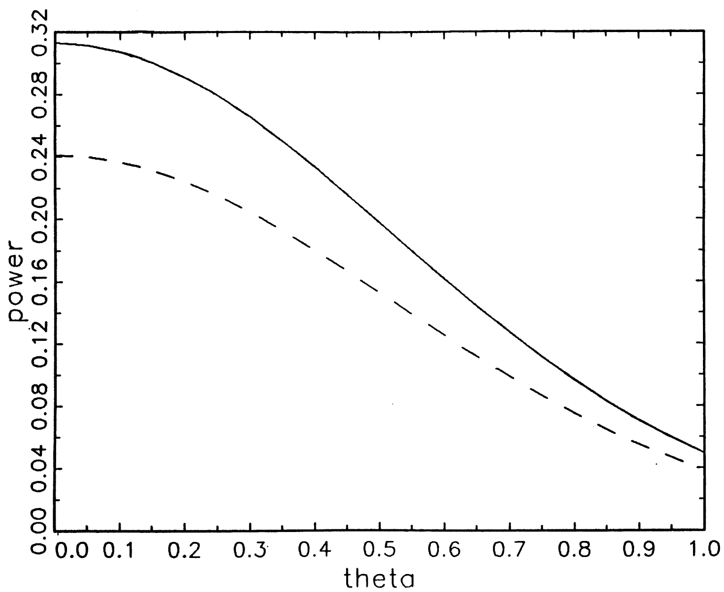

FIG. 3. *Power function as described in the title of Figure* 2, *except* $\sigma = 0.5$.

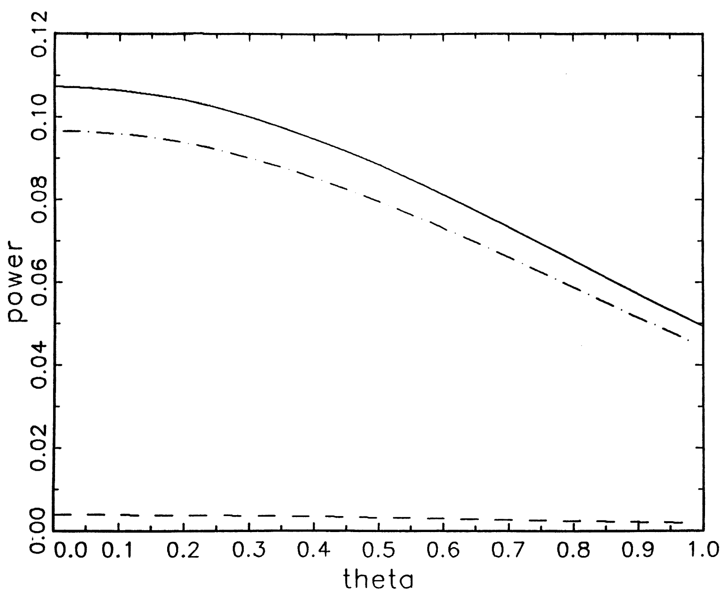

FIG. 4. *Power function as described in the title of Figure* 2, *except* $\sigma = 0.8$.

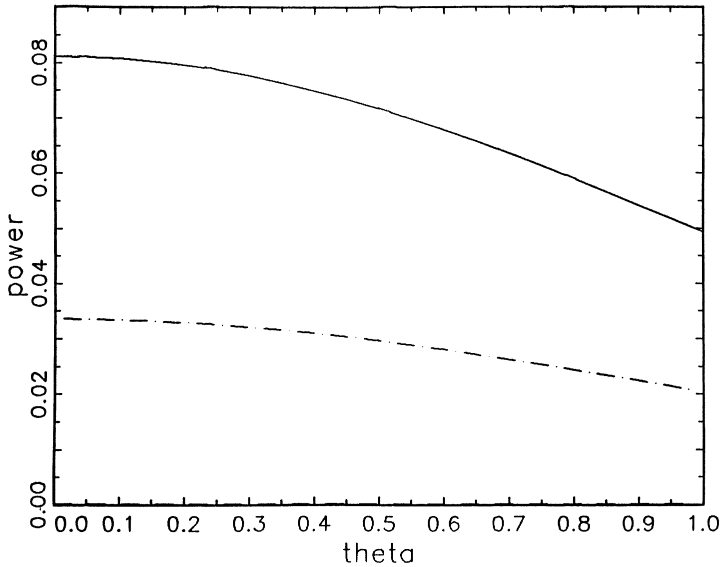

FIG. 5. *Power function as described in the title of Figure* 2, *except* $\sigma = 1.0$.

puter software Gauss which runs on an IBM personal computer. The power of C_M lies between that of C_U and C_T.

The powers of the two tests C_U and C_2 are quite similar in some cases, especially when σ is small. This is the reason why no figures are reported for $\sigma < 0.4$: the powers of C_U and C_2 are almost identical. However, in many other cases, we see a noticeable improvement in power of the unbiased test over the two one-sided tests procedure. (The two one-sided tests procedure can have very poor power when σ becomes large.) In particular, for $\sigma = 0.55$, the power at $\theta = 0$ of the unbiased test is almost twice as big as that of the two one-sided tests procedure. This range of σ may happen in bioequivalence data. For example in Westlake's data (1974, 1976), after the logarithmic transformation the estimated variance is $\frac{1}{6}(0.012) = 0.002$. In our context, however, every quantity is scaled with respect to Δ. Taking this into consideration, the estimate of σ after scaling with respect to $\Delta = \log 1.25$ is

$$\sqrt{.002}/\log 1.25 = 0.46.$$

However for the FDA's earlier recommendation, $\Delta = \log 1.2$. The corresponding estimate is

$$\sqrt{.002}/\log 1.2 = 0.56.$$

(In this discussion, we have used the logarithmic transformation log with the base 10, since Westlake did so. Obviously these estimates of σ would remain the same if we and Westlake had used ln all the way through.)

We also plotted figures corresponding to $v = 39$. The resultant pictures are similar to Figures 2–5 and are not reported.

The power of C_T is also plotted in Figures 2–5, using a dot-and-dash line. In all cases, C_T obviously has a power less than that of C_U. In many cases ($\sigma \leq 0.6$), however, the power functions of these two tests are indistinguishable. (Therefore the dot-and-dash lines disappear into the solid lines.) It so happens that these are the cases where the noticeable improvement of C_U over the two one-sided tests procedure occurs. Therefore C_T also has a noticeable improvement upon the two one-sided tests procedure in these cases. Although C_T is biased, it is bounded and it improves upon the two one-sided tests procedures in all cases, since the former contains the rejection region of the latter. The power of C_M hence is not plotted, since it is strictly between that of C_U and C_T. Therefore C_M also has noticeable improvement over C_2. These numerical results, together with the fact that C_U is unbiased, provide motivation for use of any one of the improved tests C_U, C_M or C_T rather than C_2.

