The Annals of Probability
1980, Vol. 8, No. 4, 814–827

INEQUALITIES FOR DISTRIBUTIONS WITH GIVEN MARGINALS[1]

BY ANDRÉ H. TCHEN

Carnegie-Mellon University

An ordering on discrete bivariate distributions formalizing the notion of concordance is defined and shown to be equivalent to stochastic ordering of distribution functions with identical marginals. Furthermore, for this ordering, $\int \varphi \, dH$ is shown to be H-monotone for all superadditive functions φ, generalizing earlier results of Hoeffding, Fréchet, Lehmann and others. The usual correlation coefficient, Kendall's τ and Spearman's ρ are shown to be monotone functions of H. That $\int \varphi \, dH$ is H-monotone holds for distributions on $\mathbb{R}^n$ with fixed $(n-1)$-dimensional marginals for any φ with nonnegative finite differences of order n. Some related results are obtained. Stochastic ordering is preserved under certain transformations, e.g., convolutions. A distribution on $\mathbb{R}^\infty$ is constructed, making $\max(X_1, \cdots, X_n)$ stochastically largest for all n when X_i have given one-dimensional distributions, generalizing a result of Robbins. Finally an ordering for doubly stochastic matrices is proposed.

1. Introduction and summary. For probability measures (p.m.) H on $\mathbb{R}^2$ with fixed marginals F and G, a natural notion of stochastic dominance is explored, formalizing the qualitative notion of concordance (Gini [15]), best summarized as "large values of X go with large values of Y" (Kruskal [21]). When F and G concentrate their mass on finitely many atoms, say that H' is *more concordant* than H iff H' can be obtained from H by a finite number of repairings which add mass ε at (x, y) and (x', y') while subtracting mass ε at (x', y) and (x, y') where $x' > x$ and $y' > y$, so that *large values of* X *are more often associated with large values of* Y *under* H′ *than under* H. In Section 2 it is shown that H' is more concordant than H iff H' stochastically dominates H, namely if their cdf's satisfy $H'(x, y) \geqslant H(x, y)$ for all x, y.

From this property it is easy to verify that for any real-valued superadditive function φ, i.e., satisfying $\varphi(x, y) + \varphi(x', y') \geqslant \varphi(x', y) + \varphi(x, y')$ for all $x' > x$ and $y' > y$ (cf. [12]), $\int\varphi \, dH$ is monotone in H. In Section 3 this is shown directly in the continuous, n-dimensional case under certain regularity conditions.

This result generalizes various earlier results which we now briefly review. Lehmann [23] called the distributions satisfying $H(x,y) \geqslant H(x, \infty)H(\infty,y)$ for all x,y positive quadrant dependent (p.q.d.) and showed that Kendall's τ, Spearman's ρ and the usual correlation coefficient are greater than 0—their value under $H(x, \infty)H(\infty,y)$—for p.q.d. distributions. Hoeffding and others [19], [13], [2], [8], [3] (see also [9], [1], [26], [37]) showed that the extreme bounds of $E_H(XY)$,

Received May 14, 1975; revised July 30, 1979.

[1]Prepared under the auspices of Air Force Office of Scientific Research Contract AFOSR #F44620-74-C-0068 and National Science Foundation Research Grant NSF MCS-76-091979.

AMS 1970 *subject classifications.* Primary 62E10; secondary 62H05, 62H20.

Key words and phrases. Distributions with given marginals, correlation, measures of dependence, concordance, stochastic ordering, inequalities, bounds, rearrangements.

$E_H(X-Y)^2$, $E_H|X-Y|^p$, $p \geqslant 1$ and $E_H f(|X-Y|)$ with $f'' \geqslant 0$ are attained at $\underline{H}(x,y) = \max\{0, F(x)+G(y)-1\}$ and $\overline{H}(x,y) = \min\{F(x), G(y)\}$. Hoeffding also characterized the class of p.m.'s H with marginals F and G, denoted $\Gamma(F, G)$, as the class of p.m.'s with cdf's $H(x,y)$ such that $\underline{H}(x,y) \leqslant H(x,y) \leqslant \overline{H}(x,y)$.

That the extreme bounds of $E_H\varphi(X, Y)$ are attained at $\underline{H}$ and $\overline{H}$ for superadditive functions can be derived differently: to maximize $E_H(XY)$ in the discrete case one is led to a rearrangement theorem of Hardy, Littlewood and Polya [18], page 261, via Birkhoff's [5] characterization of permutation matrices as extreme points of the convex set of doubly stochastic matrices. In view of the representation theorem of Skorokhod [33], the continuous version of the rearrangement theorem [18], page 278, shows that for general F and G the bounds of $E_H(XY)$ for $H \in \Gamma(F, G)$ are attained at $\underline{H}$ and $\overline{H}$. Lorentz [25] extended the theorem of Hardy, et al., to all bounded superadditive continuous functions and thus Lorentz [25] essentially contains the results of Hoeffding, et al. Whitt [37] first noted the connection between Hoeffding's and Hardy, Littlewood and Polya's results; Thomas Cover and the author independently rediscovered this connection and Lorentz's result.

It should also be noted that in a paper completed after the first version of this paper and published before this version, Cambanis, et al., [7], have derived the result on the H-monotonicity of $\int\varphi\, dH$ under different regularity conditions.

In $\mathbb{R}^n$ the H-monotonicity of $\int\varphi\, dH$ holds for any real-valued function φ with nonnegative finite differences of order n when H varies in a class of p.m.'s with identical $(n-1)$-dimensional marginals. The applicability of this result is extended by noting that a stronger form of stochastic ordering is preserved when taking product probabilities. Some results of Lehmann [23] can thus be generalized and it is also shown that stochastic ordering is preserved under convolutions: if H and H' have identical $(n-1)$-dimensional marginals, and similarly for K and K', with $H' \geqslant H$ and $K' \geqslant K$, then $H' * K' \geqslant H * K$, where $H * K$ denotes the cdf of the convolution of H and K. Section 4 contains these and other extensions and applications.

Let $\overline{H}(x) = \min_i F_i(x_i)$; then for any real-valued function φ on $\mathbb{R}^n$ which is only superadditive in each pair of arguments, $\int\varphi\, d\overline{H}$ attains the maximum of $\int\varphi\, dH$ over the class of p.m.'s H with one-dimensional marginals F_i, with $1 \leqslant i \leqslant n$; this was shown for continuous and bounded functions φ by Lorentz [25] in the context of rearrangements of functions, and was recently rediscovered by Schaeffer [31] in the special case where $\varphi(x) = \min x_i - \max x_i$. A simpler proof using a martingale representation and requiring weaker regularity conditions is presented.

The notion of least concordance is used to construct explicitly a p.m. on $\mathbb{R}^\infty$ which makes $\max(X_1, \cdots, X_n)$ stochastically largest for all n, where the X_i's have specified one-dimensional distributions but are otherwise arbitrarily jointly distributed; this extends the result of Robbins [30], [22], obtained for the identically distributed case.

Finally an ordering for doubly stochastic matrices is proposed (cf. Sherman [32]), and a counterexample to a conjecture of Gray, Neuhoff and Shields [16] concerning distances for stationary processes is constructed.

Throughout the rest of this paper, $H(x)$ denotes the right-continuous cdf of the p.m. H, $H^-(x)$ its left-continuous version, and $H\{A\}$ the probability of A under H; for ease of notations $H\{(x_1, \cdots, x_n)\}$ is abbreviated as $H\{x_1, \cdots, x_n\}$.

