The Annals of Statistics
1980, Vol. 8, No. 1, 190–198

A STOCHASTIC ORDERING INDUCED BY A CONCEPT OF POSITIVE DEPENDENCE AND MONOTONICITY OF ASYMPTOTIC TEST SIZES

BY YOSEF RINOTT AND MOSHE POLLAK

The Hebrew University of Jerusalem

An ordering of distributions related to a concept of positive dependence is studied and stochastic monotonicity with respect to this ordering is established for a wide class of two-sample test statistics. Asymptotic conservativeness of test sizes under certain departures from independence between samples is discussed. For example, if the observations are paired and the joint density is positive semidefinite then tests such as Kolmogorov-Smirnov, χ^2 and Cramér-von Mises, as well as a large class of linear rank tests, are shown to be asymptotically conservative.

1. Introduction. Consider the following stochastic partial ordering:

DEFINITION. Let (X, Y) and (Q, R) be bivariate random vectors. We say that X, Y are more positively dependent than Q, R (denoted by $(X, Y) \succ_{pd} (Q, R)$) if X, Y, Q and R have a common continuous marginal distribution and if $\operatorname{Cov}[h(X), h(Y)] \geqslant \operatorname{Cov}[h(Q), h(R)]$ for all functions h for which the covariances exist.

For X, Y having a common continuous marginal distribution F, let $Z^{(X, Y)}$ denote the Gaussian process in $C[0, 1]$ with zero expectation and covariance function defined for $0 \leqslant s \leqslant t \leqslant 1$ by

$$(1) \qquad c(s, t) \equiv P(F(X) < s, F(Y) > t) + P(F(X) > t, F(Y) < s).$$

Monotonicity of $P(Z^{(X, Y)} \in A)$ with respect to the ordering $\succ_{pd}$ is established (Theorem 1) for sets A in $C[0, 1]$ which are essentially closed, convex and symmetric. The main tool used in the proof of Theorem 1 is a generalization of a well-known result of Anderson (1955). Theorem 1 is applied in the context of testing the hypothesis H_0 that (X, Y) has equal marginal distributions, where $P(Z^{(X, Y)} \in A)$ arises as the asymptotic acceptance probability of tests whose statistics are essentially convex functionals of the difference between the marginal empirical distribution functions based on a sample from (X, Y). The resulting monotonicity property of asymptotic test sizes implies conservativeness (Theorem 2): if such a test is designed to be of asymptotic level α when X and Y are independent then its asymptotic size will continue to be bounded by α when the assumption of independence is replaced by the condition

$$(2) \qquad \operatorname{Cov}[h(X), h(Y)] \geqslant 0$$

Received April 1978; revised August 1978.
AMS 1970 *subject classifications.* Primary 62E10; secondary 62E20.
Key words and phrases. Stochastic ordering, positive dependence, two-sample test, conservativeness, convexity, Gaussian processes, total positivity.

for every h for which the covariance exists. Shaked (1974) formulated condition (2) as a concept of positive dependence and proved its equivalence to positive semidefiniteness of the joint distribution function. This and other conditions which imply (2) are given in Theorem 3. See Lehmann (1966) and Jogdeo (1977) (and references therein) for a discussion of other concepts of dependence.

The two-sample t-test is not contained in the class of tests under consideration. However, it illustrates the relationship between conservativeness and positive dependence and motivates the study of this relationship in nonparametric tests. Denoting the sample means and variances by $\bar{X}$, $\bar{Y}$, S_X^2, S_Y^2, consider $t = [n(\bar{X} - \bar{Y})^2/(S_X^2 + S_Y^2)]^{\frac{1}{2}}$. If

$$(X, Y) \sim \mathfrak{N}\left((\mu_X, \mu_Y), \sigma^2\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

then (up to a constant) the denominator of t^2 is the trace of a Wishart matrix, which is distributed as a weighted sum of two independent $\chi^2_{(n-1)}$ variables. It follows that (under H_0) t^2 is distributed as $2(n-1)W/[U + V(1+\rho)/(1-\rho)]$, where $W \sim \chi^2_{(1)}$ and $U, V \sim \chi^2_{(n-1)}$ are all independently distributed. Thus the size $P_{H_0}(t^2 > c_\alpha)$ of the two-sided t-test is monotonically decreasing in ρ. In particular, if the test has size α under independence ($\rho = 0$), it will be conservative (size $\leqslant \alpha$) for $\rho > 0$.