Finally, even though all the alternative tests improve upon C_2, the improvement is negligible when σ is less than 0.4 or when the maximum power is high (higher than 80 percent). In well-designed studies, σ often falls in this range. However, in situations when the variability of the observations has been underestimated, these alternative tests are valuable. See Hauck and Anderson (1996).

4. Construction of C_U: theory. We consider the canonical model based on D and S which are independent and satisfy (1.2) and (1.3). The joint probability density function of (D, S) is proportional to

$$\exp\left(-\frac{(D-\theta)^2}{2\sigma^2}\right) S^p \exp\left(\frac{-S^2}{2\sigma^2}\right), \tag{4.1}$$

where $p = v - 1$.

An unbiased test denoted by a critical function φ_U will be constructed by an approach similar to that of Hodges and Lehmann (1954). They considered testing the null hypothesis H_1 versus the alternative H_0. Somewhat surprisingly, their solution is quite different from ours.

Unbiasedness of φ_U implies that it is similar, that is,

$$E_{\theta,\sigma}\varphi_U(D, S) = \alpha \tag{4.2}$$

for $\theta = \pm\Delta$ and every $\sigma > 0$. [See, e.g., Chapter 4 of Lehmann (1986).] From (4.1), when $\theta = \Delta$, the statistic

$$R^2 = (D - \Delta)^2 + S^2 \tag{4.3}$$

is sufficient and complete for $\sigma > 0$. This and (4.2) imply that

$$E_{\theta,\sigma}(\varphi_U(D, S) \mid R = r) = \alpha \quad \text{for } \theta = \Delta \tag{4.4}$$

and for every value of σ and r.

Statements similar to those in the last paragraph can be concluded when Δ is replaced by $-\Delta$. However, the critical function φ_U to be constructed is symmetric in D, that is,

$$\varphi_U(D, S) = \varphi_U(-D, S).$$

Therefore condition (4.4) remains unchanged when Δ is replaced by $-\Delta$. Hence, it is sufficient to focus on $\theta = \Delta$, which is assumed throughout this section unless mentioned otherwise.

It is convenient to express (D, S) in terms of the polar coordinate (R, β), where the origin of the polar coordinate is $(\Delta, 0)$. Here R is defined in (4.3) and β is the angle between the D axis and the line segment joining (D, S) and $(\Delta, 0)$. (See Figure 6.) It can be calculated easily that the probability density function of (R, β) is proportional to

$$\left(\frac{r}{\sigma}\right)^{p+1} \exp\left(\frac{-r^2}{2\sigma^2}\right)(\sin\beta)^p, \qquad 0 < \beta < \pi \text{ and } r > 0.$$

This obviously implies independence of R and β. Therefore in calculating the conditional expectation (4.4), only the marginal distribution of β is involved. The probability density function of β is

$$k_p(\sin\beta)^p, \tag{4.5}$$

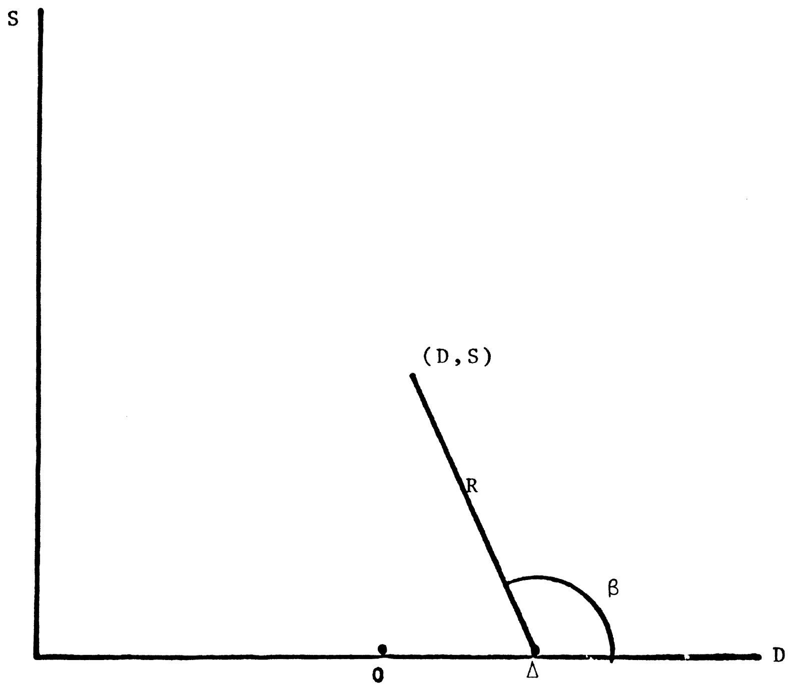

FIG. 6. *With respect to* $(\Delta, 0)$, (R, β) *is the polar coordinate of* (D, S).

where

$$(k_p)^{-1} = \int_0^{\pi} (\sin\beta)^p \, d\beta = \begin{cases} \dfrac{(p-1)(p-3)\cdots 1}{p(p-2)\cdots 2}\pi, & \text{if } p \text{ is even,} \\ \dfrac{(p-1)(p-3)\cdots 2}{p(p-2)\cdots 3}2, & \text{if } p \text{ is odd.} \end{cases}$$

We shall summarize the above discussion in the following lemma. For any critical function $\varphi(D, S)$, $\varphi^*(R, \beta)$ denotes the same function value as $\varphi(D, S)$ where (R, β) is the polar coordinate of (D, S) defined above and in Figure 6.

LEMMA 4.1. *Suppose the critical function $\varphi(D, S)$ is symmetric with respect to D. Then φ is similar if and only if*

$$k_p \int_0^{\pi} \varphi^*(r, \beta) \sin^p \beta \, d\beta = \alpha \tag{4.6}$$

for every value of r.

We shall now construct the unbiased test. In fact, we shall describe the boundary $(D(r), S(r))$ of the critical region C_U. We shall focus on the boundary points in the first quadrant. By symmetry, the boundary point in the second quadrant is $(-D(r), S(r))$ where r is the distance from $(-D(r), S(r))$ to $(-\Delta, 0)$. This relationship is mathematically described by (4.3) as well. The critical region C_U to be constructed below using Lemma 4.1 is the region between the curves $\{(-D(r), S(r))\colon r > 0\}$ and $\{(D(r), S(r)\colon r > 0\}$.

In the above terms, the two one-sided tests boundary is $\{(\pm D'(r), S'(r))\colon 0 \le r \le \Delta/\cos(\pi - \xi)\}$ where ξ satisfies

$$\begin{aligned} k_p \int_{\xi}^{\pi} (\sin\beta)^p \, d\beta &= \alpha, \\ D'(r) &= \Delta + r\cos\xi, \\ S'(r) &= r\sin\xi. \end{aligned} \tag{4.7}$$

(See the triangle in Figure 7.)

