The Annals of Statistics
1981, Vol. 9, No. 5, 1087–1095

RANK TESTS FOR BIVARIATE SYMMETRY

BY TOM SNIJDERS

University of Groningen

The problem is considered of testing symmetry of a bivariate distribution $\mathscr{L}(X, Y)$ against "asymmetry towards high X-values," subject to the restriction of invariance under the transformations $(x_i, y_i) \mapsto (g(x_i), g(y_i))$ $(1 \le i \le n)$ for increasing bijections g. This invariance restriction prohibits the common reduction to the differences $x_i - y_i$. The intuitive concept of "asymmetry towards high X-values" is approached in several ways, and a mathematical formulation for this concept is proposed. Most powerful and locally most powerful invariant similar tests against certain subalternatives are characterized by means of a Hoeffding formula. Asymptotic normality and consistency results are obtained for appropriate linear rank tests.

1. Introduction. Consider the testing problem where, on the basis of a random sample from a bivariate distribution $\mathscr{L}(X, Y)$, the null hypothesis

$$H: \mathscr{L}(X, Y) = \mathscr{L}(Y, X)$$

of bivariate symmetry has to be tested against the composite one-sided alternative that asymmetry exists, X tending to be larger than Y. A standard treatment reduces this problem to a problem of univariate symmetry by restricting attention to the difference $X - Y$. The alternative hypothesis is then formulated as "$X - Y$ is stochastically larger than $Y - X$", and a commonly used test is the Wilcoxon signed-rank test; see, e.g., Lehmann (1959) Section 6.7. The original testing problem (with the informally expressed alternative hypothesis) can, however, be considered to be invariant under the group of transformations $(x, y) \mapsto (g(x), g(y))$, where g is an increasing bijection $g: \mathbb{R} \to \mathbb{R}$. The formulation of the alternative hypothesis mentioned above and the Wilcoxon signed-rank test (for most significance levels) are not invariant under this group. In Section 2 a concept of bivariate asymmetry is developed which leads to an invariant alternative hypothesis; several equivalent formulations of this concept are given. Invariant similar size α tests are studied in the following sections. In Section 4 a Hoeffding formula is derived, and used to characterize most powerful and locally most powerful tests in this class against certain subalternatives. Section 5 gives some large sample properties of linear bivariate rank tests. An important test encountered in Sections 4 and 5 is a test using Wilcoxon scores, which can be viewed as a competitor to the Wilcoxon signed rank test.

Other papers about testing bivariate symmetry include Bell and Haller (1969), discussing the formulation of the null hypothesis and similar size α tests; Hollander (1971) and Koziol (1979), treating a test against the unrestricted alternative; and Sen (1967) and Yanagimoto and Sibuya (1976), discussing rank tests against restricted alternatives. Section 2 of the present paper is based mainly on Schaafsma (1976).

2. Bivariate asymmetry. Let $\mathscr{G}$ be the group of all increasing bijections $g: \mathbb{R} \to \mathbb{R}$, and $\mathscr{H}_0$ the group of all transformations $(x, y) \mapsto (g(x), g(y))$ with $g \in \mathscr{G}$. We wish to

Received March, 1980; revised November, 1980.

AMS 1970 *subject classifications.* Primary 62G10; secondary 62A05, 62C99, 62E20.

Key words and phrases. Nonparametric tests, bivariate symmetry and asymmetry, locally most powerful tests, asymptotic normality.

define asymmetry concepts for bivariate probability distributions, which are expressions of the intuitive concept that for the pair of random variables (X, Y), X tends to be larger than Y; and which are invariant under $\mathscr{H}_0$. Three approaches will be considered.

(i) Yanagimoto and Sibuya (1972) study a family of asymmetry concepts. For a class $\mathscr{R}$ of measurable subsets of $\mathbb{R}^2$, these authors define X to be stochastically larger than Y with respect to $\mathscr{R}$ if

$$P\{(X, Y) \in A\} \geq P\{(Y, X) \in A\} \qquad \text{for all} \quad A \in \mathscr{R}.$$

Yanagimoto and Sibuya (1972) consider many classes $\mathscr{R}$, and the relationships between the corresponding asymmetry concepts. We only consider

$$\mathscr{R}_1 = \{\{(x, y) \mid g(x) - g(y) \geq c\}\ g \in \mathscr{G}, c \in \mathbb{R}\}$$

$$\mathscr{R}_2 = \{\{(x, y) \mid f(x) \geq y\} \mid f : \mathbb{R} \to \mathbb{R} \text{ increasing continuous}\}$$

$$\mathscr{R}_3 = \{\{(x, y) \mid x \geq a, y \leq a\} \mid a \in \mathbb{R}\}$$

$$\mathscr{R}_4 = \{A \subset \mathbb{R}^2 \mid \text{if } (x_1, y_1) \in A,\ x_2 \geq x_1,\ y_2 \leq y_1, \quad \text{then} \quad (x_2, y_2) \in A\}.$$