2. Stochastic ordering as greater concordance. Consider the class of p.m.'s in $\Gamma(F, G)$, where F and G concentrate their mass at the points $x_1 < \cdots < x_I$, $y_1 < \cdots < y_J$, respectively. Suppose that H and H' are in $\Gamma(F, G)$ and that there exists a sequence $H_1 \equiv H, H_2, \cdots, H_n \equiv H'$ of p.m.'s in $\Gamma(F, G)$ such that for $i = 1, 2, \cdots, n-1$, H_{i+1} is obtained from H_i by adding mass $\varepsilon_i > 0$ at some points (x_p, y_q) and $(x_{p'}, y_{q'})$ while subtracting mass ε_i at the points $(x_p, y_{q'})$ and $(x_{p'}, y_q)$, where $p < p'$ and $q < q'$. Then H' is said to be *more concordant* than H, and we write $H' \geqslant_c H$.

THEOREM 1. *Let F and G be two discrete* p.m.'s *on* $\mathbb{R}$ *with finitely many atoms. Then for $H, H' \in \Gamma(F, G)$, $H' \geqslant_c H$ if and only if $H'(x, y) \geqslant H(x, y)$ for all x, y.*

We first introduce two lemmas and some notation. For convenience, replace $x_1, \cdots, x_I$ and $y_1, \cdots, y_J$ by $1, \cdots, I$ and $1, \cdots, J$ respectively. Define the usual lexicographic ordering on $\{1, \cdots, I\} \times \{1, \cdots, J\}$ by $(k, l) < (m, n)$ if $l < n$, or if $l = n$ and $k < m$. For fixed i, j, $H' \in \Gamma(F, G)$ and for any $K \in \Gamma(F, G)$ satisfying condition

(A) $K\{i, j\} < H'\{i, j\}$ and $K(x, y) = H'(x, y)$ for all (x, y) preceding (i, j),

define the sequence $S(K, H') = (x_l, y_l)_{l=1,\cdots,k}$ by

$$\begin{aligned} y_1 &= j \\ x_1 &= \min\{r \mid r \geqslant i+1,\ K\{r, j\} > 0\} \\ y_n &= \min\{s \mid s \geqslant j+1,\ K\{r, s\} \geqslant 0 \text{ for some } r \text{ with } i \leqslant r < x_{n-1}\} \\ x_n &= \min\{r \mid i \leqslant r,\ K\{r, y_n\} > 0\}, \end{aligned}$$

These relations define a nonempty finite sequence with $x_k = i < x_{k-1} \cdots < x_1, y_1 = j < y_2 \cdots < y_k$, and $y_k = \min\{s \mid K\{i,s\} > 0, s > j\}$. Define the partial order $\gg$ on the sequences $(x_l,y_l)_{l=1,\cdots p}$ with $x_p < x_{p-1} \cdots < x_1$ and $y_1 < y_2 \cdots < y_p$ by $(x_l,y_l)_{l=1,\cdots p} \gg (u_l,v_l)_{l=1,\cdots q}$ iff $x_p = u_q = i, y_1 = v_1 = j$ and $\cup_{l=1}^{p-1}([i,x_l) \times [j,y_{l+1})) \supset \cup_{l=1}^{q-1}([i,u_l) \times [j,v_{l+1}))$. Also say that $(x_l,y_l)_{l=1,\cdots p} \gg (u_l,v_l)_{l=1,\cdots q}$ strictly if the corresponding set inclusion is strict.

LEMMA 1.1. *Let $K \in \Gamma(F, G)$ satisfy condition* (A); *then there exists $K' \in \Gamma(F, G)$ such that $K' \geqslant_c K$, $K'(x,y) \leqslant H'(x,y)$ for all x, y, and $K'\{i,j\} = K\{i,j\} + \Delta$, where*

$$\Delta = \min\{H'\{i,j\} - K\{i,j\}, \min_{S(K,H')} K\{x_l,y_l\}\}.$$

Further $S(K',H') \gg S(K,H)$ strictly if $\Delta < H'\{i,j\} - K\{i,j\}$.

PROOF. Let $S(K,H') = (x_l,y_l)_{l=1,\cdots k}$ as above, and define $K^{(1)} \equiv K$, $K^{(l)}$ for $l = 2, 3, \cdots k$ as the measure obtained from $K^{(l-1)}$ where the mass Δ is subtracted from (x_{l-1},j) and (x_l,y_l) and added to (x_{l-1},y_l) and (x_l,j), and let $K' \equiv K^{(k)}$.

By definition of Δ, each $K^{(l)}$ is a positive measure; further, $K'\{i,j\} = K^{(k)}\{i,j\} = K\{i,j\} + \Delta$. To show that $K^{(l)}(x,y) \leqslant H'(x,y)$ for all (x,y) for each $l, k \leqslant l \leqslant 1$, assume by recurrence that $K^{(l-1)} \leqslant H'$. If (x,y) is not in $(x_{l-1}, x_l) \times (j, y_l)$, $K^{(l)}(x,y) = K^{(l-1)}(x,y)$; otherwise define $A = [i,x] \times [0,j-1]$, $B = [0,i-1] \times [0,y]$ and $C = [i,x] \times [j,y]$. By hypothesis $K^{(l)}\{A\} = K^{(l-1)}\{A\} = H'\{A\}$, and by construction and recurrence assumption $K^{(l)}\{B\} = K^{(l-1)}\{B\} \leqslant H'\{B\}$. By construction $K^{(l)}\{C\} = K\{i,j\} + \Delta \leqslant H'\{A\} + H'\{i,j\}$. Thus $K^{(l)}(x,y) = K^{(l)}\{A \cup B \cup C\} \leqslant H'\{A\}+H'\{B\}+H'\{C\} \leqslant H'(x,y)$. Finally if $\Delta < H'\{i,j\} - K\{i,j\}$, $\Delta = K\{x_p,y_p\}$ for some p, $1 \leqslant p \leqslant k$; then $S(K',H')$ contains all the points of $S(K, H')$ except (x_p,y_p) so that $S(K',H') \gg S(K,H)$ strictly.

LEMMA 1.2. *Let $H_k \in \Gamma(F, G)$ be such that $H_k \geqslant_c H$, $H_k(x, y) \leqslant H'(x, y)$ for all (x, y) and $H_k(x, y) = H'(x, y)$ for all (x, y) strictly preceding (i, j). Then there exists $H_{k+1} \geqslant_c H_k$ such that $H_{k+1} \leqslant H'(x, y)$ for all (x, y), $H_{k+1}(x, y) = H'(x, y)$ for all (x, y) strictly preceding (i, j), and $H_{k+1}(i, j) = H'(i, j)$.*

PROOF. Apply Lemma 1.1 repeatedly starting with $K = H$; the sequence K_m obtained must eventually satisfy $K_m\{i, j\} = H'\{i, j\}$ since $S(K_m, H')$ is a strictly increasing sequence with respect to $\gg$ on a finite range, of size less than 2^{IJ}.

PROOF OF THEOREM 1. It is easy to show that if $H' \geqslant_c H$, then $H'(x, y) \geqslant H(x, y)$ for all (x, y). For the converse, an explicit algorithm is provided by repeatedly applying Lemma 1.2: first with $H_0 = H$ and $(i, j) = (1, 1)$ to obtain H_1; then with H_1 and $(i, j) = (2, 1)$ to obtain H_2, etc.