The boundary of C_U is constructed inductively on regions $(r_i, r_{i+1}]$, $i = 0, 1, \ldots,$ with $r_0 = 0$ and $\lim_{i\to\infty} r_i = \infty$. It will later be shown that the construction is valid for

$$\alpha_* < \alpha < 1/2, \tag{4.8}$$

where

$$\alpha_* = k_p \int_{3\pi/4}^{\pi} (\sin\beta)^p \, d\beta.$$

Table 1 shows the values of α_* for $0 \le p \le 19$. Since α_* is decreasing in p, $\alpha_* < 0.0001$ for $v = p + 1 \ge 21$. For the FDA recommended choice $\alpha = 0.05$, the condition is met if and only if $v \ge 5$. The restriction is therefore quite minor. (See also Section 7.)

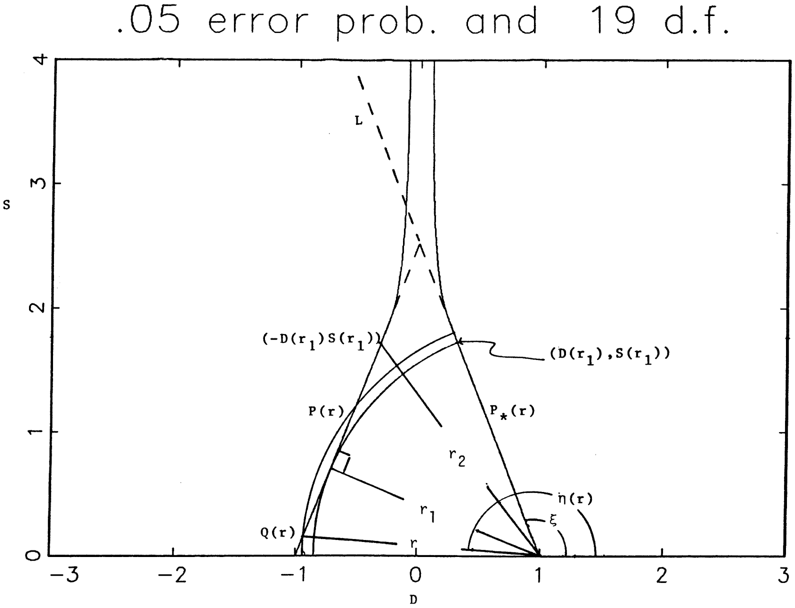

FIG. 7. *Construction of r_1, r_2 and $\eta(r)$. The dashed lines represent the boundary of the two one-sided tests corresponding to $v = 19$ and $\alpha = 0.05$. Below the points $(\pm D(r_1), S(r_1))$, this coincides with the boundary of C_U, which is represented by the solid line.*

As noted, $r_0 = 0$. Then let

$$r_1 = 2\Delta \sin(\pi - \xi);$$

that is, r_1 is the distance between $(\Delta, 0)$ and the opposite side of the triangle. Condition (4.8) guarantees that $3\pi/4 > \xi > \pi/2$, so that $\sqrt{2}\Delta < r_1 < 2\Delta$.

On this region, namely $0 < r \le r_1$, the boundary curve is the same as that for the two one-sided tests region. To explain why they agree, note first that the one-sided t test with a critical region to the left of line L (in Figure 7) has type one error α. By sufficiency and completeness of R under $\theta = \Delta$, the conditional rejection probability of the one-sided t test, given R, is α. Hence integrating the density (4.5), while β varies from ξ to the horizontal axis pointing to the negative infinity, is α. Therefore when $r < r_1$, we cannot add

TABLE 1
Values of α_* in (4.8)

$p+1=v$:	1	2	3	4	5	6	7
α_*:	0.25	0.1464	0.0908	0.0581	0.0378	0.0249	0.0166
$p+1=v$:	8	9	10	11	12	13	14
α_*:	0.0111	0.0075	0.0051	0.0034	0.0023	0.0016	0.0011
$p+1=v$:	15	16	17	18	19	20	≥ 21
α_*:	0.0008	0.0005	0.0004	0.0002	0.0002	0.0001	< 0.0001

more region than the set defined by L and the horizontal axis and still hope that (4.6) is satisfied. Consequently, C_U and C_2 have the same boundary.

However, when $r > r_1$, the situation is different. The arc from L to the horizontal axis is not entirely included in the rejection region. See Figure 7, which also shows r_2 as defined in the next paragraph.

Suppose now that the curve has been defined for $r \le r_i$. See Figure 7. Let

$$r_{i+1} = [(\Delta + D(r_i))^2 + S^2(r_i)]^{1/2}, \tag{4.9}$$

which is greater than

$$r_i = ((\Delta - D(r_i))^2 + S^2(r_i))^{1/2}$$

as long as $D(r_i)$ is never zero, which, among other things, shall be proved in Section 6.

Note that for $r_i < r < r_{i+1}$, the arc centered at $(\Delta, 0)$ with radius r intersects the left boundary of C_U. For a fixed r, let $P(r) = (-p_1(r), p_2(r))$ be the intersection point with the largest second coordinate.

Note that $P(r)$ is defined, being symmetric to $P_*(r) = (p_1(r), p_2(r))$ (in the first quadrant), which is already defined, since the distance between $P_*(r)$ and $(\Delta, 0)$ is smaller than r_i. See Figure 8.

Let

$$\eta = \eta(r) = \begin{cases} 3\pi/2 - \xi + \cos^{-1}(r_1/r), & \text{if } r < 2\Delta, \\ \pi, & \text{otherwise.} \end{cases} \tag{4.10}$$

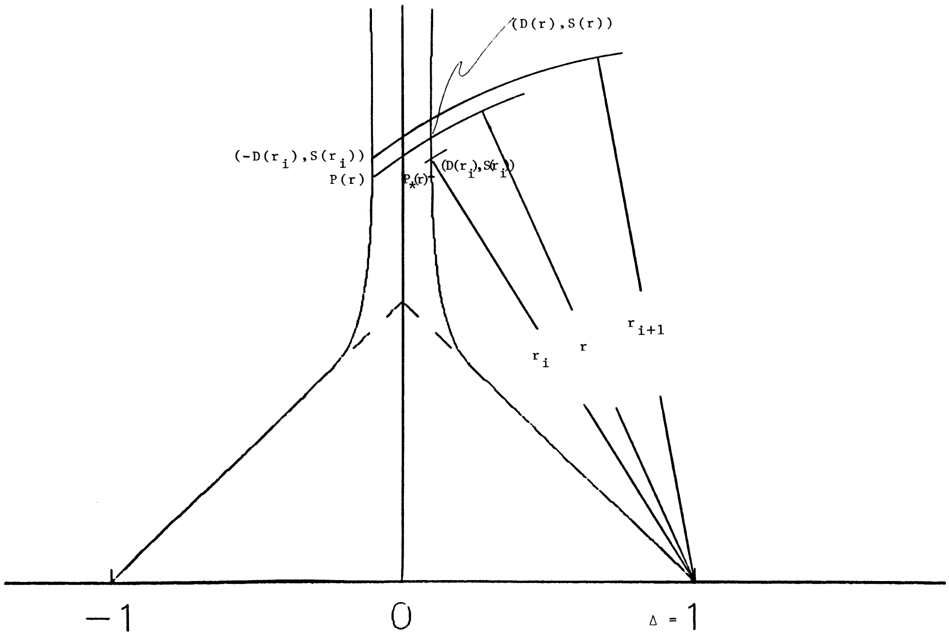

FIG. 8. *Inductive construction of the boundary of C_U.*

[For $r < 2\Delta$, η is the angle between the D axis and the vector joining $(\Delta, 0)$ and $Q(r)$ where $Q(r)$ is the lower intersection point of the left portion of the boundary and the arc with length r. See Figure 7.]

