JOURNAL OF MULTIVARIATE ANALYSIS 40, 84–93 (1992)

A Partial Ordering of Rank Densities

MICHAEL PROSCHAN

National Heart, Lung, and Blood Institute, Bethesda, Maryland 20205

AND

FREDERICK LEYSIEFFER

Florida State University

Communicated by the Editors

A function $f(\pi)$ on the set of permutations of $\{1, 2, ..., n\}$ is called arrangement increasing (AI) if it increases each time we transpose a pair of coordinates in descending order, $i < j$ and $\pi_i > \pi_j$, putting them in ascending order. We define and develop a partial ordering $\leqslant^{AI}$ on densities of rank vectors in terms of expectations of AI functions. Specially, one density g is defined to be AI-larger than another density f ($f \leqslant^{AI} g$) if the expectation under g of any AI function is at least as large as its expectation under f. We show that the uniform density is the AI-smallest AI density, and this leads to power results for tests of agreement of two rank vectors. The extreme points of the convex set of AI densities are determined, from which additional results concerning the minimum power of rank tests are shown to follow. We also give applications to ranking and selection problems.

1. INTRODUCTION AND SUMMARY

A function $f(\pi)$ on the set of permutations of $\{1, 2, ..., n\}$ is called arrangement increasing (AI) if $i < j$ and $\pi_i > \pi_j \Rightarrow f(\pi) \leqslant f(\pi_1, ..., \pi_j, ..., \pi_i, ..., \pi_n)$. Henery [4] proposes a probability model for outcomes of horse races, and in the process hints at the notion of an AI function, though he does not use this term. Sobel [8] defines the concept of an AI function, as does Lehmann [6], although Lehmann abandons it in favor of another partial ordering of permutations. The first thorough investigation of the

Received August 28, 1990.

AMS 1980 subject classifications: primary 62F07, 62H20; secondary 62G10.

Key words and phrases: arrangement increasing, permutation group, rank density, partial ordering, tests of agreement between rankings, ranking, and selection problems.

0047-259X/92 $3.00

properties of these functions is by Hollander, Proschan, and Sethuraman [5], who use the term decreasing in transposition. They define an AI function of two vectors and show that their definition is equivalent to the function being AI in the second vector after ordering the first. Marshall and Olkin [7] prefer the term arrangement increasing.

We define a partial ordering $\leqslant^{\mathrm{AI}}$ on densities of rank vectors. Specifically, if f and g are densities for the rank vectors $\mathbf{R}$ and $\mathbf{S}$, we say that f is AI-smaller than g ($\mathbf{R}$ is AI-smaller than $\mathbf{S}$) and write $f \leqslant^{\mathrm{AI}} g$ ($\mathbf{R} \leqslant^{\mathrm{AI}} \mathbf{S}$) if $E\phi(\mathbf{R}) \leqslant E\phi(\mathbf{S})\ \forall \mathrm{AI}\ \phi$. It is shown that to determine whether $f \leqslant^{\mathrm{AI}} g$ it suffices to check that $E\phi(\mathbf{R}) \leqslant E\phi(\mathbf{S})$ for certain AI functions called upper set functions. More basic functions called branch functions, whose expectations are shown to completely determine a density on S_n, are discussed. It is shown that among all AI densities, the uniform is the AI-smallest. This fact is used in Section 3 to show that a number of tests to determine whether there is a positive association between two rank vectors are unbiased against AI alternatives. The uniform density is just one extreme point of the convex set of AI rank densities. The extreme points are shown to consist of upper set densities, and these are used to find the least favorable distribution in testing situations. Other applications in ranking and selection are presented in Section 4, including power results in ranking problems concerning location and scale parameters.

