The Annals of Statistics
1991, Vol. 19, No. 3, 1626–1638

INFERENCE FOR THE CROSSING POINT OF TWO CONTINUOUS CDF'S

BY D. L. HAWKINS AND SUBHASH C. KOCHAR

University of Texas, Arlington, and Indian Statistical Institute

Let $\mathscr{F}$ denote the set of cdf's on $\mathbb{R}$ with density everywhere positive. Let $C_A = \{(F, G) \in \mathscr{F} \times \mathscr{F}$: there exists a unique $x^* \in \mathbb{R}$ such that $F(x) > G(x)$ for $x < x^*$ and $F(x) < G(x)$ for $x > x^*\}$, $C_B = \{(F, G) \in \mathscr{F} \times \mathscr{F}: (G, F) \in C_A\}$. Based on independent random samples from F and G (assumed unknown), we give distribution-free tests of H_0: $F = G$ versus the alternatives that $(F, G) \in C_A$, $(F, G) \in C_B$ or $(F, G) \in C_A \cup C_B$. Next, assuming that $(F, G) \in C_A$ (or in C_B), a point estimate of the crossing point x^* is obtained and is shown to be strongly consistent and asymptotically normal. Finally, an asymptotically distribution-free confidence interval for x^* is obtained. All inferences are based on a special criterion functional of F and G, which yields x^* when maximized (minimized) if $(F, G) \in C_A$ $[(F, G) \in C_B]$.

1. Introduction. Let $\mathscr{F}$ denote the set of cdf's on $\mathbb{R}$ with density everywhere positive on $\mathscr{S} \subset \mathbb{R}$. [We consider only $\mathscr{S} = \mathbb{R}$ or $\mathscr{S} = (0, \infty)$, assumed to be known a priori.] Let $C_A = \{(F, G) \in \mathscr{F} \times \mathscr{F}$: there exists a unique $x^* \in \mathscr{S}$ such that $F(x) > G(x)$ for $x < x^*$ and $F(x) < G(x)$ for $x > x^*\}$, and let $C_B = \{(F, G) \in \mathscr{F} \times \mathscr{F}: (G, F) \in C_A\}$. Assuming that we have independent random samples $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$ from F and G (both unknown), we first obtain distribution-free (under H_0) tests for H_0: $F = G$ versus each of H_1^A: $(F, G) \in C_A$, H_1^B: $(F, G) \in C_B$ and H_1^{AB}: $(F, G) \in C_A \cup C_B$. We then obtain point and confidence interval estimates of x^*, given that $(F, G) \in C_A$ or $(F, G) \in C_B$. The point estimate is shown to be strongly consistent and asymptotically normal. The confidence interval is asymptotically distribution-free and has endpoints which are order statistics of the combined sample. All of our inferences are based on the criterion functional (2.1).

Perhaps the most immediate application of these results is to the following problem. Suppose treatments C (e.g., control) and T (e.g., a "live" treatment) are applied (respectively) to two groups of subjects whose lifetimes $X_i \sim F$ and $Y_j \sim G$ are then observed. The hypothesis H_1^A is of interest since $(F, G) \in C_A$ means that there exists some lifetime x^* such that $1 - F(x) < 1 - G(x)$ for all $x < x^*$ and $1 - F(x) > 1 - G(x)$ for all $x > x^*$. That is, control subjects have a lower chance of survival to age $x < x^*$ than treatment subjects, but a higher chance of survival to any age $x > x^*$. An example of this particular setting, given in Doksum (1974), is discussed in Section 3.1.

Received August 1989; revised August 1990.
AMS 1980 *subject classifications*. Primary 62G10; secondary 62G05, 62G15.
Key words and phrases. Criterion functional, weak convergence, survival analysis.

All of our results are for large samples. Hence we set $N = n + m$, and for some fixed $\lambda \in (0,1)$, assume that

$$n = n_N = [N\lambda], \qquad m = m_N = N - [N\lambda] \tag{1.1}$$

(where $[s]$ denotes the integer part of s). This amounts to assuming that $n_N/N \to \lambda$, $m_N/N \to 1 - \lambda$ as $N \to \infty$. Further, let $a_N = n_N m_N/N$.

Section 2 contains the main results. All results are stated and proved assuming that $\mathscr{S} = \mathbb{R}$. Only minor modifications are required for the case $\mathscr{S} = (0, \infty)$. These are omitted for brevity. Section 3 contains some applications of our methods. First, we apply our estimation technique to the above-mentioned survival dataset considered by Doksum, in which the survival functions cross. We compare our estimation results with those of Doksum. In this process, we uncover the nonrobustness of our estimator to local shifts in the data. Second, we compare the power of our test with some competitors for testing H_1^B and special cases of it. Particularly, we make comparisons with tests recently proposed by Deshpandé and Shanubhogue (1989) for testing the special case of H_1^B where x^* is known to be the common αth quantile of F and G, with α known. Section 4 contains proofs of the theorems. The proofs of some technical lemmas required in Section 4 are given in Hawkins and Kochar (1990). Software for implementing all of our methods is available from the first author.

