THE SUBSET SELECTION PROBLEM. I. THE CLASS OF SCHUR-PROCEDURES AND ITS PROPERTIES.[1]

by

Jan F. Bjørnstad

University of California, Berkeley and Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #78-24

October 1978

[1] The research was supported by the Norwegian Research Council for Science and the Humanities.

THE SUBSET SELECTION PROBLEM. I. THE CLASS OF SCHUR-PROCEDURES AND ITS PROPERTIES.[1]

BY

Jan F. Bjørnstad

University of California, Berkeley and Purdue University

1. Introduction

An important class of multiple decision problems is concerned with the selection of good populations out of k possible populations.

There are several formulations of the selection problem. We will discuss the "subset selection approach", first considered by Paulson (1949), Seal (1955) and Gupta (1956), where the size of the selected subset is a random variable, i.e. the size is determined by the data.

The purpose of this two-part paper is to study the problem of subset selection in the location-model.

One of the problems with measuring performance and deriving optimality results for subset selection procedures has been the lack of monotonicity-results for the different risk functions that have been considered. Part I deals with this problem. In Section 2 the general subset selection problem in the location-model is considered, and the performance-criteria to be discussed are presented. In Section 3 a special class, the class of Schur-procedures, is defined. It is shown in Section 5 that this class has nice monotonicity-properties for certain types of risk functions. Part II of this two-part paper deals with minimax theory for the risk functions considered in Part I.

[1] The research was supported by the Norwegian Research Council for Science and the Humanities.

2. The Subset Selection Problem in the Location-Model. Presentation of some Performance-Criteria

$\Pi_1,\ldots,\Pi_k$ are k populations characterized by $\theta_1,\ldots,\theta_k$ respectively. Let $\underline{\theta} = (\theta_1,\ldots,\theta_k)$ and let Ω be the parameter space. Let X_i be the observation from population Π_i. $\underline{X} = (X_1,\ldots,X_k)$. Usually $\underline{X}$ is a sufficient statistic for $\underline{\theta}$, e.g. X_i is the sample mean of n observations from Π_i. We will assume that $X_1,\ldots,X_k$ are independent. X_i has density $f(x-\theta_i)$ with respect to Lebesgue measure. The ordered θ_i are denoted by

$$\theta_{[1]} \leq \cdots \leq \theta_{[k]}.$$

Let $\Pi_{(i)}$, $X_{(i)}$ correspond to $\theta_{[i]}$.

Definition 2.1. Π_i is called a best population if $\theta_i = \theta_{[k]}$, and a non-best population if $\theta_i < \theta_{[k]}$.

The concept of good and bad populations plays an important role in this work.

Definition 2.2. Π_i is said to be a good population if $\theta_i > \theta_{[k]} - \Delta$ and a bad population if $\theta_i \leq \theta_{[k]} - \Delta$. Δ is a given positive constant.

For the risk functions we will consider, two subset selection procedures are equivalent if their individual selection probabilities are the same. Therefore we can define a subset selection procedure by:

$$\psi(\underline{x}) = [\psi_1(\underline{x}),\ldots,\psi_k(\underline{x})] \tag{2.1}$$

where $\psi_i(\underline{x}) = P(\text{selecting } \Pi_i \mid \underline{X}=\underline{x})$.

The size of the selected subset is a random variable. However we will usually require that at least one population is selected, i.e. (for non-randomized procedures)

$$\sum_{i=1}^{k} \psi_i(\underline{x}) \geq 1, \quad \forall \underline{x}. \tag{2.2}$$

A correct selection (CS) is defined to be a selection that includes the best population $\Pi_{(k)}$.

The usual basic condition in the literature has been:

$$\inf_{\underline{\theta} \in \Omega} P_{\underline{\theta}}(CS|\psi) = \inf_{\underline{\theta} \in \Omega} E_{\underline{\theta}}\{\psi_{(k)}\} \geq \gamma, \tag{2.3}$$

i.e. the probability of a correct selection is guaranteed to be at least γ.

Here γ is a given positive constant, $\gamma < 1$, and $\psi_{(i)}$ corresponds to $\theta_{[i]}$, for $i = 1,\ldots,k$.

Subject to a condition like (2.3) a procedures should exclude the bad or non-best populations. One criterion for measuring how well a procedure excludes the non-best populations is given by

$$S'(\underline{\theta},\psi) = \sum_{i=1}^{k-1} E_{\underline{\theta}}\{\psi_{(i)}\}. \tag{2.4}$$

$S'(\underline{\theta},\psi)$ is the expected number of non-best populations selected when $\theta_{[k-1]} < \theta_{[k]}$.

The related criterion for excluding only the bad populations is

$$B(\underline{\theta},\psi) = \sum_{i \in I_\Delta} E_{\underline{\theta}}(\psi_i). \tag{2.5}$$

Here $I_\Delta = I_\Delta(\underline{\theta}) = \{i: \theta_i \leq \theta_{[k]}-\Delta\}$. $B(\underline{\theta},\psi)$ is the expected number of bad populations selected.

The risk function $S'(\underline{\theta},\psi)$ has equal weights on all $E_{\underline{\theta}}\{\psi_{(i)}\}$. In many situations it may be desirable to attach more weight to the worst populations than to those closest to the best. The following criterion accomplishes this and will be considered later.

$$L(\underline{\theta},\psi) = \sum_{i=1}^{k-1} \log[E_{\underline{\theta}}\{\psi_{(i)}\}]. \tag{2.6}$$

Since $\log(y)$ is a concave function of y, L has the property we want: Let

$$L_p = \sum_{\substack{i=1 \\ i \neq p}}^{k-1} \log\{E_{\underline{\theta}}\psi_{(i)}\} + \log\{E_{\underline{\theta}}\psi_{(p)} - \delta\},$$

$\delta > 0$, for $p = 1,\ldots,k-1$. Then

$$L_p < L_q$$

if $p < q$.