2. Theoretical Results

Let S_n denote the set of permutations of $\{1, 2, ..., n\}$. If $\tau \in S_n$, $\pi \in S_n$, we define $\tau \circ \pi$ to be the composition $\tau \circ \pi = \tau_\pi = (\tau_{\pi_1}, ..., \tau_{\pi_n})$. Under this operator S_n is a group with identity $\mathbf{e} = (1, 2, ..., n)$. We say that π is *better ordered* than τ, and write $\tau \leqslant^o \pi$, if π is obtainable from τ by a sequence of transpositions of out of order pairs, $i < j$ and $\tau_i > \tau_j$. We say that a function $f(\pi)$ on S_n is *arrangement increasing* (AI) if $f(\tau) \leqslant f(\pi)$ whenever $\tau \leqslant^o \pi$.

Although we have defined an AI function on S_n, we could extend the definition to any function whose domain is permutation invariant. Hollander, Proschan, and Sethuraman [5] define and give many examples of AI functions of two vectors. Each of these is AI in the second vector after ordering the first. We shall make use of the following result from Hollander, Proschan, and Sethuraman [5].

2.1. Theorem. *If f and g are AI functions then*

$$h(\pi) = \sum_{\tau \in S_n} f(\pi_{\tau^{-1}})\, g(\tau)$$

is AI.

We now define and develop properties of a partial ordering on the set of densities on S_n.

2.2. DEFINITION. Let $f(\pi)$ and $g(\pi)$ be densities for the rank vectors $\mathbf{R}$ and $\mathbf{S}$. We say that *f is AI-smaller than g* ($\mathbf{R}$ *is AI-smaller than* $\mathbf{S}$) and write $f \leqslant^{\mathrm{AI}} g$ ($\mathbf{R} \leqslant^{\mathrm{AI}} \mathbf{S}$) if $E\phi(\mathbf{R}) \leqslant E\phi(\mathbf{S})\ \forall \mathrm{AI}\ \phi$.

2.3. *Notes.* (a) This definition does not require f or g to be AI.

(b) To show that $f \leqslant^{\mathrm{AI}} g$ it suffices to verify that $E\phi(\mathbf{R}) \leqslant E\phi(\mathbf{S})$ for all AI rank densities ϕ.

(c) Let $\mathbf{e}$ denote the identity permutation $(1, 2, ..., n)$ and $\mathbf{e}'$ denote the worst ordered permutation $(n, n-1, ..., 1)$. Any rank density g satisfies

$$f(\pi) = 1(\pi = \mathbf{e}') \leqslant^{\mathrm{AI}} g(\pi) \leqslant^{\mathrm{AI}} 1(\pi = \mathbf{e}) = h(\pi).$$

For fixed $\tau \in S_n$ the set $B_\tau = \{\pi: \tau \leqslant^o \pi\}$ will be called a *branch* and the indicator function $f(\pi) = 1(\pi \in B_\tau)$ of a branch set B_τ will be called a *branch function.* Branch functions are, of course, AI. The branch sets are a determining class, as shown in the following lemma.

2.4. LEMMA. *If P and Q are 2 probability measures on S_n which agree on all branches then $P \equiv Q$.*

Proof. We prove the lemma by induction. Associate with each permutation π the integer $\pi_n \times 10^0 + \cdots + \pi_1 \times 10^{n-1}$. Number the permutations in ascending order of these associated integers, $\pi^{(1)} = (1, 2, ..., n), ..., \pi^{(n!)} = (n, n-1, ..., 1)$. $\{\pi^{(1)}\} = \{(1, 2, ..., n)\}$ is a branch, so $P(\{\pi^{(1)}\}) = Q(\{\pi^{(1)}\})$. Now assume that P and Q agree on $\{\pi^{(j)}\}$, $j \leqslant k$. Because $\pi^{(i)} \leqslant^o \pi^{(j)} \Rightarrow j \leqslant i$, the branches B_τ satisfy $B_{\pi^{(k+1)}} = \{\pi^{(k+1)}\} \cup A$, where A is a subset of the first k permutations. By the induction hypothesis $P(A) = Q(A)$, and since $P(B_{\pi^{(k+1)}}) = Q(B_{\pi^{(k+1)}})$, $P(\{\pi^{(k+1)}\}) = Q(\{\pi^{(k+1)}\})$. By induction, $P \equiv Q$. ∎