Now define $D(r)$ by

$$\left(\int_{\eta}^{\pi} + \int_{\cos^{-1}((D(r)-\Delta)/r)}^{\cos^{-1}((-p_1(r)-\Delta)/r)}\right) k_p \sin^p \beta \, d\beta = \alpha. \tag{4.11}$$

This definition guarantees that (4.6) is satisfied. Further, $S(r)$ can be found by solving $(D(r) - \Delta)^2 + S^2(r) = r^2$.

We summarize the above construction in the following theorem. The Theorem 4.1 is based on an assumption, called the *interval assumption*: the subset that consists of points in C_U with a given S, $S > 0$, forms an interval. Namely, the intersection of C_U and a horizontal line is a line segment or an empty set. This assumption has been observed to be satisfied in all our numerical studies. We believe that the violation of the interval assumption must be quite rare, if at all possible. The assumption is satisfied by other unbiased tests which are not constructed in a way described above. An example can be found in Figure 4 of Munk (1992), which is closest to violation but still satisfies the interval assumption.

THEOREM 4.1. *Assume that $\alpha_* < \alpha < 1/2$. The test with the rejection region C_U, whose boundary points $(\pm D(r), S(r))$ are constructed above, is uniformly more powerful than the α-level two one-sided tests procedure. The test is similar. Moreover if C_U satisfies the interval assumption, C_U is an α-level unbiased test.*

PROOF. It is shown in Section 6 that when $\alpha_* < \alpha < 1/2$ then $r_{i+1} > r_i$, $i = 0, 1, \ldots$, and $\lim_{i\to\infty} r_i = \infty$. Hence C_U is well defined.

The first statement of the theorem follows from the fact that C_U contains the triangle, which is the rejection region of the two one-sided tests procedure. To see this, we need focus only on the first quadrant by symmetry. Here all we need to argue is that for every r the angle ξ' between the D axis and the line segment joining $(D(r), S(r))$ and $(\Delta, 0)$ is smaller than ξ. [In terms of the earlier notation, $\xi' = \cos^{-1}((D(r) - \Delta))/r)$.] This is true, since

$$\alpha = k_p \int \varphi^*(r, \beta)(\sin\beta)^p \, d\beta \le \int_{\xi'}^{\pi} k_p (\sin\beta)^p \, d\beta$$

and by (4.7),

$$\int_{\xi}^{\pi} k_p (\sin\beta)^p \, d\beta = \alpha.$$

Now we argue for the second statement. By construction, C_U is *similar*, that is, which implies that the power function $P_{\theta,\sigma}((D, S) \in C_U)$ is α for $\theta = \pm\Delta$ and for every σ. The power function depends on θ only through $|\theta|$. By construction, C_U is symmetric about zero with respect to D, and given an S, C_U is an interval by the interval assumption. It follows that, for each σ, $P_\theta((D, S) \in$

$C_U \mid S$) is decreasing in $|\theta|$, implying the same for the unconditional power. This implies that C_U has α-level and is unbiased. $\square$

Finally, we discuss an asymptotic formula in the first quadrant. When r is large, the boundary approaches the line determined by

$$S(r) = D(r)/\tan\lambda, \tag{4.12}$$

where λ is the angle such that

$$\int_{\pi/2-\lambda}^{\pi/2+\lambda} k_p \sin^p \beta \, d\beta = \alpha. \tag{4.13}$$

The rationale for this is that when r is large, then Δ/r is small and the region approaches that for $\Delta = 0$, leading to the above line.

5. Numerical computation. The following steps indicate how to construct a numerical approximation to C_U. The approximate boundary points will still be denoted in this section as $(D(r), S(r))$.

STEP 1. Find ξ that satisfies (4.7),

$$\int_{\xi}^{\pi} k_p(\sin\theta)^p \, d\theta = \alpha$$

and verify $\xi < 3\pi/4$. If $\alpha = 0.05$, this inequality holds if and only if $p \geq 4$. See Table 1.

STEP 2. For $0 < r \leq 2\Delta\sin(\pi - \xi) = r_1$, define

$$D(r) = \Delta + r\cos\xi \quad \text{and} \quad S(r) = r\sin\xi.$$

For Steps 1 and 2, see Figure 7.

STEP 3. Let $\rho_1 = r_1$. Suppose $D(r)$ and $S(r)$ have already been defined for $0 \leq r \leq \rho_k$. We now describe how to define the next point $P(\rho_{k+1}) = (D(\rho_{k+1}), S(\rho_{k+1}))$. Choose $\varepsilon > 0$ so that $\varepsilon \leq D(\rho_k)$. (The smaller ε is, the more precise the approximation will be, but the total computation will be longer.) Generally a value of ε can be fixed throughout the computation, but one must check at each stage that $\varepsilon \leq D(\rho_k)$.) Then let

$$\rho_{k+1}^2 = \rho_k^2 + 4\varepsilon\Delta.$$

Let P denote the intersection point of the left boundary of C_U and the arc A centered at $(\Delta, 0)$ with radius ρ_{k+1}. See Figure 9. Note that P is well defined since the point P_*, symmetric to P with respect to the S axis, is well defined. To see this, let ρ_{k+1}^* denote the distance between P_* and $(\Delta, 0)$, that is,