$\mathscr{R}_1$ yields the $\mathscr{H}_0$-invariant analogue of the asymmetry concept mentioned in the introduction, that $X - Y$ is stochastically larger than $Y - X$: for every $g \in \mathscr{G}$, $g(X) - g(Y)$ is stochastically larger than $g(Y) - g(X)$. The asymmetry concept generated by $\mathscr{R}_2$ was proposed by Schaafsma (1966), who also derived some properties of this concept. $\mathscr{R}_3$ leads to comparison of the marginal distributions: the marginal distribution of X is stochastically larger than that of Y. $\mathscr{R}_4$ is the class of increasing subsets (subsets with increasing indicator function), when $\mathbb{R}^2$ is partially ordered by

$$(x_1, y_1) \overset{a}{\leq} (x_2, y_2) \qquad \text{iff} \quad x_1 \leq x_2 \quad \text{and} \quad y_1 \geq y_2. \tag{2.1}$$

It can be shown by means of a limiting argument, that $\mathscr{R}_2$ and $\mathscr{R}_4$ generate the same asymmetry concept. (Yanagimoto and Sibuya (1972) mention the classes $\mathscr{R}_3$ and $\mathscr{R}_4$, but not $\mathscr{R}_1$ and $\mathscr{R}_2$).

(ii) Schaafsma (1976) gives the following two "probability 1-concepts."

(2.2) There exists a probability distribution $\mathscr{L}(Z_1, Z_2, Z_3)$ on $\mathbb{R}^3$ with $\mathscr{L}(X, Y) = \mathscr{L}(Z_1, Z_2)$, $\mathscr{L}(Z_2, Z_3) = \mathscr{L}(Z_3, Z_2)$, $P\{Z_1 \geq Z_3\} = 1$.

(2.3) There exists a probability distribution $\mathscr{L}(Z_1, Z_2, Z_3, Z_4)$ on $\mathbb{R}^4$ with $\mathscr{L}(X, Y) = \mathscr{L}(Z_1, Z_2) = \mathscr{L}(Z_4, Z_3)$, $\mathscr{L}(Z_1, Z_4) = \mathscr{L}(Z_4, Z_1)$, $\mathscr{L}(Z_2, Z_3) = \mathscr{L}(Z_3, Z_2)$, $P\{Z_1 \geq Z_3, Z_2 \leq Z_4\} = 1$.

These definitions can be interpreted in the following way. Let Z_1 and Z_2 be scores obtained from a pair of subjects, of which one is assigned to treatment (yielding Z_1) and the other to control (yielding Z_2). Then Z_3 gives the hypothetical unobservable score of the first subject, had it been assigned to control; Z_4 gives the hypothetical unobservable score of the second subject, had it been assigned to treatment. It is immediately seen that (2.3) implies (2.2). The reverse implication is also true: let $\mathscr{L}(Z_1, Z_2, Z_3)$ be as in (2.2), and let W be a random variable such that

(a) conditionally on (Z_2, Z_3), $Z_1 - Z_3$ and W are independent;

(b) $\mathscr{L}(W \mid (Z_2, Z_3) = (z_2, z_3)) = \mathscr{L}(Z_1 - Z_3 \mid (Z_2, Z_3) = (z_3, z_2))$;

then $\mathscr{L}(Z_1, Z_2, Z_3, Z_2 + W)$ satisfies (2.3). So (2.2) and (2.3) define the same asymmetry concept for $\mathscr{L}(X, Y)$.

(iii) A quasi-axiomatic approach can be taken by formulating some properties which the asymmetry concept should have, and deriving the smallest/largest/unique set of bivariate distributions with these properties, if such a set exists. The following properties seem to be fundamental. Denote by $\mathscr{P}_{as}$ the class of all bivariate probability distributions which satisfy the asymmetry concept or are symmetric.

(2.4) If X and Y are independent and X is stochastically larger than Y, then $\mathscr{L}(X, Y) \in \mathscr{P}_{as}$.

(2.5) If $\mathscr{L}(X, Y) \in \mathscr{P}_{as}$, then the marginal distribution of X is stochastically larger than the marginal distribution of Y.

(2.6) $\mathscr{P}_{as}$ is convex (closed under mixtures).

(2.7) $\mathscr{P}_{as}$ is closed under convolutions.

Let $\mathscr{P}_a$ be the (weakly) closed convex hull of all $\mathscr{L}(X, Y)$ mentioned in (2.4) and $\mathscr{P}_m$ the class of all $\mathscr{L}(X, Y)$ for which X is stochastically larger than Y. Both $\mathscr{P}_a$ and $\mathscr{P}_m$ satisfy these four properties, and for every weakly closed $\mathscr{P}_{as}$ which satisfies them one has $\mathscr{P}_a \subset \mathscr{P}_{as} \subset \mathscr{P}_m$.

