SOME OPTIMUM PROPERTIES OF RANKING PROCEDURES[1]

By Morris L. Eaton[2]

Stanford University

1. Introduction. The purpose of this paper is to present a fairly general treatment of a class of statistical problems commonly referred to as ranking (or selection) problems. The presentation is decision theoretic and attention is restricted to the symmetric case so that certain invariance arguments are applicable. For the problem of selecting the best one of several populations, Bahadur (1950) and Bahadur and Goodman (1952) have proved that, for certain families of distributions, the natural selection procedure uniformly minimizes the risk among all symmetric procedures for a large class of loss functions. More recently, Lehmann (1966) has given an alternative proof of the above result and has indicated several other optimum properties of the natural selection procedure. (The results presented here were obtained by the author (1966) independently of the results given by Lehmann (1966).)

In Section 2, we introduce a monotonicity property (called property M) for density functions of k real variables and k real parameters. The class of densities with property M includes the class of densities considered by Lehmann (1966). This class also contains densities of practical interest which have not been considered in previous treatments of ranking problems.

The general ranking problem which was explicitly described by Bechhofer (1954) is the following: on the basis of a set of observations, we wish to partition the set of coordinate values of a k-dimensional parameter vector $\theta = (\theta_1, \cdots, \theta_k)$ into s disjoint subsets, say $\lambda_1, \cdots, \lambda_s$, such that λ_1 contains the k_1 largest components of θ, λ_2 contains the k_2 next largest components of $\theta, \cdots$, and λ_s contains the k_s smallest components of θ where $1 \leqq k_i < k$ and $\sum k_i = k$. In Section 3, we discuss the assumptions on the set of observations and introduce the loss structure. In Section 4, it is shown that if the density of the observations has property M then the natural selection procedure for the problem above: (i) is Bayes for every prior distribution which is symmetric in θ, (ii) uniformly minimizes the risk among symmetric decision rules, and (iii) is minimax and admissible.

Section 5 is devoted to showing that, in a certain sense, property M is a natural assumption for symmetric selection problems. More specifically it is shown, under regularity conditions, that if the natural selection procedure is uniformly best among symmetric decision rules, then the underlying density has property M. In Section 6, certain results on most economical decision rules

Received 14 July 1966.

[1] This research was based in part on the author's doctoral dissertation submitted to Stanford University, and was written with the partial support of the National Science Foundation, Grant GP-3837.

[2] Now at The University of Chicago.

due to Hall (1958), (1959) are extended to the ranking problem described above. Section 7 contains some specific applications of the general results.

2. A monotonicity property of some multivariate density functions. Consider a Borel subset $\mathfrak{X} \subseteq R^k$ (k dimensional Euclidean space) and let μ be a σ-finite measure on $\mathfrak{X}$. Also, let Θ be a symmetric subset of R^k and A be an arbitrary set. For a family of density functions (with respect to μ) $\{f_\alpha(x, \theta) \mid \alpha \varepsilon A\}$, we introduce a concept which is of use in the treatment of ranking problems.

DEFINITION 2.1. A family of real valued density functions $\{f_\alpha(x, \theta)\}$ is said to have *property M* if for each $\alpha \varepsilon A$ and for each i, j ($i \neq j$), $1 \leqq i, j \leqq k$, the following holds:

$$(2.1) \quad x_i \geqq x_j \quad \text{and} \quad \theta_i \geqq \theta_j \quad \text{implies that} \quad f_\alpha(x, \theta) \geqq f_\alpha(x, (i, j)\theta)$$

where $(i, j)\theta$ is the vector θ with the components θ_i and θ_j interchanged.

For a density function of one real variable and one real parameter, recall the definition of monotone likelihood ratio (MLR).

DEFINITION 2.2. A family of density functions $\{g_\alpha(z, \xi) \mid \alpha \varepsilon A\}$ is said to have a MLR if for all $\alpha \varepsilon A$

$$(2.2) \quad z_1 \geqq z_2 \quad \text{and} \quad \xi_1 \geqq \xi_2 \quad \text{implies that}$$
$$g_\alpha(z_1, \xi_1)g_\alpha(z_2, \xi_2) \geqq g_\alpha(z_1, \xi_2)g_\alpha(z_2, \xi_1).$$

PROPOSITION 2.1. *Suppose a family of density functions $\{f_\alpha(x, \theta) \mid \alpha \varepsilon A\}$ has the form*

$$(2.3) \qquad f_\alpha(x_1, \cdots, x_k; \theta_1, \cdots, \theta_k) = \textstyle\prod_{i=1}^k g_\alpha(x_i, \theta_i),$$

where g_α is a density. Then the family $\{f_\alpha \mid \alpha \varepsilon A\}$ has property M if and only if $\{g_\alpha \mid \alpha \varepsilon A\}$ has a MLR.

PROOF. The result follows directly from the definitions. □

The above result allows the construction of many densities $\{f_\alpha\}$ which have property M. However, there are multivariate densities of interest which cannot be written in the form (2.3). The following proposition gives a necessary and sufficient condition that a certain class of densities has property M.

PROPOSITION 2.2. *Let f be a positive strictly decreasing function defined on* $[0, \infty)$ *and consider*

$$(2.4) \qquad h_\Lambda(x, \theta) \equiv c(\Lambda)f((x - \theta)\Lambda(x - \theta)')$$

where θ and x are (row) vectors in R^k, $k \geqq 2$, Λ is a $k \times k$ positive definte matrix and $c(\Lambda)$ is a positive constant. Assume that $h_\Lambda(x, \theta)$ is a density on R^k. The following are equivalent:

(i) *$h_\Lambda(x, \theta)$ has property M.*

(ii) *$\Lambda = c_1 I - c_2 e'e$ where $e = (1, 1, \cdots, 1)$, $c_1 > 0$, $-\infty < c_2/c_1 < 1/k$, and I is the $k \times k$ identity matrix.*

PROOF. Assume that (ii) holds. The conditions on c_1 and c_2 simply guarantee that Λ be positive definite. For $x_i \geqq x_j$ and $\theta_i \geqq \theta_j$ ($i \neq j$), a direct computation yields

$$(x - \theta)\Lambda(x - \theta)' \leqq (x - (i,j)\theta)\Lambda(x - (i,j)\theta)'$$

when $\Lambda = c_1 I - c_2 e' e$. Since f is a decreasing function,

$$h_\Lambda(x, \theta) \geqq h_\Lambda(x, (i,j)\theta), \tag{2.5}$$

for $x_i \geqq x_j$ and $\theta_i \geqq \theta_j$. Note that we have only used that f is decreasing (not that f is strictly decreasing) for (ii) to imply (i).

