

MONOTONICITY IN SELECTION PROBLEMS: A UNIFIED APPROACH

BY ROGER L. BERGER¹ AND FRANK PROSCHAN²

North Carolina State University and Florida State University

Let $\mathbf{X} = (X_1, \dots, X_n)$ have a density $g(\mathbf{x}, \lambda)$ which is decreasing in transposition, where $\lambda = (\lambda_1, \dots, \lambda_n)$. Suppose one wishes to select a subset of $\{1, \dots, n\}$ containing the subscripts associated with the largest values of the λ_i 's. Let S be a permutation invariant selection rule which is more likely to select a subset associated with the largest values of the X_i 's. Let $A = \{i(1), \dots, i(k)\} \subset \{1, \dots, n\}$ and $B = \{j(1), \dots, j(k)\} \subset \{1, \dots, n\}$ be such that $\lambda_{i(s)} \geq \lambda_{j(s)}$, $s = 1, \dots, k$. The following five inequalities are proved for nonrandomized selection rules. ($|D|$ denotes the number of elements in D . D^c denotes the complement of D .) $P_\lambda(|S \cap A| \geq (>)m) \geq P(|S \cap B| \geq (>)m)$ for every $m \in R^1$, $P_\lambda(|S \cap A^c| \leq (<)m) \geq P(|S \cap B^c| \leq (<)m)$ for every $m \in R^1$, and $P_\lambda(S = A) \geq P_\lambda(S = B)$. Inequalities for randomized selection rules are also obtained. These generalized monotonicity properties are derived using a unified approach. The results apply to selection rules proposed under several formulations of the selection problem.

1. Introduction. In this paper we study some monotonicity properties of ranking and selection rules. Let $\mathbf{X} = (X_1, \dots, X_n)$ be a random observation with density $g(\mathbf{x}; \lambda)$, where the unknown parameter vector $\lambda = (\lambda_1, \dots, \lambda_n) \in \Lambda \subset R^n$. The general goal of a selection problem is to decide which coordinates of λ are the largest or which are larger than a value λ_0 (possibly unknown). This is accomplished by selecting $S \subset \{1, \dots, n\}$, a random set depending on \mathbf{X} , and asserting that the largest parameters are in $\{\lambda_i: i \in S\}$. The subset S may be of fixed or random size depending on the formulation of the selection problem under consideration. See, for example, Bechhofer (1954), Gupta and Sobel (1958), Lehmann (1961), Gupta (1965) and Tong (1969) for five formulations.

Gupta (1965) calls a selection rule *monotone* if $\lambda_i \geq \lambda_j$ implies $P_\lambda(i \in S) \geq P_\lambda(j \in S)$. This monotonicity property is a desirable property for a selection rule, given the goal of selecting a subset consisting of the large values of λ_i . Many authors (for example, Santner, 1975), have shown that their heuristically proposed selection rules are monotone. In this paper we generalize the above notion of monotonicity and present some other notions of monotonicity. Then we show in a unified way that for many selection problems a large class of selection rules possess these monotonicity properties.

The monotonicity properties we consider are the following. Let $A = \{a_1, \dots, a_k\}$ and $B = \{b_1, \dots, b_k\}$ denote two subsets of $\{1, \dots, n\}$ with $|A| = |B| = k$.

Received January 1983; revised July 1983.

¹ This research was supported in part by the U.S. Army Research Office Grant No. DAAG 29-82-K-0168.

² Research sponsored by the Air Force Office of Scientific Research Grant AFOSR 82-K-0007.

AMS 1980 subject classifications. Primary 62F07; secondary 62H10.

Key words and phrases. Ranking and selection, decreasing in transposition, arrangement increasing, monotonicity, comparison with a standard.