2.5. COROLLARY. *$\leqslant^{\mathrm{AI}}$ is a partial ordering of the class of rank densities.*

Proof. Clearly the symmetry and transitivity properties are satisfied, so it suffices to prove anti-symmetry. That is, that $f \leqslant^{\mathrm{AI}} g$ and $g \leqslant^{\mathrm{AI}} f \Rightarrow f \equiv g$. But this follows from Lemma 2.4 since branch functions are AI. ∎

It is difficult to verify that $f \leqslant^{\mathrm{AI}} g$ using Definition 2.2, since we would have to check that $E_g\phi \geqslant E_f\phi\ \forall$ AI rank densities ϕ. To narrow the class of AI functions we need to check, we examine the structure of the convex set $\{\phi: \phi \text{ an AI rank density}\}$.

A subset $U \subset S_n$ is called an *upper set* if $\pi \in U$ and $\pi \leqslant^o \tau \Rightarrow \tau \in U$. Note that if ϕ is any AI function and K is a constant then $\{\pi: \phi(\pi) > K\}$ is an upper set. Conversely, any upper set U may be written as $\{\pi: \phi(\pi) > K\}$ for some AI function ϕ and constant K, namely $\phi(\pi) = 1(\pi \in U)$ and $K = 0$. If U is an upper set with cardinality n_U, the function $\phi(\pi) = (1/n_U)\, 1(\pi \in U)$ will be called an *upper set density.*

2.6. THEOREM. *The extreme points of the convex set of AI densities on S_n are the upper set densities $\{\phi_U: \phi_U(\pi) = (1/n_U)\, 1(\pi \in U), U \text{ an upper set}\}$. Consequently, any AI density may be written as a convex combination of upper set densities.*

Proof. First we will show that ϕ_U are extreme. Suppose that for some $\alpha \in (0, 1)$ $\phi_U(\pi) = \alpha f(\pi) + (1-\alpha)\, g(\pi)$ for AI densities f and g. Let $\mathbf{e}$ be the identity permutation. Clearly $\phi_U(\pi) = 0 \Rightarrow f(\pi) = 0$ and $g(\pi) = 0$, while $\phi_U(\pi) > 0 \Rightarrow \alpha f(\pi) + (1-\alpha)\, g(\pi) = \phi_U(\pi) = \phi_U(\mathbf{e}) = \alpha f(\mathbf{e}) + (1-\alpha)\, g(\mathbf{e})$. Because f and g are AI, $f(\pi) \leqslant f(\mathbf{e})$ and $g(\pi) \leqslant g(\mathbf{e})$, so $f(\pi) = f(\mathbf{e})$ and $g(\pi) = g(\mathbf{e})$. Since f and g are both densities, $f \equiv g \equiv \phi_U$.

Next we show that there are no other extreme points by demonstrating that every AI density can be written as a convex combination of upper set densities. If f is any AI density, let $\pi^{(1)}, \pi^{(2)}, \ldots, \pi^{(n!)}$ be a relabeling of the permutations such that $f(\pi^{(i)}) \leqslant f(\pi^{(j)})$ for all $i < j$. Then

$$f(\pi) = f(\pi^{(1)}) + \sum_{i=2}^{n!} [f(\pi^{(i)}) - f(\pi^{(i-1)})]\, 1(f(\pi^{(i)}) \leqslant f(\pi)).$$

Hence any AI density may be written as a nonnegative combination of upper set densities, and therefore as a convex combination of upper set densities. ∎

Recall that the random variable Y is said to be *stochastically larger* than the random variable X, written $X \leqslant^{\text{st}} Y$ if $\Pr(X > t) \leqslant \Pr(Y > t)\ \forall t$. The next theorem is discussed in Marshall and Olkin [7] in a more general partial ordering setting.