Conversely, let (i) hold so that h_Λ has property M. For each i, j ($i \neq j$), $x_i \geqq x_j$ and $\theta_i \geqq \theta_j$ implies (2.5), or equivalently,

$$-2x\Lambda\theta' + \theta\Lambda\theta' \leqq -2x\Lambda[(i,j)\theta]' + [(i,j)\theta]\Lambda[(i,j)\theta]'.$$

Setting $x = 0$, $\theta = (1, 0, \cdots, 0)$, $(i,j) = (1, 2)$ yields $\lambda_{11} \leqq \lambda_{22}$; while $x = 0$, $\theta = (0, 1, 0, \cdots, 0)$, $(i,j) = (2, 1)$ yields $\lambda_{22} \leqq \lambda_{11}$ so that $\lambda_{11} = \lambda_{22}$. Similarly, all the diagonal elements of Λ are equal, say $\lambda_{ii} = \lambda$, $i = 1, \cdots, k$.

Now, setting $x = 0$, $\theta = (1, 1, 0, \cdots, 0)$, $(i,j) = (1, 3)$ yields $2(\lambda + \lambda_{12}) \leqq 2(\lambda + \lambda_{23})$; setting $x = 0$, $\theta = (0, 1, 1, 0, \cdots, 0)$, $(i,j) = (3, 1)$ yields $2(\lambda + \lambda_{23}) \leqq 2(\lambda + \lambda_{12})$ so that $\lambda_{12} = \lambda_{23}$. In a similar manner, all the off diagonal elements of Λ are shown to be equal. The proof is completed by noting that Λ is assumed to be positive definite. □

In his treatment of ranking problems, Lehmann (1966) considered densities of the form

$$f(x, \theta) = \beta(\theta)\textstyle\prod_{i=1}^{k} g(x_i, \theta_i), \tag{2.6}$$

where g is assumed to have a MLR and $\beta(\theta)$ is constant when the coordinates of θ are permuted. Of course, densities of the form (2.6) have property M. In an unpublished manuscript, Bachhofer, Kiefer and Sobel have defined a "rankability" condition for densities of the form (2.6) where g is assumed to be in the one parameter Koopman-Darmois family. Their "rankability" condition is essentially property M for the class of densities they consider. Examples of densities which cannot be written in the form (2.6) but which have property M can easily be constructed using Proposition 2.2.

3. The ranking problem: notation and assumptions. We begin with a description of the ranking problem. Consider a random observable $Z = (X, Y)$ with values in a measurable space $(\mathfrak{X} \times \mathcal{Y}, \mathcal{B}(\mathfrak{X}) \times \mathcal{B}(\mathcal{Y}))$ so that X has values in $\mathfrak{X}$ and Y has values in $\mathcal{Y}$. The space $\mathfrak{X}$ is assumed to be a symmetric Borel subset of R^k and $\mathcal{B}(\mathfrak{X})$ is the Borel field inherited from R^k while $\mathcal{Y}$ is arbitrary. It is assumed that Z has a density

$$p_\alpha(x, y; \theta)\, d\mu(x, y) \tag{3.1}$$

where μ is a σ-finite measure on $\mathcal{B}(\mathfrak{X}) \times \mathcal{B}(\mathcal{Y})$ and (θ, α) is a parameter in a set $\Theta \times A (\theta \in \Theta, \alpha \in A)$. The set Θ is assumed to be a symmetric Borel subset of R^k. For the ranking problem, think of θ as a vector of parameters we want to rank while α is a nuisance parameter. In most problems, the observation Z

represents a sufficient statistic for the parameter (θ, α) based on a sample of size n from each of k populations.

Given the above structure, the ranking problem may be described as follows: on the basis of Z, partition the set of coordinate values of the parameter $\theta = (\theta_1, \cdots, \theta_k)$ into s disjoint subsets, say $\lambda_1, \cdots, \lambda_s$, such that λ_1 contains the k_1 largest θ_i, λ_2 contains the k_2 next largest $\theta_i, \cdots$, and λ_s contains the k_s smallest θ_i where $1 \leqq k_i < k$ and $\sum_{i=1}^{s} k_i = k$. An equivalent formulation of the above problem is: partition the set $\{1, \cdots, k\}$ into s disjoint subsets, say $\gamma_1, \cdots, \gamma_s$, where γ_i has k_i elements, $1 \leqq k_i < s$, $\sum_1^s k_i = k$ and then make the obvious association between γ_i and λ_i. It is now clear that the action space for the ranking problem can be taken to be the set $\Gamma = \{\gamma\}$ of all partitions $\gamma = (\gamma_1, \cdots, \gamma_s)$ of $\{1, 2, \cdots, k\}$ where γ_i has k_i elements and the k_i are fixed, $1 \leqq k_i < k$, $\sum_1^s k_i = k$.

As in Lehmann's (1966) treatment of the ranking problem, invariance plays a central role in the treatment here. Let π denote a permutation of the set $\{1, \cdots, k\}$ and let GP be the group of such permutations. The element of GP which interchanges i and j, leaving all other members of $\{1, 2, \cdots, k\}$ fixed, is denoted by (i, j). For $(x, y) \varepsilon \mathfrak{X} \times \mathfrak{Y}$ and $\pi \varepsilon GP$, define $\pi(x, y)$ by $\pi(x, y) \equiv (\pi x, y)$ where πx is defined by $(\pi x)_i \equiv x_{\pi^{-1}i}$. With his definition, it is easy to check that $(\pi_1\pi_2)(x, y) = \pi_1(\pi_2(x, y))$ so that the group GP operates on the left of the space $\mathfrak{X} \times \mathfrak{Y}$. Similarly for $(\theta, \alpha) \varepsilon \Theta \times A$ and $\pi \varepsilon GP$, $\pi(\theta, \alpha)$ is defined by $\pi(\theta, \alpha) \equiv (\pi\theta, \alpha)$ where $(\pi\theta)_i \equiv \theta_{\pi^{-1}i}$. Also, for $\gamma = (\gamma_1, \cdots, \gamma_s) \varepsilon \Gamma$ and $\pi \varepsilon GP$, define $\pi\gamma$ by $\pi\gamma \equiv (\pi\gamma_1, \cdots, \pi\gamma_s)$ where $\pi\gamma_i$ is the image of γ_i under π. For the density p and the measure μ, the following invariance is assumed:

$$p_\alpha(x, y; \theta) = p_\alpha(\pi x, y; \pi\theta), \tag{3.2}$$

$$d\mu(x, y) = d\mu(\pi x, y). \tag{3.3}$$

A decision function φ is a measurable vector function on $\mathfrak{X} \times \mathfrak{Y}$ such that $\varphi = \{\varphi_\gamma : \gamma \varepsilon \Gamma\}$ where $0 \leqq \varphi_\gamma \leqq 1$ and $\sum_{\gamma \varepsilon \Gamma} \varphi_\gamma \equiv 1$. Let $\mathfrak{D}$ be the class of decision functions. For $\varphi \varepsilon \mathfrak{D}$ and $\pi \varepsilon GP$, define $\pi\varphi$ by $(\pi\varphi)_\gamma = \varphi_{\pi^{-1}\gamma}$. A decision function φ is *invariant* if $\pi\varphi(x, y) = \varphi(\pi x, y)$, that is, if $\varphi_{\pi^{-1}\gamma}(x, y) = \varphi_\gamma(\pi x, y)$. Let $\mathfrak{D}_I$ be the set of invariant decision functions. To introduce the loss structure of the problem, let $L_\gamma(\theta, \alpha)$ be the loss for taking action $\gamma \varepsilon \Gamma$ at the parameter point (θ, α). Before stating the assumptions on the loss functions, we need the following:

DEFINITION 3.1. If $\gamma = (\gamma_1, \cdots, \gamma_s)$ and $\gamma' = (\gamma_1', \cdots, \gamma_s')$ are elements of Γ, then γ *differs adjacently from* γ' *at* $[i, j]$ if there exists an integer β $(1 \leqq \beta < s)$ such that: (i) $i \varepsilon \gamma_\beta$, $i \varepsilon \gamma_{\beta+1}'$, (ii) $j \varepsilon \gamma_\beta'$, $j \varepsilon \gamma_{\beta+1}$, and (iii) $(i, j)\gamma' = \gamma$.