The corresponding risk function, if we want to exclude only the bad populations, is

$$\ell(\underline{\theta},\psi) = \sum_{i \in I_\Lambda} \log\{E_{\underline{\theta}}(\psi_i)\} \tag{2.7}$$

A discussion of the different risk functions is given in Part II.

We will mostly be concerned with $B(\underline{\theta},\psi)$ and $\ell(\underline{\theta},\psi)$ as the risk functions. When using these criteria we will require, instead of (2.3) that ψ satisfies the following condition.

$$\inf_{\underline{\theta} \in \Omega} R(\underline{\theta},\psi) \geq \gamma. \tag{2.8}$$

where $R(\underline{\theta},\psi)$ is the expected number of good populations selected.

3. The Class of Schur-Procedures

The notion of a Schur-concave function plays an important role in this work. To describe it we need the following definitions.

Definition 3.1. Let R^k be the k-dimensional euclidean space, and let $\underline{s}, \underline{t} \in R^k$. The ordered components are denoted by: $s_{[1]} \leq \cdots \leq s_{[k]}$, $t_{[1]} \leq \cdots \leq t_{[k]}$. We say that $\underline{s}$ is majorized by $\underline{t}$, $\underline{s} \underset{m}{<} \underline{t}$, if

$$\sum_{i=1}^{k} s_i = \sum_{i=1}^{k} t_i \tag{3.1}$$

and

$$(3.2) \qquad \sum_{j=0}^{p} t_{[k-j]} \geq \sum_{j=0}^{p} s_{[k-j]} \text{ for } p = 0,1,\ldots,k-2.$$

Definition 3.2. Let g be a real-valued function from R^k. Then g is said to be Schur-concave if

$$\underline{s} \underset{m}{<} \underline{t} \Rightarrow g(\underline{s}) \geq g(\underline{t}).$$

If $g(\underline{s}) \leq g(\underline{t})$, g is called Schur-convex.

Definition 3.3. Let A be a subset of R^k. We say that A is a Schur-concave set if the indicator function $I_A(\underline{u})$, is Schur-concave, i.e., if $\underline{u}' \underset{m}{<} \underline{u}$, $\underline{u} \in A \Rightarrow \underline{u}' \in A$.

Remarks. (from Marshall and Olkin (1974)).

(a) If g is Schur-concave or Schur-convex then g is permutation-symmetric.

(b) If g (or log g) is concave and permutation-symmetric then g is Schur-concave. So if A is a convex and permutation-symmetric set, then A is a Schur-concave set.

(c) A Schur-concave function achieves maximum at a point where the coordinates are equal.

We will assume that the marginal density $f(x-\theta)$ has monotone likelihood ratio (MLR) in x. The joint density is $h(\underline{x}-\underline{\theta}) = \prod_{i=1}^{k} f(x_i-\theta_i)$. From Marshall and Olkin (1974), $h(\underline{x})$ is Schur-concave if and only if log f is concave, i.e. f is strongly unimodal. For a discussion of strongly unimodal densities we refer to Hájek and Šidák (1967) and Ibragimov (1956).

From Lehmann (1959), p. 330, we have that log f is concave if and only if $f(x-\theta)$ has MLR in x. So our assumption of MLR is equivalent with the assumption that the joint density $h(\underline{x})$ is Schur-concave.

Marshall and Olkin (1974) deal with preservation theorems of Schur-concavity for the location-model, and is the main reference for this section and Section 5. Articles dealing with the concept of Schur-concave functions for other types of distributions are Proschan and Sethuraman (1977), Nevius, Proschan and Sethuraman (1977), and Hollander, Proschan and Sethuraman (1977).

A reasonable requirement on a procedure is that it is just.

Definition 3.3. (See Nagel (1970)). ψ is said to be just if $\psi_i(\underline{x})$ is non-decreasing in x_i and non-increasing in x_j, $j \neq i$; for $i = 1,..,k$.

Since we are dealing with the location-parameter case it is natural that a selection procedure is also translation-invariant. It is readily shown (see e.g. Berger and Gupta (1977)) that ψ is a just and translation-invariant procedure if and only if

(i) ψ_i is a function only of the (k-1)-differences $\{x_j - x_i : j \neq i\}$, and

(ii) If $x_j - x_i \leq y_j - y_i$, $\forall j \neq i$, then $\psi_i(\underline{x}) \geq \psi_i(\underline{y})$.

Let $\underline{X}^* = \{X_j - X_i : j \neq i\}$. Then $\underline{X}^*$ has a location-density with parameter $\underline{\theta}^* = \{\theta_j - \theta_i : j \neq i\}$. From well-known properties of a location-family of distributions (see Lehmann (1955) and Alam (1973)) we have the following result.

Lemma 3.1. *$\underline{X}$ has density $h(\underline{x}-\underline{\theta})$. Let ψ be a just and translation-invariate procedure. If $\theta_j' - \theta_i' \leq \theta_j - \theta_i$ for all $j \neq i$, then*

$$E_{\underline{\theta}'}(\psi_i) \geq E_{\underline{\theta}}(\psi_i) \text{ for } i = 1,\ldots,k.$$

In particular,

$$\inf_{\underline{\theta} \in \Omega} P_{\underline{\theta}}(CS|\psi) \text{ occurs when } \theta_1 = \ldots = \theta_k.$$

Remarks. Since the distribution of $\underset{\sim}{X}^*$ depends only on $\underset{\sim}{\theta}^*$, then $E_{\underset{\sim}{\theta}}(\psi_i)$ depends only on $\underset{\sim}{\theta}^*$. If we want ψ to satisfy the basic condition (2.3) with equality we must have

(3.3) $$E_{\underset{\sim}{0}}(\psi_i) = \gamma, \text{ for } i = 1,\ldots,k; \quad \underset{\sim}{0} = (0,\ldots,0).$$

We are now in a position to define the class of Schur-procedures.