COROLLARY. (Hoeffding). *$\Gamma(F, G)$ admits a maximum $\overline{H}$ and a minimum $\underline{H}$ relatively to $\geqslant_c$, corresponding to $\overline{H}(x, y) = \min\{F(x), G(y)\}$ and $\underline{H}(x, y) = \max\{0, F(x) + G(y) - 1\}$.*

PROOF. For any $H \in \Gamma(F, G)$, $\max\{0, F(x) + G(y) - 1\} \leqslant H(x, y) \leqslant \min\{F(x), G(y)\}$ for all (x, y). The corollary then follows from Theorem 1 since these two bounds are actually cdf's namely the cdf's of $(F^{-1}(U), G^{-1}(1 - U))$ and $(F^{-1}(U), G^{-1}(U))$ respectively, where U is uniformly distributed in $[0, 1]$ and F^{-1}, G^{-1} are the inverses of F and G.

3. Isotonicity properties. Call a function $\varphi : \mathbb{R}^n \to \mathbb{R}$ *n-positive* if it has non-negative finite differences of order n, namely if $\Delta^n_{x_n'x_n} \cdots \Delta^1_{x_1'x_1}\varphi \geqslant 0$ for all $x_1' \geqslant x_i$, $i = 1, \cdots, n$, where $\Delta^p_{x_p'x_p}$ is the finite difference operator

$$\Delta^p_{x_p'x_p}\varphi = \varphi(x_1, \cdots x_{p-1}, x_p', x_{p+1}, \cdots x_n) - \varphi(x_1, \cdots x_{p-1}, x_p, x_{p+1}, \cdots x_n).$$

A 1-positive function is a nondecreasing function; a 2-positive function is also called superadditive. Some examples of n-positive functions are: (1) $x \to$

$-\max(x_1, \cdots x_n)$; (2) $x \to K(x)$ where K is a cdf; (3) $x \to \varphi(x)$ where φ has nonnegative nth order partial derivative $\varphi_{x_1 \cdots x_n}$; (4) $(x_1, x_2) \to x_1 \cdot x_2$; (5) $(x_1, x_2) \to -|x_1 - x_2|^p$ for $p \geq 1$; (6) $(x_1, x_2) \to h(x_1 - x_2)$ where h is a concave function; (7) $(x_1, x_2) \to \psi[A(x_1), B(x_2)]$ where ψ is superadditive and A, B are monotone in the same direction.

Note that for $n = 2$, a consequence of Theorem 1, in the discrete, finite case, is that $\int \varphi\, dH$ is monotone with respect to $H \in \Gamma(H(x, \infty), H(\infty, y))$ if and only if φ is superadditive.

An n-positive function φ defines a positive, additive measure K on the semialgebra of rectangles in $\mathbb{R}^n$ by $K\{\Pi_1^n(x_i, x_i']\} = \Delta^n_{x_n' x_n} \cdots \Delta^1_{x_1' x_1}\varphi$; taking φ right-continuous ensures that K be σ-additive on $\mathfrak{B}(\mathbb{R}^n)$, cf. Neveu [28], page 29.

THEOREM 2. *Let φ be a bounded right-continuous n-positive function, H_1 and H_2 two* p.m.'s *on $\mathbb{R}^n$ having identical $(n-1)$-dimensional marginals. Then $\int \varphi\, dH_2 - \int \varphi\, dH_1 = (-1)^n \int [H_2^-(x) - H_1^-(x)]\, dK$ where K is the positive measure associated with φ.*

PROOF. Assume first that φ is k-positive in each subset of k variables so that it can be regarded as a cdf. Using Skorokhod's construction [33] let $(X_{i1}, \cdots, X_{in})$ be distributed as H_i, for $i = 1, 2$ respectively, and $(F_1, \cdots, F_n)$ be distributed as φ and independent of $(X_{i1}, \cdots, X_{in})$, for $i = 1, 2$. From the inclusion-exclusion formula,

$$\begin{aligned}\int \varphi\, dH_i &= P\{\cap_1^n (F_j \leq X_{ij})\} \\ &= 1 + (-1)^n P\{\cap_1^n (X_{ij} < F_j)\} \\ &\quad + (-1)^{n-1} \Sigma_{k=1}^n P\{\cap_{j \neq k}(X_{ij} < F_j)\} + \cdots \\ &= 1 + (-1)^n \int H_i^-(x)\, dK + (-1)^{n-1}[EH_i^-(\infty, F_2, \ldots F_n) \\ &\quad + EH_i^-(F_1, \infty, F_3, \cdots F_n) + \cdots \\ &\quad + EH_i^-(F_1, \cdots F_{n-1}, \infty)] + \cdots \qquad \text{for } i = 1, 2.\end{aligned}$$

This case then follows by taking differences.

If $\varphi(x)$ is n-positive only, let $\Delta(B, n)$ be the finite difference operator $\Delta(B, n)\varphi = \Delta^n_{x_n B} \cdots \Delta^1_{x_1 B}\varphi$. Since $K\{\Pi_1^n(B \wedge x_j, B \vee x_j]\} = |\Delta(B, n)\varphi(x_1, \cdots x_n)| \leq 2^n \sup|\varphi|$, where $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$, K is a bounded measure. Let $\psi(x) = K\{\Pi_1^n(-\infty, x_j]\} = \lim_{B \to -\infty} \Delta(B, n)\varphi(x)$. From the first part of the proof,

$$\int \psi\, dH_2 - \int \psi\, dH_1 = (-1)^n \int [H_2^-(x) - H_1^-(x)]\, dK.$$

Theorem 2 follows by dominated convergence, noting that φ and ψ have identical finite differences of order n.

The next corollaries give some regularity conditions to relax the boundedness condition, cf. also [7] for $n = 2$.

COROLLARY 2.1. *Let H_1 and H_2 have identical $(n-1)$-dimensional marginals and φ be a right-continuous n-positive function with associated measure K. Then*

$$\int \varphi\, dH_2 - \int \varphi\, dH_1 = (-1)^n \int [H_2^-(x) - H_1^-(x)]\, dK$$

if $H_2^-(x) - H_1^-(x)$ is nonnegative or K-integrable and either

(a) *there exists an H_1 and H_2-integrable function g bounded on compact sets, such that $|\varphi(x^B)| \leqslant g(x)$ for all B, where $x^B = (x_1^B, \cdots, x_n^B)$ and $x_j^B = x_j$ if $|x_j| < B$ and $x_j^B = B\,\mathrm{sgn}(x_j)$ otherwise, or*

(b) *φ is H_1 and H_2-integrable and such that $\sup_{\partial QB}|\varphi| \cdot F_j\{(-B, B)^C\} \to 0$ as $B \to \infty$ for $j = 1, \cdots, n$, where ∂QB is the boundary of the cube $QB = \prod_1^n(-B, B]$ and F_j are the one-dimensional marginals of H_i.*

PROOF. If (a) holds, note that $\varphi_B(x) = \varphi(x^B)$ is a bounded right-continuous n-positive function whose associated measure is the restriction to $\prod_1^n[-B, B]$ of the measure K associated with φ; the conclusion follows by dominated convergence. For (b), let $\Pi(j, B)$ be the operator on measures on $\mathbb{R}^n$ which projects the mass of the points x with $x_j \geqslant B$ onto the hyperplane $x_j = B$ and that of x with $x_j \leqslant -B$ onto the hyperplane $x_j = -B$. Then $H_{iB} = \Pi(1, B) \cdots \Pi(n, B) H_i$, $i = 1, 2$, will have the same $(n-1)$-dimensional marginals and will coincide with H_i in QB. More precisely

$$\int \varphi\, dH_i = \int \varphi\, dH_{iB} - \int_{\partial QB} \varphi\, dH_{iB} + \int_{QB^C} \varphi\, dH_i \qquad \text{for } i = 1, 2.$$

The conclusion follows by noting that

$$\left|\int_{\partial QB} \varphi\, dH_{iB}\right| \leqslant \sup_{\partial QB}|\varphi| H_i\{QB^C\} \leqslant n \sup_{\partial QB}|\varphi| \sup_j F_j\{(-B, B)^C\}.$$

Concerning the extrema attained at Hoeffding's $\underline{H}$ and $\overline{H}$ in the case $n = 2$ we have

COROLLARY 2.2. *Let $\varphi(x)$ be a continuous superadditive function; then for $H \in \Gamma(F_1, F_2)$*

(a) *$\sup \int \varphi\, dH = \int \varphi\, d\overline{H}$ if there exists a continuous function h with $h(x) \geqslant \varphi(x)$ for all x with $\int h\, dH$ finite and constant for all $H \in \Gamma(F_1, F_2)$. A similar result holds for the* inf.