The loss functions $L_\gamma(\theta, \alpha)$ are assumed to satisfy the following:

(3.4) If γ differs adjacently from γ' at $[i, j]$, then

$$L_\gamma(\theta, \alpha) \leqq L_{\gamma'}(\theta, \alpha) \quad \text{when} \quad \theta_i \geqq \theta_j,$$

$$0 \leqq L_\gamma(\theta, \alpha) = L_{\pi\gamma}(\pi\theta, \alpha). \tag{3.5}$$

For the ranking problem, (3.4) is surely natural and, in some sense, the least we could require of the loss functions, while (3.5) is suggested by the natural invariance of the problem.

REMARK. We emphasize that the group operations have been defined so that the group GP operates on the left of Γ, $\mathfrak{X}$, Θ and $\mathfrak{D}$. If the group GP operates on the left of Γ and we define $x\pi$, $\theta\pi$ and $\varphi\pi$ by $(x\pi)_i \equiv x_{\pi i}$, $(\theta\pi)_i \equiv \theta_{\pi i}$ and $(\varphi\pi)_\gamma = \varphi_{\pi\gamma}$, then GP operates on the right of $\mathfrak{X}$, Θ and $\mathfrak{D}$. Now, consider the following assumption on the loss function

$$L_\gamma(\theta, \alpha) = L_{\pi\gamma}(\theta\pi, \alpha). \tag{3.5'}$$

If (3.5′) holds, then the results established in Section 4 still hold. However, (3.5′) is a much stronger assumption than is needed to obtain the desired results. We note that the results of Karlin and Truax (1960) and Lehmann (1966) are proved under assumption (3.5′). The author is indebted to Professor C. Stein for emphasizing the distinction between operation on the right and the left and, in particular, for pointing out the above pitfall when dealing with the permutation group.

The risk function of $\varphi \varepsilon \mathfrak{D}$ is defined by

$$\rho(\varphi, \theta, \alpha) = \int \textstyle\sum_\gamma \varphi_\gamma(x, y) L_\gamma(\theta, \alpha) p_\alpha(x, y; \theta)\, d\mu(x, y). \tag{3.6}$$

Also, if F is a probability measure on $\Theta \times A (\Theta \times A$ is assumed to be a measurable space with $\mathfrak{B}(\Theta)$ the Borel sets of $\Theta)$, then the Bayes risk of $\varphi \varepsilon \mathfrak{D}$ is

$$\rho(\varphi, F) = \int \rho(\varphi, \theta, \alpha)\, dF(\theta, \alpha). \tag{3.7}$$

The terms admissible, minimax and Bayes applied to decision functions will be as defined in Blackwell and Girshick (1954).

4. The main result. Throughout this section, assumptions 3.2–3.5 are to hold. The results below establish certain optimum properties of the decision rule φ^* which ranks the vector θ according to the ranking of the observed vector X. To specify this decision rule more precisely, for each $\gamma = (\gamma_1, \cdots, \gamma_s) \varepsilon \Gamma$, let

$$B_\gamma = \{x \mid x \varepsilon \mathfrak{X}, x_{i_1} \geqq \cdots \geqq x_{i_s} \text{ for all } i_j \varepsilon \gamma_j, j = 1, \cdots, s\}. \tag{4.1}$$

For each $x \varepsilon \mathfrak{X}$, let $H(x) = \{\gamma \mid \gamma \varepsilon \Gamma,\ x \varepsilon B_\gamma\}$ and let $n(x)$ be the number of elements in the set $H(x)$ so that $n(x) \geqq 1$. The decision rule φ^* is defined by

$$\begin{aligned} \varphi_\gamma^*(x, y) &= 1/n(x) \quad &\text{if} \quad \gamma \varepsilon H(x), \\ &= 0 &\text{if} \quad \gamma \not\varepsilon H(x). \end{aligned} \tag{4.2}$$

Thus $\varphi^* = \{\varphi_\gamma^* \mid \gamma \varepsilon \Gamma\}$ is not a function of y and it is easy to see that $\varphi^* \varepsilon \mathfrak{D}_I$.

Now let $\mathcal{P}_0$ be the class of probability measures on $(\Theta \times A, \mathfrak{B}(\Theta) \times \mathfrak{B}(A))$ such that $F \varepsilon \mathcal{P}_0$ if and only if $F = F_1F_2$, where $F_1[F_2$, resp.] is a probability measure on $\Theta[A$, resp.] and F_1 is invariant under the group GP operating on Θ. Note that each $\pi \varepsilon GP$ is a measurable function on Θ to Θ since $\mathfrak{B}(\Theta)$ is assumed to be the usual Borel subsets of R^k restricted to Θ. Fix $F \varepsilon \mathcal{P}_0$ and for $\gamma \varepsilon \Gamma$ define

$r_\gamma(x, y)$ by

$$r_\gamma(x, y) = \int_A \int_\Theta L_\gamma(\theta, \alpha) p_\alpha(x, y; \theta)\, dF_1(\theta)\, dF_2(\alpha). \tag{4.3}$$

To establish the main result, it is convenient to first show that if p has property M then for each $\gamma \varepsilon \Gamma$,

$$r_\gamma(x, y) \leqq r_{\gamma'}(x, y) \quad \text{for all} \quad x \varepsilon B_\gamma, y \varepsilon \mathcal{Y}, \gamma' \varepsilon \Gamma. \tag{4.4}$$

The next two lemmas establish (4.4).