Subset A is better than B if, for some arrangements $a_{i(1)}, \dots, a_{i(k)}$ and $b_{j(1)}, \dots, b_{j(k)}$ of the elements of A and B , $\lambda_{a_{i(r)}} \geq \lambda_{b_{j(r)}}$ for every $r = 1, \dots, k$. If A is better than B , then each of the following inequalities would be desirable for a selection rule:

$$(1.1) \quad P_\lambda(|A \cap S| \geq m) \geq P_\lambda(|B \cap S| \geq m) \text{ for every } m \in R^1;$$

[in words, P_λ (at least m elements of A are selected) \geq P_λ (at least m elements of B are selected).]

$$(1.2) \quad P_\lambda(A = S) \geq P_\lambda(B = S);$$

$$(1.3) \quad P_\lambda(A^c \cap S| \leq m) \geq P_\lambda(B^c \cap S| \leq m) \text{ for every } m \in R^1.$$

Some special cases may be of particular interest. By setting $m = k$ in (1.1) we obtain $P_\lambda(A \subset S) \geq P_\lambda(B \subset S)$. Furthermore, if $k = 1$, we obtain the classical monotonicity property of Gupta (1965). By setting $m = 0$ in (1.3) we obtain $P_\lambda(A \supset S) \geq P_\lambda(B \supset S)$.

In Section 2 we present the notation we will use and the assumptions concerning $g(\mathbf{x}; \lambda)$ and S we will make. The main results, which imply inequalities (1.1)–(1.3) for nonrandomized selection rules, are proved in Section 3. The extension of these results to include additional parameters and rules based on statistics in addition to \mathbf{X} is also discussed. Dual results for selection rules designed to select the smallest rather than the largest values of the parameter are straightforward to obtain and are not stated.

2. Notation and assumptions. Let \mathcal{P} be the group of all permutations of $\{1, \dots, n\}$. Let $\pi = (\pi_1, \dots, \pi_n)$ denote a member of \mathcal{P} . For $\mathbf{x} \in R^n$, let $\mathbf{x} \circ \pi = (x_{\pi_1}, \dots, x_{\pi_n})$. *Transpositions* and *decreasing in transposition* (DT) functions on R^{2n} play a central role in our exposition. See Hollander, Proschan, and Sethuraman (1977) (HPS(1977)) for the definition of these concepts. The DT property is called arrangement increasing by Marshall and Olkin (1979) and Property M by Eaton (1967). DT functions are examples of Eaton's (1982) decreasing reflection functions.

We assume that the observation vector $\mathbf{X} = (X_1, \dots, X_n)$ has a density $g(\mathbf{x}; \lambda)$ with respect to a measure $\sigma(\mathbf{x})$, where σ satisfies $\int_A d\sigma(\mathbf{x}) = \int_A d\sigma(\mathbf{x} \circ \pi)$ for each $\pi \in \mathcal{P}$ and Borel set $A \subset R^n$. We assume that g is DT. HPS(1977) list several discrete and continuous DT densities. Eaton (1967), Hsu (1977), and Gupta and Miescke (1982) have investigated selection problems involving a DT density. They compared the operating characteristics or risk functions of different selection rules, whereas, inequalities (1.1)–(1.3) compare different operating characteristics of a single selection rule at a time.

Let \mathcal{S} be the set of all subsets of $\{1, \dots, n\}$. A *selection rule* is a function $\delta(s; \mathbf{x}): \mathcal{S} \times R^n \rightarrow [0, 1]$ satisfying i) $\sum_{s \in \mathcal{S}} \delta(s; \mathbf{x}) = 1$ for every $\mathbf{x} \in R^n$ and ii) $\delta(s; \cdot)$ is measurable for every $s \in \mathcal{S}$. When $\mathbf{X} = \mathbf{x}$ is observed, an element of \mathcal{S} is chosen according to the probability distribution on \mathcal{S} defined by $\delta(\cdot; \mathbf{x})$. This selected subset, which depends on \mathbf{X} and δ , is what we have called S . The *individual selection probabilities*, $\psi_1(\mathbf{x}), \dots, \psi_n(\mathbf{x})$, are defined by $\psi_i(\mathbf{x}) =$

$\sum_{s \in \mathcal{S}_i} \delta(s; \mathbf{x})$ where $\mathcal{S}_i = \{s \in \mathcal{S} : i \in s\}$. $\psi_i(\mathbf{x})$ is the probability $i \in S$ when $\mathbf{X} = \mathbf{x}$ is observed. If $P_\lambda(\psi_i(\mathbf{X}) \in \{0, 1\}) = 1$ for every $i \in \{1, \dots, n\}$ and $\lambda \in \Lambda$, then $\delta(\cdot; \mathbf{x})$ is a point mass with probability one and δ is called a *nonrandomized selection rule*.

