A Nonparametric, Intersection-Union Test for Stochastic Order

by

Roger L. Berger
North Carolina State University

Institute of Statistics Mimeo Series No. 1685

May, 1986

1. **Introduction.** A distribution function (cdf) $G(x)$ is said to be stochastically larger than another cdf $F(x)$ if $G(x) \leq F(x)$ for all x with strict inequality for some x. This relationship will be denoted by $G \geq^{st} F$. Stochastic ordering defines a situation in which, in a very strong sense, observations from a population with cdf G tend to be larger than observations from a population with cdf F.

To test if $G \geq^{st} F$, null and alternative hypotheses commonly considered are H_0': $G = F$ and H_a': $G \geq^{st} F$. The Mann-Whitney-Wilcoxon rank sum test (RST) is often recommended as a nonparametric test for testing H_0' versus H_a'. See for example Lehmann (1975). Lee and Wolfe (1976) study a competitor to the RST, based on a restricted maximum likelihood estimate, for testing H_0' versus H_a'.

Using the asymptotic distribution of the RST statistic, it can be shown that the power of the RST converges to one for any pair (F,G) for which $\theta = \int_{-\infty}^{\infty} F(x)dG(x) > ½$. The same is true of the Lee and Wolfe (1976) test. Here $\theta = P(X \leq Y)$ if X and Y are independent random variables with cdfs F and G respectively. If $G \geq^{st} F$ then $\theta > ½$. But there are many other pairs (F,G) for which G is not stochastically larger than F and still $\theta > ½$. If the model includes such pairs, then rejection of the null hypothesis by the RST or Lee and Wolfe test cannot be interepreted as convincing evidence that $G \geq^{st} F$ but only that $\theta > ½$. Indeed, the power of the RST converges to 0 and α (the size of the test) according as $\theta < ½$ or $\theta = ½$. Thus the RST might be more properly thought of as a test of H_0'': $\theta \leq ½$ versus H_a'': $\theta > ½$ rather than as a test of H_0' versus H_a'.

To develop a nonparametric test for which "rejection of the null hypothesis" could be interpreted as convincing evidence that $G \geq^{st} F$, one might try to derive a size α test of H_0^*: G is not stochastically larger than F

versus H_a^*: $G \geq^{st} F$. Unfortunately, as we shall show in Section 4, this is not practical. The uniformly most powerful level α test of H_o^* versus H_a^* is the "no data" test which rejects H_o^* with probability α, regardless of the data. The reason for this is that for any pair (F,G) satisfying H_a^*, there is a sequence of pairs (F,G_n), n = 1,2,..., satisfying H_o^* which converge to (F,G) in such a way that the power of any test at (F,G) is the limit of the powers at (F,G_n). Since (F,G_n) satisfies H_o^*, this limit must be at most α. The proof of this fact suggests that the difficulty may be that G(x) may be less than F(x) for most x's but G(x) may be slightly greater than F(x) for some x in the tail of the distributions. Thus, although $P(Y \geq x) = 1 - G(x) \geq 1 - F(x) = P(X \geq x)$ for most x's, G is not stochastically larger than F. In this paper we consider null and alternative hypotheses which are similar to H_o^* and H_a^* but which circumvent this troublesome tail behavior and allow the development of a reasonable level α test.

Specifically, in this paper we consider testing

$$H_o: \; F(x) = G(x) \text{ for all } x \in [a,b]$$
$$\text{or} \quad F(x) < G(x) \text{ for at least one } x \in [a,b]$$

versus

$$H_a: \; F(x) \geq G(x) \text{ for all } x \in [a,b] \text{ with strict inequality for at least one } x \in [a,b] \, .$$

Here $-\infty < a \leq b < \infty$ are fixed constants which define the interval of x values on which F(x) and G(x) are to be compared. The relationship described in H_a is the same as that which defines stochastic ordering except that the inequality is required to hold only for $x \in [a,b]$ rather than for all x. If the interval [a,b] is reasonably wide, H_o and H_a provide practical approximations to H_o^* and H_a^*. Rejection of H_o by a level α test (α small) will provide convincing evidence that $F(x) \geq G(x)$ for all $x \in [a,b]$. In many situations H_o

and H_a provide more meaningful and easily understood approximations to H_o^* and H_a^* than do the hypotheses H_o'' and H_a'' associated with RST. In this paper we consider the one sample problem in which F(x) is a known standard and a random sample is available from G(x). Research is currently being completed on the two sample problem.

In Section 2, a level α intersection-union test of H_o versus H_a is described. A table of small sample critical values and a large sample approximation for the critical values are given. Also a graphical description of the test is provided. In Section 3, the size of the test is shown to be the prescribed value. The test is shown to be the uniformly most powerful level α test among all monotone, nonrandomized, permutation invariant tests. The consistency class of the test is described. The problem of testing H_o^*: G is not stochastically larger than F versus H_a^*: $G >^{st} F$ is considered in Section 4.

2. Derivation of the test. Let F be a known cdf. Let $Y_1,\ldots,Y_n$ be a random sample from an unknown cdf G. Note that neither F nor G are required to be continuous. We will note in later results when continuity simplifies something. Let $0 < \alpha < 1$ be the desired level of the test. Define the inverse of F in the usual way, namely, $F^{-1}(p) = \inf\{x: F(x) \geq p\}$. (If $F(x) < 1$ for all x, define $F^{-1}(1) = \infty$.) Let a and b be fixed values satisfying $F^{-1}(1 - \alpha^{1/n}) \leq a \leq b < F^{-1}(1)$. Then a level α test of H_o versus H_a is given by the following.

First define constants $p_0,\ldots,p_{n-1}$ by the relationship

$$P_{p_i}(B \leq i) = \alpha, \quad i = 0,\ldots,n-1 , \tag{2.1}$$

where B is a binomial random variable with parameters p_i and n. The constant p_i also depends on n and α but that dependence will be suppressed in the notation. Table 1 provides values of p_i for n = 1,...,30, i = 0,...,n-1, and α = .01, .05, and .10. Next, for i = 0,...,n-1, define critical values $c_i = F^{-1}(p_i)$ and let $c_n = F^{-1}(1)$. Note that the restriction $F^{-1}(1 - \alpha^{1/n}) \leq a$ simply requires that $a \geq c_0$. Define integers I and J by the relationships $c_{I-1} \leq a < c_I$ and $c_{J-1} \leq b < c_J$. Note that $I \leq J$ since $a \leq b$. Let $Y_{(1)} \leq \ldots \leq Y_{(n)}$ denote the order statistics from the sample $Y_1, \ldots, Y_n$. Then the proposed test, that we will call test *T*, is defined as follows

(2.2) Reject H_o if and only if $Y_{(i)} \geq c_i$, $i = I, \ldots, J-1$ and $Y_{(J)} > b$.