2.7. THEOREM. *The following are equivalent:*

(a) $\mathbf{R} \leqslant^{\text{AI}} \mathbf{S}$

(b) $\psi(\mathbf{R}) \leqslant^{\text{st}} \psi(\mathbf{S})\ \forall \text{AI } \psi$.

(c) $\Pr(\mathbf{R} \in U) \leqslant \Pr(\mathbf{S} \in U)\ \forall$ *upper sets* U.

If f and g are rank densities we define

$$f \diamond g(\pi) = \sum_{\tau \in S_n} f(\pi_{\tau^{-1}})\, g(\tau). \tag{2.8}$$

If $\mathbf{R}$ and $\mathbf{S}$ are independent rank vectors with respective densities f and g, then $f \diamond g$ is the density of $\mathbf{R} \circ \mathbf{S}$.

2.9. PROPOSITION. *Under the operation* $\diamond$,

(a) *the set of rank densities forms a semigroup with identity element the density assigning probability* 1 *to the identity permutation* $\pi = (1, 2, ..., n)$, *and zero element the density assigning probability* $1/n!$ *to each permutation, and*

(b) *the set of AI rank densities forms a subsemigroup containing the above identity and zero elements.*

Proof. The assertions in (a) are easy to verify, and (b) follows from Theorem 2.1. ∎

2.10. THEOREM. *Let* $\mathbf{R}$ *be a rank vector with an AI density, and let* $\mathbf{U}^{(1)}$ *and* $\mathbf{U}^{(2)}$ *be rank vectors independent of* $\mathbf{R}$, *with* $\mathbf{U}^{(1)} \leqslant^{\mathrm{AI}} \mathbf{U}^{(2)}$. *Then* $\mathbf{R} \circ \mathbf{U}^{(1)} \leqslant^{\mathrm{AI}} \mathbf{R} \circ \mathbf{U}^{(2)}$ *and* $\mathbf{U}^{(1)} \circ \mathbf{R} \leqslant^{\mathrm{AI}} \mathbf{U}^{(2)} \circ \mathbf{R}$.

Proof. Let f, g_1, and g_2 be the respective densities of $\mathbf{R}$, $\mathbf{U}^{(1)}$, and $\mathbf{U}^{(2)}$. If ϕ is an AI function then

$$E\phi(\mathbf{R} \circ \mathbf{U}^{(i)}) = \sum_{\pi \in S_n} \sum_{\tau \in S_n} \phi(\pi) f(\pi_{\tau^{-1}}) g_i(\tau)$$

$$= \sum_{\tau \in S_n} \psi(\tau) g_i(\tau),$$

where $\psi(\tau) = \sum_{\pi \in S_n} \phi(\pi) f(\pi_{\tau^{-1}})$. By Theorem 2.1, $\psi(\tau)$ is AI. Because $g_1 \leqslant^{\mathrm{AI}} g_2$, $E\phi(\mathbf{R} \circ \mathbf{U}^{(1)}) \leqslant E\phi(\mathbf{R} \circ \mathbf{U}^{(2)})$. This being true for arbitrary AI ϕ, $\mathbf{R} \circ \mathbf{U}^{(1)} \leqslant^{\mathrm{AI}} \mathbf{R} \circ \mathbf{U}^{(2)}$. That $\mathbf{U}^{(1)} \circ \mathbf{R} \leqslant^{\mathrm{AI}} \mathbf{U}^{(2)} \circ \mathbf{R}$ follows from the representation

$$\mathbf{U}^{(i)} \circ \mathbf{R} = (\mathbf{R}^{-1} \circ (\mathbf{U}^{(i)})^{-1})^{-1},$$

and the facts that $\mathbf{R}$ AI $\Leftrightarrow \mathbf{R}^{-1}$ AI, and $\mathbf{R} \leqslant^{\mathrm{AI}} \mathbf{S} \Leftrightarrow \mathbf{R}^{-1} \leqslant^{\mathrm{AI}} \mathbf{S}^{-1}$. ∎

As a special case of this result we conclude that if $\mathbf{R}$ has an AI density and $\tau \leqslant^{o} \pi$, then $\mathbf{R}_\tau \leqslant^{\mathrm{AI}} \mathbf{R}_\pi$, from which we may deduce many of the stochastic rearrangement inequalities found in Chan, D'Abadie, and Proschan [2].