The asymmetry concepts generated by $\mathscr{R}_1$, $\mathscr{R}_2$ and $\mathscr{R}_4$ in (i), the asymmetry concept of (ii) and that defined by $\mathscr{P}_a$ in (iii) are identical. A "natural" setting for this result is the partial ordering of probability distributions on a partially ordered outcome space studied by Lehmann (1955) and Kamae, Krengel and O'Brien (1977). For a measurable space $\mathscr{X}$ with partial ordering $\prec$, the class of probability distributions on $\mathscr{X}$ can be partially ordered by defining $P_1 \prec P_2$ if

(2.8) $E_1 f(X) \leq E_2 f(X)$ for all bounded measurable $\prec$-increasing $f: \mathscr{X} \to \mathbb{R}$.

Kamae, Krengel and O'Brien (1977) demonstrate that if $\mathscr{X}$ is a Polish space and $\prec$ a closed partial ordering, then (2.8) is equivalent with both

(2.9) $$P_1(A) \leq P_2(A) \quad \text{for all closed } \prec\text{-increasing } A \subset \mathscr{X}$$

and

(2.10) There exists a probability distribution $\mathscr{L}(X_1, X_2)$ on $\mathscr{X}^2$ with
$$\mathscr{L}(X_1) = P_1, \quad \mathscr{L}(X_2) = P_2 \quad \text{and} \quad P\{X_1 \prec X_2\} = 1.$$

(This shows that Lehmann (1955) errs when he states that his conditions A and B are not equivalent.) This leads to the following definition and theorem.

DEFINITION 1. The probability distribution $\mathscr{L}(X, Y)$ on $\mathbb{R}^2$ is asymmetric towards high X-values if

$$Ef(X, Y) \geq Ef(Y, X)$$

for every bounded measurable $f: \mathbb{R}^2 \to \mathbb{R}$ which is increasing in the first and decreasing in the second coordinate, with strict inequality for some f.

THEOREM 1. *For probability distributions $\mathscr{L}(X, Y)$ on $\mathbb{R}^2$ with $\mathscr{L}(X, Y) \neq \mathscr{L}(Y, X)$, the following conditions are equivalent.*

(a) *$\mathscr{L}(X, Y)$ is asymmetric towards high X-values*
(b) *X is stochastically larger than Y with respect to $\mathscr{R}_1$*
(c) *X is stochastically larger than Y with respect to $\mathscr{R}_2$*
(d) *X is stochastically larger than Y with respect to $\mathscr{R}_4$*
(e) *$\mathscr{L}(X, Y)$ satisfies* (2.2) *or* (2.3)
(f) *$\mathscr{L}(X, Y) \in \mathscr{P}_a$*
(g) $\mathscr{L}(Y, X) \overset{a}{\gtrsim} \mathscr{L}(X, Y)$ (*see* (2.1) *and* (2.8)).

PROOF. It is trivial that (g) $\Leftrightarrow$ (a) $\Rightarrow$ (b). The equivalence of (a), (d) and (e) follows from the equivalence of (2.8), (2.9) and (2.10). The equivalence of (c) and (d) was mentioned in (i) above. It will now be demonstrated that (b) $\Rightarrow$ (c) and (e) $\Leftrightarrow$ (f).

(b) $\Rightarrow$ (c). A sufficient condition for (c) is that

$$P\{f(X) \geq Y\} \geq P\{f(Y) \geq X\} \tag{2.11}$$

for all $f \in \mathscr{G}$. For such f, (2.11) is not affected when f is replaced by $\bar{f}$ with $\bar{f}(x) = \min\{f(x), f^{-1}(x)\}$. This $\bar{f}$ satisfies $\bar{f}(x) \leq x$. With an approximation argument, this shows that it is sufficient to prove (2.11) for all $f \in \mathscr{G}$ with $f(x) < x$ for all x. This will be deduced from (b) by demonstrating that for every such f, there exists a $g \in \mathscr{G}$ with $g(x) - g(f(x)) \equiv 1$, implying that

$$\{f(x) \geq y\} = \{g(x) - g(y) \geq 1\}.$$

Define $g(x) = -x/f(0)$ for $f(0) \leq x < 0$, and extend g as follows. If g is defined on $[f^n(0), f^{n-1}(0))$ then define $g(x) = g(f^{-1}(x)) - 1$ for $f^{n+1}(0) \leq x < f^n(0)$. Extend to the right in a similar way. This yields a function $\mathscr{G} \in \mathscr{G}$ satisfying $g(x) - g(f(x)) \equiv 1$.