In Section 3 we show that our approach applies to tests which are not asymptotically normal (χ^2, Kolmogorov-Smirnov, Cramér-von Mises). The same approach applies also to tests which are asymptotically normal (e.g., Wilcoxon and other linear rank tests) in which case the asymptotic size depends only on the asymptotic variance. This fact was employed by Hollander, Pledger and Lin (1974) to obtain asymptotic conservativeness for the Wilcoxon test. For linear rank tests the asymptotic theory for dependent samples was developed by Sen (1967).

2. The order relation and its application to asymptotic test sizes. In this section we study the ordering $\succ_{pd}$ and its relation to the process $Z^{(X,Y)}$ and discuss applications to two-sample tests.

A functional T on $C[0, 1]$ is said to be symmetric if $T(z) = T(-z)$ for all $z \in C[0, 1]$. A set $A \subseteq C[0, 1]$ is said to be symmetric if $A = -A$. ($-A = \{z \in C[0, 1] : -z \in A\}$.) The class of closed convex and symmetric sets in $C[0, 1]$ will be denoted by $\mathcal{Q}$.

THEOREM 1. *Let* $(X, Y) \succ_{pd} (Q, R)$ *and let* $A \in \mathcal{Q}$. *Then*

$$P(Z^{(X,Y)} \in A) \geqslant P(Z^{(Q,R)} \in A).$$

PROOF. The following lemma which is needed for the proof of Theorem 1 is a generalization of Corollaries 3 and 4 of Anderson (1955).

LEMMA 1. *Let* Z_1, Z_2 *be Gaussian processes in* $C[0, 1]$ *with zero expectations and covariance functions* $c_1(s, t)$, $c_2(s, t)$ *respectively such that* $c_2(s, t) - c_1(s, t)$ *is a positive semidefinite function. Then for* $A \in \mathcal{Q}$ $P(Z_1 \in A) \geqslant P(Z_2 \in A)$.

The proof of Lemma 1 is relegated to the Appendix. We continue with the proof of Theorem 1. For $0 \leqslant s \leqslant t \leqslant 1$, the covariance functions c_1 and c_2 of the processes $Z^{(X,Y)}$ and $Z^{(Q,R)}$ are given by

$$c_1(s,t) = P(F(X) < s, F(Y) > t) + P(F(X) > t, F(Y) < s)$$
$$c_2(s,t) = P(F(Q) < s, F(R) > t) + P(F(Q) > t, F(R) < s)$$

where F is the common marginal distribution of X, Y, Q and R. Let $\mathbf{F}$ and $\mathbf{G}$ be the joint distribution functions of $(X, Y), (Q, R)$ respectively, and let $\tilde{\mathbf{F}}(s,t) = \frac{1}{2}[\mathbf{F}(s,t) + \mathbf{F}(t,s)]$, $\tilde{\mathbf{G}}(s,t) = \frac{1}{2}[\mathbf{G}(s,t) + \mathbf{G}(t,s)]$ be their symmetrizations. By Lemma 1 it suffices to prove positive semidefiniteness of the kernel $c_2(s,t) - c_1(s,t)$, or equivalently of the kernel $\tilde{\mathbf{F}}(s,t) - \tilde{\mathbf{G}}(s,t)$, since $c_2(F(s), F(t)) - c_1(F(s), F(t)) = 2[\tilde{\mathbf{F}}(s,t) - \tilde{\mathbf{G}}(s,t)]$. Note that since X and Y have equal marginals, $\operatorname{Cov}(h(X), h(Y))$ remains unchanged when the joint distribution $\mathbf{F}$ is replaced by $\tilde{\mathbf{F}}$, for any function h. For measurable h of bounded variation

$$\begin{aligned} 0 &\leqslant \operatorname{Cov}[h(X), h(Y)] - \operatorname{Cov}[h(Q), h(R)] \\ &= \int\int h(s)h(t)\,d\tilde{\mathbf{F}}(s,t) - \int\int h(s)h(t)\,d\tilde{\mathbf{G}}(s,t) \\ &= \int\int [\tilde{\mathbf{F}}(s,t) - \tilde{\mathbf{G}}(s,t)]\,dh(s)\,dh(t). \end{aligned} \tag{3}$$