We see then that composition is a kind of smoothing operation. That is, composition $\mathbf{R} \circ \mathbf{S}$ of an AI random vector $\mathbf{R}$ with **any** random vector $\mathbf{S}$ makes it AI-smaller. The AI-largest density is the identity element of the semigroup of densities of S_n, as we have seen. We next prove that the AI-smallest AI density is the zero element.

2.11. THEOREM. *The uniform density (the zero element) is the AI-smallest of the AI rank densities. That is, if f is the uniform rank density and g is any other AI rank density, then $f \leqslant^{\mathrm{AI}} g$.*

Proof. Let $\mathbf{R}$, $\mathbf{S}$, and $\mathbf{Z}$ be independent random vectors, $\mathbf{R}$ having density g, $\mathbf{S}$ having the uniform density, and $\mathbf{Z}$ having the density assigning probability 1 to the identity permutation. $\mathbf{S} \leqslant^{\mathrm{AI}} \mathbf{Z}$ (because $\mathbf{U} \leqslant^{\mathrm{AI}} \mathbf{Z}$ for any random vector $\mathbf{U}$), so $\mathbf{R} \circ \mathbf{S} \leqslant^{\mathrm{AI}} \mathbf{R} \circ \mathbf{Z}$ by Theorem 2.10. But $\mathbf{R} \circ \mathbf{S}$ is uniformly distributed and $\mathbf{R} \circ \mathbf{Z}$ has density g. ∎

3. APPLICATIONS TO TESTS OF AGREEMENT BETWEEN RANKINGS

Suppose we have two rank vectors, and we are interested in testing the degree of agreement between them. We put the first vector in ascending order, permuting the second in the same way as the first. In this way we reduce the problem to a single observation vector $\mathbf{R}$ with density f. The null hypothesis in such a test is that $\mathbf{R}$ is drawn from a specific rank density f_0 (usually uniform), and the alternative is that $\mathbf{R}$ is drawn from a density that is AI-larger than f_0. We shall use a randomized test function $\phi(\mathbf{r})$, which means having observed $\mathbf{R} = \mathbf{r}$, we reject H_0 with probability $\phi(\mathbf{r})$.

3.1. PROPOSITION. *If f_0 is any rank density and $\phi(\mathbf{R})$ is an AI test function of*

$$H_0 : f = f_0$$
$$H_1 : f_0 \leqslant^{\mathrm{AI}} f, f \neq f_0,$$

then ϕ is unbiased. In particular, any test of

$$H_0 : f \text{ uniform}$$
$$H_1 : f \text{ AI but not uniform},$$

using an AI test function is unbiased.

Proof. The first statement follows from the definitions, while the second statement follows from the first statement and Theorem 2.11.

If we are testing whether $\mathbf{R}$ has a uniform distribution, and we have a specific AI alternative density f_1 in mind then the Neyman Pearson lemma implies that the most powerful size α test of

$$H_0 : f = f_0$$
$$H_1 : f = f_1,$$

after observing $\mathbf{R} = \mathbf{r}$, is to reject H_0 with probability

$$\phi(\mathbf{r}) = \begin{cases} 1 & \text{if } f_1(\mathbf{r}) > C, \\ \lambda & \text{if } f_1(\mathbf{r}) = C, \\ 0 & \text{if } f_1(\mathbf{r}) < C, \end{cases}$$

where C and λ are suitably chosen constants. We know that $E_{f_1}\phi(\mathbf{R}) \geq \alpha$ (the inequality is strict unless $f_1 \equiv f_0$), but with Proposition 3.1 we can say much more. We conclude that $E_f\phi(\mathbf{R}) \geq \alpha$ for any AI alternative f. Thus the most powerful test against a specific AI alternative f_1 will also be unbiased against all AI alternatives.