2. Main results. For $(F, G) \in \mathscr{F}$, a fixed $\lambda \in (0,1)$ and $t \in \mathbb{R}$, let

$$\psi(t) = \int_{-\infty}^{t} [F(x) - G(x)]\, dH_\lambda(x) - \int_{t}^{\infty} [F(x) - G(x)]\, dH_\lambda(x), \tag{2.1}$$

where $H_\lambda(x) = \lambda F(x) + (1 - \lambda) G(x)$. All inferences in this paper are based on ψ. The weight function H_λ may be replaced (without destroying the essential properties of ψ) by any increasing bounded differentiable function, yielding test statistics and estimators with possibly different (better?) properties than the ones studied here. Alternatively, different functions of $(F - G)$ might be considered in the integrand, cf. Koul (1978). These variations may be investigated in a future paper.

2.1. *Hypothesis tests.* Clearly $\psi(t) = 0$ for all t under H_0. Further, it is easy to see by differentiation that under H_1^A, $\psi(t)$ is increasing in $t < x^*$ and decreasing in $t > x^*$, with $\psi(x^*) = \sup\{\psi(t)\colon t \in \mathbb{R}\} > 0$. Similarly, under H_1^B, $\psi(t)$ is decreasing in $t < x^*$, increasing in $t > x^*$ and $\psi(x^*) = \inf\{\psi(t)\colon t \in \mathbb{R}\} < 0$. These observations suggest the following test statistics. Let $\hat{F}_n(x) = n^{-1}\sum_{i=1}^{n} I(X_i \le x)$ and $\hat{G}_m(x) = m^{-1}\sum_{j=1}^{m} I(Y_j \le x)$ denote the empirical cdf's. For $t \in \mathbb{R}$, define

$$\begin{aligned} \hat{\psi}_{nm}^*(t) = \sqrt{\frac{nm}{n+m}} \Big\{ &\int_{-\infty}^{t-} \big[\hat{F}_n(x) - \hat{G}_m(x)\big]\, d\hat{H}_{nm}(x) \\ &- \int_{t}^{\infty} \big[\hat{F}_n(x) - \hat{G}_m(x)\big]\, d\hat{H}_{nm}(x) \Big\}, \end{aligned} \tag{2.2}$$

where $\hat{H}_{nm}(x) = (n+m)^{-1}\{n\hat{F}_n(x) + m\hat{G}_m(x)\}$. For testing the indicated hypotheses, we propose:

$$\text{for } H_0 \text{ versus } H_1^A\colon T_{nm}^{(A)} = \sup\{\hat{\psi}_{nm}^*(t)\colon t \in \mathbb{R}\},$$

$$\text{for } H_0 \text{ versus } H_1^B\colon T_{nm}^{(B)} = \inf\{\hat{\psi}_{nm}^*(t)\colon t \in \mathbb{R}\},$$

$$\text{for } H_0 \text{ versus } H_1^{AB}\colon T_{nm}^{(AB)} = \sup\{|\hat{\psi}_{nm}^*(t)|\colon t \in \mathbb{R}\}.$$

Since the bracketed ({ }) factor of $\hat{\psi}_{nm}^*(t)$ converges to $\psi(t)$ [see (2.3)], each of these statistics will be near 0 under H_0, but $T_{nm}^{(A)}$ will be large positive under H_1^A, $T_{nm}^{(B)}$ will be large negative under H_1^B and $T_{nm}^{(AB)}$ will be large positive under H_1^{AB}.

By a standard argument (essentially given in the proof of Theorem 1), one may show that for $F \in \mathscr{F}$ the distribution of $\hat{\psi}_{nm}^*(t)$ under H_0 is the same as if F and G were uniform cdf's. Thus all of the statistics $T_{nm}^{(A)}$, $T_{nm}^{(B)}$ and $T_{nm}^{(AB)}$ are distribution free over $\mathscr{F}$ under H_0. Of course, their distributions will be complicated functions of n and m, so asymptotic distributions are needed.

The asymptotic null distributions of these statistics are given in Theorem 1. In this direction, let $\tilde{Z} = \{\tilde{Z}(u)\colon 0 \le u \le 1\}$ denote a mean-zero Gaussian process with covariance $E\{\tilde{Z}(v)\tilde{Z}(u)\} = \frac{1}{3}(u^3 - v^3) - \frac{1}{2}(u^2 + v^2) + 2uv^2 - u^2v^2 + \frac{1}{12}$ for $0 \le v \le u \le 1$.

THEOREM 1. *Under H_0 and (1.1), as $N \to \infty$,*

(i) $$T_{n_N m_N}^{(A)} \to_{\mathscr{L}} Z_S \triangleq \sup\{\tilde{Z}(u)\colon 0 \le u \le 1\},$$

(ii) $$T_{n_N m_N}^{(B)} \to_{\mathscr{L}} Z_I \triangleq \inf\{\tilde{Z}(u)\colon 0 \le u \le 1\},$$

(iii) $$T_{n_N m_N}^{(AB)} \to_{\mathscr{L}} Z_D \triangleq \sup\{|\tilde{Z}(u)|\colon 0 \le u \le 1\}.$$

For $0 < \beta < 1$, let $Z_{S;\beta}$, $Z_{I;\beta}$ and $Z_{D;\beta}$ denote, respectively, the 100β quantiles of Z_S, Z_I and Z_D. Since $\tilde{Z} \overset{d}{=} -\tilde{Z}$ we have $Z_I \overset{d}{=} -Z_S$, so that $Z_{S;1-\beta} = -Z_{I;\beta}$. By Monte Carlo simulation of the process $\tilde{Z}$ (500 realizations using the Cholesky method on a grid of 500 points on [0, 1]), estimates of $Z_{S;\beta}$ and $Z_{D;\beta}$ were obtained. These are given in Table 1.