Definition 3.4. A subset selection procedure $\psi = (\psi_1,\ldots,\psi_k)$ is said to be a Schur-procedure if

(i) ψ is just and translation-invariant.

(ii) ψ_i is the same Schur-concave function of $\underset{\sim}{x}_i^* = \{x_j - x_i : j \neq i\}$, $\forall i$, i.e. $\psi_i(\underset{\sim}{x}) = \psi'(\underset{\sim}{x}_i^*)$ for some Schur-concave function ψ': $R^{k-1} \to R$, for $i = 1,\ldots,k$.

Consider now the case where ψ is a non-randomized procedure, i.e.

$$\psi_i(\underset{\sim}{x}) = I_{A_i}(\underset{\sim}{x}), \quad A_i \subset R^k.$$

Then ψ is a Schur-procedure if

$$\psi_i(\underset{\sim}{x}) = I_B(\underset{\sim}{x}_i^*)$$

for some $B \subset R^{k-1}$ and

(3.4) B is a monotone decreasing set, i.e. if $\underset{\sim}{u} \in B$ and $v_j \leq u_j$, $j = 1,\ldots,k-1$, then $\underset{\sim}{v} \in B$

and

(3.5) B is a Schur-concave set.

As the following observation indicates, many reasonable procedures are Schur-procedures. Consider the class C discussed by Seal (1955), which can be described as follows.

Let $X'_{[1]} \leq \cdots \leq X'_{[k-1]}$ be the ordered $\{X_j : j \neq i\}$. Let $\underline{c} = (c_1,\ldots,c_{k-1}) \in R^{k-1}$, $c_i \geq 0$ and $\sum_{i=1}^{k-1} c_i = 1$. The procedure $\psi^{\underline{c}}$ is defined by

$$\psi_i^{\underline{c}} = 1 \text{ iff } \sum_{j=1}^{k-1} c_j X'_{[j]} \leq d(\underline{c}). \tag{3.6}$$

Here $d(\underline{c})$ is determined such that (2.3) holds with equality. From Lemma 3.1 and (3.3), $d(\underline{c})$ is determined by

$$P\left(\sum_{j=1}^{k-1} c_j Y_{[j]} - Y_0 \leq d(\underline{c})\right) = \gamma \tag{3.7}$$

where $Y_0, Y_1, \ldots, Y_{k-1}$ are i.i.d. with density $f(y)$, and $Y_{[1]} \leq \cdots \leq Y_{[k-1]}$ are the ordered $(Y_1,\ldots,Y_{k-1})$. Note that condition (2.2) implies that $\gamma \geq 1/k$ and $d(\underline{c}) \geq 0$.

Then

$$\mathcal{C} = \{\psi^{\underline{c}} : \sum_{i=1}^{k} c_i = 1\}.$$

All procedures in $\mathcal{C}$ are just and translation-invariant.

<u>Lemma 2.3.</u> *Let*

$$\mathcal{C}_0 = \{\psi^{\underline{c}} \in \mathcal{C} : c_1 \leq \cdots \leq c_{k-1}\}.$$

Assume $\psi^{\underline{c}} \in \mathcal{C}$. *Then*

$$\psi^{\underline{c}} \in \mathcal{C}_0 \iff \psi^{\underline{c}} \textit{ is a Schur-procedure.}$$

<u>Proof</u>. Let $A = \{\underline{y} \in R^{k-1} : \sum_{j=1}^{k-1} c_j y_{[j]} \leq d(\underline{c})\}$. $\psi_i^{\underline{c}}(\underline{x}) = I_A(\underline{x}_i^*)$. Clearly A is a monotone decreasing set.

($\Rightarrow$) We must show that A is a Schur-concave set. Now, assume $\underline{y} \in A$ and $\underline{y}' \underset{m}{<} \underline{y}$.

$$\sum_{j=1}^{k-1} c_j y_{[j]} = c_1 \sum_{j=1}^{k-1} y_{[j]} + (c_2-c_1)\sum_{j=2}^{k-1} y_{[j]} + \ldots + (c_i-c_{i-1})\sum_{j=i}^{k-1} y_{[j]}$$
$$+ \ldots + (c_{k-1}-c_{k-2})y_{[k-1]}$$
$$\geq c_1 \sum_{j=1}^{k-1} y'_{[j]} + (c_2-c_1)\sum_{j=2}^{k-1} y'_{[j]} + \ldots + (c_i-c_{i-1})\sum_{j=i}^{k-1} y'_{[j]}$$
$$+ \ldots + (c_{k-1}-c_{k-2})y'_{[k-1]}$$
$$= \sum_{j=1}^{k-1} c_j y'_{[j]}.$$

Hence $\underline{y}' \in A$.

($\Leftarrow$) Assume $c_j < c_{j'}$ for some $j' < j$. We shall show that A is not a Schur-concave set. Clearly there exists i such that $c_i < c_{i-1}$. Let $\underline{y}' \underset{m}{<} \underline{y}$ and such that $y_{[j]} = y_j$, $y'_{[j]} = y'_j$, $\forall j$, and

(i) $\sum_{j=1}^{k-1} c_j y_j = d(\underline{c})$

(ii) $y'_j = y_j$ for $j \neq i, i-1$.