The latter expression is a quadratic form in the kernel $\tilde{\mathbf{F}}(s,t) - \tilde{\mathbf{G}}(s,t)$ and it follows that the kernel is positive semidefinite. □

COROLLARY 1. *It follows from* (3) *that* $(X, Y) \succ_{pd} (Q, R)$ *if and only if* $\tilde{\mathbf{F}}(s,t) - \tilde{\mathbf{G}}(s,t)$ *is a positive semidefinite kernel.*

We will now be concerned with testing the hypothesis H_0 that the random vector (X, Y) has equal marginal distributions (which are assumed to be continuous) and with formulating the relation between $Z^{(X,Y)}$ and two-sample tests. Under H_0 let F be the common continuous marginal distribution of X, Y. We shall consider test statistics T_n based on a sample (X_i, Y_i), $i = 1, \cdots, n$ of i.i.d. random vectors distributed as (X, Y). F_n^X and F_n^Y will denote the empirical distribution functions of the X_i and Y_i samples respectively. We consider the uniform (0, 1) variables defined by $X_i^* = F(X_i)$ and $Y_i^* = F(Y_i)$, $i = 1, \cdots, n$. Let $F_n^{X^*}$ and $F_n^{Y^*}$ denote the smoothed versions of the empirical distribution functions of the X_i^* and Y_i^* samples respectively in the same way as $G_n(t, \omega)$ is a smoothed version of $F_n(t, \omega)$ in Billingsley (1968, page 104). For $t \in [0, 1]$ consider the process $Z_n(t) = n^{\frac{1}{2}}(F_n^{X^*}(t) - F_n^{Y^*}(t))$ in $C[0, 1]$. The following lemma can be proved by a method similar to that of the proof of Theorem 13.1 of Billingsley (1968), or by Theorem 2 of Dudley (1966).

LEMMA 2. *Under* H_0 $Z_n \to_{\mathcal{D}} Z^{(X,Y)}$ *(*Z_n *converges weakly to the process* $Z^{(X,Y)}$ *in* $C[0, 1]$ *as* $n \to \infty$.

When X and Y are independent $c(s,t) = 2s(1-t)$, $0 \leqslant s \leqslant t \leqslant 1$, and the process $Z^{(X,Y)}$ (which in this case is $2^{\frac{1}{2}} \times$ a Brownian bridge) will be denoted by Z^{IND}.

ASSUMPTION 1. A sequence (T_n, C_n) of test statistics T_n and acceptance regions C_n is said to satisfy Assumption 1 if, under H_0, for any common continuous marginal distribution function F of X, Y there exists a set $A_F \in \mathcal{C}$ such that

$$(4) \qquad P(T_n \in C_n) \to_{n\to\infty} P(Z^{(X,Y)} \in A_F).$$

LEMMA 3. *Let T_n be a sequence of test statistics satisfying: under H_0, for any common continuous marginal distribution function F of X, Y*

(A) *T_n can be represented as $T_n = T_F(n^{\frac{1}{2}}(F_n^{X^*} - F_n^{Y^*})) + o_P(1)$ where T_F is a continuous convex and symmetric functional on $C[0, 1]$ and $o_P(1) \to 0$ in probability $(n \to \infty)$.*

(B) *The distribution function $P(T_F(Z^{(X,Y)}) \leqslant t)$ is continuous at $t = k$.*

Let $C_n = \{T_n \leqslant k\}$. Then the sequence (T_n, C_n) satisfies Assumption 1 with $A_F = \{z \in C[0, 1] : T_F(z) \leqslant k\}$.

PROOF. Since T_F is continuous Lemma 2 implies $T_n \to_{\mathfrak{D}} T_F(Z^{(X,Y)})$. Therefore by (B) $P(T_n \leqslant k) \to P(T_F(Z^{(X,Y)}) \leqslant k)$, implying (4). $A_F \in \mathcal{C}$ since T_F is also convex and symmetric. $\square$

A variety of tests satisfying Assumption 1 will be discussed in Section 3.