Some comparisons of the power of these tests with that of competing tests is given in Section 3.2.

TABLE 1
Approximate critical values for tests

β	$Z_{S;\beta}$	$Z_{D;\beta}$
0.90	0.499	0.504
0.95	0.574	0.587
0.99	0.743	0.745

2.2. *Point estimation of* x^*. We consider estimating x^* when it is known (or assumed) that $(F, G) \in C_A$. Since $(F, G) \in C_A$ if and only if $(G, F) \in C_B$, the case $(F, G) \in C_B$ may be trivially reduced to the case $(G, F) \in C_A$ by reversing the labels F and G. Condition (1.1) is assumed throughout this section, for some fixed $\lambda \in (0, 1)$.

For $t \in \mathbb{R}$, let $\hat{\psi}_{nm}(t) = \sqrt{(n+m)/nm}\,\hat{\psi}^*_{nm}(t)$, $\hat{\psi}_N(t) = \hat{\psi}_{n_N m_N}(t)$. Then it is proved in Hawkins and Kochar (1990) that regardless of $F, G \in \mathscr{F}$,

$$\sup_{t \in \mathbb{R}} \left| \hat{\psi}_N(t) - \psi(t) \right| \to 0 \quad \text{a.s as } N \to \infty. \tag{2.3}$$

If $(F, G) \in C_A$, then as noted above, $\psi(t)$ has, regardless of $\lambda \in (0, 1)$, a global maximizer at $t = x^*$. In view of (2.3), it is natural to estimate x^* by any value, $\hat{x}^*_N$ say, which maximizes $\hat{\psi}_N(t)$. However, it is easily checked, writing $Z_{(1)} \le Z_{(2)} \le \cdots \le Z_{(N)}$ as the order statistics of the samples $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$ combined, that

$$\begin{aligned}\hat{\psi}_N(t) = N^{-1}\Bigg\{ &\sum_{\{k:\, Z_{(k)} < t\}} \left[\hat{F}_{n_N}(Z_{(k)}) - \hat{G}_{m_N}(Z_{(k)})\right] \\ &\sum_{\{k:\, Z_{(k)} \ge t\}} \left[\hat{F}_{n_N}(Z_{(k)}) - \hat{G}_{m_N}(Z_{(k)})\right]\Bigg\},\end{aligned} \tag{2.4}$$

and hence that $\hat{\psi}_N(t)$ is a left-continuous step function with jumps at the $Z_{(k)}$'s. Thus the maximum of $\hat{\psi}_N(t)$ is attained in the set $\{Z_{(1)}, \dots, Z_{(N)}\} \triangleq O_N$ or at $t = Z_{(N)^+}$. Although (2.3) implies that, as $N \to \infty$, the maximizer becomes unique, there is a positive probability of multiple maxima for each finite N. So, to make our estimate $\hat{x}^*_N$ well defined for each N, we set

$$\hat{x}^*_N = \min\left\{t \colon \hat{\psi}_N(t) = \max\left[\hat{\psi}_N(s) \colon s \in \mathbb{R}\right]\right\}. \tag{2.5}$$

The following result gives the strong consistency of $\hat{x}^*_N$.

THEOREM 2. *If* $(F, G) \in C_A$, *then* $\hat{x}^*_N \to x^*$ *a.s. as* $N \to \infty$.

The next result says that $\hat{x}^*_N$ is asymptotically normally distributed with mean x^* and variance which, as might be expected, depends dramatically upon how fast F and G are changing near the crossing point x^*. Put $p^* = F(x^*) = G(x^*)$ and $f(x) = F'(x)$, $g(x) = G'(x)$.

THEOREM 3. *If* $(F, G) \in C_A$, *then*

$$a_N^{1/2}(\hat{x}^*_N - x^*) \to_{\mathscr{L}} N\left(0, p^*(1-p^*)/\{f(x^*) - g(x^*)\}^2\right) \quad \textit{as } N \to \infty.$$

We see that estimating the asymptotic variance of $\hat{x}^*_N$ involves estimating densities. This problem is essentially circumvented in Section 2.3 by the confidence interval for x^*, which requires no such estimation.

The quality of the approximation in Theorem 3 was studied via Monte Carlo simulation. Generally, the results indicate that the approximation improves as $n = m$ increases and deteriorates as $|f(x^*) - g(x^*)|$ decreases. Apparently, the rate of convergence in Theorem 3 depends on $|f(x^*) - g(x^*)|$. Although we did not investigate this issue theoretically, some illumination of it is provided by the proof of the theorem.

2.3. *Interval estimation of* x^*. In this section we give an asymptotically distribution-free confidence interval for x^*, assuming that $(F, G) \in C_A$. Such is, of course, provided by Theorem 3 if consistent estimates of $f(x^*)$ and $g(x^*)$ [or of $f(x^*) - g(x^*)$] are available. Our interval does not require such estimates, but only certain quantities computed as by-products of the computation of $\hat{x}_N^*$. Our method for obtaining this interval is an adaptation of the method outlined in Serfling (1980), page 103, for obtaining a confidence interval for a specified quantile.