Note that if I = J then the set i = I,...,J-1 is empty and test *T* simply rejects H_0 if and only if $Y_{(J)} > b$. But this corresponds to the atypical situation in which the interval [a,b] is very short and $c_{J-1} \leq a \leq b < c_J$. Typically the interval is wide so that H_0 and H_a are similar to the stochastic ordering hypotheses H_0^* and H_a^* from Section 1. In all the remaining discussion we will assume that $I < J$. The arguments can be easily modified to handle the I = J case.

Values for p_i, that are used to define the critical values $c_i = F^{-1}(p_i)$, are given in Table 1 for n = 1,...,30. For larger values of n, $p_0 = 1 - \alpha^{1/n}$ and $p_{n-1} = (1 - \alpha)^{1/n}$ are easily calculated. For other values of i, a normal approximation to the binomial distribution (with continuity correction) gives reasonable approximations of p_i as follows. The approximation is based on

$$P_{p_i}\left(\frac{B - np_i}{\sqrt{np_i(1-p_i)}} \leq \frac{i + \frac{1}{2} - np_i}{\sqrt{np_i(1-p_i)}}\right) = \alpha \approx P_{p_i}\left(\frac{B - np_i}{\sqrt{np_i(1-p_i)}} \leq z_\alpha\right)$$

where z_α is the lower $(100\alpha)^{th}$ percentile of a standard normal distribution. So setting $z_\alpha = (i+\frac{1}{2}-np_i)/(np_i(1-p_i))^{\frac{1}{2}}$ and solving for p_i yields

$$p_i = \frac{i + ½ + ¼z_\alpha^2 + \sqrt{z_\alpha^2(-(i + ½)^2/n + i + ½ + ¼z_\alpha^2)}}{n + z_\alpha^2} \quad .$$

Table 2 gives the actual and approximate values of p_i for n = 30 and α = .01, .05, and .10. The approximation is quite good, especially for i near ½n. The approximation is within .005 of the actual value for $1 \le i \le 28$ for $\alpha = .10$.

The rejection region for test *T* can be related to the empirical cdf, $G_n(x) = n^{-1}\Sigma_{i=1}^n I(Y_{(i)} \le x)$, of the sample $Y_1,\ldots,Y_n$. ($I(\cdot)$ is the indicator function.) The condition $Y_{(i)} \ge c_i$ is equivalent to $G_n(x) \le (i-1)/n$ for all x $< c_i$. Similarly $Y_{(J)} > b$ is equivalent to $G_n(x) \le (J-1)/n$ for all $x \le c_J$. These conditions define a step-shaped region as shown in Figures 1a and 1b. Test *T* rejects H_0 if and only if the graph of the empirical cdf lies entirely in the shaded, step-shaped region. The rejection region extends to $-\infty$, as in Figure 1a, if $a \ge c_1$ and, hence, $I \ge 2$. In this case the rejection region puts no restriction on $Y_{(1)}$. But if $c_0 \le a < c_1$, then the rejection region has a finite lower endpoint, as in Figure 1b.

Test *T* was originally derived using the intersection-union (IU) method of test construction. This method was used as early as 1952 by Lehmann. More recently the method has been used by a variety of authors including Gleser (1973), Sasabuchi (1980), Cohen and Marden (1983), Berger (1984a) and Berger (1984b). The method has not always been identified as the IU method.

The IU method of test construction can be used if the null hypothesis can be conveniently expressed as a union of sets. In our problem, we express H_0 as $(\cup_{a \le x \le b} \mathcal{G}_x) \cup \mathcal{G}_=$ where, for each $x \in [a,b]$, $\mathcal{G}_x = \{G: G(x) > F(x)\}$ and $\mathcal{G}_= = \{G: G(x) = F(x) \text{ for all } x \in [a,b]\}$. The IU method prescribes that each of the hypotheses H_{0x}: $G \in \mathcal{G}_x$ as well as $H_{0=}$: $G \in \mathcal{G}_=$ be tested with a level

α test and H_o is rejected if and only if each of the individual tests rejects its hypothesis. That is the rejection region for the test of H_o is the intersection of the rejection regions for each of the tests of H_{ox} and $H_{o=}$.

The usual test of H_{ox} is based on the statistic B_x = the number of sample values Y_i that are less than or equal to x. The hypothesis H_{ox} is rejected if B_x is sufficiently small. In particular, if $c_{i-1} \leq x < c_i$, then $p_{i-1} \leq F(x) < p_i$. Thus, from (2.1) we see that the level α test of H_{ox} rejects H_{ox} if and only if $B_x \leq i - 1$. This is equivalent to $Y_{(i)} > x$. Now, if $a < c_i \leq b$, the requirement that $Y_{(i)} > x$ for all $x \in [c_{i-1},c_i)\cap[a,b]$ is equivalent to $Y_{(i)} \geq c_i$. If $c_{i-1} \leq b < c_i$, the requirement that $Y_{(i)} > x$ for all $x \in [c_{i-1}, c_i)\cap[a,b]$ is equivalent to $Y_{(i)} > b$. By the definition of I and J, the only intervals of the form $[c_{i-1},c_i)$ which intersect $[a,b]$ are $[c_{I-1},c_I),\ldots,[c_{J-1},c_J)$. Thus the intersection over all $x \in [a,b]$ of the rejection regions of the tests of the hypotheses H_{ox} is the rejection region described in (2.2). According to the IU method, $H_{o=}$ must also be tested with a level α test. But note that for any $x \in [a,b]$, the test which rejects if $Y_{(i)} > x$ is also a level α test for $\bar{H}_{ox}$: $G \in \bar{G}_x$ where $\bar{G}_x = \{G: G(x) \geq F(x)\}$. Since $G_{o=} \subset \bar{G}_x$, this test which was already included as the test for H_{ox} is also a level α test of $H_{o=}$. Thus, no additional test for $H_{o=}$ need be included and, indeed, the inclusion of a different test for $H_{o=}$ would only serve to reduce the power of the resulting test.

Some properties of the test T that we have developed using the IU methodology will be discussed in the next section.

3. Power properties of test. In this section we will discuss three properties of the power function of test T. First we will show that test T is a level α test and give conditions under which the size is exactly α. Next we will show that test T is a uniformly most powerful (UMP) level α test in the class of nonrandomized, monotone, permutation invariant tests. Finally we will describe the consistency class for test *T*.

3.1 Size of test. Test *T* is a level α test of H_o versus H_a. In Theorem 3.1, a condition is given under which the size of the test is exactly the specified value α. This property is important because sometimes the use of level α tests in constructing a test using the IU method will yield an overall test with size much less than α. Thus the constructed test will be overly conservative. Note, in particular, that the condition in Theorem 3.1 that ensures a size α test is satisfied if F is a continuous cdf and [a,b] is wide enough that $I < J$.