(b) *$\int \varphi\, d\underline{H} \leqslant \int \varphi\, dH \leqslant \int \varphi\, d\overline{H}$ if $\varphi(X)$ are uniformly integrable for all X with distribution in $\Gamma(F_1, F_2)$.*

PROOF. We proceed by rearranging any H locally to approximate $\overline{H}$ inside $\Gamma(F_1, F_2)$. Let $QB = (-B, B)^2$ and let H_B be the measure which is H outside QB and H reordered inside $QB: H_B\{A\} = H\{A \cap QB^C\} + \overline{H}_B\{A \cap QB\}$ for all $A \subset \mathbb{R}^2$, where $\overline{H}_B$ is the maximal subprobability sub-cdf

$$\overline{H}_B(x) = \min\{H\{(-B, x_1] \times (-B, B]\}, H\{(-B, B] \times (-B, x_2]\}\}$$

for x in QB. Clearly $H_B \in \Gamma(F_1, F_2)$ and H_B converges weakly to $\overline{H}$ as $B \to \infty$. Note that $\int \varphi\, dH_B \geqslant \int \varphi\, dH$ for all B so that (b) holds. To prove (a), assume $\int \varphi\, dH$

finite or else the theorem is trivial. Since $h - \varphi \geqslant 0, 0 \leqslant \int(h - \varphi)\,dH \leqslant \lim \int(h - \varphi)\,dH_B$; it is then easy to see that $\int \varphi\, dH \leqslant \limsup \int \varphi\, dH_B \leqslant \int \varphi\, d\overline{H}$.

COROLLARY 2.3. *For any convex nonnegative function* ψ *defined on* $\mathbb{R}$, $\int \psi(x_1 - x_2)\,d\overline{H} \leqslant \int \psi(x_1 - x_2)\,dH \leqslant \int \psi(x_1 - x_2)\,dH$ *for all* $H \in \Gamma(F_1, F_2)$.

PROOF. The inf case follows from Corollary 2.2. For any $Z(u)$ defined on $[0, 1]$ with Lebesgue measure, let $\overline{Z}(u) = K^{-1}(u)$ and $\underline{Z}(u) = K^{-1}(1 - u)$, where K is the cdf of Z and $K^{-1}(t)$ is the left-continuous inverse of K. Given $H \in \Gamma(F_i, i = 1, 2)$ let (X, Y) be distributed as H, cf. [33]. Then $(\overline{X}, \overline{Y})$ and $(\overline{X}, \underline{Y})$ are distributed as $\overline{H}$ and $\underline{H}$ respectively. Let $C = \sup\{s \mid F_1(s) \leqslant 1 - F_2(s)\}$. For $B > |C|$ let $X_B(u) = X(u)$ if $|X| < B$ and $X_B(u) = C$ otherwise. The cdf F_{1B} of X_B is parallel to that of X in $(-B, B)$ and similarly for Y_B so that by the convexity of ψ, $|\psi(\overline{X}_B - \underline{Y}_B)| \leqslant |\psi(0)| + |\psi(\overline{X} - \underline{Y})|$ for all B. Note that $(\overline{X}_B - \underline{Y}_B)$ converges almost surely to $(\overline{X} - \underline{Y})$ since for nondecreasing functions, convergence at all points of continuity is equivalent to convergence in the Lévy metric. By Fatou's lemma and Theorem 2 applied to $\Gamma(F_{1B}, F_{2B})$, $E\psi(X - Y) \leqslant \liminf E\psi(X_B - Y_B) \leqslant \liminf E\psi(\overline{X}_B - \underline{Y}_B) = E\psi(\overline{X} - \underline{Y})$.

In the following examples, H, H_1 and H_2 are in $\Gamma(F_1, F_2)$ with $H_1(x,y) \leqslant H_2(x,y)$ for all x,y.

EXAMPLE 1. For any nondecreasing right-continuous F_2-integrable function $f(\cdot)$,

$$\int_{-\infty}^{t-0} E_{H_1}[f(Y) \mid X = x]\,dF_1(x) \geqslant \int_{-\infty}^{t-0} E_{H_2}[f(Y) \mid X = x]\,dF_1(x) \qquad \text{for all } t.$$

This is obtained for $\varphi(x,y) = I_{(-\infty,t)}(x) \cdot f(y)$ in Corollary 2.1. In particular if $F_1 = F_2$ is the uniform distribution on $[0, 1]$ the regression function $m(x) = E_H[Y \mid X = x]$ must satisfy $\int_0^u m(x)\,dx \geqslant u^2/2$ for all u (cf. Vitale and Pipkin [36]).

EXAMPLE 2. (cf. Bass [2], Dall'Aglio [8], Vallender [35]).

$$E_{\overline{H}}|X - Y|^p \leqslant E_{H_2}|X - Y|^p \leqslant E_{H_1}|X - Y|^p \leqslant E_{\underline{H}}|X - Y|^p \quad \text{for } p \geqslant 1.$$

Since $f_B(x,y) = |x^B - y^B|^p$ increases to $|x - y|^p$ as $B \uparrow \infty$, this follows from Theorem 2 and the monotone convergence theorem. For instance, if X and Y have symmetric distributions of the same location-scale family, with means and variances $m_i, \sigma_i^2, i = 1, 2$,

$$(m_1 - m_2)^2 + (\sigma_1 - \sigma_2)^2 \leqslant E_H(X - Y)^2 \leqslant (m_1 - m_2)^2 + (\sigma_1 + \sigma_2)^2.$$

EXAMPLE 3. (cf. Hardy, Littlewood and Polya [18], page 278). If $\int xy\,d\underline{H}$ and $\int xy\,d\overline{H}$ exist finitely, then $\int xy\,d\underline{H} \leqslant \int xy\,dH_1 \leqslant \int xy\,dH_2 \leqslant \int xy\,d\overline{H}$. By Corollary 2.3, writing $|xy| = \exp[\log|x| - \log|1/y|]$, all $\int xy\,dH$ are finite. Let $f_B(x,y) = x^B \cdot y^B$; f_B is superadditive and bounded, so the result follows from Theorem 2 and the dominated convergence theorem.

EXAMPLE 4. (cf. Day [10], page 941). If φ is superadditive, continuous and such that $|\varphi(x, y)| \leqslant k(1 + |x| + |y|)^p$, then

$$-\infty \leqslant \int\varphi \, d\underline{H} \leqslant \int\varphi \, dH_1 \leqslant \int\varphi \, dH_2 \leqslant \int\varphi \, d\overline{H} < \infty.$$

This follows from Corollary 2.1 since $(1 + |x| + |y|)^p \leqslant 3^{p-1}(1 + |x|^p + |y|^p)$.