(iii) $y'_i < y_i$

Then
$$\sum_{j=1}^{k-1} c_j y_j - \sum_{j=1}^{k-1} c_j y'_j = (c_i - c_{i-1}) \left(\sum_{j=1}^{k-1} y_j - \sum_{j=i}^{k-1} y'_j\right) < 0.$$

Hence $\underline{y}' \notin A$. Q.E.D.

Remark. The problem of choosing a rule from the class C was considered by Seal (1955,1957) and Gupta (1956), for the case of normal populations. Two rules were proposed. They correspond respectively to $c_1=\ldots=c_{k-1}=1/(k-1)$ and $c_1=\ldots=c_{k-2}=0$, $c_{k-1}=1$. Let us call them ψ^a and ψ^m ("a" for average, "m" for maximum). Hence

(3.8) $$\psi_i^a = 1 \text{ iff } X_i \geq (1/k-1) \sum_{j \neq i} X_j - c$$

(3.9) $$\psi_i^m = 1 \text{ iff } X_i \geq \max_{1\leq j\leq k} X_j - d.$$

ψ^a was proposed by Seal (1955), and ψ^m was proposed by Gupta (1956), Seal (1957) and also by Paulson (1949). ψ^a, ψ^m have since been the main contenders for this problem. We see that both ψ^a, ψ^m are in C_0, and are therefore Schur-procedures.

4. Two Fundamental Convexity Lemmas

The monotonicity-results for the different risk criteria are based on two lemmas that deal with convexity-properties for a certain sum of functions.

Lemma 4.1. *Let* g *be a real-valued, Schur-concave function from* R^{k-1}, *non-increasing in each component. Define* G: $R^k \to R$ *by:*

(4.1) $$G(\underline{u}) = \sum_{i=1}^{k-1} g(\underline{u}_i^*)$$

where

(4.2) $$u_i^* = \{u_{[j]}-u_{[i]}: \; j \neq i\} \text{ and } u_{[1]} \leq \cdots \leq u_{[k]}.$$

Let $\underline{v} \underset{m}{<} \underline{u}$ *and* $v_{[i]} \geq u_{[i]}$ *for* $i = 1,\ldots,k-1$ (*i.e.* $\sum_{i=1}^{k} v_i = \sum_{i=1}^{k} u_i$ *and* $v_{[i]} \geq u_{[i]}$ *for* $i = 1,\ldots,k-1$). *Then*

$$G(\underline{u}) \leq G(\underline{v})$$

Proof. Let $\underline{v} \underset{m}{<} \underline{u}$. Since $G(\underline{u})$ only depends on $(u_{[1]},\ldots,u_{[k]})$, G is permutation-symmetric and we can assume that

$$u_1 \leq \cdots \leq u_k \text{ and } v_1 \leq \cdots \leq v_k.$$

From Hardy, Littlewood and Polya (1952, p. 47) there exists a finite sequence $\underline{u}^0 \underset{m}{>} \underline{u}^1 \underset{m}{>} \cdots \underset{m}{>} \underline{u}^\ell$ such that $\underline{u}^0 = \underline{u}$, $\underline{u}^\ell = \underline{v}$ and for each i, $\underline{u}^i$,

$\underline{u}^{i+1}$ differ in two coordinates only. Now $u_k > v_k$ and $u_i \leq v_i$ for $i \leq k-1$. Hence we can assume $u_k^i > u_k^{i+1}$ for $i = 0,\ldots,\ell-1$. We may therefore assume without loss of generality that

$$u_k > v_k \text{ and } u_p < v_p \text{ for some } p \leq k-1,$$

and

$$u_i = v_i \text{ for } i \neq p,k, \text{ and } u_k + u_p = v_k + v_p.$$

Let now $i \leq k-1$, $i \neq p$. Then

$$\sum_{j\neq i}(u_j-u_i) = \sum_{j\neq i}(v_j-v_i). \tag{4.3}$$

Since $u_k > v_k$ we have

$$\begin{aligned}
&\sum_{j=n}^{k}(u_j-u_i) > \sum_{j=n}^{k}(v_j-u_i) = \sum_{j=n}^{k}(v_j-v_i) \text{ for } n > p\\
&\sum_{j=n}^{k}(u_j-u_i) = \sum_{j=n}^{k}(v_j-v_i) \qquad \text{for } n \leq p.
\end{aligned} \tag{4.4}$$

From (4.3) and (4.4) it follows that $\underline{v}_i^* \underset{m}{<} \underline{u}^*_i$, which implies that

$$g(\underline{u}_i^*) \leq g(\underline{v}_i^*) \text{ for } i \leq k-1, \quad i \neq p. \tag{4.5}$$

Since g is non-increasing in each component we have that

$$g(\underline{u}_p^*) \leq g(\underline{v}_p^*). \tag{4.6}$$

(4.5) and (4.6) $\Rightarrow G(\underline{u}) \leq G(\underline{v})$. Q.E.D.

Remark. A natural question is whether $G(\underline{u})$ is in fact Schur-concave. The reason why we need $v_{[i]} \geq u_{[i]}$ for $i \leq k-1$ is that we can then assume in the proof that $u_k^i > u_k^{i+1}$ for all $i = 0,1,\ldots,\ell-1$. This is essential in proving Lemma 4.1. Of course this does not mean that this condition is necessary. However the counterexample given below shows that G is not in general Schur-concave if g is a Schur-concave function, non-increasing in each component.