THEOREM 3.1. Test *T* is a level α test of H_o. That is,

$$\sup_{G \varepsilon H_o} P_G(\text{reject } H_o) \leq \alpha \ . \tag{3.1}$$

If in addition, for some j satisfying $I \leq j < J$, there exists an x such that $F(x) = p_j$ then test *T* has size exactly α, that is (3.1) is true with equality.

PROOF. Fix $G \varepsilon H_o$. Either $G \varepsilon \mathcal{G}_x$ for some $x \varepsilon [a,b]$ or $G \varepsilon \mathcal{G}_=$. If $G \varepsilon \mathcal{G}_x$ then, as explained in Section 2, the test will reject only if $Y_{(i)} > x$ where i satisfies $c_{i-1} \leq x < c_i$. Thus, $P_G(\text{reject } H_o) \leq P_G(Y_{(i)} > x) \leq \alpha$. If $G \varepsilon \mathcal{G}_=$, then, as also explained in Section 2, for any $x \varepsilon [a,b]$ the test for H_{ox} is also a level α test for $H_{o=}$. Thus, $P_G(\text{reject } H_o) \leq P_G(Y_{(i)} > x) \leq \alpha$ and (3.1) is verified.

To prove the equality in (3.1) note that the definition of c_j as $F^{-1}(p_j)$, the right continuity of F, and the existence of an x with $F(x) = p_j$ together imply that $F(c_j) = p_j$ and $c_{j+1} > c_j$. The definitions of I and J imply that $a < c_j \leq b$. Fix an $x^* > b$ and for any ν satisfying $0 < \nu < 1-p_j$ define $G_\nu(x)$

as the cdf of the probability measure which puts mass $p_j + \nu$ on the point c_j and mass $1 - p_j - \nu$ on the point x^*. For every ν, $0 < \nu < 1-p_j$, $G_\nu \varepsilon H_o$ because $G_\nu(c_j) = p_j + \nu > p_j = F(c_j)$ and $c_j \varepsilon [a,b]$. Consider the inequalities which define the rejection region in (2.2). If G_ν is the true distribution, the inequalities corresponding to subscripts i satisfying $I \leq i \leq j$ are true with probability one. The inequalities for subscripts i satisfying $j + 1 \leq i \leq J$ are true if and only if $Y_{(j+1)} = x^*$ (recall $c_{j+1} > c_j$), that is, if at most j of the Y_i's equal c_j. But B, the random variable that counts the number of Y_i's equal to c_j, has a binomial distribution with success probability $p_j + \nu$. Combining (2.1), (3.1), and the fact the $G_\nu \varepsilon H_o$ we have

$$\alpha \geq \sup_{G\varepsilon H_o} P_G(\text{reject } H_o) \geq \overline{\lim_{\nu\downarrow 0}} P_{G_\nu}(\text{reject } H_o)$$

$$= \overline{\lim_{\nu\downarrow 0}} P_{p_j+\nu}(B \leq j) = P_{p_j}(B \leq j) = \alpha .$$

Thus the equality in (3.1) is proved. □

3.2 *Optimality property.* Test T is a UMP level α test of H_o in the class of nonrandomized, monotone, permutation invariant tests. A nonrandomized test is called *monotone* if $(y_1,\ldots,y_n)$ is in rejection region of the test and $y_i^* \geq y_i$, $i = 1,\ldots,n$, imply that $(y_1^*,\ldots,y_n^*)$ is in the rejection region of the test. Monotone tests have monotone power functions in that if $G \geq^{st} G^*$ then $P_G(\text{reject } H_o) \geq P_{G^*}(\text{reject } H_o)$ for any monotone test. Optimality properties such as this for IU tests in other problems have been discussed by Lehmann (1952), Berger (1982), and Cohen and Marden (1983). It is interesting to note that the following results does not require test T to have size exactly α. The distribution F may be discrete and the test may have size less than α and still the test will be UMP in this class of tests.

THEOREM 3.2. Test T is a UMP level α test of H_0 in the class of nonrandomized, monotone, permutation invariant tests.

PROOF: We will show that any monotone, permutation invariant, level α rejection region is a subset of the rejection region for test T defined in (2.2). Hence test T is UMP. We will show this by showing that any monotone, permutation invariant rejection region which is not a subset of the region in (2.2) has size greater than α.

Consider a monotone, permutation invariant rejection region which is not a subset of the region in (2.2). Let $(y_1,\dots,y_n)$ be a point in this rejection region which does not satisfy all the inequalities in (2.2). Fix j, $I \le j \le J$ such that the inequality in (2.2) involving $y_{(j)}$ is not satisfied. That is $y_{(j)} < c_j$, if $j < J$, or $y_{(j)} \le b$ if $j = J$. Set $y^* = \max(y_{(j)},a)$. Recall that $a < c_I$ and $b < c_J$ so that, regardless of which subscript j we are considering, $F(y^*) < p_j$. Choose $\nu > 0$ so that $p_j - \nu > F(y^*)$. Fix $x^* > \max(y_{(n)},b)$. Note that $x^* > y^*$. Let G_ν denote the cdf of the probability distribution that puts mass $p_j - \nu$ on y^* and $1 - p_j + \nu$ on x^*. Then $G_\nu \in H_0$ since $a \le y^* \le b$ and $G_\nu(y^*) = p_j - \nu > F(y^*)$. Let B be the random variable which counts the number of Y_i's equal to y^*. Under G_ν, B has a binomial distribution with success probability $p_j - \nu < p_j$.

Now consider sample points (n-vectors) consisting only of x^*'s and y^*'s. Since $x^* > \max(y_{(n)},b) \ge y^* \ge y_{(j)}$, by the monotonicity and permutation invariance of the rejection region, all such sample points with at most j y^*'s are in this rejection region. Thus for this rejection region, by (2.1) we have

$$\sup_{G \in H_0} P_G(\text{reject } H_0) \ge P_{G_\nu}(\text{reject } H_0) \ge P_{G_\nu}(B \le j)$$

$$= P_{p_j - \nu}(B \le j) > P_{p_j}(B \le j) = \alpha.$$

This rejection region has size greater than α as was to be shown. □

A nonparametric upper confidence bound for the cdf G might be used to construct a level α test of H_o. A test which rejected H_o only if the $100(1-\alpha)\%$ upper confidence bound was less than or equal to F(x) for all $x \varepsilon [a,b]$ would be a level α test of H_o. But the usual upper confidence bounds, such as those described by Sandford (1985), would produce a nonrandomized, monotone, permutation invariant test. Thus, by Theorem 3.2, test *T* is a more powerful test than a test constructed in this way.

Test *T* is not UMP among all level α tests. Test *T* is biased in that $P_G(\text{reject } H_o) < \alpha$ if $G(x) = F(x)$ for a nondegenerate interval of x values, even if $G \varepsilon H_a$. On the other hand, the randomized "no data" test than simply rejects H_o with probabibility α, regardless of the data, has $P_G(\text{reject } H_o) = \alpha$ for all G. Thus the "no data" test has higher power than test *T* for some $G \varepsilon H_a$.