$$(D(\rho_{k+1}^*) - \Delta)^2 + S^2(\rho_{k+1}^*) = (\rho_{k+1}^*)^2$$

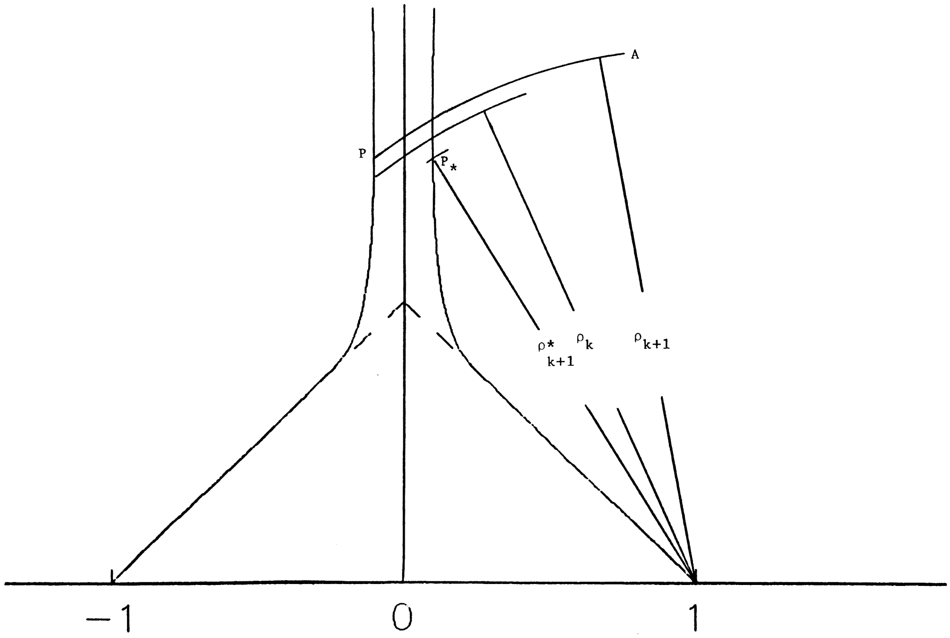

FIG. 9. *Inductive construction of the boundary of C_U.*

where $P_* = (D(\rho^*_{k+1}), S(\rho^*_{k+1}))$. We now show that $\rho^*_{k+1} < \rho_k$, which implies that P_* is defined and hence that P is also. Note that $P = (-D(\rho^*_{k+1}), S(\rho^*_{k+1}))$ and the distance between P and $(\Delta, 0)$ is ρ_{k+1}. Hence

$$(-D(\rho^*_{k+1}) - \Delta)^2 + (S(\rho^*_{k+1}))^2 = \rho^2_{k+1}.$$

Putting the last two displayed equations together, we have

$$\begin{aligned}(\rho^*_{k+1})^2 &= (D(\rho^*_{k+1}) - \Delta)^2 + S^2(\rho^*_{k+1})\\ &= (D(\rho^*_{k+1}) - \Delta)^2 + \rho^*_{k+1} - (-D(\rho^*_{k+1}) - \Delta)^2\\ &= \rho^2_{k+1} - 4\Delta D(\rho^*_{k+1}) < \rho^2_k,\end{aligned}$$

assuming $D(\rho^*_{k+1}) > \varepsilon$ as well. If $D(\rho^*_{k+1})$ is not larger than ε, we have to start with a smaller ε.

Now we define the point $(D(\rho_{k+1}), S(\rho_{k+1}))$ to be on the arc A such that (4.11) is satisfied; that is,

$$\left(\int_\eta^\pi + \int_{\cos^{-1}((D(\rho_{k+1})-\Delta)/\rho_{k+1}}^{\cos^{-1}((-D(\rho^*_{k+1})-\Delta)/\rho_{k+1})}\right) k_p(\sin\beta)^p \, d\beta = \alpha \tag{5.1}$$

and

$$(D(\rho_{k+1}) - \Delta)^2 + S^2(\rho_{k+1}) = \rho^2_{k+1}, \tag{5.2}$$

where $\eta = \eta(\rho_{k+1})$ is defined in (4.10).

Equation (5.1) can be used to solve for $D(\rho_{k+1})$, since $D(\rho^*_{k+1})$ is known. Due to the monotonicity of the left-hand side of (5.1) in $D(\rho_{k+1})$, this can be done without much effort. Using $D(\rho_{k+1})$ and (5.2), one can then solve for $S(\rho_{k+1})$ easily.

Connect the points by line segments. Hence for $\rho_k < r \le \rho_{k+1}$ define

$$D(r) = D(\rho_k) + \frac{r - \rho_k}{\rho_{k+1} - \rho_k}(D(\rho_{k+1}) - D(\rho_k))$$

and

$$S(r) = S(\rho_k) + \frac{r - \rho_k}{\rho_{k+1} - \rho_k}(S(\rho_{k+1}) - S(\rho_k)).$$

This completes the definition of $(D(r), S(r))$ for $\rho_k < r \le \rho_{k+1}$.

STEP 4. Terminate the calculation at some (large) ρ_{k+1} when the points are close to the asymptote described by (4.12) and (4.13). For a larger r, the boundary in the first quadrant is then the asymptote.

6. Existence of C_U. The construction in Section 4 relies on the facts that $r_{i+1} > r_i$, $i = 0, 1, \dots$, and $\lim_{i\to\infty} r_i = \infty$. The following geometric lemmas are crucial to the proof of these facts.

Let $O = (a, 0)$ with $a > 0$ be a point on the horizontal axis. See Figure 10. For a fixed point $A = (x, y)$ in the first quadrant, let $B = (0, h_1)$, $h_1 > 0$ be the point on the vertical axis such that

$$\overline{OB} = \overline{OA},$$

where, for example, $\overline{OB}$ denotes the distance between O and B. It is assumed that

$$\overline{OA} > a \tag{6.1}$$

so that point B exists. Let $C = (-x, y)$ denote the symmetric point of A with respect to the vertical axis and let $D = (0, h_2)$, $h_2 > 0$, denote the point on the vertical axis such that $\overline{OD} = \overline{OC}$. Also shown in Figure 10 are $E = (0, y)$, $\alpha = \angle BOC$, $\beta = \angle DOB$ and $\gamma = \angle AOD$.

Figure 10 illustrates only the most interesting of the possible geometric configurations for this construction, where $x \le a$ and $\alpha > 0$. All configurations have $\gamma > 0$, as verified in Lemma 6.1.