REMARK 1. By an argument similar to that of the proof of Lemma 1 (see Appendix) we have $T_F(Z^{(X,Y)}) = \sup_{n=1,2,\ldots}\{L_n(Z^{(X,Y)})\}$ where L_n are continuous affine support functionals. $L_n(Z^{(X,Y)})$ are jointly Gaussian and so by an argument of Rinott (1976) $H_F(t) = P(T_F(Z^{(X,Y)}) \leqslant t)$ is log concave. Hence condition (B) of Lemma 3 holds for all k except possibly at the largest lower bound of the support of H_F.

REMARK 2. The asymptotic size of a sequence of tests (T_n, C_n) satisfying (4) is $P(Z^{(X,Y)} \in A_F^c)$ where A^c = complement of A. Thus Theorem 1 can be interpreted as monotonicity of asymptotic test sizes in the relation $\succ_{pd}$. In particular we obtain conservativeness under departure from independence:

THEOREM 2. *Let (T_n, C_n) be a sequence satisfying Assumption 1 and suppose that $P(Z^{\text{IND}} \in A_F^c) \leqslant \alpha$ (asymptotic level α under independence) for every continuous F. Then $P(Z^{(X,Y)} \in A_F^c) \leqslant \alpha$ under H_0 for any joint distribution of (X, Y) having a common continuous marginal F provided that under H_0*

$$(5) \qquad \operatorname{Cov}(h(X), h(Y)) \geqslant 0$$

for any function h for which the covariance exists.

PROOF. Let (X, Y) satisfy (5) and let (Q, R) be independent such that X, Y, Q and R have the same (marginal) distribution F. Then by (5) $(X, Y) \succ_{pd} (Q, R)$ implying by Theorem 1 that $P(Z^{(X,Y)} \in A_F^c) \leqslant P(Z^{(Q,R)} \in A_F^c) = P(Z^{\text{IND}} \in A_F^c) \leqslant \alpha$. $\square$

The following theorem is useful for verifying the condition (5).

THEOREM 3. *Let the random vector (X, Y) have a joint distribution $\mathbf{F}$ with equal marginals and let $\tilde{\mathbf{F}}(x, y) = \frac{1}{2}[\mathbf{F}(x, y) + \mathbf{F}(y, x)]$ be the symmetrization of $\mathbf{F}$. In order* *for* $\mathrm{Cov}(h(X), h(Y)) \geq 0$ *to hold for all measurable h for which the covariance exists any of the following three conditions is sufficient. Condition* (i) *is also necessary.*

(i) $\tilde{\mathbf{F}}(x, y)$ *is positive semidefinite.*

(ii) $\tilde{\mathbf{F}}(x, y)$ *has a density function* $\tilde{\mathbf{f}}(x, y)$ *which is positive semidefinite.*

(iii) *X and Y are conditionally* i.i.d. (i.e., *there exists a random R such that conditional on R the variables X and Y are* i.i.d. *R may be regarded as taking values on the space of univariate distribution functions*).

PROOF. The equivalence to condition (i) is an immediate generalization of Theorem 2.3 of Shaked (1974). If a density exists, equivalence to (ii) follows since $\mathrm{Cov}(h(X), L(Y)) = \int\int \tilde{\mathbf{f}}(x, y) g(x) g(y) dx dy$ where $g(x) = h(x) - Eh(X)$, and this is a quadratic form in $\tilde{\mathbf{f}}$. Note that if f is TP_∞ (totally positive) then it is positive semidefinite (Karlin (1968)). The sufficiency of (iii) is a special case of a result of Dykstra, Hewett and Thompson (1973). A short proof is:

$$Eh(X)h(Y) = EE(h(X)h(Y)|R) = E\left[E(h(X)|R)^2\right]$$
$$\geq (E(h(X))^2 = Eh(X)Eh(Y). \qquad \square$$

EXAMPLE 1. Let (X, Y), (Q, R) have a bivariate normal distribution with a common marginal $F = N(\mu, \sigma^2)$, and correlations ρ_1, ρ_2 respectively, $0 \leq \rho_2 < \rho_1 \leq 1$. Then $(X, Y) \succ_{pd} (Q, R)$.