LEMMA 4.1. *Let $p_\alpha(x, y; \theta)$ have property M for each $y \varepsilon \mathcal{Y}$. Consider $\gamma = (\gamma_1, \cdots, \gamma_s) \varepsilon \Gamma$ and suppose that $\gamma' \varepsilon \Gamma$ differs adjacently at $[i, j]$ from $\gamma'' \varepsilon \Gamma$ and $i \varepsilon \gamma_\beta$ and $j \varepsilon \gamma_\delta$. If $\beta < \delta$, then*

$$r_{\gamma'}(x, y) \leqq r_{\gamma''}(x, y) \quad \textit{for} \quad x \varepsilon B_\gamma, y \varepsilon \mathcal{Y}. \tag{4.5}$$

PROOF. Let $\Theta_0 = \{\theta \mid \theta_i = \theta_j\}$, $\Theta_1 = \{\theta \mid \theta_i > \theta_j\}$ and $\Theta_2 = \{\theta \mid \theta_i < \theta_j\}$ so that

$$\begin{aligned} &r_{\gamma''}(x, y) - r_{\gamma'}(x, y) \\ &\qquad = \textstyle\sum_{l=0}^{2} \int_A \int_{\Theta_l} [L_{\gamma''}(\theta, \alpha) - L_{\gamma'}(\theta, \alpha)] p_\alpha(x, y; \theta)\, dF_1(\theta)\, dF_2(\alpha). \end{aligned} \tag{4.6}$$

The invariance assumptions imply that $L_{\gamma''}(\theta, \alpha) = L_{\gamma'}(\theta, \alpha)$ for $\theta \varepsilon \Theta_0$ and

$$\begin{aligned} &\int_A \int_{\Theta_2} [L_{\gamma''}(\theta, \alpha) - L_{\gamma'}(\theta, \alpha)] p_\alpha(x, y; \theta)\, dF_1(\theta)\, dF_2(\alpha) \\ &\qquad = -\int_A \int_{\Theta_1} [L_{\gamma''}(\theta, \alpha) - L_{\gamma'}(\theta, \alpha)] p_\alpha(x, y; (i, j)\theta)\, dF_1(\theta)\, dF_2(\alpha). \end{aligned} \tag{4.7}$$

Thus we can write (4.6) as

$$\begin{aligned} &r_{\gamma''}(x, y) - r_{\gamma'}(x, y) \\ &= \int_A \int_{\Theta_1} [L_{\gamma''}(\theta, \alpha) - L_{\gamma'}(\theta, \alpha)][p_\alpha(x, y; \theta) - p_\alpha(x, y; (i, j)\theta)]\, dF_1(\theta)\, dF_2(\alpha). \end{aligned} \tag{4.8}$$

Since γ' differs adjacently at $[i, j]$ from γ'', $\theta \varepsilon \Theta_1$ and (3.4) imply that $L_{\gamma''}(\theta, \alpha) - L_{\gamma'}(\theta, \alpha) \geqq 0$. Also, if $x \varepsilon B_\gamma$ then $x_i \geqq x_j$ since $\beta < \delta$. Consequently,

$$p_\alpha(x, y; \theta) - p_\alpha(x, y; (i, j)\theta) \geqq 0 \quad \text{for} \quad x \varepsilon B_\gamma, \theta \varepsilon \Theta_1.$$

Since the integrand in (4.8) is non-negative on the range of integration (for $x \varepsilon B_\gamma$), (4.8) is non-negative for $x \varepsilon B_\gamma$. □

LEMMA 4.2. *If $p_\alpha(x, y; \theta)$ has property M for each $y \varepsilon \mathcal{Y}$, then for each $\gamma \varepsilon \Gamma$,*

$$r_\gamma(x, y) \leqq r_{\gamma'}(x, y) \quad \textit{for all} \quad x \varepsilon B_\gamma, \gamma' \varepsilon \Gamma, y \varepsilon \mathcal{Y}. \tag{4.9}$$

PROOF. The essence of the proof is the construction of a sequence in Γ, $\gamma^{(0)}$, $\gamma^{(1)}, \cdots, \gamma^{(n)}$ (for a fixed γ and γ') such that $\gamma^{(0)} = \gamma'$, $\gamma^{(n)} = \gamma$ and $r_{\gamma^{(i)}}(x, y) \leqq r_{\gamma^{(i-1)}}(x, y)$ for $x \varepsilon B_\gamma$, $y \varepsilon \mathcal{Y}$, and $i = 1, 2, \cdots, n$. To accomplish this, fix $\gamma \varepsilon \Gamma$ and consider the function P_γ on Γ to Γ defined as follows: write $\gamma = (\gamma_1, \cdots, \gamma_s)$ and let $\gamma' = (\gamma_1', \cdots, \gamma_s') \varepsilon \Gamma$ be such that $\gamma \neq \gamma'$. Let j be the smallest index such that $\gamma_j \neq \gamma_j'$ and let $C = \{i_1, \cdots, i_\beta\}$ be the elements in γ_j which are not in γ_j'. Note that $C \cap \gamma_i'$ is empty for $i = 1, 2, \cdots, j$. Let

ν be the smallest index such that $C \cap \gamma_i'$ is not empty and define i^* to be the smallest element in $C \cap \gamma_\nu'$. Since $\nu > j$, there exists a smallest element $j^* \varepsilon \gamma_{\nu-1}'$ such that $j^* \not\varepsilon \gamma_j$ and $j^* \varepsilon \gamma_\tau$ where $\tau > j$. The function P_γ is then defined by

$$(4.10) \qquad P_\gamma(\gamma') = (i^*, j^*)\gamma', \quad \text{if} \quad \gamma' \neq \gamma, \qquad = \gamma, \quad \text{if} \quad \gamma' = \gamma.$$

It follows easily from the definition of P_γ that for each $\gamma' \varepsilon \Gamma$, there exists a finite positive integer n (depending on γ') such that $P_\gamma^n(\gamma') \equiv P_\gamma(\cdots(P_\gamma(\gamma'))\cdots) = \gamma$. Also, if $\gamma' \neq \gamma$, then $P_\gamma(\gamma')$ differs adjacently from γ' at $[i^*, j^*]$. Setting $\gamma^{(0)} = \gamma'$ and $\gamma^{(i)} = P_\gamma(\gamma^{(i-1)})$ for $i = 1, 2, \cdots, n(\gamma')$ where $n(\gamma')$ is the smallest n such that $P_\gamma^{(n)}(\gamma') = \gamma$, it is easy to show that the conditions of Lemma 4.1 are satisfied for each pair $(\gamma^{(i)}, \gamma^{(i-1)})$. Thus $r_{\gamma^{(i)}}(x, y) \leqq r_{\gamma^{(i-1)}}(x, y)$ for $x \varepsilon B_\gamma$, $y \varepsilon \mathcal{Y}$ which completes the proof. □

Now that (4.4) has been established, the result below follows easily.