Let us first state a result that gives necessary and sufficient conditions for a permutation-symmetric function to be Schur-concave. Suppose $h: R^m \to R$ is permutation-symmetric and differentiable. Then, from Ostrowski (1952), we have that h is Schur-concave if and only if

$$(4.7) \qquad \left(\frac{\partial h(\underline{x})}{\partial x_i} - \frac{\partial h(\underline{x})}{\partial x_j}\right)(x_i - x_j) \leq 0 \quad \forall i \neq j \text{ and } \forall (x_1,\ldots,x_m).$$

(The inequality is reversed for Schur-convex functions.)

As a counterexample to Schur-concavity of G, let $k = 3$ and $g(x_1,x_2) = \Phi(-x_1-x_2)$, where Φ is the distribution function of the N(0,1)-distribution. g is Schur-concave and decreasing in each component. G depends only on $(u_{[1]},u_{[2]},u_{[3]})$ so we can assume $u_1 < u_2 < u_3$. Then

$$G(\underline{u}) = \Phi(2u_1-u_2-u_3) + \Phi(2u_2-u_1-u_3).$$

(Note: G corresponds here to $S'(\underline{\theta},\psi^a)$, ψ^a given by (3.8), with $\gamma = 1/2$ and normal populations.)

It is readily seen that

$$\frac{\partial G(\underline{u})}{\partial u_1} < \frac{\partial G(\underline{u})}{\partial u_2} .$$

So (4.7) is not satisfied, and G is not Schur-concave in $\underline{u}$.

The next question we want to ask is: What kind of assumptions on g do we need for G to be Schur-concave? The following result gives sufficient conditions.

<u>Lemma 4.2</u>. *Let* g *be a real-valued function from* R^{k-1}, *non-increasing in each component, admitting partial derivatives. Assume that* g *is permutation-symmetric and concave.* G *is defined by* (4.1). *Then* $G(\underline{u})$ *is Schur-concave in* $\underline{u}$.

<u>Proof</u>. Let $\underline{v} \underset{m}{<} \underline{u}$. We can assume $v_1 \leq \ldots \leq v_k$ and $u_1 \leq \ldots \leq u_k$. It is enough to consider the case where $v_p + v_q = u_p + u_q$, $v_q < u_q$ for some $p < q$ and $v_i = u_i$ for $i \neq p,q$. As in the proof of Lemma 4.1 we see that

$$g(\underline{u}_i^*) \leq g(\underline{v}_i^*) \text{ for } i \leq k-1, \quad i \neq p,q$$

(since $\underline{v}_i^* \underset{m}{<} \underline{u}_i^*$ for $i \neq p,q$) and

$$g(\underline{u}_p^*) \leq g(\underline{v}_p^*).$$

So result is proven for $q = k$.

Next, assume that $q < k$. It remains to show

$$g(\underline{u}_p^*) + g(\underline{u}_q^*) \leq g(\underline{v}_p^*) + g(\underline{v}_q^*). \tag{4.8}$$

Let $\underline{e}_i \in R^{k-1}$ denote the vector $(1,..1,2,1,..,1)$ where the i'th component is 2 and the other components are 1. Define for $y \in R$, and fixed $\underline{u},\underline{v}$, the vectors $\underline{y}$, $\underline{z} \in R^{k-1}$ by

$$\underline{y} = \underline{y}(y) = \underline{u}_p^* - y\underline{e}_{q-1}$$

$$\underline{z} = \underline{z}(y) = \underline{u}_p^* + y\underline{e}_p.$$

Let

$$t(y) = g(\underline{y}) + g(\underline{z}).$$

We see that

$$t(0) = g(\underline{u}_p^*) + g(\underline{u}_p^*),$$

$$t(y^0) = g(\underline{v}_p^*) + g(\underline{v}_p^*),$$

where $y^0 = u_q - v_q = v_p - u_p$. Since g is concave it follows that t is concave in y. Let $y^1 = (u_q - u_p)/2$. Now $y^1 \geq y^0$, so to show (4.8) it is therefore sufficient to show that $t'(y^1) = 0$.

Let now $g_i = \partial g/\partial y_i$; $i = 1,\dots,k-1$. Then

$$t'(y) = -\sum_{\substack{i=1\\ i\neq q-1}}^{k-1} g_i(\underline{y}) - 2g_{q-1}(\underline{y}) + \sum_{i=1}^{k-1} g_i(\underline{z}) + 2g_p(\underline{z}).$$

Let $a = (1/2)(u_p + u_q)$. Then

$$\underline{y}(y^1) = (u_1,\dots,u_{p-1},u_{p+1},\dots,0,u_{q+1},\dots,u_k) - a(1,\dots,1)$$

$$\underline{z}(y^1) = (u_1,\dots,u_{p-1},0,u_{p+1},\dots,u_{q-1},u_{q+1},\dots,u_k) - a(1,\dots,1).$$

Let π be a permutation of $(1,\dots,k)$ such that πi is the new position of element i, i.e. $\pi(1,\dots,k) = (\pi^{-1}1,\dots,\pi^{-1}k)$. Then the permutation $\pi\underline{y}$ of $\underline{y}$ is defined by $(\pi\underline{y})_i = y_{\pi^{-1}i}$. Let π_0 be the permutation defined by:

$$\pi_0^{-1} i = i \qquad \text{for } i \leq p-1,\ \ i \geq q$$

$$\pi_0^{-1} i = i-1 \quad \text{for } p+1 \leq i \leq q-1$$

$$\pi_0^{-1} p = q-1.$$

We see that $\underline{z}(y^1) = \pi_0\underline{y}(y^1)$. Since g is permutation-invariant, the partial derivatives satisfy

$$g_{\pi^{-1}i}(\underline{y}) = g_i(\pi\underline{y}) \quad \forall(\pi,i).$$

It follows that

$$g_i(\underline{z}) = g_{\pi_0^{-1}i}(\underline{y}) \text{ when } y = y^1.$$

Hence

$$t'(y^1) = -\sum_{\substack{i=1\\ i\neq q-1}}^{k-1} g_i(\underline{y}) - 2g_{q-1}(\underline{y}) + \sum_{\substack{i\leq p-1\\ i\geq q}} g_i(\underline{y}) + 2g_{q-1}(\underline{y}) + \sum_{i=p+1}^{q-1} g_{i-1}(\underline{y}) = 0.$$

Q.E.D.