EXAMPLE 5. (cf. Rinott [29]). Let $\varphi(x, y) = \log P_m(x, y)$ where $P_m(x, y) = \log(x^m y^m + a_1 x^{m-1} y^{m-1} + \cdots + a_m)$, with $a_i \geqslant 0$. If $X \geqslant 0$, $Y \geqslant 0$ and if $\log X$ and $\log Y$ are integrable, then

$$-\infty < E_{\underline{H}}\varphi(X, Y) \leqslant E_{H_1}(X, Y) \leqslant EH_2\varphi(X, Y) \leqslant E_{\overline{H}}\varphi(X, Y) < \infty.$$

Since P_m is TP_2, (cf. [20], page 101), φ is superadditive; further, since $c \log(xy) \leqslant \log P_m(x, y) \leqslant c' \log(x^m y^m)$ for some constants c and c', it follows that $|\log P_m(x^B, y^B)| \leqslant m(|c| + |c'|)(|\log x| + |\log y|)$ so that Corollary 2.1 applies.

EXAMPLE 6. (cf. Robbins [30], [22]). $E_{\overline{H}}[\max(X, Y)]^p \leqslant E_{H_2}[\max(X, Y)]^p \leqslant E_{H_1}[\max(X, Y)]^p \leqslant E_{\underline{H}}[\max(X, Y)]^p$, when $p \geqslant 1$ is odd and $X, Y \in L^p$ or when $p \geqslant 2$ is even and X, Y are nonnegative. Under these conditions $-[\max(x, y)]^p$ is superadditive; for an n-dimensional extension concerning the extreme bounds, see Section 4.3.

4. Some extensions and applications.

4.1. *Invariance of stochastic ordering under transformations.* Note that the condition $\int\varphi \, dH' \geqslant \int\varphi \, dH$ for all continuous bounded k-positive functions $\varphi : \mathbb{R}^k \to \mathbb{R}$ is equivalent to the condition that H and H' have identical $(k-1)$-dimensional marginals with $H'(x) \geqslant H(x)$ for all $x \in \mathbb{R}$ when k is even and $H'(x) \leqslant H(x)$ for all $x \in \mathbb{R}^k$ when k is odd. The direct part is verified using a continuous approximation to $\prod_1^k I_{(-\infty, a_i)}$ and the converse part follows from Theorem 2.

THEOREM 3. *Let H_{i1}, H_{i2} be* p.m.'s *on $\mathbb{R}^{k_i}$ for $i = 1, \cdots, n$ and let H_j denote the product measure $H_{1j} \times \cdots \times H_{nj}$ for $j = 1, 2$. Suppose $\int\varphi_i \, dH_{i1} \leqslant \int\varphi_i \, dH_{i2}$ for each $i = 1, \cdots, n$, for all continuous bounded k_i-positive functions $\varphi_i : \mathbb{R}^{k_i} \to \mathbb{R}$; then $\int\varphi \, dH_1 \leqslant \int\varphi \, dH_2$ for all continuous bounded functions $\varphi : \mathbb{R}^{k_1 + \cdots + k_n} \to \mathbb{R}$ which are k_i-positive in each block of k_i coordinates separately.*

PROOF. Since k_i-positivity in block i is preserved when integrating φ with respect to other coordinates, it suffices to note that the integrals of φ with respect to the sequence of measures $(H_1, H_{11} \times H_{21} \cdots \times H_{n-1,1} \times H_{n2}, H_{11} \times H_{21} \cdots \times H_{n-2,1} \times H_{n-1,2} \times H_{n2}, \cdots, H_2 = H_{12} \times H_{22} \cdots \times H_{n2})$ form an increasing sequence.

A class of blockwise k_i-positive functions is as follows: let $r_1, \cdots, r_q$, with $q \leqslant \min_i k_i$, be functions mapping $\mathbb{R}^{k_1 + k_2 \cdots + k_n}$ into $\mathbb{R}$ such that for each $j = 1, 2, \cdots, q$, r_j is a function of at most one variable from each block only. If each r_j is nondecreasing in each coordinate, then for any p-positive $\psi : \mathbb{R}^p \to \mathbb{R}$ with $p \leqslant q$, $\psi(r_1, \cdots, r_p)$ is separately blockwise k_i-positive.

Lehmann [23] called a probability distribution H on $\mathbb{R}^2$ positive quadrant dependent (p.q.d) if $H(x,y) \geqslant H(x,\infty)\cdot H(\infty\cdot y)$ for all x,y, and showed that if $r_1, r_2 : \mathbb{R}^n \to \mathbb{R}$ are monotone in the same direction separately in their kth coordinates for each $k = 1,\cdots,n$, then $(r_1(X_1,\cdots,X_n), r_2(Y_1,\cdots,Y_n))$ has a p.q.d. distribution when (X_i, Y_i), for $i = 1,\cdots,n$, are independent with p.q.d. distributions. An easily verified generalization is given in

COROLLARY 3.1. *Suppose* $H_{i1}, H_{i2}, i = 1,\cdots,n$ *are* p.m.'s *on* $\mathbb{R}^q$ *satisfying* $\int\varphi\, dH_{i1} \leqslant \int\varphi\, dH_{i2}$ *for all bounded continuous q-positive functions* φ; *for* $j = 1,\cdots,q$ *let* $r_j : \mathbb{R}^n \to \mathbb{R}$ *be monotone in each coordinate with the number of* r_j's *decreasing in their* pth *coordinate an even number, for each p. Then* $\int\varphi(r)\, dH_1 \leqslant \int\varphi(r)\, dH_2$ *for all bounded continuous q-positive functions* φ, *where* H_j *is the product measure* $H_{1j} \times\cdots\times H_{nj}$, *for* $j = 1, 2$.

EXAMPLE 1. Stochastic ordering is preserved under convolutions. Let the p.m.'s H_{i1} and H_{i2} on $\mathbb{R}^q$ have identical $(q-1)$-dimensional marginals, for each $i = 1,\cdots n$, with $H_{i1}(x) \leqslant H_{i2}(x)$ for all $x \in \mathbb{R}^q$ and $i = 1,\cdots,n$. Then $H_{11}*H_{21}\cdots *H_{n1}(x) \leqslant H_{12}*H_{22}\cdots *H_{n2}(x)$ for all $x \in \mathbb{R}^q$, where $*$ denotes the convolution operation.

Suppose Z, Z_1, Z_2 have the same distribution and are independent of (U, V). For any $a > 0$, $b > 0$ (resp. $b < 0$), the distribution of (aZ, bZ) is maximal so that the cdf of $(U + aZ, V + bZ)$ dominates that of $(U + aZ_1, V + bZ_2)$. Choosing (Z_1, Z_2) bivariate normal, U normal and $V \equiv 0$ yields the following special case of a result of Slepian [34], cf. [17], page 805.

EXAMPLE 2. (Slepian). Let H_{ρ_j} for $j = 1, 2$ be standardized bivariate normal distributions with $\rho_1 \leqslant \rho_2$; then $H_{\rho_1}(x, y) \leqslant H_{\rho_2}(x, y)$ for all x, y.

The following is obtained using a continuous approximation to $r_1(u, v) = r_2(u, v) = \operatorname{sgn}(v - u)$ and generalizes Corollary 1 of Lehmann [23].

COROLLARY 3.2. *Kendall's* τ, *Spearman's* ρ *and the quadrant measure q of Blomquist are monotone functions of H: if* H_1 *and* H_2 *have the same marginals and* $H_1(x, y) \leqslant H_2(x, y)$ *for all x, y, then* $\tau_{H_1} \leqslant \tau_{H_2}$, $\rho_{H_1} \leqslant \rho_{H_2}$ *and* $q_{H_1} \leqslant q_{H_2}$.