(e) $\Leftrightarrow$ (f). Let $\mathscr{P}_e$ be the class of all bivariate probability distributions satisfying (2.3); it must be proved that $\mathscr{P}_e = \mathscr{P}_a$. It is clear that $\mathscr{P}_e$ is closed and convex and satisfies (2.4), so $\mathscr{P}_a \subset \mathscr{P}_e$. The proof of $\mathscr{P}_e \subset \mathscr{P}_a$ will only be sketched, without giving the topological details. It follows from (2.3) that the extreme points of $\mathscr{P}_e$ are precisely the distributions concentrated in some point (x, y) with $x \geq y$. These distributions satisfy the condition mentioned in (2.4); so $\mathscr{P}_e$, being the closed convex hull of its set of extreme points, is a subset of $\mathscr{P}_a$. □

PROPOSITION 1. *If $\mathscr{L}(X, Y)$ is asymmetric towards high X-values, then $P\{X \geq c\} \geq P\{Y \geq c\}$, with strict inequality for some c.*

PROOF. Theorem 1 implies the existence of a $\mathscr{L}(Z_1, Z_2, Z_3, Z_4)$ as mentioned in (2.3). As $\mathscr{L}(X, Y) \neq \mathscr{L}(Y, X)$, one has

$$0 < P\{Z_1 > Z_3 \quad \text{or} \quad Z_2 < Z_4\} \leq P\{Z_1 > Z_3\} + P\{Z_2 < Z_4\} = 2P\{Z_1 > Z_3\}.$$

But $P\{Z_1 \geq Z_3\} = 1$ and $P\{Z_1 > Z_3\} > 0$ imply that there exists a c with $P\{X \geq c\} = P\{Z_1 \geq c\} > P\{Z_3 \geq c\} = P\{Y \geq c\}$. □

3. Formulation of the testing problem. A random sample $(X_1, Y_1), \cdots, (X_n, Y_n)$ is drawn from the bivariate probability distribution $\mathscr{L}(X, Y)$. The null hypothesis of symmetry

$$H : \mathscr{L}(X, Y) = \mathscr{L}(Y, X)$$

is to be tested against the alternative hypothesis (see Definition 1)

$$A : \mathscr{L}(X, Y) \quad \text{is asymmetric towards high } X\text{-values.}$$

The level of significance is denoted by α.

Denote the group of transformations, generated by $\mathscr{H}_0$ (as defined at the beginning of Section 2) and the transformation $(x, y) \mapsto (-y, -x)$, by $\mathscr{H}$. It can be proved that $\mathscr{H}$ is the class of all $\overset{a}{\leq}$-increasing bijections $h : \mathbb{R}^2 \to \mathbb{R}^2$ which have an $\overset{a}{\leq}$-increasing inverse and which leave the null hypothesis H invariant. The testing problem is invariant under $\mathscr{H}$ when the action of $h \in \mathscr{H}$ on the outcome space $(\mathbb{R}^2)^n$ is defined by $((x_1, y_1), \cdots, (x_n, y_n)) \mapsto (h(x_1, y_1), \cdots, h(x_n, y_n))$. Of course, the testing problem is also invariant under permutations of the indices $1, 2, \cdots, n$.

NOTATION. Let $\tilde{U}_i = \max\{X_i, Y_i\}$, $\tilde{V}_i = \min\{X_i, Y_i\}$ and $\tilde{Z}_i = \text{sign}(X_i - Y_i)$. Let $U = ((U_1, V_1), \cdots, (U_n, V_n))$ be the permutation of $((\tilde{U}_1, \tilde{V}_1), \cdots, (\tilde{U}_n, \tilde{V}_n))$ for which $U_i \leq U_{i+1}$ and $V_i \leq V_{i+1}$ if $U_i = U_{i+1}$, and let $Z = (Z_1, \cdots, Z_n)$ be the corresponding permutation of $(\tilde{Z}_1, \cdots, \tilde{Z}_n)$. Let R_i and S_i be the ranks of U_i and V_i, respectively, in the combined sample $X_1, Y_1, X_2, \cdots, Y_n$ when ordered increasingly and let $R = ((R_1, S_1), \cdots, (R_n, S_n))$; use mid-ranks in case of ties.

(R, Z) is a maximal invariant statistic under $\mathscr{H}_0$ and permutation of the coordinates. Yanagimoto and Sibuya (1976) show that a test is similar-size α if and only if it is of size α conditionally on U. Hence the class Φ_0 of all permutation-and $\mathscr{H}_0$-invariant similar-size α tests consists of all tests $\varphi = \varphi(Z, R)$ which are of size α conditionally on R.

In this paper, attention will be concentrated on the case where the marginal distributions $\mathscr{L}(X)$ and $\mathscr{L}(Y)$ are continuous. For the case where $\mathscr{L}(X)$ and $\mathscr{L}(Y)$ can assume only finitely many values so that the data can be represented in a square contingency table, the reader is referred to Schaafsma (1966, Chapter 10) and Snijders (1979, Section 9.5).

4. A Hoeffding formula with some applications. For this testing problem, just as for other nonparametric testing problems, one can derive a Hoeffding formula for the distribution under the alternative and deduce from this the form of most powerful and locally most powerful rank tests; see Hoeffding (1951), Hájek and Šidák (1967, Section II.4) and Witting and Nölle (1970, Section 3.6). This produces the following results.