Remark. The condition on g in Lemma 4.2 implies that g is Schur-concave.

5. Properties of Schur-Procedures

From Marshall and Olkin (1974) we have the following two results.

Lemma 5.1. *Let* $Y_1,\dots,Y_m$ *have a joint density* $h(\underset{\sim}{y}-\underset{\sim}{\mu})$, *where* $h(\underset{\sim}{y})$ *is Schur-concave. Assume* $v(\underset{\sim}{y})$ *is a Schur-concave function in* $\underset{\sim}{y}$ *and bounded. Then*

$$E_{\underset{\sim}{\mu}} v(\underset{\sim}{Y}) = \int v(\underset{\sim}{y})h(\underset{\sim}{y}-\underset{\sim}{\mu})d\nu(\underset{\sim}{y}) \text{ is Schur-concave in } \underset{\sim}{\mu}.$$

Lemma 5.2. *Let* $\phi_y(u)$ *be linear and increasing in* u *for all* y. *Assume* $U_1,\dots,U_m,Y$ *are independent, and that* U_i *have a common density* f(u) *and* log f(u) *is concave. Then the joint density of* $Y_i = \phi_Y(U_i)$, $i = 1,\dots,m$ *is Schur-concave.*

Applying Lemma 5.1 and Lemma 5.2 we obtain the first interesting result about Schur-procedures.

Theorem 5.1. *Let* $X_1,\dots,X_k$ *be independent.* X_i *has density* $f(x-\theta_i)$ *and* f *has* MLR *in* x. *Let* ψ *be a Schur-procedure. Then*

$$E_{\underset{\sim}{\theta}}(\psi_i) \text{ is Schur-concave in } \underset{\sim}{\theta}^* = \{\theta_j-\theta_i:\ j \neq i\}.$$

Proof. ψ_i is a Schur-concave function of $\underset{\sim}{X}^* = \{X_j-X_i:\ j \neq i\}$. $\underset{\sim}{X}^*$ has density $h(\underset{\sim}{x}^*-\underset{\sim}{\theta}^*)$ where $h(\underset{\sim}{y})$, $\underset{\sim}{y} \in R^{k-1}$, is the density of $\underset{\sim}{Y}^* = \{Y_j-Y_i:\ j \neq i\}$ and $Y_1,\dots,Y_k$ are i.i.d. with density f(y). From Lemma 5.2, $h(\underset{\sim}{y})$ is Schur-concave, since log f is concave. The result now follows from Lemma 5.1. Q.E.D.

Consider now the risk function $S'(\underset{\sim}{\theta},\psi)$ defined by (2.4). One of the main results for Schur-procedures is a monotonicity-result for $S'(\underset{\sim}{\theta},\psi)$.

Theorem 5.2. *Assume the conditions of Theorem 5.1 hold. Let* ψ *be a Schur-procedure and let*

(5.1) $$\underline{\theta}' \underset{m}{\leq} \underline{\theta} \text{ and } \theta'_{[i]} \geq \theta_{[i]} \text{ for } i = 1,\ldots,k-1.$$

Then

(5.2) $$S'(\underline{\theta},\psi) \leq S'(\underline{\theta}',\psi).$$

Proof. Let g be defined by

$$g(\underline{\theta}_i^*) = E_{\underline{\theta}}\{\psi_{(i)}\}.$$

$\underline{\theta}_i^*$ is defined in (4.2). Then

$$S'(\underline{\theta},\psi) = \sum_{i=1}^{k-1} g(\underline{\theta}_i^*).$$

From Theorem 5.1 and Lemma 3.1 we see that the assumptions in Lemma 4.1 are satisfied and result follows. Q.E.D.

As mentioned in Section 3 a nice property of Schur-concave functions is that they achieve their maximum at a point where all components are equal. $S'(\underline{\theta},\psi)$ is not quite Schur-concave, but by applying Theorem 5.2 we can show a similar result for $S'(\underline{\theta},\psi)$ over certain subsets of Ω.

Theorem 5.3. *Assume the conditions of Theorem 5.1 hold and that* ψ *is a Schur-procedure. Let* $\delta \geq 0$ *and define the slippage-set*

(5.3) $$\Omega_k(\delta) = \{\underline{\theta} \in \Omega\colon \theta_{[k]}-\theta_{[k-1]} \geq \delta + \sum_{i=1}^{k-2} (\theta_{[k-1]}-\theta_{[i]})\}.$$

Then

$$\sup_{\underline{\theta} \in \Omega_k(\delta)} S'(\underline{\theta},\psi) = S'(\underline{\theta}^{\delta},\psi)$$

where

(5.4) $$\theta^{\delta}_{[k]} = \delta,\ \theta^{\delta}_{[i]} = 0 \text{ for } i = 1,\ldots,k-1.$$