3.3 Consistency. In this section, consistency properties of test *T* are described. Let $\gamma = \inf_{a \le x \le b}[F(x) - G(x)]$. We show that the asymptotic value of $P_G(\text{reject } H_o)$ depends on the value of γ. An interpretation of the parameter γ is given at the end of this section. The following two theorems relate the power of the test to γ.

THEOREM 3.3. For test *T*, if $\gamma \le 0$ then $P_G(\text{reject } H_o) \le \alpha$, for every sample size n, and hence, $\lim_{n\to\infty} P_G(\text{reject } H_o) \le \alpha$.

THEOREM 3.4. For test *T*, if $\gamma > 0$ then $\lim_{n\to\infty} P_G(\text{reject } H_o) = 1$.

PROOF OF THEOREM 3.3. If $\gamma < 0$, $G(x) > F(x)$ for some $x \varepsilon [a,b]$. Thus $G \varepsilon H_o$, and by Theorem 3.1, $P_G(\text{reject } H_o) \le \alpha$. If $\gamma = 0$ and $G(x) = F(x)$ for some $x \varepsilon [a,b]$, then as explained in Section 2, the test which rejects if $Y_{(i)} > x$ (where $c_{i-1} \le x < c_i$) is a level α test of $\bar{H}_{ox}: G \varepsilon \bar{G}_x = \{G: G(x) \ge F(x)\}$.

So also for this type of G we have $P_G(\text{reject } H_o) \le P_G(Y_{(1)} \ge x) \le \alpha$. The following argument handles the remaining case, $\gamma = 0$ and $G(x) < F(x)$ for all $x \in [a,b]$.

Since $[a,b]$ is compact, there exist $x_o \in [a,b]$ and a sequence of values $x_m \in [a,b]$ such that $\lim_{m\to\infty} x_m = x_o$ and $\lim_{m\to\infty} [F(x_m) - G(x_m)] = 0$. Since $F(x_o) \ge G(x_o)$, the right continuity of F and G imply that x_m must approach x_o from below and

$$(3.2) \qquad G(x_o^-) = \lim_{x \uparrow x_o} G(x) = \lim_{x \uparrow x_o} F(x) = F(x_o^-) .$$

Note that

$$(3.3) \qquad 0 < F(a) \le F(x_o^-) = G(x_o^-) \le G(x_o) < F(x_o) \le F(b) < 1.$$

Define a cdf H(x) by

$$(3.4) \qquad H(x) = \begin{cases} G(x)/G(x_o^-) & x < x_o \\ 1 & x \ge x_o \end{cases} .$$

For any $\nu \in (0,1)$ define $G_\nu(x) = (1 - \nu)G(x) + \nu H(x)$.

First we will show that for any $\nu \in (0,1)$, $G_\nu \in H_o$ and hence, by Theorem 3.1,

$$(3.5) \qquad P_{G_\nu}(\text{reject } H_o) \le \alpha .$$

Recall that $F(x_o^-) = G(x_o^-) < 1 = H(x_o^-)$. Thus

$$\lim_{m\to\infty} [F(x_m) - G_\nu(x_m)] = F(x_o^-) - (1 - \nu)G(x_o^-) - \nu H(x_o^-)$$

$$= \nu (G(x_o^-) - 1) < 0 .$$

Thus $G_\nu(x_m) > F(x_m)$ for some x_m, verifying that $G_\nu \in H_o$ and (3.5).

Now we will show that

$$(3.6) \qquad \lim_{\nu \downarrow 0} P_{G_\nu}(\text{reject } H_o) \geq P_G(\text{reject } H_o)$$

which with (3.5) will complete the proof that $P_G(\text{reject } H_o) \leq \alpha$. Let $Y = (Y_1,\ldots,Y_n)$ be a random sample from G. Let $Z_1,\ldots,Z_n$ be a random sample from H. Fix $\nu \in (0,1)$. Let $I_1,\ldots,I_n$ be independent Bernoulli random variables with $P(I_i = 1) = \nu$. All the Y's, Z's, and I's are mutually independent. Define $Y_i^* = Z_i I_i + Y_i(1 - I_i)$. That is $Y_i^* = Z_i$ if $I_i = 1$ and $Y_i^* = Y_i$ if $I_i = 0$. Then $Y^* = (Y_1^*,\ldots,Y_n^*)$ is a random sample from the cdf G_ν. Let A = the rejection region for test T and $B = I_1+\ldots+I_n$. Note that B and Y are independent and $Y = Y^*$ if $B = 0$. Using Theorem 3.1 and the fact that $G_\nu \in H_o$ we have

$$\alpha \geq P_{G_\nu}(\text{reject } H_o) = P(Y^* \in A) \geq P(Y^* \in A,\ B = 0)$$

$$= P(Y \in A,\ B = 0) = P(Y \in A)P(B = 0) = P_G(\text{reject } H_o)\cdot(1 - \nu)^n .$$

Taking the limit as $\nu \downarrow 0$ yields (3.6). □

SKETCH OF PROOF OF THEOREM 3.4. A proof of Theorem 3.4 can be based on the graphical representation of the rejection region described in Section 2.

Let [a] denote the greatest integer less than or equal to a. Then the Weak Law of Large Numbers can be used to show that for any $v \in (0,1)$, $\lim_{n\to\infty} p_{[nv]} = v$. This convergence is uniform for $v \in [F(a), F(b)]$. These facts can be used to show that the step-shaped boundary, with vertices at the points $(F^{-1}(p_i), i/n)$, converges to F(x) uniformly on the interval [a,b]. On the other hand, by the Glivenko-Cantelli Theorem, the empirical cdf converges to G(x) uniformly on [a,b] with probability one. Thus, if $\gamma = \inf_{a\leq x\leq b}[F(x) - G(x)] > 0$, for large n the boundary is greater than $F(x) - \gamma/2 \geq G(x) + \gamma/2$ for all $x \in [a,b]$. Test T will reject H_o if

$G_n(x) < G(x) + \gamma/2$ for all $x \in [a,b]$ and this event has probability approaching one as $n \to \infty$. □

Theorems 3.3 and 3.4 say that test T is a level α test of H_o^+: $\gamma \leq 0$ versus H_a^+: $\gamma > 0$ and the test is consistent against all points in H_a^+. Thus the parameter γ plays the same role for test T that the parameter $\theta = P(X \leq Y)$ plays for the RST. The parameter $\gamma = \inf_{a \leq x \leq b} [F(x) - G(x)]$ may be interpreted as follows:

$\gamma > 0$ if and only if $P(X \leq x) > P(Y \leq x)$ for all $x \in [a,b]$
and $P(X < x) > P(Y < x)$ for all $x \in [a,b]$

$\gamma \leq 0$ if and only if $P(X \leq x) \leq P(Y \leq x)$ for some $x \in [a,b]$
or $P(X < x) \leq P(Y < x)$ for some $x \in [a,b]$.