THEOREM 4.1. *Suppose* $p_\alpha(x, y; \theta)$ *has property M for all* $y \varepsilon \mathcal{Y}$. *If* $F \varepsilon \mathcal{P}_0$, *then* φ^* *is Bayes for F, that is,*

$$(4.11) \qquad \rho(\varphi^*, F) = \inf_{\varphi \varepsilon \mathfrak{D}} \rho(\varphi, F).$$

PROOF. For each $x \varepsilon \mathfrak{X}$, let $\gamma(x)$ be a fixed element of $H(x)$. Since $x \varepsilon B_{\gamma(x)}$, Lemma (4.2) implies that $r_{\gamma(x)}(x, y) \leqq r_{\gamma'}(x, y)$ for all $\gamma' \varepsilon \Gamma$ and $y \varepsilon \mathcal{Y}$. Also, if $\gamma' \varepsilon H(x)$, then $r_{\gamma(x)}(x, y) = r_{\gamma'}(x, y)$ for all $y \varepsilon \mathcal{Y}$. Thus,

$$(4.12) \qquad \textstyle\sum_{\gamma \varepsilon \Gamma} \varphi_\gamma^*(x, y) r_\gamma(x, y) = (1/n(x)) \sum_{\gamma \varepsilon H(x)} r_\gamma(x, y) = r_{\gamma(x)}(x, y)$$

so that for every $\varphi \varepsilon \mathfrak{D}$,

$$(4.13) \qquad \textstyle\sum_{\gamma \varepsilon \Gamma} \varphi_\gamma^*(x, y) r_\gamma(x, y) = r_{\gamma(x)}(x, y) \leqq \sum_{\gamma' \varepsilon \Gamma} \varphi_{\gamma'}(x, y) r_{\gamma'}(x, y).$$

Consequently,

$$(4.14) \qquad \rho(\varphi^*, F) = \textstyle\int \sum_{\gamma \varepsilon \Gamma} \varphi_\gamma^*(x, y) r_\gamma(x, y)\, d\mu(x, y) \leqq \int \sum_{\gamma \varepsilon \Gamma} \varphi_\gamma(x, y) r_\gamma(x, y)\, d\mu(x, y) = \rho(\varphi, F)$$

for all $\varphi \varepsilon \mathfrak{D}$. □

The following is a generalization of the result presented by Lehmann (1966) which was proved in its original form by Bahadur (1950) and Bahadur and Goodman (1952).

THEOREM 4.2. *If* $p_\alpha(x, y; \theta)$ *has property M for all* $y \varepsilon \mathcal{Y}$, *then for all* $(\theta, \alpha) \varepsilon \Theta \times A$

$$(4.15) \qquad \rho(\varphi^*, \theta, \alpha) \leqq \rho(\varphi, \theta, \alpha) \quad \textit{for all} \quad \varphi \varepsilon \mathfrak{D}_I.$$

PROOF. Fix θ and α and let F_2 be the probability measure on A which puts mass 1 at α. Let F_1 be the probability measure on Θ which puts mass $1/k!$ at

$\pi\theta$ for each $\pi \varepsilon GP$. Then $F_0 = F_1 F_2 \varepsilon \mathcal{P}_0$ so that $\rho(\varphi^*, F_0) \leqq \rho(\varphi, F_0)$. However, if φ is invariant then $\rho(\varphi, F_0) = \rho(\varphi, \theta, \alpha)$. Since $\varphi^* \varepsilon \mathfrak{D}_I$, we have $\rho(\varphi^*, F_0) = \rho(\varphi^*, \theta, \alpha)$. □

Note that if Theorem 4.2 is assumed to hold then Theorem 4.1 can easily be proved, so that the conclusions of the two theorems are equivalent. From Theorem 4.2, we conclude that φ^* is both minimax and admissible within the class $\mathfrak{D}_I$ of invariant decision functions. Since the group GP is finite, it follows that φ^* is both minimax and admissible in $\mathfrak{D}$ (see Blackwell and Girshick (1954) Chapter 8). Thus the following theorem holds.

THEOREM 4.3. *If $p_\alpha(x, y; \theta)$ has property M for all $y \varepsilon \mathcal{Y}$, then φ^* is minimax and admissible.*

5. A partial converse. In this section we consider a random variable $X = (X_1, \cdots, X_k)$ taking values in a symmetric Borel subset $\mathfrak{X}$ of R^k. It is assumed that X has a density $p(x; \theta)\, d\mu(x)$ where μ is a σ-finite measure on $\mathfrak{X}$ and $\theta \varepsilon \Theta$ is a symmetric Borel subset of R^k. Assume further that for each $\pi \varepsilon GP$,

$$p(\pi x; \pi\theta) = p(x; \theta), \tag{5.1}$$

$$d\mu(\pi x) = d\mu(x). \tag{5.2}$$

Since the density p satisfies (5.1), to verify that p has property M, it is sufficient to verify Definition 2.1 for $i = 1$ and $j = 2$. Also, if $x_1 = x_2$ or if $\theta_1 = \theta_2$, then (5.1) implies that

$$p(x; \theta) = p(x; (1, 2)\theta) \tag{5.3}$$

so it is only necessary to verify Definition 2.1 for $x_1 > x_2$ and $\theta_1 > \theta_2$.

Now, consider the following ranking problem: On the basis of X, decide whether $\theta_1 > \theta_2$ or whether $\theta_2 > \theta_1$. For this two-action problem, let $L_i(\theta)$ be the loss for taking action i for $i = 1, 2$, and assume that

$$L_i(\theta) < L_j(\theta) \quad \text{if} \quad \theta_i > \theta_j, \tag{5.4}$$

$$0 \leqq L_i(\theta) = L_{\pi i}(\pi\theta) \tag{5.5}$$

for $i, j = 1, 2 (i \neq j)$ where π is an element of $GP(2)$—the subgroup of GP which acts only on $\{1, 2\}$, leaving $\{3, 4, \cdots, k\}$ fixed. Writing $Z = ((X_1, X_2), Y)$ where $Y = (X_3, \cdots, X_k)$ and $\alpha = (\theta_3, \cdots, \theta_k)$, the above problem is a special case of the ranking problem considered in Section 3. For this problem, let $\varphi^* = (\varphi_1^*, \varphi_2^*)$ be the decision function as defined by (4.2) so that φ^* is only a function of (X_1, X_2). If we assume that $p(x, \theta)$ has property M, then Theorem 4.2 holds. To provide a partial converse, we now show that if the conclusion of Theorem 4.2 holds for the problem considered above then, under regularity conditions, the density p has property M.

To the set $\mathfrak{X} \subset R^k$ assign the metric topology of R^k and assume that: (i) for each $\theta \varepsilon \Theta$, $p(\cdot; \theta)$ is a continuous function on $\mathfrak{X}$, and that (ii) the measure μ assigns positive measure to each non-empty open subset of $\mathfrak{X}$. Let $\mathfrak{D}_I^{(2)}$ be the class of decision rules for the problem above which are functions of (X_1, X_2) only and which are invariant under $GP(2)$.

THEOREM 5.1. *Suppose the density* p *satisfies* (5.1) *and the measure* μ *satisfies* (5.2). *If* $\rho(\varphi^*, \theta) \leqq \rho(\varphi, \theta)$ *for all* $\varphi \varepsilon \mathfrak{D}_I^{(2)}$, *then* p *has property* M.