LEMMA 6.1. *The condition*

$$x^2 + y^2 > a^2 \tag{6.2}$$

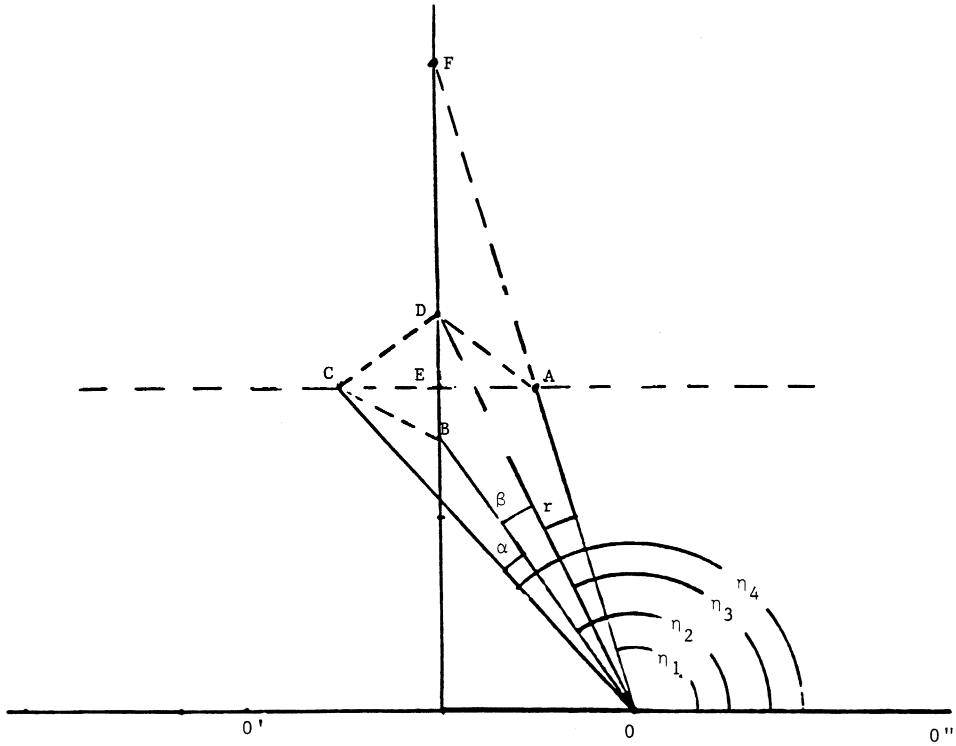

FIG. 10. *Illustration of the construction in Lemmas* 6.1–6.3.

implies, in the above construction, $\gamma > 0$. (*That is, the segment* DO *is located counterclockwise from* AO.)

PROOF. If $x \geq a$, the lemma is trivial. Assume $x < a$ below. Extend the line OA to meet the vertical axis at F. Algebraic manipulation shows that the assumpiton (6.2) is equivalent to $(\overline{O'C})^2 + (\overline{CO})^2 > (\overline{O'O})^2$, where O' is the symmetric point of O with respect to the vertical line. This implies $\angle O'CO < \pi/2$. It follows that the point F' on the y-axis, such that $\angle OCF' = \pi/2$, lies below F. Since OCD is an equilateral triangle, F' is above D. Hence F is above D and $\gamma > 0$. □

The other important geometric fact is contained in Lemma 6.2.

LEMMA 6.2. *In the above construction* $\angle DOC < \angle AOB$.

PROOF. Examination of the construction displayed in Figure 10 and some algebraic calculation show that $\overline{ED} = \overline{EB}$ if and only if $\angle EOO' = \pi/4$ and $x < a$.

If $\overline{ED} \le \overline{EB}$ then $\angle EOO' \le \pi/4$ and B falls below the segment CO so that $DOC \subset AOB$ by Lemma 6.1. It follows trivially that $\angle DOC < \angle AOB$.

If $\overline{EB} < \overline{ED}$ then $\overline{CB} < \overline{CD} = \overline{AD}$. Consider the triangles COB and DOA. They then have $\overline{CO} = \overline{DO}$, $\overline{OB} = \overline{OA}$ and $\overline{CB} < \overline{DA}$. Hence $\alpha < \gamma$. It follows that $\angle DOC = \alpha + \beta < \gamma + \beta = \angle AOB$. □

These two geometric lemmas combine to yield the following analytic result. Here, let $\eta_1 \doteq \angle O''OA$, $\eta_2 \doteq \angle O''OB$, $\eta_3 \doteq \angle O''OD$, $\eta_4 = \angle O''OC$.

LEMMA 6.3. *Assume* (6.1) *and* (6.2). *Then*

$$\int_{\eta_1}^{\eta_2} \sin^p \beta \, d\beta > \int_{\eta_3}^{\eta_4} \sin^p \beta \, d\beta. \tag{6.3}$$

PROOF. $\eta_2 - \eta_1 > \eta_4 - \eta_3$ by Lemma 6.2, and $\eta_1 < \eta_3$ by Lemma 6.1. If $\alpha \le 0$, then $\eta_1 < \eta_3 < \eta_4 \le \eta_2$ so the conclusion of this lemma follows trivially. If $\alpha > 0$, make use of the fact that $\sin\beta$ is unimodal and symmetric about $\pi/2$ and $\eta_4 - \alpha = \eta_2 > \pi/2$ and $\eta_3 > \pi/2$ to write

$$\int_{\eta_3}^{\eta_4} (\sin\beta)^p d\beta < \int_{\eta_3-\alpha}^{\eta_4-\alpha} (\sin\beta)^p d\beta = \int_{\eta_3-\alpha}^{\eta_2} (\sin\beta)^p d\beta < \int_{\eta_1}^{\eta_2} (\sin\beta)^p d\beta,$$

which establishes the lemma. □

We can now establish the validity of the construction of C_U, as follows.

THEOREM 6.1. *When $\alpha_* < \alpha < 1/2$, the quantities r_i defined in the construction of C_U satisfy $r_i < r_{i+1}$, $i = 0, 1, \ldots$, and $\lim_{i\to\infty} r_i = \infty$. Hence C_U exists. The corresponding test is unbiased if the interval assumption is satisfied.*

PROOF. We will suppose the theorem is false and obtain a contradiction. Hence, suppose

$$r^* = \lim_{i\to\infty} r_i < \infty. \tag{6.4}$$

Since D is continuous, in the case of (6.3) $D(r^*) = 0$. In the case of (6.4) there must be a sequence $\{r_i\}$ increasing to r^* so that $\lim_{i\to\infty} D(r_i)$ and $\lim S(r_i)$ exist. So, for notational convenience in this case, let $D(r^*) = \lim_{i\to\infty} D(r_i)$, $S(r^*) = \lim_{i\to\infty} S(r_i)$.

In either case, there is an $r^{**} < r^*$ such that

$$(-D(r^{**}) - \Delta)^2 + S(r^{**})^2 = (r^*)^2. \tag{6.5}$$

(Expressed in terms of the construction in Section 4, $p(r^*) = (-D(r^{**}), S(r^{**}))$.) See Figure 10 with $O = (\Delta, 0) = (a, 0)$, $A = (D(r^{**}), S(r^{**})) = (x, y)$, $C = (-D(r^{**}), S(r^{**}))$, $\overline{DO} = r^*$ and $\overline{BO} = r^{**}$. Here we assume that there exists B on the vertical axis so that (6.5) holds. To argue that this is the case, it suffices to establish (6.1), which together with (6.2) shall be established at the end of the proof. Consequently, Lemma 6.3 is applicable.

By continuity from (4.11),

$$\alpha = \left(\int_{\eta(r^*)}^{\pi} + \int_{\eta_3}^{\eta_4}\right) \sin^p \beta \, d\beta, \tag{6.6}$$

where $\eta(r^*)$ is as defined in (2.9), $\eta_3 = \cos^{-1}(-\Delta/r^*)$ and $\eta_4 = \cos^{-1}([-\Delta - D(r^{**})]/r^*)$.