If F is continuous, then each of the inequalities involving $<$ can be dropped in the above interpretation. As mentioned in Section 1, we believe that in many situations the hypothesis H_a^+: $\gamma > 0$ provides a more meaningful approximation to H_a^*: $G >^{st} F$ than does H_a'': $\theta > ½$, the alternative hypothesis associated with the RST.

4. UMP test of H_o^* versus H_a^*. There is no satisfactory level α test for testing H_o^*: G is not stochastically larger than F versus H_a^*: $G >^{st} F$. For any level α test of H_o^*, P_G(reject H_o^*) $\leq \alpha$ for *every* cdf G including all $G \in H_a^*$. This means that the randomized "no data" test that rejects H_o^* with probability α, regardless of the data, and has P_G(reject H_o^*) $= \alpha$ for all G is the UMP level α test of H_o^* versus H_a^*. This fact is proved in the following theorem.

THEOREM 4.1. Consider testing H_o^* versus H_a^*. Let G be any cdf. Then there exists a sequence of cdf's G_m, $m = 1,2,\ldots$, such that $G_m \in H_o^*$ for every m and, for any (possibly randomized) level α test of H_o^*,

(4.1) $$\alpha \geq \lim_{m\to\infty} P_{G_m}(\text{reject } H_o^*) \geq P_G(\text{reject } H_o^*) \ .$$

PROOF: The proof is similar to the proof of Theorem 3.3. For each $m = 1,2,\ldots$, fix a value x_m such that $F(x_m) < 1/m$ and let H_m be a cdf such that $H_m(x) = 1$ for $x \geq x_m$. Let $G_m = (1 - 1/m)G + (1/m)H_m$. Then $G_m(x_m) = (1 - 1/m)G(x_m) + 1/m \geq 1/m > F(x_m)$. So G_m is not stochastically larger than F. Let $y = (y_1,\ldots,y_n)$ denote a sample point and let $\phi(y)$ denote the test function for any level α test. That is $\phi(y)$ is the probability H_o^* is rejected given that y is observed. Let $Y = (Y_1,\ldots,Y_n)$ be a random sample from G, let $Z_m = (Z_{1m},\ldots,Z_{nm})$ be a random sample from H_m and let $I_{1m},\ldots,I_{nm}$ be Bernoulli random variables with success probability 1/m. Assume all the Y's, Z's and I's are mutually independent. Let $Y_m^* = (Y_{1m}^*,\ldots,Y_{nm}^*)$ where $Y_{im}^* = Z_{im}I_{im} + Y_i(1 - I_{im})$. Note that Y_m^* is a random sample from G_m. Let $B_m = I_{1m} +\ldots+ I_{nm}$ and note that Y and B_m are independent and $Y = Y_m^*$ if $B_m = 0$. Thus the conditional distribution of Y_m^* given $B_m = 0$ is the same as the conditional distribution of Y given $B_m = 0$ and this is just the unconditional distribution of Y. Thus we have for every $m = 1,2,\ldots$,

$$P_{G_m}(\text{reject } H_o^*) = E\,\phi(Y_m^*) = E(\phi(Y_m^*)|B_m=0)P(B_m=0) + E(\phi(Y_m^*)|B_m>0)P(B_m>0)$$

(4.2)

$$\geq E(\phi(Y_m^*)|B_m=0)P(B_m=0) = E\phi(Y)(1-1/m)^n = P_G(\text{reject } H_o^*)(1-1/m)^n.$$

Since $G_m \in H_o^*$ for every $m = 1,2,\ldots$, $P_{G_m}(\text{reject } H_o^*) \leq \alpha$ for every m. Thus taking limits as $m \to \infty$ in (4.2) yields (4.1). □

In the above proof, since $F(x_m) < 1/m$, we must have $\overline{\lim}_{m\to\infty} x_m \le \inf\{x\colon F(x) > 0\}$. That is, the x_m are in the left tail of F. This suggests why we were able to develop a reasonable test of H_o versus H_a. These hypotheses ignore the tails of F.

REFERENCES

Berger, R. L. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics* 24, 295-300.

Berger, R. L. (1984a). Testing for the same ordering in several groups of means. *Design of Experiments: Ranking and Selection* 241-249. Ed. T. J. Santner and A. C. Tamhane. Marcel Dekker, New York.

Berger, R. L. (1984b). Testing whether one regression function is larger than another. *Comm. Statist. A - Theory Methods* 13, 1793-1810.

Cohen, A. and Marden, J. (1983). Hypothesis tests and optimality properties in discrete multivariate analysis. *Studies in Econometrics, Time Series, and Multivariate Statistics* 379-405. Academic Press, New York.

Gleser, L. J. (1973). On a theory of intersection-union tests. *IMS Bulletin* 2, 233. Abstract.

Lee, Y. J. and Wolfe, D. A. (1976). A distribution-free test for stochastic ordering. *J. Amer. Statist. Assoc.* 71, 722-727.

Lehmann, E. L. (1952). Testing multiparameter hypotheses. *Ann. Math. Statist.* 23, 541-552.

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks.* Holden-Day, Oakland.

Sandford, M. D. (1985). Nonparametric one-sided confidence intervals for an unknown distribution function using censored data. *Technometrics* 27, 41-48.

Sasabuchi, S. (1980). A test of a multivariate normal mean with composite hypotheses determined by linear inequalities. *Biometrika* 67, 429-439.

TABLE 1

Critical values p_i for $\alpha = .10$

n

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	.9000	.6838	.5358	.4377	.3690	.3187	.2803	.2501	.2257	.2057	.1889	.1746	.1623	.1517	.1423
1		.9487	.8042	.6795	.5839	.5103	.4526	.4062	.3684	.3368	.3102	.2875	.2678	.2507	.2356
2			.9655	.8574	.7534	.6668	.5962	.5382	.4901	.4496	.4152	.3855	.3598	.3372	.3173
3				.9740	.8878	.7991	.7214	.6554	.5994	.5517	.5108	.4753	.4443	.4170	.3928
4					.9791	.9074	.8304	.7603	.6990	.6458	.5995	.5590	.5234	.4920	.4640
5						.9826	.9212	.8531	.7896	.7327	.6823	.6377	.5982	.5631	.5317
6							.9851	.9314	.8705	.8124	.7595	.7118	.6691	.6309	.5965
7								.9869	.9392	.8842	.8308	.7813	.7363	.6954	.6585
8									.9884	.9455	.8952	.8458	.7995	.7568	.7178
9										.9895	.9505	.9043	.8584	.8149	.7744
10											.9905	.9548	.9120	.8691	.8280
11												.9913	.9583	.9185	.8782
12													.9919	.9613	.9241
13														.9925	.9640
14															.9930