PROOF. Let $\mathbf{F}_\rho$ and $\mathbf{f}_\rho$ be the joint distribution and density of $\mathfrak{N}\left(\begin{pmatrix}\mu\\ \mu\end{pmatrix}, \sigma^2\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)$. By Corollary 1 it suffices to prove that the function $\mathbf{F}_{\rho_1}(s, t) - \mathbf{F}_{\rho_2}(s, t)$ is positive definite.

For $t_1 < t_2 < \cdots < t_n$ let D_ρ be the matrix $D_\rho = (\mathbf{F}_\rho(t_i, t_j))_{n \times n}$. By a result of Slepian (1962) $d\mathbf{F}_\rho(s, t)/d\rho = \sigma^2 \mathbf{f}_\rho(s, t)$. In matrix notation we thus have $dD_\rho/d\rho = \sigma^2(\mathbf{f}_\rho(t_i, t_j))_{n \times n}$ which is totally positive (see Karlin (1968)) and hence positive definite for $0 < \rho < 1$. Thus for any $x \in R^n$

$$\frac{d}{d\rho}(x' D_\rho x) = x'\left(\sigma^2 \mathbf{f}_\rho(t_i, t_j)\right)_{n \times n} x > 0 \qquad \text{for} \quad x \neq 0,$$

implying $x'(D_{\rho_1} - D_{\rho_2})x = x' D_{\rho_1} x - x' D_{\rho_2} x > 0$, so that $D_{\rho_1} - D_{\rho_2}$ is positive definite. It follows that $\mathbf{F}_{\rho_1}(s, t) - \mathbf{F}_{\rho_2}(s, t)$ is positive definite. $\square$

From Theorem 1 it now follows that if $(X, Y) \sim \mathfrak{N}\left(\begin{pmatrix}\mu\\ \mu\end{pmatrix}, \sigma^2\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)$ then the asymptotic size $P(Z^{(X, Y)} \in A_F^c)$ of a sequence of tests satisfying Assumption 1 is monotone in $0 \leq \rho \leq 1$. Note also that for $\rho \geq 0$ conditions (i), (ii) and (iii) of Theorem 3 hold. Conservativeness in the sense of Theorem 2 follows.

Examples of distributions satisfying condition (ii) can be found in Karlin (1968). Further examples of distributions which satisfy conditions of Theorem 3 are listed by Shaked (1974, 1977). These examples include bivariate versions of the F, χ^2, logistic and exponential distributions.

We now provide a simple example of the order relation $\succ_{pd}$ designed to illustrate the underlying concept of dependence.

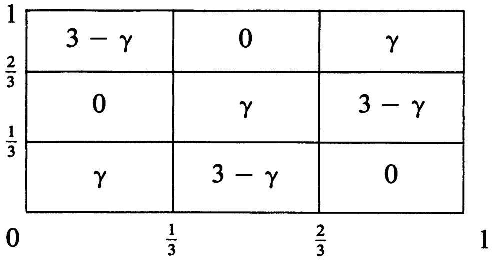

FIG. 1

EXAMPLE 2. Consider the bivariate density $f_\gamma(x, y)$ $0 \leqslant x, y \leqslant 1$ having values $0 \leqslant \gamma \leqslant 3$, $3 - \gamma$, and 0, as described in Figure 1. Let (X, Y) have density $f_\gamma(x, y)$ (note that $f_\gamma(x, y)$ is not symmetric, i.e., X and Y are not interchangeable) and let (Q, R) have density $f_{\gamma'}$. Then $\gamma > \gamma'$ implies $(X, Y) >_{pd} (Q, R)$.