PROOF. As noted earlier, it is sufficient to show p has property M for $i = 1$ and $j = 2$. Let $\varphi \varepsilon \mathfrak{D}_I^{(2)}$ and write

$$(5.6)\quad \begin{aligned}\rho(\varphi, \theta) - \rho(\varphi^*, \theta) &= \int \textstyle\sum_{i=1}^2 L_i(\theta)[\varphi_i(x) - \varphi_i^*(x)]p(x;\theta)\, d\mu(x) \\ &= \textstyle\sum_{j=0}^2 \int_{B_j} \sum_{i=1}^2 L_i(\theta)[\varphi_i(x) - \varphi_i^*(x)]p(x;\theta)\, d\mu(x),\end{aligned}$$

where $B_0 = \{x \mid x_1 = x_2\}$, $B_1 = \{x \mid x_1 > x_2\}$ and $B_2 = \{x \mid x_1 < x_2\}$. From the invariance of φ and φ^*, we have $\varphi_i(x) = \varphi_i^*(x) = \frac{1}{2}$ on the set B_0 for $i = 1, 2$. Using the invariance assumptions, it follows directly that

$$(5.7)\quad \begin{aligned}&\int_{B_2} \textstyle\sum_{i=1}^2 L_i(\theta)[\varphi_i(x) - \varphi_i^*(x)]p(x;\theta)\, d\mu(x) \\ &\quad = \int_{B_1} \{L_1(\theta)[\varphi_2(x) - \varphi_2^*(x)] + L_2(\theta)[\varphi_1(x) - \varphi_1^*(x)]\} p(x; (1,2)\theta)\, d\mu(x).\end{aligned}$$

Noting that $\varphi_1 = 1 - \varphi_2$, $\varphi_1^* = 1$ on B_1, we can write (5.6) as

$$(5.8)\quad \begin{aligned}&\rho(\varphi, \theta) - \rho(\varphi^*, \theta) \\ &\quad = \int_{B_1} [L_2(\theta) - L_1(\theta)][1 - \varphi_1(x)][p(x;\theta) - p(x;(1,2)\theta)]\, d\mu(x).\end{aligned}$$

Now, fix $x^0 \varepsilon B_1$ and let N_δ be an open sphere of radius $\delta > 0$ centered at x^0 such that $N_\delta \subset B_1$ (N_δ is a sphere in the metric space $\mathfrak{X}$). Choosing the decision rule φ so that $\varphi_1 = 1$ on $B_1 - N_\delta$, $\varphi_1 = \frac{1}{2}$ on N_δ such that $\varphi \varepsilon \mathfrak{D}_I^{(2)}$, (5.8) yields

$$(5.9)\quad \int_{N_\delta} [L_2(\theta) - L_1(\theta)][p(x,\theta) - p(x,(1,2)\theta)]\, d\mu(x) \geqq 0,$$

where the assumption that

$$\rho(\varphi, \theta) \geqq \rho(\varphi^*, \theta) \quad \text{for all} \quad \varphi \varepsilon \mathfrak{D}_I^{(2)}$$

has been used. Fixing θ^0 so that $\theta_1^0 > \theta_2^0$, (5.4) and (5.9) yield

$$(5.10)\quad \int_{N_\delta} [p(x,\theta^0) - p(x,(1,2)\theta^0)]\, d\mu(x) \geqq 0.$$

Letting $\delta \to 0$ and using the assumptions on p and μ, we conclude that

$$(5.11)\quad p(x^0;\theta^0) \geqq p(x^0;(1,2)\theta^0).$$

Combining this with (5.3) yields the desired result. □

6. Most economical decision rules. In this section we extend some results of Hall (1958), (1959), concerning most economical decision rules, to include the class of densities with property M. Consider a sequence of random observables $Z^{(i)} = (X^{(i)}; Y^{(i)})$, $i = 1, 2, \cdots$, taking values in a measurable space $(\mathfrak{X}^{(i)} \times \mathfrak{Y}^{(i)}, \mathfrak{B}(\mathfrak{X}^{(i)}) \times \mathfrak{B}(\mathfrak{Y}^{(i)}))$ where $\mathfrak{X}^{(i)}$ is a symmetric Borel subset of R^k. Assume that $Z^{(i)}$ has a density $p_\alpha^{(i)}(x, y; \theta)\, d\mu^{(i)}(x, y)$ where $\mu^{(i)}$ is a σ-finite measure on the sample space of $Z^{(i)}$ and $(\theta, \alpha) \varepsilon \Theta \times A$. As usual, Θ is a symmetric Borel subset of R^k. Then densities $p^{(n)}$ and the measures $\mu^{(n)}$ are assumed to satisfy (3.2) and (3.3), respectively. In most applications, $Z^{(n)}$ is a sufficient statistic for (θ, α) based on n observations from each of k populations.

Consider the ranking problem described in Section 3 with action space $\Gamma = \{\gamma \mid \gamma \varepsilon \Gamma\}$. For $\gamma \varepsilon \Gamma$, let

$$W_\gamma = \{\theta \mid \theta_{i_1} > \theta_{i_2} > \cdots > \theta_{i_s} \text{ for all } i_j \varepsilon \gamma_j, j = 1, \cdots, s\}, \tag{6.1}$$

and assume that

$$\Theta = \textstyle\bigcup_{\gamma \varepsilon \Gamma} W_\gamma . \tag{6.2}$$

Let the loss functions $L_\gamma(\theta, \alpha)$ be defined by

$$\begin{aligned} L_\gamma(\theta, \alpha) &= 0, \quad \text{if} \quad \theta \varepsilon W_\gamma, \\ &= 1, \quad \text{if} \quad \theta \not\varepsilon W_\gamma, \end{aligned} \tag{6.3}$$

so that $\{L_\gamma\}$ satisfy (3.4) and (3.5). Corresponding to the notation used in Sections 3 and 4, let $\mathfrak{D}^{(n)}$ be the class of decision functions on $Z^{(n)}$ and $\mathfrak{D}_I^{(n)}$ be the class of invariant decision functions on $Z^{(n)}$. Also, let $\varphi^{*(n)}$ be defined on $Z^{(n)}$ by (4.2) so that $\varphi^{*(n)} \varepsilon \mathfrak{D}_I^{(n)}$. For any $\varphi^{(n)} \varepsilon \mathfrak{D}^{(n)}$, the risk function is

$$\rho(\varphi^{(n)}, \theta, \alpha) = 1 - \int \varphi_\gamma^{(n)}(x, y) p_\alpha(x, y; \theta) \, d\mu(x, y) \quad \text{for} \quad \theta \varepsilon W_\gamma . \tag{6.4}$$

Setting

$$P_{\theta,\alpha}^{(n)}(CD \mid \varphi^{(n)}) = \int \varphi_\gamma^{(n)}(x, y) p_\alpha(x, y; \theta) \, d\mu(x, y) \quad \text{for} \quad \theta \varepsilon W_\gamma, \tag{6.5}$$

$P_{\theta,\alpha}^{(n)}(CD \mid \varphi^{(n)})$ represents the probability of making a correct decision at the parameter point (θ, α) when using $\varphi^{(n)} \varepsilon \mathfrak{D}^{(n)}$.