We consider selection rules $\delta(s; \mathbf{x})$ which satisfy (2.1) and (2.3) for every i and $j \in \{1, \dots, n\}$, $\mathbf{x} \in R^n$, and $\pi \in \mathcal{P}$:

$$(2.1) \quad x_j \geq x_i \text{ implies } \psi_j(\mathbf{x}) \geq \psi_i(\mathbf{x}),$$

$$(2.2) \quad \psi_{\pi_i}(\mathbf{x}) = \psi_i(\mathbf{x} \circ \pi).$$

Rules satisfying (2.1) have been called “natural” in some of the selection literature (for example, Eaton, 1967). Gupta and Miescke (1982) have shown that for problems involving exponential families, selection rules satisfying (2.1) form an essentially complete class among all rules satisfying (2.2) for many loss functions. The permutation invariance assumption (2.2) is standard and reasonable in light of the permutation invariance of the density g and measure σ .

Lemmas 2.1 and 2.2 will be used to prove the monotonicity results in Section 3.

LEMMA 2.1. *If A is better than B with $k = |A| = |B|$, then there exist vectors λ^* and λ^{**} such that a) $\{\lambda_i : i \in A\} = \{\lambda_{n-k+1}^*, \dots, \lambda_n^*\}$, b) $\{\lambda_i : i \in B\} = \{\lambda_{n-k+1}^{**}, \dots, \lambda_n^{**}\}$, and c) λ^{**} is a transposition of λ^* .*

PROOF. We will define λ^* and λ^{**} and then show they have the required characteristics. Let λ^1 be the elements of $\{\lambda_i : i \in A^c \cap B^c\}$ in an arbitrary but fixed order. Let λ^2 be the elements of $\{\lambda_i : i \in A^c \cap B^c\}$ arranged in nondecreasing order. Let λ^3 be defined like λ^2 but using $A \cap B^c$. Let λ^4 be defined like λ^1 but using $A \cap B$. The two vectors are $\lambda^* = (\lambda^1, \lambda^2, \lambda^3, \lambda^4)$ and $\lambda^{**} = (\lambda^1, \lambda^3, \lambda^2, \lambda^4)$.

Clearly, a) and b) are true by the definition of λ^* and λ^{**} . Let $r = |A^c \cap B|$. Note that $r = |A \cap B^c|$. To show that λ^{**} is a transposition of λ^* , it suffices to show that $\lambda_{n-k+i}^* \geq \lambda_{n-k-r+i}^*$, $i = 1, \dots, r$; for, if this is true then λ^{**} can be obtained from λ^* by the sequence of r simple transpositions which switch λ_{n-k+i}^* and $\lambda_{n-k-r+i}^*$, $i = 1, \dots, r$.

To verify that $\lambda_{n-k+i}^* \geq \lambda_{n-k-r+i}^*$, $i = 1, \dots, r$, fix i . Let $t = |A \cap B \cap \{j : \lambda_j \geq \lambda_{n-k-r+i}^*\}|$. At least $t + r - i + 1$ elements of B are greater than or equal to $\lambda_{n-k-r+i}^*$ because the coordinates of λ^2 are in nondecreasing order. Since A is better than B , corresponding to each of these there must be an element in A which is greater than or equal to $\lambda_{n-k-r+i}^*$. The definition of t implies $|A \cap B^c \cap \{j : \lambda_j \geq \lambda_{n-k-r+i}^*\}| \geq r - i + 1$. Since the elements of λ^3 are in nondecreasing order, $\lambda_{n-k+j}^* \geq \lambda_{n-k-r+i}^*$, $j = i, \dots, r$. In particular $\lambda_{n-k+i}^* \geq \lambda_{n-k-r+i}^*$, as was to be shown. \square

Let $I^D(\cdot)$ denote the indicator function of the set D . Let \mathbf{v}' be the transpose of the row vector \mathbf{v} . And let $\psi(\mathbf{x}) = (\psi_1(\mathbf{x}), \dots, \psi_n(\mathbf{x}))$.