PROOF. First note that the marginals are all uniform (0, 1). For a given function h denote $b_i = E(h(X)|(i-1)/3 \leqslant X \leqslant i/3)$ $i = 1, 2, 3$. Then $Eh(X)h(Y) = (\gamma/9)(b_1^2 + b_2^2 + b_3^2) + ((3-\gamma)/9)(b_1b_2 + b_1b_3 + b_2b_3)$ which is increasing in γ. Therefore $\gamma > \gamma'$ implies $\operatorname{Cov}(h(X), h(Y)) \geqslant \operatorname{Cov}(h(Q), h(R))$. □

REMARK 3. Replacing 0, $3 - \gamma$ in Figure 1 by $\eta > 0$, $\xi > 0$, respectively, where $\eta + \xi = 3 - \gamma$ we obtain a family of densities associated with random variables which are equivalent to (X, Y) with respect to the order relation $>_{pd}$. In particular the case $\eta = \xi$ corresponds to the symmetrization of f_γ.

3. Kolmogorov-Smirnov, Cramér-von Mises, Pearson χ^2, Wilcoxon and linear rank tests. In this section we depict the class of tests considered in Section 2 by specifying some examples of well-known tests for which Assumption 1 holds. The notation used below is defined in Section 2.

The Kolmogorov-Smirnov test. This test rejects H_0 when $T_n = \sup_{-\infty<t<\infty} n^{\frac{1}{2}}|F_n^X(t) - F_n^Y(t)| > k$. This statistic is invariant in the sense that $T_n(X_1, \cdots, X_n, Y_1, \cdots, Y_n) = T_n(g(X_1), \cdots, g(X_n), g(Y_1), \cdots, g(Y_n))$ for every increasing function g. Therefore $T_n = \sup_{0\leqslant t\leqslant 1} n^{\frac{1}{2}}|F_n^{X^*}(t) - F_n^{Y^*}(t)| + o_P(1)$. Thus (A) of Lemma 3 holds with $T_F(Z) = \sup_{0\leqslant t\leqslant 1}|Z(t)|$. (Here T_F does not depend on F because of invariance.) Since $P(T_F(Z^{(X,Y)}) = 0) = 0$ condition (B) holds here (and in the following examples by a similar argument) for all k by virtue of Remark 1. By Lemma 3 Assumption 1 holds.

Cramér-von Mises-type statistics. Set $C_n = \{T_n^{(r)} \leqslant k\}$ where

$$T_n^{(r)} = \int \left(n^{\frac{1}{2}}|F_n^X(t) - F_n^Y(t)|\right)^r d\left(\frac{F_n^X(t) + F_n^Y(t)}{2}\right), \qquad r \geqslant 1.$$

When X_i and Y_i are uniform (0, 1) then by a lemma of Kiefer (1959, page 424) $T_n^{(r)} - \int_0^1 (n^{\frac{1}{2}}|F_n^X(t) - F_n^Y(t)|)^r dt$ converges to zero in probability as $n \to \infty$. (Although Kiefer's lemma is formulated for the case of independent samples, it remains valid with essentially the same proof for any joint distribution of (X_i, Y_i) with uniform (0, 1) marginals.) Therefore, since $T_n^{(r)}$ is invariant, $T_n^{(r)} = \int_0^1 (n^{\frac{1}{2}}|F_n^{X^*}(t) - F_n^{Y^*}(t)|)^r dt + o_P(1)$ implying Assumption 1 by Lemma 3.

Pearson χ^2. This is an example of a test which is not invariant under monotone transformations of the observations.

Given $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_n$, $N_i^X = n[F_n^X(t_{i+1}) - F_n^X(t_i)]$ is the number of X's observed in the interval $(t_i, t_{i+1}]$. Define N_i^Y analogously. If the data are categorized into cells $(t_i, t_{i+1}]$, $i = 1, \cdots, k$, where $-\infty \leqslant t_1 < t_2 < \cdots < t_{k+1} \leqslant \infty$ then the χ^2 statistic can be expressed as

$$T_n = \Sigma_{i=1}^k \frac{(N_i^X - N_i^Y)^2}{e_i}$$

where $e_i = N_i^X + N_i^Y$. (Empty cells are allowed by the convention $0/0 = 0$. See Remark 4 below.) H_0 is rejected when $T_n > c_\alpha$ where c_α is the $(1 - \alpha)$-percentile of the $\chi^2_{(k-1)}$ distribution. Note that

$$T_n = \Sigma_{i=1}^k \frac{n\{[F_n^X(t_{i+1}) - F_n^X(t_i)] - [F_n^Y(t_{i+1}) - F_n^Y(t_i)]\}^2}{F_n^X(t_{i+1}) - F_n^X(t_i) + F_n^Y(t_{i+1}) - F_n^Y(t_i)}.$$