If P_0 and P_1 are bivariate distributions with $P_1 \ll P_0$ and $p = dP_1/dP_0$, while P_0 is symmetric, then

$$P_1\{Z = z \mid R = r\} = c_r E_0\{\textstyle\prod_{i=1}^n p(X_i, Y_i) \mid (Z, R) = (z, r)\}, \tag{4.1}$$

where $c_r = 2^{-k} P_0\{R = r\}/P_1\{R = r\}$ with $k = \#\{i \mid r_i \neq s_i\}$. The Fundamental Lemma of Neyman and Pearson shows that (4.1) can be used as a conditional test statistic for the most powerful Φ_0 test against the simple alternative that $\mathscr{L}(X, Y) = P_1$. It can be verified that for some $P_1 \in \mathscr{P}_a$, this yields the conditional sign test.

For the characterization of locally most powerful Φ_0 tests, consider a family $\{P_\Delta \mid 0 \leq \Delta \leq d\}$ of bivariate distributions of which P_0 is symmetric, and with $P_\Delta \ll P_0$ and $p_\Delta = dP_\Delta/dP_0$ for all Δ. Suppose that

$$\dot{p} = \lim_{\Delta \downarrow 0}(p_\Delta - 1)/\Delta$$

exists a.e. (P_0). The test statistic for the most powerful Φ_0 test against P_Δ is T_Δ defined by

$$T_\Delta(z, r) = E_0\{\textstyle\prod_{i=1}^n p_\Delta(X_i, Y_i) \mid (Z, R) = (z, r)\}.$$

Let φ be the conditionally size α test based on T defined by

$$T(z, r) = E_0\{\textstyle\sum_{i=1}^n \dot{p}(X_i, Y_i) \mid (Z, R) = (z, r)\}.$$

If interchange of expectation and differentiation is permitted, then at $\Delta = 0$ we have $dT_\Delta/d\Delta = T$; if, moreover, φ and T_Δ are such that $0 < \varphi(z, r) < 1$ and $T(z', r) = T(z, r)$ imply that $T_\Delta(z', r) = T_\Delta(z, r)$ for all Δ, then there exists a $\delta > 0$ such that φ is uniformly most powerful Φ_0 against the alternative that $\mathscr{L}(X, Y) \in \{P_\Delta \mid 0 < \Delta < \delta\}$.

For most families $\{P_\Delta \mid 0 \leq \Delta < d\}$, the test statistic T is hard to compute. If X and Y are independent under P_Δ for all Δ, then $\dot{p}(x, y) = p_1(x) - p_2(y)$ and the conditional expectation in the definition of $T(z, r)$ can be evaluated as the expectation of a function of univariate order statistics. This gives test functions which are the same as locally optimal test functions for the univariate two sample problem with equal sample sizes; only the critical values are different (and depend on r). For example, the LMP-Φ_0 test against the subalternative that X and Y are independent logistic random variables differing only in location uses a test statistic with Wilcoxon scores

$$\textstyle\sum_{i=1}^n Z_i(R_i - S_i). \tag{4.2}$$

The LMP-Φ_0 test against the subalternative that X and Y are independent normal random variables with $EX > EY$ and $\operatorname{Var} X = \operatorname{Var} Y > 0$ uses a test statistic with normal scores

$$\textstyle\sum_{i=1}^n Z_i(EV_{2n(R_i)} - EV_{2n(S_i)}), \tag{4.3}$$

where $V_{2n(j)}$ is distributed as the jth order statistic in a sample of size $2n$ from the standard normal distribution. Tests like these may be regarded as variants of the corresponding two sample tests, which are robust against dependence within the pairs.

5. Asymptotic normality and consistency. In this section it will be assumed that all bivariate distributions under consideration have continuous marginal distributions. A Hájek-type theorem of conditional asymptotic normality will be derived. It may be noted that Sen (1967) derived a Chernoff-Savage type theorem of conditional asymptotic normality. An extra index n indicates the sample size and Q_{ni} will denote the rank of X_i in $(X_1, Y_1, X_2, \cdots, Y_n)$.

The ranks will be normalized by defining $R^*_{ni} = R_{ni}/(2n+1)$ and similarly for S_{ni} and Q_{ni}. In the sequel, "almost surely" or "a.s." will mean "with probability 1 according to the distribution of $((X_1, Y_1), (X_2, Y_2), \cdots)$," and "a.e." will mean "almost everywhere according to Lebesgue measure on (0, 1)"; $\|\cdot\|_2$ denotes the L_2-norm for functions on (0, 1).

We consider a test statistic of the form

$$T_n = n^{-1/2} \textstyle\sum_{i=1}^n Z_i\{f_n(R^*_{ni}) - f_n(S^*_{ni})\}, \tag{5.1}$$

and assume that the score functions $f_n : (0, 1) \to \mathbb{R}$ satisfy the condition

C: The functions f_n are constant on the intervals with end points $j/2n$. There exists a square integrable and a.e. continuous $f : (0, 1) \to \mathbb{R}$ with $f_n \to f$ a.e. and $\|f_n - f\|_2 \to 0$.

Hájek and Šidák (1967, Section V.1) contains several sufficient conditions for C. E.g., the test statistics (4.2) and (4.3) can be expressed in this way. The following proposition is fundamental to this section; note that no independence assumptions between the X-sequence and the Y-sequence are made.