Proof. Let $\underline{\theta} \in \Omega_k(\delta)$. We can assume $\theta_{[k-1]}-\theta_{[1]} > 0$; otherwise the theorem is trivial. Also, $S'(\underline{\theta},\psi)$ is permutation-symmetric so we can let $\theta_1 \leq \cdots \leq \theta_k$. From Lemma 3.1 it follows that $S'(\underline{\theta},\psi)$ is non-increasing in θ_k. Therefore we may take

$$\theta_k = \theta_{k-1} + \delta + \sum_{i=1}^{k-2} (\theta_{k-1}-\theta_i).$$

Let now $\theta_0 = \theta_{k-1}$. $\underline{\theta}^0 = (\theta_0,\ldots,\theta_0,\theta_0+\delta)$. Then $\underline{\theta}^0 \underset{m}{<} \underline{\theta}$ and $\theta^0_{[i]} \geq \theta_{[i]}$ for $i \leq k-1$. From Theorem 5.2 it follows that

$$S'(\underline{\theta},\psi) \leq S'(\underline{\theta}^0,\psi) = S'(\underline{\theta}^\delta,\psi). \qquad \text{Q.E.D.}$$

Remark. There is one Schur-procedure, ψ^m given by (3.9), that has been shown to achieve its maximum at $\underline{\theta}^\delta$ over the set $\Omega(\delta) = \{\underline{\theta}: \theta_{[k]}-\theta_{[k-1]} \geq \delta\}$ $\delta \geq 0$. More precisely, Gupta (1965) showed that

$$\sup_{\underline{\theta} \in \Omega(\delta)} S'(\underline{\theta},\psi^m) = S'(\underline{\theta}^\delta,\psi^m). \tag{5.5}$$

(5.5) is not true in general for Schur-procedures. Consider for example ψ^a, given by (3.8), and let $\delta = 0$. Then it is readily seen that

$\sup_{\underline{\theta} \in \Omega(\delta)} S'(\underline{\theta},\psi^a) > S'(\underline{\theta}^\delta,\psi^a)$ if $\gamma < (k-2)/(k-1)$.

We shall next consider the corresponding problem for the risk

$$B(\underline{\theta},\psi) = E_{\underline{\theta}}(\text{number of bad populations selected}\,|\,\psi)$$

(see Definition 2.2 and (2.5)).

Theorem 5.4. *Assume the conditions of Theorem 5.1 hold and that* ψ *is a Schur-procedure. Let*

$$\Omega_p(\Delta) = \{\underline{\theta} \in \Omega: \theta_{[k]}-\theta_{[p]} < \Delta \ \&\ \theta_{[k]}-\theta_{[p-1]} \geq \Delta + \sum_{i=1}^{p-2}(\theta_{[p-1]}-\theta_{[i]})\}$$

for $p = 1,\ldots,k-1$.

Let

(5.6) $$\Omega_1 = \bigcup_{p=1}^{k} \Omega_p(\Delta).$$

Then

$$\sup_{\underline{\theta} \in \Omega_1} B(\underline{\theta},\psi) = B(\underline{\theta}^{\Delta},\psi)$$

where $\underline{\theta}^{\Delta}$ *is defined by* (5.4) *for* $\delta = \Delta$.

Proof. Let $\underline{\theta} \in \Omega_1$. We can assume $\theta_1 \leq \ldots \leq \theta_k$. Let first $\underline{\theta} \in \Omega_k(\Delta)$. Then $B(\underline{\theta},\psi) = S'(\underline{\theta},\psi)$. From Theorem 5.3 it follows that

$$B(\underline{\theta},\psi) \leq B(\underline{\theta}^{\Delta},\psi).$$

Now let $\underline{\theta} \in \Omega_1 - \Omega_k(\Delta)$, i.e. $\theta_k - \theta_p < \Delta$ and $\theta_k - \theta_{p-1} \geq \Delta + \sum_{i=1}^{p-2} (\theta_{p-1} - \theta_i)$, for some p, $1 \leq p \leq k-1$. In this case

$$B(\underline{\theta},\psi) = \sum_{i=1}^{p-1} E_{\underline{\theta}}(\psi_i).$$

Let $\underline{\theta}'$ be defined by

$$\theta_k' = \theta_k,\ \theta_i' = \theta_i, \text{ for } i \leq p-1 \text{ and } \theta_i' = \theta_{p-1} \text{ for } i \geq p,..,k-1.$$

Clearly, from Lemma 3.1

$$B(\underline{\theta},\psi) \leq \sum_{i=1}^{p-1} E_{\underline{\theta}'}(\psi_i) \leq \sum_{i=1}^{k-1} E_{\underline{\theta}'}(\psi_i) = B(\underline{\theta}',\psi)$$

Now $\underline{\theta}' \in \Omega_k(\Delta)$, so the result follows from Theorem 5.3. Q.E.D.

Remarks. (a) As expected, since B has fewer terms than S' when $\theta_{[k]} - \theta_{[k-1]} < \Delta$ we get a stronger result for B than for S' in the sense that $\Omega_1 \supset \Omega_k(\Delta)$.

(b) Ω_1 consists of the cases where the good populations have "slipped" from the bad populations. Also Ω_1 contains the "classical" slippage-set $\{\underline{\theta} \in \Omega\colon \theta_{[1]} = .. = \theta_{[k-1]} \ \&\ \theta_{[k]} - \theta_{[k-1]} \geq \Delta\}$.

(c) In Part II, Theorems 5.3 and 5.4 are applied to derive a certain optimal procedure which will be minimax with respect to slippage sets of the type Ω_1 and $\Omega_k(\Delta)$.

The last risk functions we will consider are $L(\underline{\theta},\psi)$ and $\ell(\underline{\theta},\psi)$ defined by (2.6) and (2.7). As mentioned in Section 2, $L(\underline{\theta},\psi)$ and $\ell(\underline{\theta},\psi)$ are in many cases more appropriate criteria, since they place more weights on the worst populations.