To define our interval, let $\hat{p}_N^* = \hat{H}_N(\hat{x}_N^*)$ $[\hat{H}_N(x) = \hat{H}_{n_N m_N}(x)]$, and let z_α denote the $100(1-\alpha)$ quantile of $N(0,1)$. Define random sequences $\{K'_{1N}\}$ and $\{K'_{2N}\}$ for $N \geq 1$ by

$$\begin{aligned} N^{-1}K'_{1N} &= \hat{p}_N^* - a_N^{-1} z_\alpha \big[\hat{p}_N^*(1-\hat{p}_N^*)\big]^{1/2} \hat{\phi}_N, \\ N^{-1}K'_{2N} &= \hat{p}_N^* + a_N^{-1} z_\alpha \big[\hat{p}_N^*(1-\hat{p}_N^*)\big]^{1/2} \hat{\phi}_N, \end{aligned} \tag{2.6}$$

where

$$\hat{\phi}_N = 2/U_N, \quad U_N = 2 \sum_{i \in A_N} \hat{d}_{Ni} \Big\{ \hat{\psi}_N\big(Z_{([N\hat{p}_N^*])}\big) - \hat{\psi}_N\big(Z_{(i)}\big) \Big\},$$

$$A_N = \{ i : |i/N - \hat{p}_N^*| \leq \Delta_N \},$$

$$\hat{d}_{Ni} = (i/N - \hat{p}_N^*)^2 \Big/ \sum_{j \in A_N} (j/N - \hat{p}_N^*)^4$$

and $\{\Delta_N\}$ is a sequence of constants satisfying

$$\Delta_N = O(N^{-1/4+\delta}), \quad \text{some } \delta > 0. \tag{2.7}$$

Finally, for $j = 1, 2$, let $K_{jN} = K'_{jN}$ if K'_{jN} is an integer and $= [K'_{jN}] + 1$, otherwise. Then our asymptotic $100(1 - 2\alpha)\%$ confidence interval is $I_N(\alpha) = [Z_{(K_{1N})}, Z_{(K_{2N})}]$.

THEOREM 4. *If* $(F, G) \in C_A$, *then under* (1.1) *and for any* $\{\Delta_N\}$ *satisfying* (2.7), *as* $N \to \infty$,

(i) $P\{x^* \in I_N(\alpha)\} \to 1 - 2\alpha$,

(ii) $a_N^{1/2} \left\{ Z_{(K_{2N})} - Z_{(K_{1N})} - 2a_N^{-1/2} z_\alpha \dfrac{[p^*(1-p^*)]^{1/2}}{g(x^*) - f(x^*)} \right\} = o_p(1).$

TABLE 2
Performance of confidence interval $I_N(0.025)$, $F = N(0, 4)$, $G = N(0, 1)$

	Coverage probability		Length	
n	$I_N(0.025)$	$J_N(0.025)$	$I_N(0.025)$	$J_N(0.025)^*$
20	0.841	0.919	4.12	3.11
50	0.923	0.941	2.91	1.97
100	0.934	0.951	1.70	1.39

*Length $= 2\,(1.96)a_N^{-1/2}\{p^*(1-p^*)/[f(x^*)-g(x^*)]^2\}^{1/2}$.

Part (i) says that $I_N(\alpha)$ is asymptotically distribution free. Part (ii) says that the length of $I_N(\alpha)$ is asymptotically that of the $100(1-2\alpha)\%$ confidence interval, say $J_N(\alpha)$, $\hat{x}_N^* \pm a_N^{-1/2} z_\alpha [p^*(1-p^*)]^{1/2}/|g(x^*)-f(x^*)|$, which derives from Theorem 3 if p^*, $g(x^*)$ and $f(x^*)$ are known.

Table 2 gives Monte Carlo estimates (based on 1000 trials) of the coverage probability and average length of $I_N(0.025)$ (with $\Delta_N = N^{-1/4}$) for a typical $(F, G) \in C_A$ and $m = n$. These are compared with the length and Monte Carlo–estimated coverage probability of the interval $J_N(0.025)$. (Note that both intervals have nominal 95% coverage probabilities.) Generally, we see that $I_N(0.025)$ compares well with $J_N(0.025)$ if $n \geq 50$. Further Monte Carlo results, given in Hawkins and Kochar (1990), indicate that the coverage probability of $I_N(\alpha)$ deteriorates as $|f(x^*) - g(x^*)|$ decreases.

3. Some applications and evaluation

3.1. *Life distributions.* We first apply our procedures to the survival problem of Doksum (1974) noted in Section 1. The data [due to Bjerkdal (1960)] on $n = 65$ control and $m = 60$ treated (with tubercle bacilli) guinea pigs are displayed in Doksum's paper. From the graph of the empirical cdf's, the cdf's F and G of the control and treatment group lifetimes apparently cross once at about 160; the value of $T_{nm}^A = 2.07$ is significant at the 0.01 level. In this situation, where crossing is apparent, primary interest would be in estimating the crossing point x^*. We obtained $\hat{x}_N^* = 114$ and (using $\Delta_N = N^{-1/4}$) the approximate 95% confidence interval $I_N(0.025) = [52, 160]$. This value of $\hat{x}_N^*$ is surprising in light of the empirical cdf plot. By comparison, using Doksum's method (which indirectly estimates x^*), we obtain from his Figure 2 the estimate 130 (approximately) for x^* and 90% confidence interval $(-\infty, 250]$. (His confidence interval has no lower limit.)