PROPOSITION 2. *Let $X_1, X_2, \cdots$ and $Y_1, Y_2, \cdots$ be two sequences of independent identically distributed random variables with distribution functions F and G, respectively, and let $H = \frac{1}{2}(F + G)$. Then*

$$n^{-1} \textstyle\sum_{i=1}^n \{f(H(X_i)) - f_n(Q^*_{ni})\}^2 \to 0 \qquad \text{a.s.}$$

PROOF. Ties which may occur contain, a.s., at most two observations. Such small ties do not affect the result and will be ignored.

The indicator function of the set A is denoted by I_A, and composition of functions by $\circ$. The following three statements are equivalent for functions g_n, g with $\|g\|_2 < \infty$ and $g_n \to g$ a.e.; for a proof see Witting and Nölle (1970, page 178).

$$\|g_n - g\|_2 \to 0 \tag{5.2}$$

$$\limsup_{n\to\infty} \|g_n\|_2 \le \|g\|_2 \tag{5.3}$$

$$\lim_{m\to\infty} \sup_n \|g_n \times (I_{(m,\infty)} \circ g_n)\|_2 = 0 \tag{5.4}$$

Lemma V. 1.6.b. of Hájek and Šidák (1967) implies that the function defined by

$$\tilde{f}_n(t) = n \int_{(i-1)/n}^{i/n} f(x)\,dx \qquad \text{for} \quad (i-1)/n < t \le i/n$$

satisfies $\tilde{f}_n \to f$ a.e. and $\|\tilde{f}_n - f\|_2 \to 0$. Also

$$\begin{aligned} & n^{-1} \textstyle\sum_{i=1}^n \{f_n(Q^*_{ni}) - \tilde{f}_n(Q^*_{ni})\}^2 \\ & \le n^{-1} \textstyle\sum_{j=1}^{2n} \{f_n(j/(2n+1)) - \tilde{f}_n(j/(2n+1))\}^2 \\ & = 2\|\tilde{f}_n - f_n\|_2^2 \to 0. \end{aligned}$$

This implies that in the following, it may be assumed that $f_n = \tilde{f}_n$.

Let $X_{n(1)}, \cdots, X_{n(n)}$ be the ordered sample $X_1, \cdots, X_n$ and let $Q^*_{n(1)} \cdots Q^*_{n(n)}$ be the ordered normalized ranks Q^*_{ni}. Define the step functions

$$A_n(t) = f_n(Q^*_{n(i)}), \qquad B_n(t) = f(H(X_{n(i)})) \qquad \text{for} \quad (i-1)/n < t \le i/n.$$

Let F_n and $(2n/(2n+1))H_n$ be the empirical distribution functions of $X_1, \cdots, X_n$ and of the combined sample $X_1, \cdots, X_n, Y_1, \cdots, Y_n$, respectively. Define the inverse of a distribution function D by

$$D^{-1}(t) = \inf\{x \mid D(x) \ge t\}.$$

These definitions entail

$$\begin{aligned} A_n &= f_n \circ H_n \circ F_n^{-1} \\ B_n &= f \circ H \circ F_n^{-1} \\ n^{-1} \textstyle\sum_{i=1}^n \{f(H(X_i)) - f_n(Q^*_{ni})\}^2 &= \| A_n - B_n \|^2 \\ \sup_t |D^{-1}(t) - t| &= \sup_t |D(t) - t|. \end{aligned} \tag{5.5}$$

The Glivenko-Cantelli Theorem implies that

$$\sup_t |F_n(t) - F(t)| \to 0, \qquad \sup_t |H_n(t) - H(t)| \to 0 \qquad \text{a.s.}$$

With the a.e: continuity of f and the assumption $f_n = \tilde{f}_n$, this implies that a.s.

$$A_n \to f \circ H \circ F^{-1}, \qquad B_n \to f \circ H \circ F^{-1} \qquad \text{a.e.} \tag{5.6}$$

The triangle inequality and (5.5) show that in order to prove the proposition, it is sufficient to prove that (i) $\| A_n - f \circ H \circ F^{-1} \|_2 \to 0$ and (ii) $\| B_n - f \circ H \circ F^{-1} \|_2 \to 0$ almost surely.

(i) It follows from (5.6) and the implication (5.4) $\Rightarrow$ (5.2) that it is sufficient to prove that

$$\lim_{m\to\infty} \sup_n \| A_n \times (I_{(m,\infty)} \circ A_n) \|_2 \to 0 \qquad \text{a.s.} \tag{5.7}$$

The latter result follows from

$$\begin{aligned} \| A_n \times (I_{(m,\infty)} \circ A_n) \|_2^2 &= n^{-1} \textstyle\sum_{i=1}^n f_n^2 \, (Q^*_{ni}) I_{(m,\infty)} \circ f_n(Q^*_{ni}) \\ &\le 2 \| f_n \times (I_{(m,\infty)} \circ f_n) \|_2^2, \end{aligned}$$

together with condition C and the implication (5.2) $\Rightarrow$ (5.4) applied to f_n.