Let β be a fixed real number $(0 < \beta < 1)$ and let $\mathfrak{D}_\beta^{(n)}$ be those decision rules in $\mathfrak{D}^{(n)}$ such that

$$\inf_{(\theta,\alpha)\varepsilon\Theta\times A} P_{\theta,\alpha}^{(n)} (CD \mid \varphi^{(n)}) \geqq \beta. \tag{6.6}$$

Following Hall (1958), (1959) we make the following:

Definition 6.1. The smallest integer, say m, such that $\mathfrak{D}_\beta^{(n)}$ is non-empty is called the *most economical sample size.* If $\varphi^{(m)} \varepsilon \mathfrak{D}_\beta^{(m)}$, then $\varphi^{(m)}$ is a *most economical decision rule.*

In some important ranking problems, $\mathfrak{D}_\beta^{(n)}$ is empty for all n. However, if a finite most economical sample size exists we have the following:

Theorem 6.1. *Suppose $p_\alpha^{(n)}(x, y; \theta)$ has property M for each $n = 1, 2, \cdots$ and each $y \varepsilon \mathcal{Y}^{(n)}$. If m is the most economical sample size, then $\varphi^{*(m)} \varepsilon \mathfrak{D}_\beta^{(m)}$*

Proof. Since $\mathfrak{D}_\beta^{(m)}$ is non-empty,

$$\sup_{\varphi\varepsilon\mathfrak{D}_\beta^{(m)}} [\inf_{(\theta,\alpha)\varepsilon\Theta\times A} P_{\theta,\alpha}^{(m)} (CD \mid \varphi)] \geqq \beta \tag{6.7}$$

so that

$$\inf_{\varphi\varepsilon\mathfrak{D}^{(m)}} [\sup_{(\theta,\alpha)\varepsilon\Theta\times A} \rho(\varphi, \theta, \alpha)] \leqq 1 - \beta, \tag{6.8}$$

since $\mathfrak{D}_\beta^{(m)} \subset \mathfrak{D}^{(m)}$. From Theorem 4.3, $\varphi^{*(m)}$ is minimax so that

$$\sup_{(\theta,\alpha)\varepsilon\Theta\times A} \rho(\varphi^{*(m)}, \theta, \alpha) \leqq 1 - \beta \tag{6.9}$$

which yields

$$\inf_{(\theta,\alpha)\varepsilon\Theta\times A} P_{\theta,\alpha}^{(m)} (CD \mid \varphi^{*(m)}) \geqq \beta. \tag{6.10}$$

Hence $\varphi^{*(m)} \varepsilon \mathfrak{D}_\beta^{(m)}$. □

It is emphasized that the only property of $\varphi^{*(n)}$ used in the proof of the above theorem is that $\varphi^{*(n)}$ is minimax for each n. Thus if it has been shown that $\varphi^{*(n)}$ is minimax for each n for the problem of this section, then $\varphi^{*(m)}$ is most economical whether or not $p_\alpha^{(n)}(x, y; \theta)$ has property M. This remark is used explicitly in some of the examples of Section 7 to conclude that $\varphi^{*(m)}$ is most economical.

7. Examples. In this section we consider specific applications of the results previously established.

7.1. *Ranking main effects in analysis of variance.* Consider observations

$$y_{ij} = \beta_i + \xi_{ij} + \epsilon_{ij} \tag{7.1}$$

where $i = 1, \cdots, K, j = 1, \cdots, J$. The vector of parameters $\beta = (\beta_1, \cdots, \beta_K)$ is assumed to lie in a symmetric subset of R^K and $\delta \equiv (1/J) \sum_{j=1}^J \xi_{ij}$ is assumed to be independent of i. The random variables ϵ_{ij} have a joint normal distribution with mean 0 and covariance matrix $\Sigma = \sigma^2[(1 - \rho)I + \rho e'e]$ where I is the $KJ \times KJ$ identity matrix, e is the vector of 1's, σ^2 is the common variance of the ϵ_{ij} and $\rho(-1/(KJ - 1) < \rho < 1)$ is the correlation between any two different ϵ_{ij}.

Consider the problem of ranking the β_i with action space $\Gamma = \{\gamma\}$ and loss functions $L_\gamma(\beta)$ which satisfy (3.4), (3.5) and $L_\gamma(\beta_1 + c, \cdots, \beta_K + c) = L_\gamma(\beta_1, \cdots, \beta_K)$ for real numbers c. To transform the problem so that the results of Section 4 are applicable, let $Y = (y_{1,1}, \cdots, y_{1,J}, \cdots, y_{K,1}, \cdots, y_{K,J})$ and let $Z = (Z_1, Z_2) = YA$ where A is a $KJ \times KJ$ column orthogonal matrix, Z_1 is $1 \times K$ such that $Z_{1,i} = (1/J) \sum_{j=1}^J y_{ij}$ for $i = 1, \cdots, K$ and Z_2 is $1 \times K(J - 1)$. From the structure of Σ and the column orthogonality of A, it is easy to show that Z_1 and Z_2 are independent, Z_1 has a normal distribution with mean vector θ ($\theta_i = \beta_i + \delta$, $i = 1, \cdots, K$) and a $K \times K$ covariance matrix $\Sigma_1 = (\sigma^2/J)[(1 - \rho)I + J\rho e'e]$, and Z_2 has a normal distribution with mean vector ξ which does not depend on β and a diagonal covariance matrix Σ_2. Since the loss functions $L_\gamma(\beta)$ are translation invariant, an equivalent problem is to rank the parameter vector θ with loss functions $L_\gamma(\theta_1, \cdots, \theta_K) = L_\gamma(\beta_1 + \delta, \cdots, \beta_K + \delta)$. However, the problem is now in the form described in Section 3 so that the results of Section 4 are directly applicable. That the density of Z_1 has property M follows from Proposition 2.2.

It is clear that if σ^2 is unknown, then no most economical sample size can exist. However, if σ^2 and ρ are assumed to be known and if the components of θ are sufficiently separated, then a most economical sample size exists and the decision rule φ^* based on Z_1 is most economical.

Many authors have considered variants of the above ranking problem. For example, see Bechhofer (1954), Dunnett (1960), Paulson (1949) and Seal (1955).

7.2. *Ranking variances in normal populations.* In this example we consider observations Y_{ij}, $i = 1, \cdots, k$, $j = 1, \cdots, n$, where Y_{ij} is $N(\mu_i, \sigma_i^2)$ for $j = 1, \cdots, n$. The problem is to rank the unknown variances with loss functions $L_\gamma(\sigma_1^2, \cdots, \sigma_k^2)$ which depend only on the unknown variances. The sufficient statistic for this problem is $Z = (X, W)$ where $W_i = (1/n)\sum_{j=1}^n Y_{ij}$ and $X_i = \sum_{j=1}^n (Y_{ij} - W_i)^2$ for $i = 1, \cdots, k$. The problem is clearly invariant under translations of the vector W by $b \varepsilon R^k$, and any invariant decision rule will be a function of X only. But for such invariant decision functions, the results of Section 4 are directly applicable since the density of X has property M.