In trying to uncover the source of our apparently "bad" estimate $\hat{x}_N^*$, we noticed that the guinea pig data contain several ties. Since our assumptions about F and G imply that ties occur with probability 0, we decided to break the ties at random by replacing X_i by $X_i^* = X_i + U_i/1000$, Y_j by $Y_j^* = Y_j + W_j/1000$, where $U_1, \ldots, U_n, W_1, \ldots, W_m$ are iid $U(0, 1)$ variables. The resulting estimates are much more appealing: $\hat{x}_N^* = 181.0$, $I_N(0.025) = [114, 291]$. This dramatic change surprised us until we noticed from (2.4) and (2.5) that $\hat{x}_N^*$ will clearly be nonrobust to small shifts in the data. One might possibly derive the influence function of $\hat{x}_N^*$; we expect the local shift sensitivity to be $+\infty$.

3.2. *Power comparisons for H_1^B and related hypotheses.* To the authors' knowledge, there are no competing tests aimed specifically at testing H_0 versus any of H_1^A, H_1^B or H_1^{AB}. Of course, certain omnibus tests, such as the Kolmogorov–Smirnov test, may be viewed as competitors in a general sense. Further, Deshpandé and Shanubhogue (1989) have devised two tests for H_0 versus H_1^{DS} (say), which is a special case of H_1^B in which it is assumed that $F(x^*) = G(x^*) = \alpha$ for some known α (i.e., F and G have common α-quantile, say ξ_α). Hence it is of interest to compare the power of our test for H_0 versus H_1^B to that of the Kolmogorov–Smirnov test and the Deshpandé–Shanubhogue tests.

The first Deshpandé–Shanubhogue test is based on

$$T_\alpha = \int_{-\infty}^{\infty} \hat{G}_m(x) J_\alpha\left(\frac{n}{n+1}\hat{F}_n(x)\right) d\hat{F}_n(x) + \int_{-\infty}^{\infty} \hat{F}_n(x)\left[1 - J_\alpha\left(\frac{m}{m+1}\hat{G}_m(x)\right)\right] d\hat{G}_m(x),$$

where $J_\alpha(u)$ is the indicator function of $(0, \alpha]$. T_α is an estimate of

$$\theta^* = \int_{-\infty}^{\xi_\alpha} G(x)\,dF(x) + \int_{\xi_\alpha}^{\infty} F(x)\,dG(x),$$

which equals $\frac{1}{2}$ under H_0 and strictly exceeds $\frac{1}{2}$ under H_1^{DS}. Thus H_0 is rejected in favor of H_1^{DS} if T_α is "large." The second Deshpandé–Shanubhogue statistic is the following modification of the Mood (1954) scale statistic:

$$M_\alpha = \sum_{i=1}^{N} \{i - (N+1)\alpha\}^2 W_{Ni},$$

where W_{Ni} equals 1 if $Z_{(i)}$ is an X observation and equals 0, otherwise.

Another special case of H_1^B arises in the context of the classical scale problem. Consider H_1^S: $F(x) = F^0(x - \xi_\alpha)$, $G(x) = F^0(\theta(x - \xi_\alpha))$, $\theta < 1$, where F^0 is an unknown increasing cdf satisfying $F^0(0) = \alpha$. Then under H_1^S, $(F, G) \in C_B$ with $x^* = \xi_\alpha$, making H_1^S a special case of H_1^{DS} if α is known and a special case of H_1^B regardless of α. Our test of H_0 versus H_1^B may thus be viewed as a competitor of the scale tests due to Mood (1954), Ansari and Bradley (1960) and others. Of course, since these latter tests use the information about the known common α-quantile and are aimed directly at scale alternatives, they may be expected to have higher power against H_1^S than our test.

Table 3 gives the results of a small Monte Carlo study comparing the power of our test (HK) based on T_{nm}^B with that of the tests of Kolmogorov and Smirnov (KS), DS (DST denotes the test based on T_{nm}^α, DSM the test based on M_α), Mood (M) and Ansari and Bradley (AB). All powers are estimates based on 1000 Monte Carlo trials at the nominal 0.05 significance level, with $n = m$.

TABLE 3
Power comparisons for H_0 versus H_1^B special cases
(nominal significance level is 0.05)

			Estimated power					
F	**G**	**n**	**HK**	**KS**	**DST**	**DSM**	**M**	**AB**
$N(0,1)$	$N(0,4)$	20	0.152	0.122	0.468	0.638	0.798	0.728
		50	0.546	0.398	0.941	0.981	0.990	0.969
		100	0.945	0.782	0.999	1.000	1.000	1.000
Exp*	Weib	20	0.237	0.188	0.301	0.484	0.645	0.580
		50	0.624	0.537	0.792	0.864	0.944	0.900
		100	0.956	0.883	0.987	0.995	0.996	0.992

*F = Exponential(mean $= 2/\pi^{1/2}$), G = Weibull($\alpha = 2$, $\lambda = 1$); $x^* = 2/\pi^{1/2} \doteq 1.13$; mean($F$) = mean($G$).