Let X_i, Y_i have a common continuous marginal F. Then

$$T_n = \Sigma_{i=1}^k \frac{\{n^{\frac{1}{2}}[F_n^{X^*}(F(t_{i+1})) - F_n^{Y^*}(F(t_{i+1}))] - n^{\frac{1}{2}}[F_n^{X^*}(F(t_i)) - F_n^{Y^*}(t_i))]\}^2}{F_n^X(t_{i+1}) - F_n^X(t_i) + F_n^Y(t_{i+1}) - F_n^Y(t_i)} + o_P(1)$$

(a term is understood to be zero if its denominator is zero). Since under H_0 the denominator of the ith term in this sum converges a.s. to $Q_i^F = 2[F(t_{i+1}) - F(t_i)]$, it follows that Lemma 3 holds with

$$T_F(Z^{(X,Y)}) = \Sigma_{i=1}^k [Z^{(X,Y)}(F(t_{i+1})) - Z^{(X,Y)}(F(t_i))]^2 / Q_i^F.$$

REMARK 4. Unlike the previous examples the χ^2 test cannot under independence attain asymptotic size α for all F. If $F(t_i)$, $i = 1, \cdots, k+1$ are not all distinct then some of the cells will be empty, and the statistic will in reality have fewer degrees of freedom than $k - 1$. In this case the actual asymptotic size will be strictly less than α and in general $P(Z^{\text{IND}} \in A_F^c) \leqslant \alpha$.

Linear rank tests. In order to define general linear rank statistics let $\xi_1, \cdots, \xi_{2n}$ denote combined ordered sample of X's and Y's and set $\eta_i = 1$ or -1 if ξ_i belongs to the sample of X's or Y's respectively. A linear rank statistic is defined by

$$T_n = \left[\Sigma_{i=1}^{2n} J\left(\frac{i}{2n+1}\right)\eta_i\right] / n^{\frac{1}{2}}$$

where J is a function defined on (0, 1). $J(t) = t$ corresponds to the Wilcoxon statistic; $J(t) = \Phi^{-1}(t)$ where Φ is the normal cdf corresponds to the Van der Waerden statistic. One can obtain under certain conditions on J (for details, see Theorem 5.1 of Sen (1967)) that, under H_0, as $n \to \infty$.

$$T_n \to_{\mathfrak{D}} \int Z^{(X,Y)}(t) dJ(t) \sim N(0, \sigma^2) \qquad \text{where}$$

(6) $$\sigma^2 = 4\int\int_{-\infty<x<y<\infty} F(x)(1 - F(y)) dJ(F(x)) dJ(F(y)) - 2\,\mathrm{Cov}(J(F(X)), J(F(Y))),$$

where (under H_0) F denotes the common marginal distribution of X and Y. Hence Assumption 1 holds with $C_n = \{|T_n| \leqslant k\}$. (Note that when J is not of bounded variation the functional $T(z) = \int_0^1 z(t) dJ(t)$ is not continuous and therefore Lemma 3 does not hold.)

REMARK 5. When (6) holds it follows that a necessary and sufficient condition for conservativeness is that $\mathrm{Cov}(J(F(X)), J(F(Y))) \geqslant 0$.

APPENDIX

Under the conditions of Lemma 1, Corollary 3 of Anderson (1955) states that for any $t_1, \cdots, t_k$ in [0, 1] and any convex and symmetric (with respect to the origin) set A in R^k

(7) $$P((Z_1(t_1), \cdots, Z_1(t_k)) \in A) \geqslant P((Z_2(t_1), \cdots, Z_2(t_k)) \in A).$$

An extension of this result to some special infinite-dimensional sets is given in Corollary 4 of Anderson (1955). Lemma 1 of Section 2 is a generalization of (7) to any closed, convex and symmetric set in $C[0, 1]$.