n

i	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
0	.1340	.1267	.1201	.1141	.1087	.1038	.0994	.0953	.0915	.0880	.0848	.0817	.0789	.0763	.0739
1	.2222	.2102	.1995	.1898	.1810	.1729	.1656	.1588	.1526	.1469	.1415	.1366	.1319	.1276	.1236
2	.2996	.2837	.2694	.2565	.2448	.2340	.2242	.2152	.2069	.1991	.1920	.1853	.1791	.1733	.1678
3	.3712	.3519	.3344	.3186	.3042	.2910	.2789	.2678	.2575	.2480	.2392	.2309	.2232	.2160	.2093
4	.4389	.4164	.3960	.3775	.3607	.3452	.3310	.3180	.3059	.2947	.2842	.2745	.2655	.2570	.2490
5	.5035	.4781	.4550	.4340	.4149	.3973	.3812	.3663	.3525	.3397	.3277	.3166	.3062	.2965	.2874
6	.5654	.5374	.5118	.4886	.4673	.4477	.4297	.4131	.3976	.3833	.3700	.3575	.3459	.3349	.3247
7	.6250	.5945	.5667	.5413	.5180	.4966	.4768	.4586	.4416	.4258	.4111	.3974	.3845	.3725	.3611
8	.6822	.6496	.6198	.5925	.5673	.5442	.5228	.5029	.4845	.4673	.4513	.4364	.4224	.4092	.3968
9	.7371	.7027	.6712	.6421	.6152	.5905	.5675	.5462	.5264	.5080	.4907	.4746	.4594	.4452	.4319
10	.7896	.7539	.7208	.6902	.6618	.6356	.6112	.5885	.5674	.5477	.5293	.5120	.4958	.4806	.4663
11	.8394	.8028	.7686	.7367	.7071	.6795	.6538	.6299	.6076	.5867	.5671	.5488	.5316	.5154	.5001
12	.8862	.8494	.8145	.7817	.7509	.7222	.6954	.6703	.6468	.6249	.6043	.5849	.5667	.5496	.5334
13	.9290	.8932	.8582	.8249	.7933	.7637	.7358	.7097	.6852	.6623	.6407	.6204	.6013	.5832	.5662
14	.9663	.9333	.8994	.8661	.8341	.8038	.7752	.7482	.7228	.6989	.6764	.6552	.6352	.6163	.5985
15	.9934	.9683	.9371	.9049	.8731	.8425	.8133	.7856	.7594	.7347	.7114	.6894	.6686	.6489	.6303
16		.9938	.9701	.9405	.9098	.8794	.8500	.8218	.7951	.7697	.7456	.7229	.7013	.6809	.6616
17			.9942	.9717	.9436	.9142	.8851	.8568	.8297	.8038	.7791	.7557	.7335	.7124	.6924
18				.9945	.9731	.9463	.9183	.8903	.8631	.8368	.8117	.7878	.7650	.7433	.7227
19					.9947	.9744	.9488	.9219	.8950	.8688	.8434	.8191	.7958	.7736	.7524
20						.9950	.9756	.9511	.9253	.8994	.8740	.8495	.8259	.8032	.7816
21							.9952	.9766	.9532	.9283	.9034	.8789	.8551	.8322	.8101
22								.9954	.9776	.9551	.9312	.9071	.8834	.8603	.8380
23									.9956	.9785	.9568	.9338	.9105	.8875	.8652
24										.9958	.9794	.9585	.9362	.9137	.8914
25											.9960	.9801	.9600	.9385	.9166
26												.9961	.9808	.9614	.9406
27													.9962	.9815	.9627
28														.9964	.9821
29															.9965

TABLE 1 - Continued

Critical values p_i for α = .05

n

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	.9500	.7764	.6316	.5271	.4507	.3930	.3482	.3123	.2831	.2589	.2384	.2209	.2058	.1926	.1810
1		.9747	.8647	.7514	.6574	.5818	.5207	.4707	.4291	.3942	.3644	.3387	.3163	.2967	.2794
2			.9830	.9024	.8107	.7287	.6587	.5997	.5496	.5069	.4701	.4381	.4101	.3854	.3634
3				.9873	.9236	.8468	.7747	.7108	.6551	.6066	.5644	.5273	.4947	.4657	.4398
4					.9898	.9372	.8712	.8071	.7486	.6965	.6502	.6091	.5726	.5400	.5108
5						.9915	.9466	.8889	.8313	.7776	.7288	.6848	.6452	.6096	.5774
6							.9927	.9536	.9023	.8500	.8004	.7547	.7130	.6750	.6404
7								.9936	.9590	.9127	.8649	.8190	.7760	.7364	.7000
8									.9943	.9632	.9212	.8772	.8343	.7939	.7563
9										.9949	.9667	.9281	.8873	.8473	.8091
10											.9953	.9695	.9340	.8960	.8583
11												.9957	.9719	.9389	.9033
12													.9961	.9740	.9432
13														.9963	.9758
14															.9966