Large-sample critical values are used in all cases. Setting 1 is a scale alternative with $\alpha = 0.50$, for which M and AB are specifically designed. The power results reflect this fact. Setting 2 satisfies $(F, G) \in C_B$, but is not a scale alternative (although F and G have the same mean). Predictably, DST and DSM have higher power than HK, since they use knowledge of α (here $\alpha \doteq 0.72$) which HK does not require. Surprisingly, M and AB have higher powers than the other tests, even for this nonscale alternative. Comparing HK with KS, we see that HK has slightly higher power than KS, although for $n = 20$, KS is slightly conservative, making comparison difficult.

4. Proofs of the theorems.

PROOF OF THEOREM 1. For $0 < p < 1$ put

$$\hat{\psi}_{nm}^{00}(p) = \sqrt{\frac{nm}{n+m}} \left\{ \int_{-\infty}^{F^{-1}(p)} \left[\hat{F}_n(x) - \hat{G}_m(x) \right] dF(x) + \int_{F^{-1}(p)}^{\infty} \left[\hat{G}_m(x) - \hat{F}_n(x) \right] dF(x) \right\},$$

$$\hat{\psi}_{nm}^{0}(p) = \sqrt{\frac{nm}{n+m}} \left\{ \int_{-\infty}^{F^{-1}(p)-} \left[\hat{F}_n(x) - \hat{G}_m(x) \right] d\hat{H}_{nm}(x) + \int_{F^{-1}(p)}^{\infty} \left[\hat{G}_m(x) - \hat{F}_n(x) \right] d\hat{H}_{nm}(x) \right\}. \tag{4.1}$$

Then $T_{nm}^{(A)} = \sup\{\hat{\psi}_{nm}^0(p)\colon 0 < p < 1\}$, $T_{nm}^{(B)} = \inf\{\hat{\psi}_{nm}^0(p)\colon 0 < p < 1\}$ and $T_{nm}^{(AB)} = \sup\{|\hat{\psi}_{nm}^0(p)|\colon 0 < p < 1\}$.

We first deal with $\hat{\psi}^{00}_{nm}(p)$. Via the transformations $u = F(x)$, $U_i = F(X_i)$, $V_i = F(Y_i)$, and defining

$$\tilde{W}_N(u) = (n_N m_N/N)^{1/2}\left\{n_N^{-1}\sum_{i=1}^{n_N}[I(U_i \le u) - u] - m_N^{-1}\sum_{j=1}^{m_N}[I(V_i \le u) - u]\right\}, \qquad 0 \le u \le 1,$$

we have

$$\hat{\psi}^{00}_{n_N m_N}(p) = \int_0^p \tilde{W}_N(u)\,du - \int_p^1 \tilde{W}_N(u)\,du, \qquad 0 \le p \le 1. \tag{4.2}$$

Now it is classical that if $\tilde{W}_N = \{\tilde{W}_N(u): 0 \le u \le 1\}$, then

$$\tilde{W}_N \to_w \tilde{W}^0 \quad \text{as } N \to \infty, \tag{4.3}$$

where $\tilde{W}^0$ denotes the Brownian bridge process and $\to_w$ denotes weak convergence of measures on $D\,([0,1])$; see Billingsley (1968). Further, by (4.2),

$$\hat{\psi}^{00}_{n_N m_N}(p) = \big(T\tilde{W}_N\big)(p), \qquad 0 \le p \le 1, \tag{4.4}$$

where $T\colon D([0,1]) \to D([0,1])$ is defined by

$$(Th)(p) = \int_0^p h(u)\,du - \int_p^1 h(u)\,du.$$

One may verify that T is Skorohod continuous (i.e., $h_N \to_S h$ implies $Th_N \to_S Th$). Put $\hat{\Phi}^{00}_N = \{\hat{\psi}^{00}_{n_N m_N}(p): 0 \le p \le 1\}$. Then, by (4.3) and (4.4),

$$\hat{\Phi}^{00}_N = T\tilde{W}_N \to_w T\tilde{W}^0 \quad \text{as } N \to \infty. \tag{4.5}$$

Now put $\hat{\Phi}^0_N = \{\hat{\psi}^0_{n_N m_N}(p): 0 \le p \le 1\}$. Then [writing $\hat{H}_N(x) = \hat{H}_{n_N m_N}(x)$]

$$\begin{aligned}\sup_{0\le p\le 1}\left|\hat{\psi}^0_{n_N m_N}(p) - \hat{\psi}^{00}_{n_N m_N}(p)\right| &\le \sup_{0\le p\le 1}\left|\int_0^{p-}\tilde{W}_N(u)\,d\big(\hat{H}_N(F^{-1}(u)) - u\big)\right| \\ &\quad + \sup_{0\le p\le 1}\left|\int_p^1\tilde{W}_N(u)\,d\big(\hat{H}_N(F^{-1}(u)) - u\big)\right|.\end{aligned} \tag{4.6}$$