Now consider the general problem of testing

$$H_0: f = f_0 \text{ not necessarily uniform}$$

$$H_1: f \neq f_0, f \text{ AI}.$$

We may wish to have good power against certain alternatives, and this motivates the choice of test function ϕ. Having chosen a test function ϕ, we may be interested in finding the least favorable AI alternative density, that f which minimizes the power $E_f\phi(\mathbf{R})$. We have seen that if ϕ is AI the least favorable distribution is the uniform, but we may not wish to use an AI test function.

Enumerate the $n!$ permutations and think of an AI function as a point in $R^{n!}$. Let $a_i = \phi(\boldsymbol{\pi}^{(i)})$ and $x_i = f(\boldsymbol{\pi}^{(i)})$, $i = 1, 2, \ldots, n!$ We must find a point $\mathbf{x}$ minimizing $\sum_{i=1}^{n!} a_i x_i$ over all $\mathbf{x}$ representable as $x_i = f(\boldsymbol{\pi}^{(i)})$, $i = 1, 2, \ldots, n!$, f an AI density. In linear programming terms we must minimize the *objective function* $\sum_{i=1}^{n!} a_i x_i$ over the set F of *feasible* $\mathbf{x}$ (points which satisfy the AI condition). Any feasible $\mathbf{x}$ may be written as a convex combination of the extreme points of F, namely the upper set densities by Theorem 2.6. Hence the search for an *optimal solution* (one which satisfies the AI constraints and minimizes the objective function) may be confined to upper set densities.

4. Applications to Ranking and Selection Problems

Our next application concerns attempting to find the smallest and/or largest group of parameters from a continuous density $g(\boldsymbol{\theta}, \mathbf{x})$, either in order or without respect to order. This is very general. We may wish to completely order the parameters, find a subset containing the smallest j or largest k parameters, find a subset containing the smallest j and largest k parameters, etc. We will use the empirical rank order $\mathbf{R}$ of the X's and assert that the parameters are ordered the same way. Assuming that g

satisfies $g(\boldsymbol{\theta}_\pi, \mathbf{x}_\pi) = g(\boldsymbol{\theta}, \mathbf{x})\ \forall \pi \in S_n$, to evaluate whether we are successful we may without loss of generality suppose that $\theta_1 < \cdots < \theta_n$. Then it is logical to measure success in terms of whether $\mathbf{R}$ belongs to a specified upper set, such as any of the ones below.

4.1. EXAMPLES. (a) $\{\pi: \pi_1 = 1, \pi_2 = 2, ..., \pi_j = j\}$, where $j \leqslant n$.

(b) $\{\pi: \pi_1 \leqslant k, \pi_2 \leqslant k, ..., \pi_j \leqslant k\}$, where $j \leqslant k \leqslant n$.

(c) $\{\pi: \pi_1^{-1} \leqslant k, \pi_2^{-1} \leqslant k, ..., \pi_j^{-1} \leqslant k\}$, where $j \leqslant k \leqslant n$, and π_j^{-1} is the jth component of π^{-1}.

(d) The analogs of (a)–(c) above with $(\pi_1, ..., \pi_j)$ replaced by $(\pi_{j+1}, ..., \pi_n)$ and $(\pi_1^{-1}, ..., \pi_j^{-1})$ replaced by $(\pi_{j+1}^{-1}, ..., \pi_n^{-1})$. For example, the analog of (a) would be $\{\pi: \pi_{j+1} = j+1, \pi_{j+2} = j+2, ..., \pi_n = n\}$.

For example, suppose we have a problem of *partial ranking*, as discussed by Critchlow [3]; we wish to specify the smallest j parameters in order, and we do so on the basis of the smallest j ranks, in order. Then being successful means observing a rank vector $\mathbf{R}$ belonging to the upper set in Example 4.1(a). If we are not concerned about the order of the j smallest parameters, we might assert that they are among the populations yielding the k smallest observed rankings, $k \geqslant j$, and then being successful means that $\mathbf{R}$ belongs to the upper set in Example 4.1(b). On the other hand 4.1(c) says that the parameters we decide are the j smallest are actually ranked no worse than kth. Many other upper sets might be used as criteria for success, and each has an analog if we are attempting to find the largest instead of smallest parameters, by 4.1(d).