4.2 *The maximal distribution of* $max(X_1,\cdots, X_n)$. This section extends to the nonidentically distributed case a result of Robbins [30], cf. also Lai and Robbins [22]. Let F_i, $i \in \mathbb{N}$ be a sequence of p.m.'s on $\mathbb{R}$; then for any sequence of random variables X_i, $i \in \mathbb{N}$ such that X_i is distributed as F_i for each $i \in \mathbb{N}$, the following algebraic inequality holds:

$$\max\{0, 1 - \Sigma_1^n[1 - F_i(x_i)]\} \leqslant P[X_i \leqslant x_i, i = 1,\cdots, n] \leqslant \min_{1\leqslant i\leqslant n} F_i(x_i)$$

for all $x_i \in \mathbb{R}$ and all n.

The upper bound is attained at $X_i = F_i^{-1}(U)$, $i \in \mathbb{N}$, where U is uniformly distributed on [0, 1]. The lower bound is not in general a cdf for $n \geqslant 3$, as shown by the following counter-example: if $F_i(x) = xI_{[0,1]}(x) + I_{(1,\infty)}(x)$ for $i = 1, 2, 3$, then

$$\Delta^3_{1,.5}\Delta^2_{1,.5}\Delta^1_{1,.5}\max\{0, 1 - \Sigma_1^3[1 - F_i(x_i)]\} = -.5 < 0.$$

Theorem 4 states that there exists a fixed distribution on $\mathbb{R}^\infty$ achieving the lower bound when all x_i's are equal; it makes $\max_{1 \leqslant i \leqslant n} X_i$ stochastically largest, uniformly for all n.

THEOREM 4. *Given a sequence $F_i, i \in \mathbb{N}$ of distributions on $\mathbb{R}$ there exists a sequence of random variables X_i, $i \in \mathbb{N}$ achieving $P[\max_{1 \leqslant i \leqslant n} X_i \leqslant x] = \max\{0, 1 - \Sigma_1^n[1 - F_i(x)]\}$ for all $x \in \mathbb{R}$ and all $n \in \mathbb{N}$.*

PROOF. Assume first that all F_i, $i \in \mathbb{N}$ are atomless. Let $X_1 = F^{-1}(U)$ and set $M_1 = X_1$, $K_1 = F_1$. Define recursively

$$X_n = F_n^{-1}[1 - K_{n-1}(M_{n-1})], \qquad M_n = \max_{1 \leqslant i \leqslant n} X_i = \max(X_n, M_{n-1}),$$

where K_n is the distribution of M_n. Since the cdf of (X_n, M_{n-1}) is minimal in its Hoeffding class, assuming by induction that M_{n-1} achieves the lower bound, $P[\max_{1 \leqslant i \leqslant n} X_i \leqslant x] = \max\{0, K_{n-1}(x) + F_n(x) - 1\} = \max\{0, \Sigma_1^n[1 - F_i(x)]\}$. In the general case, for each i, $k \in \mathbb{N}$, let F_i^k be the convolution of F_i with a normal distribution with mean 0 and variance k^{-1}.

Construct X_i^k, $i \in \mathbb{N}$ as above and let P_k denote its distribution on $\mathbb{R}^\infty$. By a simple extension to the countably infinite case of Billingsley [4], page 41, Example 6, P_k, $k \in \mathbb{N}$ is a tight family so that there exists a subsequence k' and a measure P_∞ such that $P_{k'}$ converges weakly to P_∞. Let $Y_i^{k'}$, $i \in \mathbb{N}$ and Y_i^∞, $i \in \mathbb{N}$ be the processes defined on [0, 1], with distributions $P_{k'}$ and P_∞, obtained from Skorokhod's construction [33] so that $Y_i^{k'} \to Y_i^\infty$ a.s. for each $i \in \mathbb{N}$. Thus $\max\{Y_1^{k'}, \cdots Y_n^{k'}\}$ converges to $\max\{Y_1^\infty, \cdots Y_n^\infty\}$ a.s., and its cdf converges for all x in D_n, the set of points of continuity of the cdf of the limit:

$$P[\max_{1 \leqslant i \leqslant n} Y_i^\infty \leqslant x] = \lim_{k'} P[\max_{1 \leqslant i \leqslant n} Y_i^{k'} \leqslant x] \qquad \text{for all} \quad x \in D_n.$$

On the other hand, for all $x \in \cap_1^n C_i$, where C_i is the set of points of continuity of F_i,

$$\lim_{k'} P[\max_{1 \leqslant i \leqslant n} Y_i^{k'} \leqslant x] = \lim_{k'} \max\{0, 1 - \Sigma_1^n[1 - F_i^{k'}(x)]\}$$
$$= \max\{0, 1 - \Sigma_1^n[1 - F_i(x)]\}.$$

The following corollary results from the identity

$$EM^p = \int_0^\infty[P(M > t) + (-1)^p P(M < -t)]pt^{p-1}\,dt.$$

COROLLARY 4.1. *Let X_i have distribution F_i for each i, $i \in \mathbb{N}$, and let $M_n = \max_{1 \leqslant i \leqslant n} X_i$, $m_n = \min_{1 \leqslant i \leqslant n} X_i$, $a_n = \inf\{t | \Sigma_1^n[1 - F_i(t)] \leqslant 1\}$ and $b_n = \inf\{t | \Sigma_1^n F_i(t) \geqslant 1\}$. Then, for all odd integers p, or all integers p if all X_i are nonnegative, for all $n \in \mathbb{N}$ and all $x \in \mathbb{R}$*

(a) $1 - \min_{1 \leqslant i \leqslant n} F_i(x) \leqslant P[M_n > X] \leqslant \min\{1, \Sigma_1^n[1 - F_i(x)]\}$

(b) $\max_{1 \leqslant i \leqslant n} F_i(x) \leqslant P[m_n \leqslant x] \leqslant \min\{1, \Sigma_1^n F_i(x)\}$

(c) $\sup EM_n^p = a_n^p + \int_{a_n}^\infty pt^{p-1}[n - \Sigma_1^n F_i(t)]\, dt$

$\inf Em_n^p = \int_0^\infty pt^{p-1}[1 - \min_i F_i(t)]\, dt - \int_0^\infty pt^{p-1} \min_i F_i(-t)\, dt$

(d) $\sup Em_n^p = \int_0^\infty pt^{p-1}[1 - \max_i F_i(t)]\, dt - \int_0^\infty pt^{p-1} \max_i F_i(t)\, dt$

$\inf Em_n^p = b_n^p - \int_{-\infty}^{b_n} pt^{p-1} \Sigma_1^n F_i(t)\, dt$

where the bounds are attained at two distributions on $\mathbb{R}^\infty$.

4.3. *Superadditive functions in* $\mathbb{R}^n$. Call a function $\varphi : E \subset \mathbb{R}^n \to \mathbb{R}$ superadditive if it is a superadditive function of any two coordinates considered separately. For instance, a convex function is superadditive but not necessarily n-positive for $n \geqslant 3$; an example is $\varphi = (xyz)^{-1}$ on $\mathbb{R}^{+3}$. Theorem 5 was proven for bounded continuous functions in the context of rearrangements of functions by Lorentz [25]. We present a simpler proof using a martingale representation and requiring weaker regularity conditions. See also Schaefer [31] for a special case.

THEOREM 5. (Lorentz). *Let* $F_1, \cdots, F_n$ *be n probability measures on* $\mathbb{R}$ *and let* $\bar{H}(x) = \min_i F_i(x)$. *Then for any continuous superadditive function* $\varphi : \mathbb{R}^n \to \mathbb{R}$, $\int \varphi\, d\bar{H} = \sup_{H \in \Gamma(F_1, \cdots, F_n)} \int \varphi\, dH$ *if one of the following conditions holds*:

(a) $\varphi \leqslant h$ *for some continuous function h such that* $\int h\, dH$ *is finite and constant for all* $H \in \Gamma(F_1, \cdots, F_n)$.