Further, we may [upon assuming with no loss of generality under H_0 that F is the $U(0,1)$ distribution] repeat the steps of Csörgő and Révész (1981), page 187 [beginning at line 3, with their $B_n^2(y)$ replaced by our $\tilde{W}_N(u)$], to obtain that the quantity in (4.6) is $o_p(1)$. [This will require using (4.3) and its implication that $\{\tilde{W}_N\}$ is tight.] Expressions (4.5) and (4.6) thus imply that

$$\hat{\Phi}^0_N \to_w T\tilde{W}^0 \quad \text{as } N \to \infty. \tag{4.7}$$

Now (i), (ii) and (iii) all follow from (4.7) by the Skorohod continuity of the "sup", "inf" and "sup| · |" functionals, if we verify that $T\tilde{W}^0 =_d \tilde{Z}$. That $T\tilde{W}^0$ is Gaussian is immediate, as is the fact that $E[(T\tilde{W}^0)(p)] = 0$, $0 \le p \le 1$. The desired covariance structure follows by writing

$$T\tilde{W}^0(p) =_d \int_0^p \left[\tilde{W}(u) - u\tilde{W}(1)\right] du - \int_p^1 \left[\tilde{W}(u) - u\tilde{W}(1)\right] du,$$

where $\tilde{W} = \{\tilde{W}(u)\colon 0 \le u \le 1\}$ is the Wiener process, and using formulas such as (10), page 133 in Hoel, Port and Stone (1972) to evaluate $\operatorname{cov}(T\tilde{W}^0(p), T\tilde{W}^0(p'))$. □

PROOF OF THEOREM 2. We claim that

$$\psi(\hat{x}_N^*) \to \psi(x^*) \quad \text{a.s. implies that } \hat{x}_N^* \to x^* \text{ a.s.} \tag{4.8}$$

Given this claim, the result follows by noting that

$$\left|\psi(\hat{x}_N^*) - \psi(x^*)\right| \le \left|\psi(\hat{x}_N^*) - \hat{\psi}_N(\hat{x}_N^*)\right| + \left|\hat{\psi}_N(\hat{x}_N^*) - \psi(x^*)\right|$$

and observing that the first term on the right side is $o(1)$ a.s. by (2.3), while the second term equals

$$\left|\sup_t \hat{\psi}_N(t) - \sup_t \psi(t)\right| \le \sup_t \left|\hat{\psi}_N(t) - \psi(t)\right| = o(1) \quad \text{a.s.},$$

again by (2.3).

To prove (4.8), we claim that ψ has the following property:

$$\forall\, \varepsilon > 0, \exists\, \eta > 0 \text{ such that, for every } x, \tag{4.9}$$

$$|x - x^*| \ge \varepsilon \text{ implies } |\psi(x) - \psi(x^*)| \ge \eta.$$

Clearly property (4.9) gives (4.8), since it implies that $\psi(x_N) \to \psi(x^*)$ ensures that $x_N \to x^*$ for any sequence $\{x_N\}$.

To prove that ψ has property (4.9), we follow the argument of Parzen (1962), Theorem 3.A. If (4.9) did not hold then there would exist an $\varepsilon > 0$ and a sequence $\{x_k\}$ such that

$$|\psi(x^*) - \psi(x_k)| < k^{-1} \tag{4.10}$$

and

$$|x^* - x_k| \ge \varepsilon. \tag{4.11}$$

Now since $\psi(t)$ is decreasing in $t > x^*$ and increasing in $t < x^*$, $|x_k| \to \infty$ clearly makes (4.10) impossible, so we can assume that $|x_k| \le M$, say, for all k. Thus $\{x_k\}$ contains a convergent subsequence $x_{k_j} \to \tilde{x}$, with $|x^* - \tilde{x}| \ge \varepsilon$. But then (4.10) implies that $\psi(\tilde{x}) = \psi(x^*)$, contradicting the fact that x^* is the unique maximizer of ψ. □

PROOF OF THEOREM 3. First, write by Taylor's theorem for some $\tilde{x}_N$ between $\hat{x}_N^*$ and x^* [noting that $F(x^*) - G(x^*) = 0$],

$$(4.12)\quad a_N^{1/2}\{F(\hat{x}_N^*) - G(\hat{x}_N^*)\} = [f(\tilde{x}_N) - g(\tilde{x}_N)]a_N^{1/2}(\hat{x}_N^* - x^*).$$

We then claim that

$$(4.13)\quad \begin{aligned} a_N^{1/2}\{F(\hat{x}_N^*) - G(\hat{x}_N^*)\} &= a_N^{1/2}\left\{\hat{G}_{m_N}(x^*) - G(x^*)\right\} \\ &\quad - a_N^{1/2}\left\{\hat{F}_{n_N}(x^*) - F(x^*)\right\} + o_p(1). \end{aligned}$$

Further, since $(F, G) \in C_A$, for $x < x^*$ we have

$$0 < F(x) - G(x) = F(x^*) - G(x^*) + [f(x^*) - g(x^*)](x - x^*) + o(x - x^*)^2,$$

which implies that $f(x^*) < g(x^*)$. Thus, since the two terms on the right side of (4.13) are independent and trivially asymptotically normal by the iid central limit theorem, the result follows from (4.12), (4.13) and Slutsky's theorem (using Theorem 2).

The tough part is establishing (4.13), for which we need the following technical results, proved in Hawkins and Kochar (1990).