n

i	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
0	.1707	.1616	.1533	.1459	.1391	.1329	.1273	.1221	.1173	.1129	.1088	.1050	.1015	.0981	.0950
1	.2640	.2501	.2377	.2264	.2161	.2067	.1981	.1902	.1829	.1761	.1698	.1640	.1585	.1534	.1486
2	.3438	.3262	.3103	.2958	.2826	.2706	.2595	.2493	.2398	.2310	.2229	.2153	.2082	.2016	.1953
3	.4166	.3956	.3767	.3594	.3437	.3292	.3159	.3036	.2923	.2817	.2719	.2627	.2542	.2461	.2386
4	.4844	.4605	.4389	.4191	.4010	.3844	.3691	.3549	.3418	.3296	.3182	.3076	.2977	.2884	.2796
5	.5483	.5219	.4978	.4758	.4556	.4370	.4198	.4039	.3891	.3754	.3626	.3506	.3394	.3289	.3190
6	.6090	.5803	.5540	.5300	.5078	.4874	.4685	.4510	.4347	.4195	.4054	.3921	.3797	.3680	.3570
7	.6666	.6360	.6078	.5819	.5580	.5359	.5155	.4964	.4787	.4622	.4468	.4323	.4187	.4060	.3939
8	.7214	.6892	.6594	.6319	.6064	.5828	.5609	.5405	.5214	.5036	.4870	.4714	.4567	.4429	.4299
9	.7733	.7399	.7088	.6799	.6531	.6281	.6048	.5832	.5629	.5439	.5262	.5095	.4938	.4790	.4651
10	.8222	.7881	.7560	.7261	.6980	.6719	.6475	.6246	.6032	.5832	.5643	.5466	.5300	.5143	.4994
11	.8679	.8336	.8010	.7703	.7414	.7142	.6887	.6649	.6424	.6214	.6016	.5829	.5654	.5488	.5331
12	.9097	.8762	.8437	.8125	.7829	.7550	.7287	.7039	.6806	.6586	.6379	.6184	.6000	.5825	.5661
13	.9469	.9154	.8836	.8525	.8227	.7943	.7673	.7418	.7176	.6949	.6734	.6530	.6338	.6156	.5984
14	.9773	.9501	.9203	.8901	.8604	.8318	.8044	.7784	.7536	.7301	.7079	.6869	.6669	.6480	.6301
15	.9968	.9787	.9530	.9247	.8959	.8676	.8401	.8137	.7884	.7644	.7416	.7199	.6993	.6797	.6611
16		.9970	.9799	.9555	.9286	.9012	.8740	.8475	.8220	.7976	.7743	.7521	.7309	.7107	.6915
17			.9972	.9810	.9578	.9322	.9059	.8798	.8543	.8297	.8060	.7834	.7617	.7411	.7213
18				.9973	.9819	.9599	.9354	.9102	.8851	.8605	.8367	.8138	.7918	.7707	.7505
19					.9974	.9828	.9618	.9383	.9141	.8899	.8662	.8432	.8209	.7995	.7789
20						.9976	.9836	.9635	.9410	.9177	.8944	.8715	.8492	.8275	.8067
21							.9977	.9843	.9650	.9434	.9210	.8985	.8763	.8547	.8337
22								.9978	.9850	.9665	.9457	.9241	.9023	.8808	.8598
23									.9979	.9856	.9678	.9478	.9269	.9058	.8850
24										.9980	.9862	.9690	.9497	.9295	.9091
25											.9980	.9867	.9702	.9515	.9319
26												.9981	.9872	.9712	.9531
27													.9982	.9876	.9722
28														.9982	.9880
29															.9983

TABLE 1 – Continued

Critical values p_i for $\alpha = .01$

n

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	.9900	.9000	.7846	.6838	.6019	.5358	.4821	.4377	.4005	.3690	.3421	.3187	.2983	.2803	.2644
1		.9950	.9411	.8591	.7779	.7057	.6434	.5899	.5440	.5044	.4698	.4395	.4128	.3891	.3679
2			.9967	.9580	.8944	.8269	.7637	.7068	.6563	.6117	.5723	.5373	.5062	.4783	.4532
3				.9975	.9673	.9153	.8577	.8018	.7500	.7029	.6604	.6222	.5878	.5567	.5285
4					.9980	.9732	.9292	.8791	.8290	.7817	.7378	.6976	.6609	.6274	.5969
5						.9983	.9773	.9392	.8947	.8496	.8060	.7651	.7271	.6920	.6597
6							.9986	.9803	.9467	.9068	.8656	.8254	.7871	.7512	.7177
7								.9987	.9826	.9525	.9163	.8785	.8412	.8053	.7713
8									.9989	.9845	.9572	.9241	.8892	.8543	.8205
9										.9990	.9859	.9610	.9305	.8981	.8654
10											.9991	.9872	.9642	.9360	.9056
11												.9992	.9882	.9669	.9406
12													.9992	.9890	.9693
13														.9993	.9898
14															.9993

n

i	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
0	.2501	.2373	.2257	.2152	.2057	.1969	.1889	.1815	.1746	.1682	.1623	.1568	.1517	.1468	.1423
1	.3488	.3316	.3160	.3018	.2888	.2768	.2658	.2557	.2462	.2375	.2293	.2217	.2146	.2079	.2016
2	.4305	.4099	.3912	.3741	.3583	.3439	.3305	.3181	.3066	.2959	.2859	.2766	.2679	.2596	.2519
3	.5029	.4796	.4583	.4387	.4207	.4041	.3887	.3745	.3612	.3488	.3372	.3264	.3162	.3066	.2976
4	.5690	.5434	.5199	.4983	.4783	.4598	.4426	.4267	.4118	.3979	.3849	.3727	.3613	.3505	.3403
5	.6299	.6025	.5772	.5538	.5321	.5120	.4933	.4758	.4595	.4443	.4300	.4166	.4039	.3920	.3808
6	.6866	.6577	.6309	.6060	.5829	.5613	.5412	.5224	.5048	.4884	.4729	.4584	.4447	.4317	.4195
7	.7393	.7094	.6814	.6553	.6309	.6082	.5868	.5669	.5482	.5306	.5140	.4984	.4837	.4699	.4567
8	.7883	.7578	.7290	.7020	.6766	.6528	.6304	.6094	.5897	.5711	.5535	.5370	.5214	.5066	.4927
9	.8335	.8029	.7737	.7460	.7199	.6953	.6721	.6502	.6295	.6100	.5916	.5742	.5578	.5422	.5274
10	.8749	.8448	.8156	.7876	.7610	.7358	.7119	.6892	.6678	.6476	.6284	.6102	.5930	.5767	.5612
11	.9122	.8832	.8546	.8267	.7999	.7743	.7499	.7267	.7047	.6837	.6639	.6450	.6271	.6101	.5939
12	.9446	.9178	.8904	.8632	.8366	.8109	.7862	.7626	.7401	.7186	.6982	.6787	.6602	.6426	.6258
13	.9713	.9481	.9228	.8968	.8708	.8454	.8207	.7969	.7740	.7521	.7312	.7113	.6922	.6741	.6568
14	.9905	.9731	.9512	.9272	.9025	.8777	.8532	.8295	.8065	.7844	.7631	.7428	.7233	.7047	.6868
15	.9994	.9910	.9746	.9539	.9312	.9075	.8838	.8603	.8375	.8152	.7938	.7732	.7533	.7343	.7161
16		.9994	.9915	.9760	.9564	.9347	.9121	.8893	.8668	.8447	.8232	.8024	.7824	.7631	.7445
17			.9994	.9920	.9773	.9586	.9379	.9162	.8944	.8727	.8513	.8306	.8104	.7909	.7720
18				.9995	.9924	.9784	.9606	.9408	.9200	.8990	.8780	.8574	.8373	.8177	.7987
19					.9995	.9928	.9794	.9624	.9434	.9235	.9032	.8830	.8630	.8435	.8245
20						.9995	.9931	.9804	.9640	.9458	.9266	.9071	.8875	.8682	.8492
21							.9995	.9934	.9812	.9655	.9480	.9295	.9107	.8917	.8730
22								.9996	.9937	.9820	.9669	.9500	.9322	.9140	.8956
23									.9996	.9940	.9827	.9682	.9519	.9347	.9170
24										.9996	.9942	.9834	.9694	.9537	.9370
25											.9996	.9944	.9840	.9705	.9553
26												.9996	.9946	.9845	.9715
27													.9996	.9948	.9851
28														.9997	.9950
29															.9997