LEMMA 2.2. *Assume the density of $g(\mathbf{x}; \lambda)$ is DT and the selection rule δ satisfies (2.1) and (2.2). Fix $m \in R^1$. Then the function $K(\mathbf{v}, \lambda)$ defined on $R^n \times$*

Λ by $K(\mathbf{v}, \lambda) = E_\lambda I^{(m, \infty)}(\psi(\mathbf{X})\mathbf{v}')$ is DT. The result is also true if $[m, \infty)$ is replaced by (m, ∞) .

PROOF. That $H(\mathbf{v}, \mathbf{x}) = \psi(\mathbf{x})\mathbf{v}'$ is DT is easily verified using (2.1) and (2.2). That $H^*(\mathbf{v}, \mathbf{x}) = I^{(m, \infty)}(\psi(\mathbf{x})\mathbf{v}')$ is DT follows from 5., page 724 of HPS(1977). The Composition Theorem 3.3 of HPS (1977) yields that $K(\mathbf{v}, \lambda)$ is DT. The proof for (m, ∞) is similar. \square

3. Monotonicity properties. In this section we prove the monotonicity properties (1.1)–(1.3) for nonrandomized selection rules. Properties (1.1) and (1.3) are immediate consequences of the following general theorem regarding (possibly randomized) selection rules.

THEOREM 3.1. *Assume the density $g(\mathbf{x}; \lambda)$ is DT and the selection rule δ satisfies (2.1) and (2.2). For any $D \subset \{1, \dots, n\}$ let $\mathbf{v}_D = (I^D(1), \dots, I^D(n))$. Let $A \subset \{1, \dots, n\}$, $B \subset \{1, \dots, n\}$, and $m \in R^1$. If A is better than B then $P_\lambda(\psi(\mathbf{X})\mathbf{v}'_A > m) \geq P_\lambda(\psi(\mathbf{X})\mathbf{v}'_B > m)$. The result is also true if $>$ is replaced by \geq .*

PROOF. Let π^* (π^{**}) denote the permutation such that $\lambda \circ \pi^*(\lambda \circ \pi^{**}) = \lambda^*(\lambda^{**})$ where $\lambda^*(\lambda^{**})$ is defined in Lemma 2.1. Then $\mathbf{v}_A \circ \pi^* = \mathbf{v}_B \circ \pi^{**} = (0, \dots, 0, 1, \dots, 1)$, a vector of $n - k$ zeros followed by k ones. Since λ^{**} is a transposition of λ^* (Lemma 2.1) and $K(\mathbf{v}, \lambda)$ is DT (Lemma 2.2), we obtain

$$\begin{aligned} P_\lambda(\psi(\mathbf{X})\mathbf{v}'_A \geq m) &= K(\mathbf{v}_A, \lambda) = K(\mathbf{v}_A \circ \pi^*, \lambda \circ \pi^*) = K(\mathbf{v}_B \circ \pi^{**}, \lambda \circ \pi^*) \\ &\geq K(\mathbf{v}_B \circ \pi^{**}, \lambda \circ \pi^{**}) = K(\mathbf{v}_B, \lambda) = P_\lambda(\psi(\mathbf{X})\mathbf{v}'_B \geq m). \end{aligned}$$

The $>$ result follows from the (m, ∞) part of Lemma 2.2. \square

The function $\psi(\mathbf{x})\mathbf{v}'_D$ is the conditional expected value of $|S \cap D|$ given $\mathbf{X} = \mathbf{x}$. The conclusion of Theorem 3.1 can be restated as $E_\lambda(|S \cap A| \mid \mathbf{X})$ is stochastically larger than $E_\lambda(|S \cap B| \mid \mathbf{X})$ if A is better than B . This implies other inequalities such as $E_\lambda(S \cap A) \geq E_\lambda(S \cap B)$. If δ is a nonrandomized selection rule these and related results are more simply stated as follows.