Since $r^{**} > r^*$, the first integral on the right-hand side of (6.6) is bounded above by

$$\int_{\eta(r^{**})}^{\pi} k_p \sin^p \beta \, d\beta. \tag{6.7}$$

By Lemma 6.3, the second integral is strictly less than

$$\int_{\eta_1}^{\eta_2} k_p \sin^p \beta \, d\beta, \tag{6.8}$$

where $\eta_1 = \cos^{-1}((D(r^{**}) - \Delta)/r^{**})$ and $\eta_2 = \cos^{-1}(-\Delta/r^{**})$. Therefore (6.6) is strictly less than (6.7) plus (6.8), which equals α. This shows $\alpha < \alpha$, a contradiction. Hence $r^* = \infty$.

Finally we shall prove (6.1) and (6.2), using $\alpha > \alpha^*$. We compare the rejection region $O'VO$ of the α-level two one-sided tests procedure to the corresponding region $O'V'O$ of α^*. See Figure 11.

Since $\angle V'OO' = \pi/4$, the vertex V' has the cordinate $(0, a)$. It is then obvious that $S(r_1) > a$ and $\overline{JO} > a$, where $J = (D(r_1), D(r_1))$ is on $\overline{VO}$. Therefore point A should be in region T, which is a subset of the first quadrant outside the bigger triangle and the circle passing J centered at O. The reason is that point A is on the upper boundary beyond the point J and the boundary can only be in T. [The intersecting point on the lower boundary corresponds not to A but to $\eta^*(r^*)$, which results in the first integral on the right-hand side of (6.6).] Now all the points in T satisfy (6.1) and (6.2) and so does A. □

7. An approximately unbiased test for $\boldsymbol{\alpha < \alpha^*}$. Up to this point, it has been assumed that $\alpha > \alpha^*$. Although most practical problems satisfy the condition, it may be of interest to discuss briefly what results are available when the condition fails.

For some cases, it is possible to construct tests which are exactly unbiased. See Figure 12 for $\alpha = 0.25$ and $\nu = 1$. (Here $\alpha^* = \alpha$.) In general, for $\alpha \leq \alpha^*$ we do not have existence results. However, Munk (1992) has an algorithm to construct numerically unbiased tests which dominate the two one-sided tests procedure. Note that there is a simple proof that a dominating test exists, as in Hwang and Liu (1992).

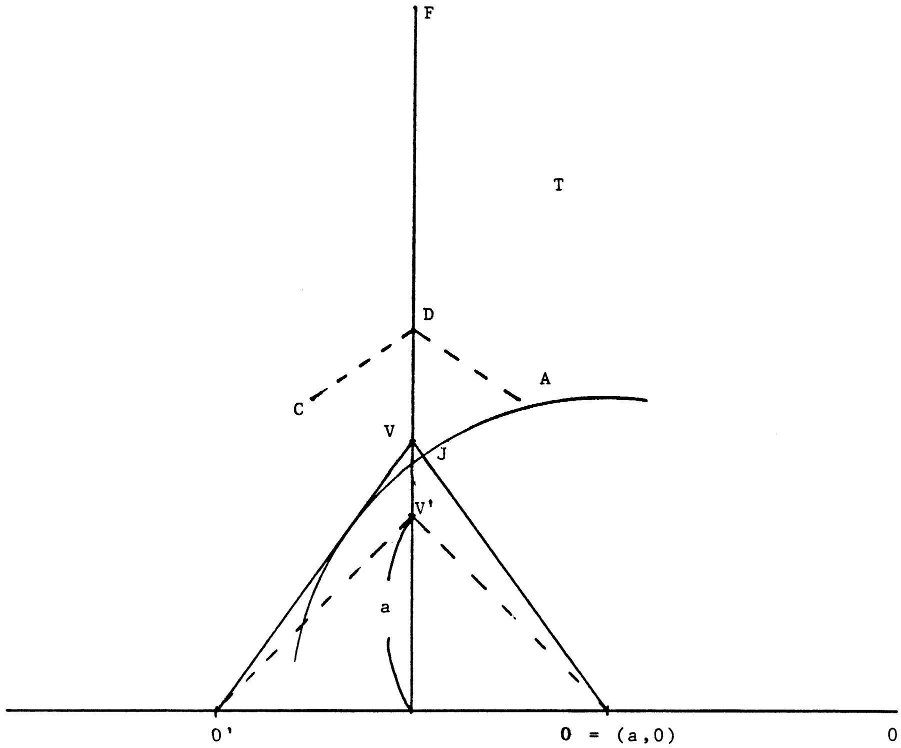

FIG. 11. *This and Figure 10 are used to establish that the intersection of the boundary with the vertical axis leads to a contradiction.*

8. Comparing our numerical results with those of Diletti, Hauschke and Steinijans. A referee pointed out that there seems to be inconsistency between our numerical results and that of Diletti, Hauschke and Steinijans (1991). We therefore explain briefly here by an example which serves to connect our results with theirs and also serves to illustrate our notation.

As in Diletti, Hauschke and Steinijans (1991), let $\overline{X}_T$ and $\overline{X}_R$ represent the sample averages corresponding to the logarithmically transformed characteristics of the test treatment and the reference treatment. The typical unbiased estimate of θ, the difference of the characteristics corresponding to the test treatment and the reference treatment, is then $\overline{X}_T - \overline{X}_R$, which is assumed to be normally distributed with varance $2\sigma_0^2/n$ where n is the total number of subjects and σ_0^2 is the within subject variance of the logarithmically transformed characteristic. Note that σ_0 is roughly equal to the coefficient of variation (CV) of the untransformed characteristic as specified in all the figures of Diletti, Hauschke and Steinijans (1991). Define