(b) $\{\varphi(X),\ X$ *distributed as* $H \in \Gamma(F_1, \cdots, F_n\}$ *is uniformly integrable.*

PROOF. Suppose the F_i's have compact support. For a p-tuple $(x_1, \cdots, x_p)$ let $(\bar{x}_1, \cdots, \bar{x}_p)$ denote its rearrangement in increasing order; recall (Lorentz [25]) that given n p-tuples $(x_1^1, \cdots, x_p^1), \cdots, (x_1^n, \cdots, x_p^n)$, for any superadditive function φ, the maximum of $\Sigma_{i=1}^p \varphi(x_i^1, x_i^2, \cdots, x_i^n)$ over all the rearrangements of the p-tuples is attained at $(\bar{x}_1^1, \cdots, \bar{x}_p^1), \cdots, (\bar{x}_1^n, \cdots, \bar{x}_p^n)$. Let $X_1, X_2, \cdots, X_n$ be jointly distributed as H and defined on [0, 1] with Lebesgue measure, according to Skorokhod's construction [33]. Thus, if the distribution of $(X_1, \cdots, X_n)$ is concentrated on p atoms $(x_i^1, \cdots, x_i^n)_{i=1,\cdots,p}$ of mass $1/p$, $E\varphi(\bar{X}_1, \cdots, \bar{X}_n) \geqslant E\varphi(X_1, \cdots, X_n)$, where $\bar{X}_i(\omega) = F_i^{-1}(\omega)$ for each i, $i = 1, \cdots, n$. In the general case, let $x_{i,k}^m = 2^m \cdot E\{X_i I_{[k2^{-m}, (k+1)2^{-m}]}\}$ and

$$X_i^m(\omega) = \Sigma_{k=0}^{2^m-1} x_{i,k}^m \cdot I_{[k2^{-m}, (k-1)2^{-m}]}(\omega), \qquad \text{for all } i = 1, \cdots, n.$$

$X_1^m, X_2^m, \cdots, X_n^m$ are step functions and bounded martingales converging a.s. to $X_1, \cdots, X_n$ respectively, cf. [6], page 94. Call $\overline{X_i^m}$, $i = 1, \cdots, n$ the reorderings of $X_i^m : \overline{X_i^m} = \overline{X^m}_{i,k}$ on $[k2^{-m}, (k+1)2^{-m}]$ $i = 1, \cdots, n$. $\overline{X_i^m}$ and X_i^m have the same distribution and $\overline{X_i^m}(\omega)$ is the inverse of the cdf of X_i^m, hence $\overline{X_i^m}(\omega) \to F_i^{-1}(\omega)$ a.s., so that in the bounded case the theorem follows by bounded convergence. The weaker regularity conditions (a) and (b) are obtained as in Corollary 2.2.

4.4. *Majorizations and doubly stochastic matrices.* Recall a result of Hardy, Littlewood and Polya [18], page 45: say that $x \in \mathbb{R}^n$ majorizes $y \in \mathbb{R}^n$ iff $\Sigma_1^k x_{(i)} \geqslant \Sigma_1^k y_{(i)}$ for each k, $k = 1, \cdots, n$ and $\Sigma_1^n x_{(i)} = \Sigma_1^n y_{(i)}$, where $x_{(i)}$ and $y_{(i)}$ are the coordinates of x and y rearranged in decreasing order. Then x majorizes y if $y = xP$ for some doubly stochastic matrix P. An ordering for doubly stochastic matrices is proposed in the next theorem (cf. Sherman [32]).

THEOREM 6. *Let P and Q be two $n \times n$ doubly stochastic matrices such that*
(a) $\Sigma_{i=1}^{k} p_{ij}$ *and* $\Sigma_{i=1}^{k} q_{ij}$ *are decreasing in j for* $k = 1, 2, \cdots, n-1$;
(b) $\Sigma_{i \leq r} \Sigma_{j \leq s} p_{ij} \leq \Sigma_{i \leq r} \Sigma_{j \leq s} q_{ij}$ *for all* $r, s = 1, 2, \cdots, n-1$.
Then for all $x \in \mathbb{R}^n$ with $x_1 \geq x_2, \cdots \geq x_n$, xQ majorizes xP.

PROOF. Condition (a) is a necessary and sufficient condition to ensure that P and Q are order-preserving, cf. [27], page 24. The theorem is then obtained from Theorem 2 with $\varphi(i, j) = x_i I_{\{1, \cdots, k\}}(j)$ or from Theorem 1, verifying that xQ will majorize xP when $q_{ij} = p_{ij}$ for all i, j, except that for $u' > u$ and $v' > v$, $q_{u'v'} - p_{u'v'} = q_{uv} - p_{uv} = \varepsilon > 0$ and $q_{uv'} - p_{uv'} = q_{u'v} - p_{u'v} = -\varepsilon$.

4.5. $\bar{\rho}$ *distance between stationary processes.* Gray, Neuhoff and Shields [16] have investigated a generalization of Ornstein's $\bar{d}$ distance, called the $\bar{\rho}$ distance, with applications to communication theory; we indicate here new lower bounds for the $\bar{\rho}$ distance and construct a counterexample to a conjecture of Gray, et al.

For $X = (X_n)_{n=1}^{\infty}$, and $Y = (Y_n)_{n=1}^{\infty}$ two stationary processes on $\mathbb{R}$ with a metric ρ, let μ_X and μ_Y be their distributions on $\mathcal{B}(\mathbb{R}^\infty)$. Let $x^n = (x_1, \cdots, x_n)$ and define the metric ρ_n on $\mathbb{R}^n$ by $\rho_n(x^n, y^n) = (1/n)\Sigma_1^n \rho(x_i, y_i)$. Let $\bar{\rho}_n(\mu_X, \mu_Y) = \sup_n \bar{\rho}_n(X^n, Y^n)$, where $\bar{\rho}_n = \inf_{P_n \in \Gamma(\mu_{X^n}, \mu_{Y^n})} E_{P_n} \rho_n(X^n, Y^n)$. According to [16] an exact lower bound for $\bar{\rho}$ is

$$\bar{\rho}_1 = \inf_{P_1 \in \Gamma(\mu_{X^1}, \mu_{Y^1})} E_{P_1} \rho(X^1, Y^1).$$

The measure $\bar{\rho}_1$ is called the Lévy-Wasserstein (cf. [24], [14], [11], page 469) distance between the fixed one-dimensional marginal distributions of $(X_n)_{n=1}^{\infty}$ and $(Y_n)_{n=1}^{\infty}$. This lower bound is attained when $(X_n)_{n=1}^{\infty}$ and $(Y_n)_{n=1}^{\infty}$ are both i.i.d. processes. For $\rho(x,y) = |x - y|$ and $\rho^* = (x - y)^2$, Gray, et al. obtained the lower bounds

$$\bar{\rho}(X, Y) \geq \int_0^1 |F_x^{-1}(t) - F_y^{-1}(t)|\, dt \tag{1}$$

$$\bar{\rho}^*(X, Y) \geq \frac{1}{2\pi} \int_{-\pi}^{\pi} |f_X^{\frac{1}{2}}(\lambda) - f_Y^{\frac{1}{2}}(\lambda)|^2 d\lambda \tag{2}$$

where $F_x(F_y)$ is the cdf of $X^1(Y^1)$, and $f_X(\lambda)$ and $f_Y(\lambda)$ are the spectral densities of X and Y. (That ρ^* and $\bar{\rho}^*$ are not metrics does not invalidate any of the previous results.) The lower bound in (2) is attained for Gaussian processes, and Gray, et al., conjecture that it is attained only for Gaussian processes.