Suppose that we have a familiy $g_{\boldsymbol{\theta},\lambda}$ of continuous densities with location parameter vector $\boldsymbol{\theta}$ and scale parameter $\lambda > 0$. That is, there is an exchangeable random vector $\mathbf{U}$ whose distribution does not depend on $(\boldsymbol{\theta}, \lambda)$ such that $(\lambda U_1 + \theta_1, ..., \lambda U_n + \theta_n)$ has density $g_{\boldsymbol{\theta},\lambda}$. Let $\mathbf{X}$ be from $g_{\boldsymbol{\theta},\lambda}$, and for $\lambda' > 0$ let

$$\mathbf{X}' = \alpha \mathbf{X} + \boldsymbol{\theta}' - \alpha \boldsymbol{\theta}, \tag{4.2}$$

where $\alpha = \lambda'/\lambda$. Then $\mathbf{X}'$ has density $g_{\boldsymbol{\theta}',\lambda'}$. Let $\mathbf{R}$ be the rank vector associated with $\mathbf{X}$ and $\mathbf{R}'$ be the rank vector associated with $\mathbf{X}'$.

4.3. THEOREM. *Under the above hypotheses* $\mathbf{R} \leqslant^{\mathrm{AI}} \mathbf{R}'$ *if either*

(a) $\lambda'/\lambda \leqslant 1$, $\boldsymbol{\theta}' = \boldsymbol{\theta}$, *and* $\theta_1 < \cdots < \theta_n$, *or*

(b) $\lambda' = \lambda$ *and* $\theta_1' - \theta_1 < \cdots < \theta_n' - \theta_n$.

Proof. It is clear from representation (4.2) that if either of conditions (a) or (b) is satisfied then $\mathbf{R} \leqslant^{o} \mathbf{R}'$ with probability 1, hence $\mathbf{R} \leqslant^{\mathrm{AI}} \mathbf{R}'$. ∎

This says that the probability of successfully ordering the θ's (as defined above in terms of upper sets) increases as the scale parameter decreases. This is useful in connection with so-called least favorable configurations. For example, suppose we are trying to find the largest location parameter, and suppose that $\theta_1 < \theta_2 < \cdots < \theta_n$. Let $\delta = \theta_n - \theta_{n-1}$. For fixed θ_{n-1} and $\theta_n = \theta_{n-1} + \delta$, the values of $\theta_1, ..., \theta_{n-2}$ which minimize the probability of correctly selecting θ_n as the largest are known as the least favorable configuration. It is intuitively clear that the least favorable configuration is given by $\theta_1 = \cdots = \theta_{n-2} = \theta_{n-1}$. To deduce this from our work above, note that if $\boldsymbol{\theta}'$ is any other ordered vector with $\theta'_{n-1} = \theta_{n-1}$, $\theta'_n = \theta_n$, then $\boldsymbol{\theta}' - \boldsymbol{\theta}$ is increasing. We can conclude that the probability of success (as measured by our upper set criteria), is smaller with this least favorable configuration than with any other configuration.

We say that a density g_λ has scale parameter vector $\boldsymbol{\lambda}$ if there is an exchangeable random vector U whose distribution does not depend on $\boldsymbol{\lambda}$ such that $(\lambda_1 U_1, ..., \lambda_n U_n)$ has density g_λ. Let g_λ be a family of continuous densities of random vectors with positive components, and suppose $\boldsymbol{\lambda}$ is a scale parameter vector, $\lambda_i > 0 \; \forall \, i \leqslant n$. Let $\mathbf{R}$ be the rank vector associated with an observation vector from g_λ, and $\mathbf{R}'$ be the rank vector associated with an observation vector from $g_{\lambda'}$. In a manner similar to the proof of Theorem 4.3 (or by taking logs and using Theorem 4.3) we may prove the following.