From Theorem 4.3, we have that the decision rule φ^* is minimax within the class of rules invariant under translations of the vector W. However, invoking a general theorem due to Kiefer (1957), we have that φ^* is minimax in the class of all decision functions for each sample size n. From the results in Section 6, it follows that if a most economical sample size exists, then φ^* is most economical in the class of all decision rules. This example has been treated in detail by Bechhofer and Sobel (1954) and in somewhat less detail by Hall (1959) and Lehmann (1966).

7.3. *Ranking correlation coefficients.* We now discuss the problem of ranking correlation coefficients in bivariate normal populations. Consider observations $Y_j^{(i)}$, $i = 1, \cdots, k$; $j = 1, \cdots, n + 1$, where $Y_j^{(i)}$ has a bivariate normal distribution with mean vector μ_i and covariance matrix Σ_i for $j = 1, \cdots, n + 1$ and $i = 1, \cdots, k$. For population i, reduce to the sufficient statistic (X_i, V_i) where $X_i = (1/n + 1)\sum_{j=1}^{n+1} Y_j^{(i)}$ and $V_i = \sum_{j=1}^{n+1} Y_j^{(i)\prime} Y_j^{(i)} - X_i' X_i$ so that X_i has a normal distribution $N(\mu_i, 1/(n + 1)\Sigma_i)$ and V_i has a Wishart distribution with n degrees of freedom and expectation $n\Sigma_i$.

Setting $\rho_i = \sigma_{12,i}/(\sigma_{11,i}\sigma_{22,i})^{\frac{1}{2}}$ we consider the problem of ranking these correlation coefficients ρ_i with loss functions $L_\gamma(\rho_1, \cdots, \rho_k)$, $\gamma \varepsilon \Gamma$ which satisfy the assumptions of Section 3. Introducing the transformations

$$(X_i, V_i) \rightarrow (X_i D_i + b_i, D_i V_i D_i) \tag{7.2}$$

and

$$(\mu_i, \Sigma_i) \rightarrow (\mu_i D_i + b_i, D_i \Sigma_i D_i) \tag{7.3}$$

where D_i is a 2×2 diagonal matrix with positive diagonal elements and $b_i \varepsilon R^2$, it is straightforward to verify that the above problems remains invariant under the group of such transformations on each population. Setting $r_i = v_{12,i}/(v_{11,i}v_{22,i})^{\frac{1}{2}}$, the maximal invariant under the transformation (7.2) is r_i and under (7.3) the maximal invariant is ρ_i. Thus all decision functions for the above ranking problem which are invariant under (7.2) will be functions only of $(r_1, r_2, \cdots, r_k)$.

However, if we now regard $(r_1, \cdots, r_k)$ as the observation vector the results obtained in Sections 4 and 6 are applicable since the joint density

$f(r_1, \cdots, r_k; \rho_1, \cdots, \rho_k)$ has property M. In particular, the decision rule φ^* based on $(r_1, \cdots, r_k)$ is minimax within the class of rules invariant under (7.2). It then follows from a result due to Kiefer (1957) that φ^* is, in fact, minimax within the class of all decision rules. Thus if the k populations are sufficiently different (in terms of their correlation coefficients) so that a most economical sample size exists, then φ^* based on $(r_1, \cdots, r_k)$ is most economical within the class of all decision functions.

8. Acknowledgment. My sincere thanks go to my dissertation advisor, Dr. Ingram Olkin. His suggestions, criticism and enthusiasm have been a source of constant encouragement in the preparation of this report.

REFERENCES

[1] Anderson, T. W. (1958). *Introduction to Multivariate Statistical Analysis*. Wiley, New York.

[2] Bahadur, R. (1950). On a problem in the theory of k populations. *Ann. Math. Statist.* **21** 362–375.

[3] Bahadur, R. and Goodman, L. (1952). Impartial decision rues and sufficient statistics. *Ann. Math. Statist.* **23** 553–562.

[4] Bahadur, R. and Robbins, H. (1950). The problem of the greater mean. *Ann. Math. Statist.* **21** 469–487.

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

[6] Bechhofer, R. E. and Sobel, M. (1954). A single sample multiple decision procedure for ranking variances of normal populations. *Ann. Math. Statist.* **25** 273–289.

[7] Bechhofer, R. E., Dunnett, C. W. and Sobel, M. (1954). A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance. *Biometrika* **41** 170–176.

[8] Bechhofer, R. E., Elmaghraby, S. A. and Morse, N. (1959). A single-sample multiple decision procedure for selecting the multinomial event which has the largest probability. *Ann. Math. Statist.* **30** 102–119.

[9] Bechhofer, R. E., Kiefer, J. and Sobel, M. Sequential ranking procedures with special reference to Koopman-Darmois populations. Unpublished manuscript (to be published in *Statistical Research Monographs*, sponsored jointly by the Institute of Mathematical Statistics and The University of Chicago).

[10] Blackwell, D. and Girshick, M. A. (1954). *Theory of Games and Statistical Decisions*. Wiley, New York.

[11] Dunnett, C. W. (1960). On selecting the largest of k normal populations means. *J. Roy. Statist. Soc. B* **22** 1–40.

[12] Eaton, M. L. (1966). Some optimal properties of ranking procedures with applications in multivariate analysis. Stanford University Technical Report No. 22, Stanford University.

[13] Gupta, S. S. and Sobel, M. (1957). On a statistic which arises in selection and ranking problems. *Ann. Math. Statist.* **28** 957–967.

[14] Gupta, S. S. (1963). Selection and ranking procedures and order statistics for the binomial distribution. Stanford University Technical Report No. 66, Stanford University.

[15] Hall, W. J. (1958). Most economical multiple decision rules. *Ann. Math. Statist.* **29** 1079–1094.

[16] Hall, W. J. (1959). The most economical character of some Bechhofer and Sobel decision rules. *Ann. Math. Statist.* **30** 964–969.

[17] KARLIN, S. (1956). Decision theory for Polya type distributions. Case of two actions, I. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 115–129. University of California Press.

[18] KARLIN, S. and RUBIN, H. (1956). The theory of decision procedures for distributions with monotone likelihood ratio. *Ann. Math. Statist.* **27** 272–300.

[19] KARLIN, S. and TRUAX, D. (1960). Slippage problems. *Ann. Math. Statist.* **31** 296–324.

[20] KIEFER, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. *Ann. Math. Statist.* **28** 573–601.

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

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

[23] LEHMANN, E. L. (1966). On a theorem of Bahadur and Goodman. *Ann. Math. Statist.* **37** 1–6.

[24] PAULSON, E. (1949). A multiple decision procedure for certain problems in analysis of variance. *Ann. Math. Statist.* **20** 95–98.

[25] PRATT, J. W. (1955). Some results in the decision theory of one parameter multivariate Polya type distributions. Stanford University Technical Report No. 37, Stanford University.

[26] RIZVI, M. H. (1963). Ranking and selection problems of normal populations using the absolute values of their means: fixed sample size case. Technical Report No. 31, Department of Statistics, University of Minnesota.

[27] SEAL, K. C. (1955). On a class of decision procedures for ranking means of normal populations. *Ann. Math. Statist.* **26** 387–398.