Corollary 2.2 shows that whenever $-\rho(x, y)$ is a superadditive function, a sharp lower bound for $\bar{\rho}$ is

$$\bar{\rho}_1 = \int_0^1 \rho\left[F_x^{-1}(t), F_y^{-1}(t)\right] dt. \tag{3}$$

Thus a different lower bound for $\bar{\rho}^*$ is

$$\bar{\rho}^*(X, Y) \geq \int_0^1 |F_x^{-1}(t) - F_y^{-1}(t)|^2\, dt. \tag{4}$$

Comparing (2) and (4), observe that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}|f_X^{\frac{1}{2}}(\lambda) - f_Y^{\frac{1}{2}}(\lambda)|^2\, d\lambda \geqslant r_X(0) + r_Y(0) - 2(r_X(0)\cdot r_Y(0))^{\frac{1}{2}}$$
$$\geqslant (\sigma_X - \sigma_Y)^2$$

where r_X is the covariance function of X, σ_X^2 is the variance of F and similarly for $r_Y(t)$ and σ_Y^2. Equality holds if and only if f_x is proportional to f_Y. Note also that $\int_0^1|F_x^{-1}(t) - F_y^{-1}(t)|^2\, dt \geqslant (\sigma_X - \sigma_Y)^2$ with equality iff F_x^{-1} is proportional to F_y^{-1}, i.e., if F_x and F_y differ by a scale factor only. These remarks lead to the following counterexample to the conjecture of [16]: let U_n be i.i.d. random variables uniformly distributed on [0, 1], let F be any cdf, $A_n = F^{-1}(U_n)$, $B_n = 2F^{-1}(U_n)$. Then $\bar{\rho}^*(A, B) = \bar{\rho}_i^* = \int_0^1[F^{-1}(t) - 2F^{-1}(T)]^2\, dt = \sigma_X^2$. Verify also that since $f_X(\lambda) = \sigma_X^2$,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}|f_X^{\frac{1}{2}}(\lambda) - f_Y^{\frac{1}{2}}(\lambda)|^2 d\lambda = (\sigma_X - 2\sigma_X)^2 = \sigma_X^2.$$

Acknowledgments. I wish to thank Professor Thomas Cover for his many invaluable discussions, also Professors S. Resnick, H. Robbins, C. Stein and A. Veinott for their comments, and the referee for his constructive contributions.

REFERENCES

[1] BARTFAI, P. (1970). Uber die Entfernung der Irrfahrtswege. *Studia Sci. Math. Hungar.* **5** 41–49.

[2] BASS, J. (1955). Sur la compatibilité des functions de répartition. *C. R. Acad. Sci. Paris Ser. A–B* **240** 839–841.

[3] BERTINO, S. (1966). Una generalizzazione della dissomiglianza di C. Gini tra variabili casuali semplici. *Giorn. Ist. It. Attuari* **29** 155–178.

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

[5] BIRKHOFF, G. (1946). Tres observaciones sobre el algebra lineal. *Univ. Nac. Tucuman. Revista Ser.* **A5** 147–151 (*MR* **8**(1947) 561).

[6] BREIMAN, L. (1968). *Probability*. Addison Wesley, Reading.

[7] CAMBANIS, S., SIMONS, G. and STOUT, W. (1976). Inequalities for $Ek(X, Y)$ when the marginals are fixed. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **36** 285–294.

[8] DALL'AGLIO, G. (1956). Sugli estremi dei momenti delle funzioni di ripartizione dopia. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* **3** 33–74.

[9] DALL'AGLIO, G. (1972). Fréchet classes and compatibility of distribution functions. *Symp. Math.* **9** 131–150.

[10] DAY, P. W. (1972). Rearrangement inequalities. *Canad. J. Math.* **XXIV**, 930–943.

[11] DOBRUSHIN, R. L. (1970). Prescribing a system of random variables by conditional distributions. *Theor. Probability Appl.* **15** 458–486.

[12] FAN, KY (1967). Subadditive functions on a distributive lattice and an extension of Szaz's inequality. *J. Math. Anal. Appl.* **18** 262–268.

[13] FRÉCHET, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. de Lyon* **14** 53–77.

[14] FRÉCHET, M. (1957). Sur la distance de deux lois de probabilité. *Publ. Inst. Statist. Univ. Paris* **6** 185–198.

[15] GINI, C. (1915–6). Sul criterio di concordanza tra due caratteri. *Atti del Reale Instituto Veneto di Scienze, Lettere ed Arti, Series* 8, **75** 309–331.

[16] GRAY, R. M., NEUHOFF, D. L. and SHIELDS, P. C. (1975). A generalization to Ornstein's $\bar{d}$ distance with applications to information theory. *Ann. Probability* **3** 315–328.

[17] GUPTA, S. S. (1963). Probability integrals of multivariate normal and multivariate t. *Ann. Math. Statist.* **34** 792–828.

[18] HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1952). *Inequalities*. Cambridge Univ. Press.

[19] HOEFFDING, W. (1940). Masstabinvariante Korrelations-theorie. *Schr. Math. Inst. Univ. Berlin* **5** 181–233.

[20] KARLIN, S. (1968). *Total Positivity, Vol.* 1. Stanford Univ. Press.

[21] KRUSKAL, W. H. (1958). Ordinal measures of association. *J. Amer. Statist. Assoc.* **53** 814–861.

[22] LAI, T. L. and ROBBINS, H. (1976). Maximally dependent random variables. *Proc. Nat. Acad. Sci. U.S.A.* **73** 286–288.

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

[24] LÉVY, P. (1950). Distance de deux variables aléatoires et distance de deux de probabilité. In *Generalités sur les probabilités*. (Fréchet, ed.) Gauthiers-Villars, Paris.

[25] LORENTZ, G. G. (1953). An inequality for rearrangements. *Amer. Math. Monthly* **60** 176–179.

[26] MALLOWS, C. L. (1972). A note on asymptotic joint normality. *Ann. Math. Statist.* **43** 508–515.

[27] MARSHALL, A. W. and OLKIN, I. (1973). Doubly stochastic matrices and majorization. Tech. Report No. 70, Stanford Univ.

[28] NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability*. Holden Day, San Francisco.

[29] RINOTT, Y. (1973). Multivariate majorization and rearrangement inequalities with some applications to probability and statistics. *Israel J. Math.* **15** 60–77.

[30] ROBBINS, H. (1975). The maximum of identically distributed random variables. *I.M.S. Bull.*, March (abstract) 75–76.

[31] SCHAEFFER, M. (1976). Note on the k-dimensional Jensen inequality. *Ann. Probability* **4** 502–504.

[32] SHERMAN, S. (1952). On a conjecture concerning doubly stochastic matrices. *Proc. Amer. Math. Soc.* **3** 511–513.

[33] SKOROKHOD, A. V. (1956). Limit theorems for stochastic processes. *Theor. Probability Appl.* **1** Section 3.1.1.

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

[35] VALLENDER, S. S. (1973). Calculation of the Wasserstein distance between probability distributions on the line. *Theor. Probability Appl.* **18** 784–786.

[36] VITALE, R. A. and PIPKIN, A. C. (1976). Conditions on the regression function when both variables are uniformly distributed. *Ann. Probability* **4** 869–873.

[37] WHITT, W. (1976). Bivariate distributions with given marginals. *Ann. Statist.* **4** 1280–1289.

DEPARTMENT OF STATISTICS
CARNEGIE-MELLON UNIVERSITY
SCHENLEY PARK
PITTSBURGH, PENNSYLVANIA