4.4. THEOREM. *Under the above hypotheses* $\lambda'_1/\lambda_1 \leqslant \cdots \leqslant \lambda'_n/\lambda_n \Rightarrow \mathbf{R} \leqslant^{\mathrm{AI}} \mathbf{R}'$.

As with Theorem 4.3, this may be interpreted in terms of the probability of successfully ordering parameters—this time scale parameters. As above, this may be used in connection with the least favorable configuration of scale parameters. For example, suppose that we are trying to find the largest scale parameter and that $\lambda_1 < \cdots < \lambda_n$. Let $\delta = \lambda_n/\lambda_{n-1}$. For fixed λ_{n-1} and $\lambda_n = \delta\lambda_{n-1}$, the least favorable configuration of the other scale parameters is $\lambda_1 = \cdots = \lambda_{n-2} = \lambda_{n-1}$.

4.5. THEOREM. *Let* $\mathbf{X}$ *be multivariate normal with mean vector* $\boldsymbol{\theta}$, $\theta_1 < \theta_2 < \cdots < \theta_n$, *and covariance matrix*

$$\Sigma = \begin{pmatrix} \sigma^2 & \tau & \cdots & \tau \\ \tau & \sigma^2 & \cdots & \tau \\ \vdots & \vdots & \ddots & \vdots \\ \tau & \tau & \cdots & \sigma^2 \end{pmatrix},$$

$0 \leqslant \tau/\sigma^2 \leqslant 1$. *Let* $\mathbf{R} = \operatorname{rank}(\mathbf{X})$. *For fixed* $\boldsymbol{\theta}$ *and* σ, $\mathbf{R}$ *becomes AI-larger as* τ *increases to* σ^2.

Proof. Let $(Z, \mathbf{Y})$ be mutually independent, Z univariate normal with mean 0 and variance τ, and $\mathbf{Y}$ multivariate normal with mean $\boldsymbol{\theta}$ and covariance matrix $(\sigma^2 - \tau) I$. Then $\mathbf{X} =^D (Z + Y_1, Z + Y_2, \ldots, Z + Y_n)$. Now $\text{Rank}(\mathbf{X}) = \text{Rank}(\mathbf{Y})$, and increasing τ decreases the variance of the Y's. By Theorem 4.3(a) this makes $\text{Rank}(\mathbf{Y})$ AI-larger. ∎

This says that the probability of successfully ranking the means of a multivariate normal vector with this covariance structure increases as the common covariance increases. This may be seen explicitly in some of the formulas of Bechhofer [1] in which the observations are independent sample means with equal sample sizes and common variance. The scale parameter in this case is the standard deviation of $\bar{X}$, and one implication is that the larger the sample size, the greater the probability of success.

REFERENCES

[1] BECHHOFER, R. (1954). A single sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.* **25** 16–39.

[2] CHAN, W., D'ABADIE, C., AND PROSCHAN, F. (1987). Stochastic rearrangement inequalities. *J. Multivariate Anal.* **23** 257–275.

[3] CRITCHLOW, D. (1985). *Metric Methods for Analyzing Partially Ranked Data.* Lecture Notes in Statistics, Vol. 34, Springer-Verlag, New York.

[4] HENERY, R. (1981). Permutation probabilities as models for horse races. *J. Roy. Statist. Soc. Ser. B* **43** 86–91.

[5] HOLLANDER, M., PROSCHAN, F., AND SETHURAMAN, J. (1977). Functions decreasing in transposition and their applications in ranking problems. *Ann. Statist.* **5** 722–733.

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

[7] MARSHALL, A., AND OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic Press, New York.

[8] SOBEL, M. (1954). On a generalization of an inequality of Hardy, Littelewood, and Pólya. *Proc. Amer. Math. Soc.* **5** 596–602.