LEMMA 1. *If* $(F, G) \in C_A$ *and* (1.1) *holds, then as* $N \to \infty$,

$$\text{(i)}\quad \begin{aligned} &N^{1/2}\left[\hat{F}_{n_N}(\hat{x}_N^*) - F(\hat{x}_N^*)\right] \\ &\qquad - N^{1/2}\left[\hat{F}_{n_N}(x^*) - F(x^*)\right] = o_p(1), \end{aligned}$$

$$\text{(ii)}\quad \begin{aligned} &N^{1/2}\left[\hat{G}_{m_N}(\hat{x}_N^*) - G(\hat{x}_N^*)\right] \\ &\qquad - N^{1/2}\left[\hat{G}_{m_N}(x^*) - G(x^*)\right] = o_p(1), \end{aligned}$$

$$\text{(iii)}\quad N^{1/2}\left|\hat{F}_{n_N}(\hat{x}_N^*) - \hat{G}_{m_N}(\hat{x}_N^*)\right| = O_p(N^{-1/2}).$$

Given these results (4.13) part is immediate, since upon multiplying the expression in part (i) by (-1) and adding to that in part (ii), we get

$$\begin{aligned} &N^{1/2}[F(\hat{x}_N^*) - G(\hat{x}_N^*)] + N^{1/2}\left[\hat{G}_{m_N}(\hat{x}_N^*) - \hat{F}_{n_N}(\hat{x}_N^*)\right] \\ &\qquad - N^{1/2}\left[\hat{G}_{m_N}(x^*) - G(x^*)\right] + N^{1/2}|\hat{F}_{n_N}(x^*) - F(x^*)\big] = o_p(1), \end{aligned}$$

which, in view of part (iii) and the fact that $a_N = O(N^{1/2})$, is the same as (4.13). □

PROOF OF THEOREM 4. The whole thing rests on the following technical lemma [proved in Hawkins and Kochar (1990)], which is an adaptation of Serfling (1980), page 104, expression (1).

LEMMA 2. *Under the conditions of Theorem* 4,

$$\text{(a)} \qquad Z_{(K_{1N})} - \left\{ \hat{x}_N^* - z_\alpha a_N^{-1/2} \frac{[\hat{p}_N^*(1-\hat{p}_N^*)]^{1/2}}{g(x^*) - f(x^*)} \right\} = o_p(a_N^{-1/2}),$$

$$\text{(b)} \qquad Z_{(K_{2N})} - \left\{ \hat{x}_N^* + z_\alpha a_N^{-1/2} \frac{[\hat{p}_N^*(1-\hat{p}_N^*)]^{1/2}}{g(x^*) - f(x^*)} \right\} = o_p(a_N^{-1/2}).$$

Given this result, part (ii) follows easily. Regarding part (i), we have

$$P\{x^* \notin I_N(\alpha)\} = P\{Z_{(K_{1N})} > x^*\} + P\{Z_{(K_{2N})} < x^*\}.$$

But by Lemma 2(a),

$$P\{Z_{(K_{1N})} > x^*\} = P\left\{ a_N^{1/2}(\hat{x}_N^* - x^*) + o_p(1) > z_\alpha \frac{[\hat{p}_N^*(1-\hat{p}_N^*)]^{1/2}}{g(x^*) - f(x^*)} \right\} \to \alpha$$

as $N \to \infty$, the convergence holding by Theorem 3. It similarly follows that $P\{Z_{(K_{2N})} < x^*) \to \alpha$, giving part (i). □

REMARKS. It has been pointed out to us by an Associate Editor that Theorems 1 and 3 may also be proved by the statistical differential method, following Gill (1989). We do not attempt this here or in Hawkins and Kochar (1990), where only standard results are utilized.

Acknowledgment. The authors thank the referee who pointed out two technical errors in an earlier version of the paper. They also thank the University of Iowa Department of Statistics and Actuarial Science for providing them visiting appointments, during which this work was completed.

REFERENCES

ANSARI, A. R. and BRADLEY, R. A. (1960). Rank-sum tests for dispersions. *Ann. Math. Statist.* **31** 1174–1189.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BJERKDAL, T. (1960). Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli. *American Journal of Hygiene* **72** 130–148.

CSÖRGŐ, M. and RÉVÉSZ, P. (1981). *Strong Approximations in Probability and Statistics*. Academic, New York.

DESHPANDÉ, J. V. and SHANUBHOGUE, A. (1989). Tests for the two-sample nonparametric dispersion alternative. Technical Report, Univ. California, Santa Barbara.

DOKSUM, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Ann. Statist.* **2** 267–277.

GILL, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method. I. *Scand. J. Statist.* **16** 97–128.

HAWKINS, D. L. and KOCHAR, S. C. (1990). Inference for the crossing point of two continuous cdf's. Technical Report 179, Dept. Statistics and Actuarial Science, Univ. Iowa.

HOEL, P. G., PORT, S. C. and STONE, C. J. (1972). *Introduction to Stochastic Processes*. Houghton Mifflin, New York.

KOUL, H. L. (1978). A class of tests for testing new is better than used. *Canad. J. Statist.* **6** 249–271.

MOOD, A. M. (1954). On the asymptotic efficiency of certain nonparametric two-sample tests. *Ann. Math. Statist.* **25** 514–522.

PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TEXAS, ARLINGTON
ARLINGTON, TEXAS 76015

MATHEMATICAL STATISTICS DIVISION
INDIAN STATISTICAL INSTITUTE
NEW DELHI 110016
INDIA