TABLE 2

Comparison of actual and approximate critical values p_i for n = 30

i	α=.10 actual	α=.10 approx.	α=.05 actual	α=.05 approx.	α=.01 actual	α=.01 approx.
0	.0739	.0802	.0950	.1110	.1423	.1798
1	.1236	.1282	.1486	.1601	.2016	.2283
2	.1678	.1715	.1953	.2044	.2519	.2724
3	.2093	.2124	.2386	.2460	.2976	.3137
4	.2490	.2515	.2796	.2857	.3403	.3530
5	.2874	.2894	.3190	.3239	.3808	.3905
6	.3247	.3264	.3570	.3609	.4195	.4267
7	.3611	.3625	.3939	.3969	.4567	.4617
8	.3968	.3979	.4299	.4321	.4927	.4957
9	.4319	.4326	.4651	.4665	.5274	.5287
10	.4663	.4667	.4994	.5002	.5612	.5608
11	.5001	.5003	.5331	.5332	.5939	.5921
12	.5334	.5334	.5661	.5655	.6258	.6226
13	.5662	.5660	.5984	.5973	.6568	.6523
14	.5985	.5981	.6301	.6284	.6868	.6813
15	.6303	.6297	.6611	.6590	.7161	.7095
16	.6616	.6608	.6915	.6890	.7445	.7370
17	.6924	.6914	.7213	.7184	.7720	.7637
18	.7227	.7215	.7505	.7472	.7987	.7897
19	.7524	.7512	.7789	.7753	.8245	.8149
20	.7816	.7802	.8067	.8028	.8492	.8393
21	.8101	.8087	.8337	.8296	.8730	.8628
22	.8380	.8365	.8598	.8556	.8956	.8853
23	.8652	.8636	.8850	.8807	.9170	.9068
24	.8914	.8899	.9091	.9048	.9370	.9271
25	.9166	.9152	.9319	.9278	.9553	.9460
26	.9406	.9392	.9531	.9492	.9715	.9632
27	.9627	.9616	.9722	.9688	.9851	.9784
28	.9821	.9815	.9880	.9857	.9950	.9907
29	.9965	.9967	.9983	.9977	.9997	.9987

REJECTION REGION IN TERMS OF EMPIRICAL CDF

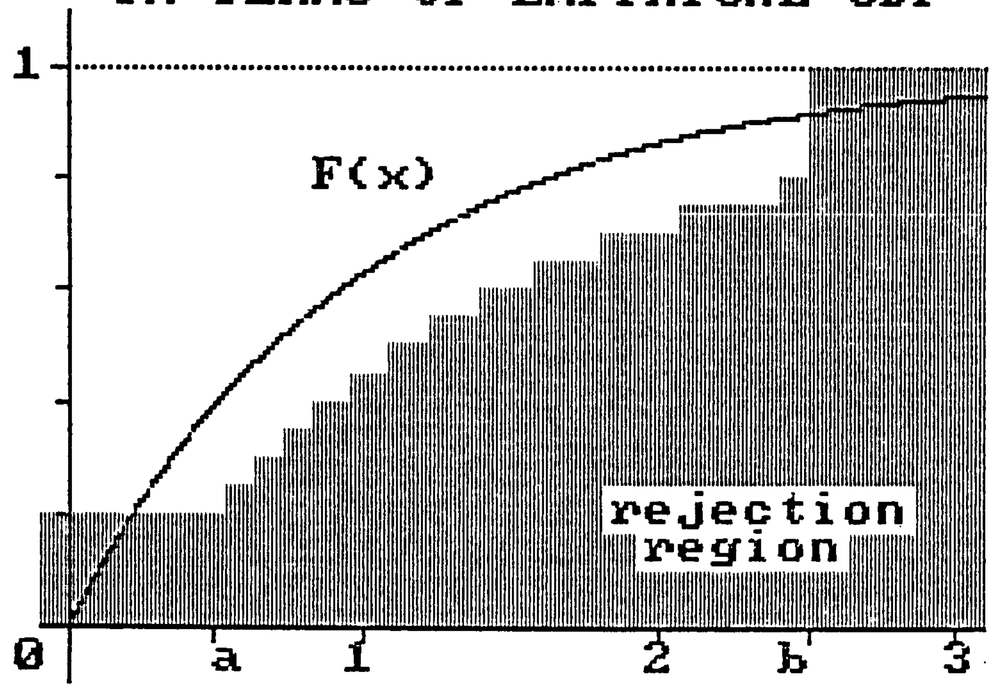

Figure 1a

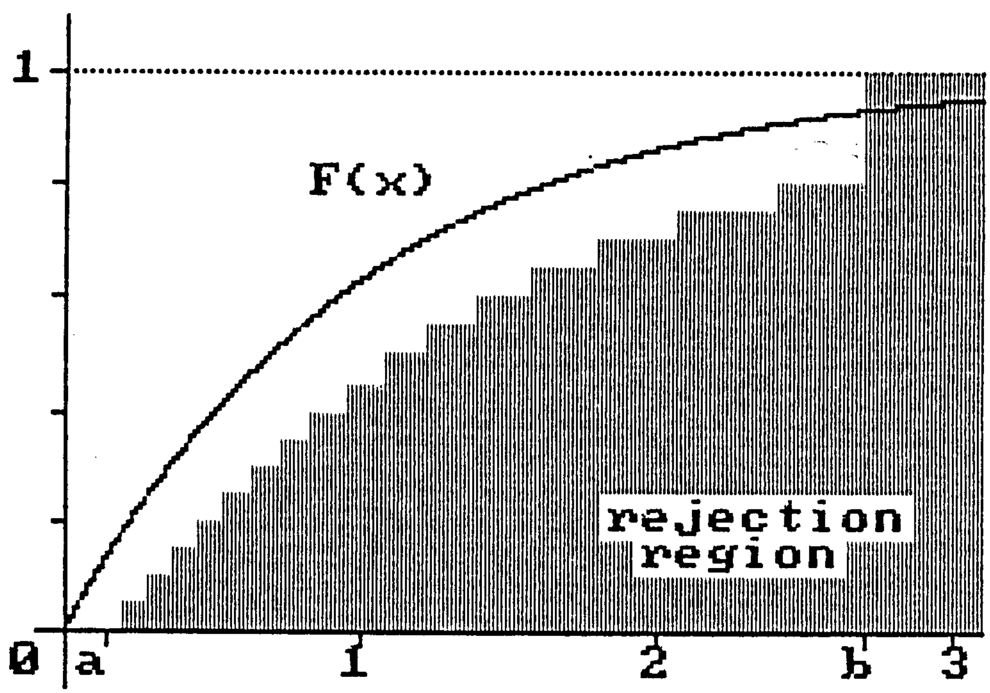

Figure 1b

F(x)=1-exp(-x) alpha=.1 n=20