$$D = \frac{\overline{X}_T - \overline{X}_R}{\ln 1.25}.$$

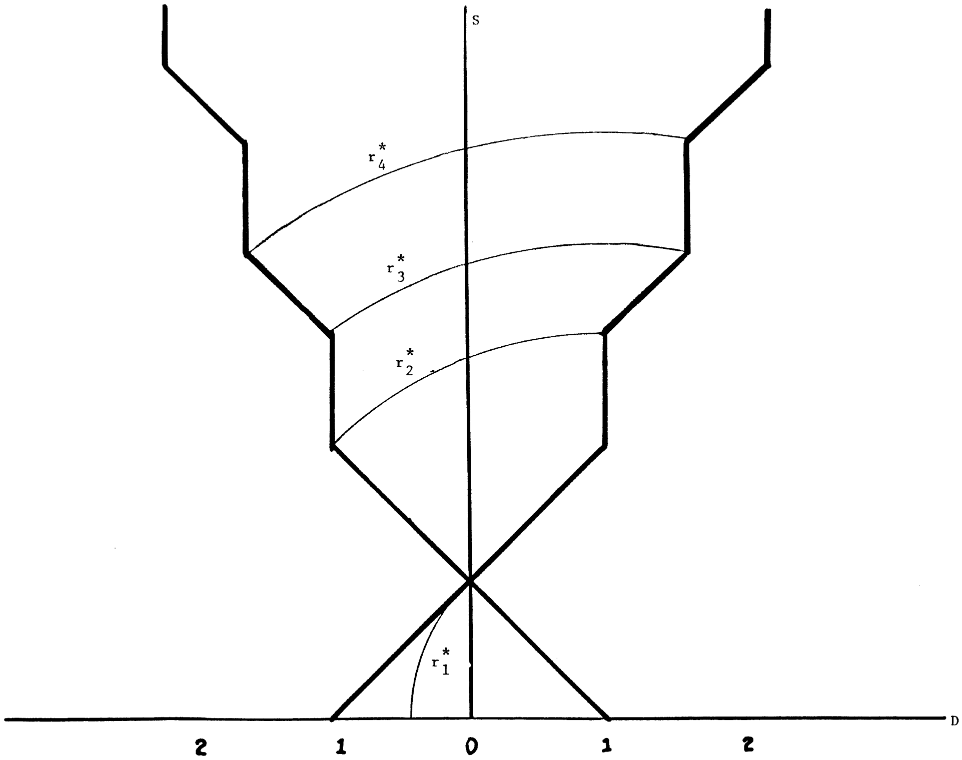

FIG. 12. *Rejection region of an unbiased test when $\nu = 1$ and $\alpha = \alpha^* = 0.25$. The arcs with radii r_1^*, r_2^*, r_3^* and r_4^* all lie within the rejection region.*

Then $D \sim N(\theta, \sigma^2)$, where

$$\sigma^2 = \frac{2\sigma_0^2}{n(\ln 1.25)^2}, \tag{8.1}$$

which gives the canonical form (1.2) and (1.3). The corresponding hypothesis tested is then (1.1) with $\Delta = 1$.

As an example, let $n = 21$ and under the 2×2 crossover design model with subject and period effects, the degrees of freedom are $v = 19$. Also consider

$$\sigma_0 = 0.2892,$$

which interests us since it leads, by (8.1), to

$$\sigma = \sqrt{\frac{2(0.2892)^2}{21(\ln(1.25))^2}} = 0.4.$$

With this σ and v, the maximum power according to Figure 4, is about 0.57. Using Figure 2d of Dilletti, Hauschke and Steinijans (1991), for $n = 21$ and

CV = 0.25, the maximum power is about 0.69. The discrepancy is due to the fact that our CV is 0.2892 whereas theirs is 0.25. By considering their Figure 2a for the case with CV = 20% and the approximate maximum power 0.89, we may linearly extrapolate to give an approximate maximum power for $CV = 0.2892$ to be 0.53, which is close to our answer 0.57.

Acknowledgements. The authors thank Mr. Y. B. Hu at Tsing Hua University and Dipl.-Math. L. Pralle at Göttingen University for their help with computing. They also thank Arthur Cohen and Holger Dette for helpful discussions and are especially indebted to Yannis Yatracos for supplying the key idea leading to Figures 10 and 11 and the current proof of Lemma 6.2. Some of the results were independently obtained by A. Munk and can be found in his thesis at the Institute of Mathematical Stochastics, Göttingen University. A. Munk is further indebted to the Studienstiftung des Deutschen Volkes and Diskrete Strukturen in der Mathematik at Bielefeld University for their support.

REFERENCES

ANDERSON, S. and HAUCK, W. W. (1983). A new procedure for testing equivalence in comparative bioavailability and other clinical trials. *Comm. Statist. Theory Methods* **12** 2263–2692.

BERGER, R. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics* **24** 295–300.

BERGER, R. and HSU, J. (1996). Bioequivalence trials, intersection–union tests and equivalence confidence sets (with discussion). *Statist. Sci.* **11** 283–319.

DILETTI, E., HAUSCHKE, D. and STEINIJANS, V. W. (1991). Sample size determination for bioequivalence assessment by means of confidence intervals. *International Journal of Clinical Pharmacology, Therapy and Toxicology* **29** 1–8.

EC-GCP (1993). Biostatistical methodology in clinical trials in applications for marketing authorization for medical products. CPMP Working Party on Efficacy of Medical Products, Commission of the European Communities, Brussels.

FDA (1992). Draft recommendation on statistical procedures for bioequivalence studies using the standard two-treatment crossover design. Center for Drug Evaluation and Research, Food and Drug Administration, Rockville, MD.

FRICK, H. (1987). On level and power of Anderson and Hauck's procedure for testing equivalence in comparative bioavailability. *Comm. Statist. Theory Methods* **16** 2771–2778.

HAUCK, W. W. and ANDERSON, S. (1984). A new statistical procedure for testing equivalence in two-group comparative bioavailability trials. *J. Pharmacokinetics and Biopharmaceutics* **12** 83–91.

HAUCK, W. W. and ANDERSON, S. (1996). Comments on "Bioequivalence trials, intersection–union tests and equivalence confidence sets" by R. Berger and J. Hsu (1996). *Statist. Sci.* **11** 303.

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

HSU, J., HWANG, J. T. G., LIU, H. K. and RUBERG, S. J. (1994). Confidence intervals associated with tests for bioequivalence. *Biometrika* **81** 103–114.

HWANG, J. T. G. and LIU, H. K. (1992). Confidence intervals associated with tests for bioequivalence. Cornell Statistical Center Technical Report.

LEHMANN, E. L. (1986). *Testing Statistical Hypotheses*, 2nd ed. Wiley, New York.

MUNK, A. (1992). Aquivalenz und Intervalltests in exponentiellen Familien. Ph.D. dissertation, Institut für Mathematische Stochastik, Univ. Göttingen.

MUNK, A. (1993). An improvement on commonly used tests in bioequivalence assessment. *Biometrics* **49** 1225–1230 [see also correspondence in *Biometrics* **50** 884–886 (1994)].

ROCKE, D. M. (1984). On testing for bioequivalence. *Biometrics* **40** 225–230 [see also correspondence in *Biometrics* **41** 561–563 (1985)].

SCHUIRMANN, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of coverage bioavailability. *J. Pharmacokinetics and Biopharmaceutics* **15** 657–680.

SCHUIRMANN, D. J. (1996). Comments on "Bioequivalence trials, intersection–union tests and equivalence confidence sets" by R. Berger and J. Hsu (1996). *Statist. Sci.* **11** 312–313.

WESTLAKE, W. J. (1974). The use of balanced incomplete block designs in comparative bioavailability trials. *Biometrics* **30** 319–327.

WESTLAKE, W. J. (1976). Symmetrical confidence intervals for bioequivalence trials. *Biometrics* **32** 741–744.

L. D. BROWN
DEPARTMENT OF STATISTICS
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104

J. T. G. HWANG
DEPARTMENT OF MATHEMATICS
WHITE HALL
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853-7901
E-MAIL: hwang@math.cornell.edu

A. MUNK
FAKULTÄT FÜR MATHEMATIK
RUHR-UNIVERSITÄT BOCHUM
UNIVERSITÄTSSTR 150
44780 BOCHUM
GERMANY
E-MAIL: axel.munk@rz.ruhr-uni-bochum.de