(ii) It follows from (5.6) and the implication (5.3) $\Rightarrow$ (5.2) that it is sufficient to prove that

$$\| B_n \|_2^2 = n^{-1} \textstyle\sum_{i=1}^n f^2 \circ H(X_i) \to \| f \circ H \circ F^{-1} \|_2^2 < \infty$$

almost surely. The latter result follows from the law of large numbers and

$$\| f \circ H \circ F^{-1} \|_2^2 = \int (f^2 \circ H) \, dF \le 2 \int (f^2 \circ H) \, dH = 2 \| f \|_2^2 < \infty. \square \tag{5.8}$$

From Proposition 2, asymptotic approximations for the conditional null distribution of T_n, for the critical value and for T_n itself will be deduced. Denote the vector of ordered ranks $((R_{n1}, S_{n1}), \cdots, (R_{nn}, S_{nn}))$ by $R_{(n)}$. The null variance of T_n conditionally on $R_{(n)}$ is

$$\hat{\sigma}_n^2(R_{(n)}) = n^{-1} \textstyle\sum_{i=1}^n \{f_n(R^*_{ni}) - f_n(S^*_{ni})\}^2.$$

Proposition 2, combined with the strong law of large numbers and the triangle inequality, implies that a.s.

$$\hat{\sigma}_n^2(R_{(n)}) \to \sigma^2 = E\{f(F(X)) - f(G(Y))\}^2 < \infty. \tag{5.9}$$

THEOREM 2. *Let $\mathscr{L}(X, Y)$ be symmetric, and let F be the marginal distribution function of X and Y. Then the distribution of T_n conditionally on $R_{(n)}$ is asymptotically normal,*

(5.10) $$\mathscr{L}(T_n \mid R_{(n)}) \to \mathscr{N}(0, \sigma^2) \quad \text{almost surely}$$

with

$$\sigma^2 = E\{f(F(X)) - f(F(Y))\}^2 = \mathrm{Var}\{f(F(X)) - f(F(Y))\}.$$

PROOF. For $\sigma^2 = 0$, (5.9) means that $E\{T_n^2 \mid R_{(n)}\} \to 0$ a.s., which implies (5.10). Now let $\sigma^2 > 0$. Conditionally on $R_{(n)}$ the Z_i are independent with $P\{Z_i = 1 \mid R_{(n)}\} = P\{Z_i = -1 \mid R_{(n)}\} = \frac{1}{2}$ for $R_{ni} \neq S_{ni}$. Define

$$\sigma_{ni}^2 = \{f_n(R_{ni}^*) - f_n(S_{ni}^*)\}^2.$$

The Central Limit Theorem of Lindeberg-Feller implies that for (5.10) it is sufficient that (5.9) holds and

(5.11) $$n^{-1} \sum_{i=1}^n \sigma_{ni}^2 I_{(\epsilon n, \infty)}(\sigma_{ni}^2) \to 0 \quad \text{for all } \epsilon > 0, \quad \text{a.s.}$$

From (5.7) and the similar result for the Y-sequence it can be deduced that (5.11) holds. □

COROLLARY. *Let $c_n(r_{(n)})$ be the critical value of the conditional size α test based on T_n. Let F and G be the marginal distribution functions of $\mathscr{L}(X, Y)$ and define $H = \frac{1}{2}(F + G)$. Then*

$$c_n(R_{(n)}) \to \sigma u_\alpha \quad \text{almost surely,}$$

with

$$\sigma^2 = E\{f(H(X)) - f(H(Y))\}^2 < \infty.$$

PROOF. The probability distribution of $R_{(n)}$ is the same for $\mathscr{L}(X, Y)$ as for $\frac{1}{2}(\mathscr{L}(X, Y) + \mathscr{L}(Y, X))$. Apply Theorem 2 to the latter (symmetric) bivariate distribution. □

From this corollary and (5.9) it can be concluded that the sequence of exact conditional tests based on T_n is under the null hypothesis equivalent to the sequence of approximate unconditional tests rejecting for

(5.12) $$T_n > u_\alpha \hat{\sigma}_n(R_{(n)}).$$

The test (5.12) is, of course, much easier to carry out.

THEOREM 3. *Both the conditional test based on T_n and the test rejecting for* (5.12) *are consistent against all distributions $\mathscr{L}(X, Y)$ whose marginal distribution functions F and G satisfy $Ef(H(X)) > Ef(H(Y))$, where $H = \frac{1}{2}(F + G)$.*

PROOF. It follows from (5.8) that $Ef^2(H(X)) < \infty$ and $Ef^2(H(Y)) < \infty$. Proposition 2, the law of large numbers and the Cauchy-Schwarz inequality imply that a.s.

$$n^{-1/2} T_n \to E\{f(H(X)) - f(H(Y))\} < \infty.$$

So $T_n \to \infty$ a.s. if $Ef(H(X)) > Ef(H(Y))$. But the corollary and (5.9) imply that the critical values converge to a finite limit a.s. □

It follows from Theorem 3 and Proposition 1 that if f is strictly increasing (as is the case, e.g., for Wilcoxon (4.2) and normal (4.3) scores), then the two tests based on T_n are consistent against all $\mathscr{L}(X, Y)$ which are asymmetric towards high X-values.