THEOREM 3.2. *Under the assumptions of Theorem 3.1, if δ is a nonrandomized subset selection rule then all the following are true:*

- (3.1) $P_\lambda(|S \cap A| > m) \geq P_\lambda(|S \cap B| > m)$
- (3.2) $P_\lambda(|S \cap A| \geq m) \geq P_\lambda(|S \cap B| \geq m)$
- (3.3) $P_\lambda(|S \cap A^c| \leq m) \geq P_\lambda(|S \cap B^c| \leq m)$
- (3.4) $P_\lambda(|S \cap A^c| < m) \geq P_\lambda(|S \cap B^c| < m)$
- (3.5) $P_\lambda(S = A) \geq P_\lambda(S = B)$.

PROOF. If δ is nonrandomized, $\psi(\mathbf{X})\mathbf{v}'_D = |S \cap D|$. Thus (3.1) and (3.2) are just the inequalities from Theorem 3.1. If A is better than B , it is easily verified

that B^c is better than A^c . Thus (3.3) and (3.4) follow from (3.1) and (3.2) respectively.

To prove (3.5), consider the function $H^{**}(\mathbf{v}, \mathbf{x}) = I^{(-\infty, 0]}(\psi(\mathbf{x})(\mathbf{1} - \mathbf{v})')$ where $\mathbf{1}$ is a vector of n ones. Arguing as in Lemma 2.2, H^{**} can be shown to be DT. By Theorem 3.3 and the Composition Theorem of HPS(1977) the $E_\lambda H^{**}(\mathbf{v}, \mathbf{X})H^*(\mathbf{v}, \mathbf{X})$ is also DT (H^* is from the proof of Lemma 2.2). But for a nonrandomized selection rule with $m = |D|$ in H^* , $E_\lambda H^{**}(\mathbf{v}_D, \mathbf{X})H^*(\mathbf{v}_D, \mathbf{X}) = P_\lambda(S = D)$. So arguing as in the proof of Theorem 3.1, (3.5) is obtained. \square

In many problems the model depends not only on λ but also on another parameter ν and the selection rule depends not only on \mathbf{X} but also on another statistic \mathbf{Y} . If the probability model and assumptions are extended as in Eaton (1967, Section 3), the results of this paper, in particular (3.1)–(3.5), continue to hold. For example, in the comparison of n treatments with a standard, Gupta and Sobel's (1958) proposed selection rule will satisfy (3.1)–(3.5) if the same number of observations are taken on each treatment. For this application, one would set $\nu = (\lambda_0, \sigma^2)$, the control mean and common variance, and $\mathbf{Y} = (\bar{X}_0, S^2)$, the estimate of ν .

REFERENCES

- BECHHOFFER, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.* **25** 16–39.
- EATON, M. L. (1967). Some optimum properties of ranking procedures. *Ann. Math. Statist.* **38** 124–137.
- EATON, M. L. (1982). A review of selected topics in multivariate probability inequalities. *Ann. Statist.* **10** 11–43.
- GUPTA, S. S. (1965). On some multiple decision (selection and ranking) rules. *Technometrics* **7** 225–245.
- GUPTA, S. S. and MIESCKE, K. J. (1982). Sequential selection procedures—a decision theoretic approach. Tech. Report 82-6. Dept. of Statist., Purdue Univ., West Lafayette, Indiana.
- GUPTA, S. S. and SOBEL, M. (1958). On selecting a subset which contains all populations better than a standard. *Ann. Math. Statist.* **29** 235–244.
- HOLLANDER, M., PROSCHAN, F., and SETHURAMAN, J. (1977). Functions decreasing in transposition and their applications in ranking problems. *Ann. Statist.* **5** 722–733.
- Hsu, J. C. (1977). On some decision-theoretic contributions to the problem of subset selection. Ph.D. Thesis (Mimeo Ser. No. 491). Dept. of Statist., Purdue Univ., West Lafayette, Indiana.
- LEHMANN, E. L. (1961). Some Model I problems of selection. *Ann. Math. Statist.* **32** 990–1012.
- MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic, New York.
- SANTNER, T. J. (1975). A restricted subset selection approach to ranking and selection problems. *Ann. Statist.* **3** 334–349.
- TONG, Y. L. (1969). On partitioning a set of normal populations by their locations with respect to a control. *Ann. Math. Statist.* **40** 1300–1324.

INSTITUTE OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
BOX 5457
RALEIGH, NORTH CAROLINA 27650-5457

DEPARTMENT OF STATISTICS
THE FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306