PROOF OF LEMMA 1. Since $C[0, 1]$ is separable and A is closed and convex it follows that $A = \cap_{n=1}^{\infty} C_n$ where C_n are half spaces, i.e., $C_n = \{z \in C[0, 1] : \int_0^1 z(t) dg_n(t) \leqslant \lambda_n\}$ where g_n is a function of bounded variation. Since A is symmetric we also have $A = \cap_{n=1}^{\infty} D_n$ where $D_n = \{z \in C[0, 1] : |\int_0^1 z(t) dg_n(t)| \leqslant \lambda_n\}$. We first show that for $A_k = \cap_{n=1}^{k} D_n$

(8) $$P(Z_1 \in A_k) \geqslant P(Z_2 \in A_k).$$

Consider the Gaussian variables $V_j^{(m)} = \int_0^1 Z_m(t) dg_j(t)$, $m = 1, 2; j = 1, \cdots, k$. We have

(9) $$P(Z_m \in A_k) = P\big(\big(V_1^{(m)}, \cdots, V_k^{(m)}\big) \in B_k\big), \qquad m = 1, 2$$

where $B_k = \Pi_{j=1}^{k}[-\lambda_j, \lambda_j] \subseteq R^k$. For $i, j = 1, \cdots, k$ set

$$\Psi_{ij} = \mathrm{Cov}\big(V_i^{(2)}, V_j^{(2)}\big) - \mathrm{Cov}\big(V_i^{(1)}, V_j^{(1)}\big) = \int_0^1\int_0^1 [c_2(s, t) - c_1(s, t)] dg_i(s) dg_j(t).$$

Since $c_2(s, t) - c_1(s, t)$ is positive semidefinite it follows that the matrix $(\Psi_{ij})_{k \times k}$ is positive semidefinite. By Corollary 3 of Anderson (1955)

$$P\big((V_1^{(1)}, \cdots, V_k^{(1)}) \in B_k\big) \geqslant P\big((V_1^{(2)}, \cdots, V_k^{(2)}) \in B_k\big).$$

Now (9) implies (8). Since the sequence A_k decreases to A we have $P(Z_m \in A_k) \to P(Z_m \in A)$ as $k \to \infty$ and Lemma 1 follows. ▯

REFERENCES

[1] ANDERSON, T. W. (1955). The integral of a symmetric unimodal function. *Proc. Amer. Math. Soc.* **6** 170–176.

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

[3] DUDLEY, R. (1966). Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.* **10** 109–126.

[4] DYKSTRA, R. L., HEWETT, J. E. AND THOMPSON, W. A. (1973). Events which are almost independent. *Ann. Statist.* **1** 674–681.

[5] HOLLANDER, M., PLEDGER, G. AND LIN, P. (1974). Robustness of the Wilcoxon test to a certain dependency between samples. *Ann. Statist.* **2** 177–181.

[6] JOGDEO, K. (1977). Association and probability inequalities. *Ann. Statist.* **5** 495–504.

[7] KARLIN, S. (1968). *Total Positivity.* Stanford Univ. Press.

[8] KIEFER, J. (1959). *K*-sample analogues of the Kolmogorov-Smirnov and Cramer-von Mises tests. *Ann. Math. Statist.* **30** 420–447.

[9] LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37** 1137–1153.

[10] RINOTT, Y. (1976). On convexity of measures. *Ann. Probability* **4** 1020–1026.

[11] SEN, P. K. (1967). Nonparametric tests for multivariate interchangeability, part 1: problems of location and scale in bivariate distributions. *Sankhyá Ser. A* **29** 351–372.

[12] SHAKED, M. (1974). Some concepts of positive dependence for exchangeable bivariate distributions. Technical Report, Depart. Statist., Univ. Rochester.

[13] SHAKED, M. (1977). A concept of positive dependence for exchangeable random variables. *Ann. Statist.* **5** 505–515.

[14] SLEPIAN, D. (1962). The one sided barrier problem for Gaussian noise. *Bell System Tech. J.* **41** 463–501.

DEPARTMENT OF STATISTICS
THE HEBREW UNIVERSITY OF JERUSALEM
JERUSALEM, ISRAEL