THEOREM 4. *Let $\mathscr{L}(X, Y)$ be symmetric, let F be the marginal distribution function of X and Y, and*

$$\tilde{T}_n = n^{-1/2} \sum_{i=1}^n \{f(F(X_i)) - f(F(Y_i))\}.$$

Then $T_n - \tilde{T}_n \to 0$ in probability.

PROOF. Let W_{ni}^* be the normalized rank of Y_i in $(X_1, Y_1, X_2, \cdots, Y_n)$ and let $U_{(n)} = ((U_{n1}, V_{n1}), \cdots, (U_{nn}, V_{nn}))$. The triangle inequality yields

$$[E\{(T_n - \tilde{T}_n)^2 \mid U_{(n)}\}]^{1/2} = [n^{-1} \sum_{i=1}^n \{f_n(Q_{ni}^*) - f_n(W_{ni}^*) - f(F(X_i)) + f(F(Y_i))\}^2]^{1/2}$$
$$\le [n^{-1} \sum_{i=1}^n \{f_n(Q_{ni}^*) - f(F(X_i))\}^2]^{1/2}$$
$$+ [n^{-1} \sum_{i=1}^n \{f_n(W_{ni}^*) - f(F(Y_i))\}^2]^{1/2}.$$

According to Proposition 2, the right-hand sides converges to 0 a.s. This implies that

$$P\{| T_n - \tilde{T}_n| > \epsilon \mid U_{(n)}\} \to 0 \qquad \text{a.s.}$$

for every $\epsilon > 0$, which again implies that $T_n - \tilde{T}_n \to 0$ in probability. □

Theorem 4 can be helpful when one wishes to apply to the test based on T_n well-known general methods for obtaining asymptotic relative efficiencies against contiguous alternatives, as given by Hájek and Šidák (1967, Section VII.2.1) or Witting and Nölle (1970, Section 4.3). This yields results which are also obtained by Sen (1967). An interesting special case is that the test using normal scores (4.3) or Van der Waerden scores is asymptotically uniformly most powerful (with respect to contiguous alternatives) for testing H against the subalternative that $\mathscr{L}(X, Y)$ is a bivariate normal distribution with $EX > EY$ and Var X = Var Y.

Acknowledgment. I would like to thank Willem Schaafsma for arousing my interest in this subject, and for many illuminating discussions.

REFERENCES

BELL, C. B. and HALLER, H. SMITH (1969). Bivariate symmetry tests: parametric and non-parametric. *Ann. Math. Statist.* **40** 259–269.

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests.* Academic, New York.

HOEFFDING, WASSILY (1951). "Optimum" nonparametric tests. *Proc. 2nd Berkeley Symp. Math. Stat. Prob.* 83–92.

HOLLANDER, MYLES (1971). A nonparametric test for bivariate symmetry. *Biometrika* **58** 203–212.

KAMAE, T., KRENGEL, U. and O'BRIEN, G. L. (1977). Stochastic inequalities on partially ordered spaces. *Ann. Probability* **5** 899–912.

KOZIOL, JAMES, A. (1979). A test for bivariate symmetry based on the empirical distribution function. *Comm. Statist.-Theor. Meth. A* **8** 207–221.

LEHMANN, E. L. (1955). Ordered families of distributions. *Ann. Math. Statist.* **26** 399–419.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

SCHAAFSMA, WILLEM (1966). *Hypothesis Testing Problems With the Alternative Restricted by a Number of Inequalities.* Noordhoff, Groningen.

SCHAAFSMA, WILLEM (1976). Bivariate symmetry and asymmetry. Report TW-170, Math. Inst. Univ. Groningen.

SEN, PRANAB KUMAR (1967). Nonparametric tests for multivariate interchangeability. Part 1: problems of location and scale in bivariate distributions. *Sankhyā A* **29** 351–372.

SNIJDERS, T. A. B. (1979). *Asymptotic Optimality Theory for Testing Problems with Restricted Alternatives.* Mathematical Centre, Amsterdam.

WITTING, H. and NÖLLE, G. (1970). *Angewandte Mathematische Statistik*, Teubner, Stuttgart.

YANAGIMOTO, TAKEMI and SIBUYA, MASAAKI (1972). Stochastically larger component of a random vector. *Ann. Inst. Statist. Math.* **24** 259–269.

YANAGIMOTO, TAKEMI and SIBUYA, MASAAKI (1976). Tests of symmetry of a bivariate distribution. *Sankhyā A* **38** 105–115.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF GRONINGEN
P.O. BOX 800
GRONINGEN, THE NETHERLANDS