Let us consider the following class K of non-randomized procedures. $\psi \in K$ if and only if ψ is just, translation-invariant and satisfies

(5.7) $$\psi_i(\underline{x}) = I_A(\underline{x}_i^*) \text{ for } i = 1,\dots,k$$

where A is a permutation-symmetric convex set. Since ψ is just, A is also a monotone decreasing set.

K is called the class of non-randomized, convex procedures and is a subclass of the Schur-procedures.

ψ^a and ψ^m are both convex procedures. Our aim is to show that $L(\underline{\theta},\psi)$ is Schur-concave for the normal case, if $\psi \in K$. More precisely:

Theorem 5.5. *Assume* $X_1,\dots,X_k$ *are independent.* X_i *is* $N(\theta_i,\sigma^2)$ *where* σ^2 *is known. Let* $\psi \in K$. *Then*

$$L(\underline{\theta},\psi) = \sum_{i=1}^{k-1} \log\{E_{\underline{\theta}}\psi_{(i)}\}$$

is Schur-concave in $\underline{\theta}$.

As an immediate consequence we get the following result.

Corollary 5.1. *Assume the conditions of Theorem 5.5 hold. Let*

(5.8) $$\Omega(\delta) = \{\underline{\theta} \in \Omega: \theta_{[k]}-\theta_{[k-1]} \geq \delta\},\ \delta \geq 0.$$

Then

$$\sup_{\underline{\theta} \in \Omega(\delta)} L(\underline{\theta},\psi) = L(\underline{\theta}^\delta,\psi),$$

and

$$\sup_{\underline{\theta} \in \Omega} \ell(\underline{\theta},\psi) = \ell(\underline{\theta}^\Delta,\psi).$$

Here $\underline{\theta}^{\delta}$, $\underline{\theta}^{\Lambda}$ *are defined by* (5.4).

Remark. In Part II, Corollary 5.1 is used to derive an optimal procedure for $L(\underline{\theta},\psi)$ and $\ell(\underline{\theta},\psi)$. The procedure is minimax for ℓ with respect to the whole parameter space Ω.

In order to prove Theorem 5.5 we will apply Lemma 4.2. First we need the following basic result, which follows from Kanter (1977).

Lemma 5.3. *Let* $\underline{Y} = (Y_1,\ldots,Y_m)$ *have a density* $h(\underline{y}-\underline{\mu})$ *with respect to the Lebesgue-measure* ν. *Assume that* log h *is permutation-symmetric and concave. Let* A *be a permutation-symmetric convex subset of* R^m. *Then*

$$g(\underline{\mu}) = P_{\underline{\mu}}(\underline{Y} \in A) = \int_A h(\underline{y}-\underline{\mu})d\nu(\underline{y})$$

is log-concave and permutation-symmetric in μ.

Remarks. (a) If log h is concave and permutation-symmetric then h is Schur-concave. Hence the condition on h is stronger in Lemma 5.3 than in Lemma 5.1.

(b) Let $h(\underline{y}) = \prod_{i=1}^{m} f(y_i)$. Then log h is concave if and only if h is Schur-concave.

A class of log-concave densities is given in the following result.

Lemma 5.4. *Let* $h(\underline{x}) = f(\underline{x}\Lambda\underline{x}')$ *where* f *is a decreasing and log-concave function with continuous second-order derivative.* Λ *is non-negative definite. Then* **log** h *is concave.*

Proof. Let $\Lambda = (\lambda_{ij})$, $H = \log h$, $\ell = \log f$; ℓ', ℓ'' are first- and second-order derivatives of ℓ. Then

$$\ell' < 0,\ \ell'' \leq 0.$$

Let $H_{ij} = \frac{\partial^2 H}{\partial x_i \partial x_j}$. H is concave if $-(H_{ij})$ is non-negative definite. It is readily shown that

$$H_{ij}(\underline{x}) = 2\cdot \ell'(\underline{x}\Lambda\underline{x}')\cdot\lambda_{ij} + 4\ell''(\underline{x}\Lambda\underline{x}')(\sum_p x_p\lambda_{ip})(\sum_q x_q\lambda_{jq}).$$

Now

$$\sum_{i,j} y_i y_j (\sum_p x_p\lambda_{ip})(\sum_q x_q\lambda_{jq}) = (\sum_i \sum_p y_i x_p \lambda_{ip})^2 \geq 0.$$

Since Λ is non-negative definite it follows that $-(H_{ij})$ is non-negative definite. Q.E.D.

Remark. Lemma 5.4 shows that if $\underline{Y}$ is $N_m(0,\Sigma)$, Σ is positive definite, then $\underline{Y}$ has a log-concave density.

Proof of Theorem 5.5. Let g: $R^{k-1} \to R$ be defined by

$$g(\underline{\theta}_i^*) = \log\{E_{\underline{\theta}}\psi_{(i)}\}.$$

(g does not depend on i.) From Lemma 5.4 and the remark above we have that the density of $\underline{X}_i^* = \{X_j - X_i : j \neq i\}$ is log-concave and permutation-symmetric. From Lemma 5.3 it follows that g is concave and permutation-symmetric, since ψ_i is given by (5.7). From Lemma 3.1 g is non-decreasing in each component. From a general theorem of analytic integrals for the exponential family of distributions (first proved by Sverdrup (1953), see also Lehmann (1959), p. 52), it follows that g has partial derivatives.

The result now follows by applying Lemma 4.2. Q.E.D.

Acknowledgment. This paper is a part of the author's Ph.D. dissertation written under the supervision of Professor Erich L. Lehmann. I would also like to thank Professor Lucien LeCam and Professor Kjell Doksum for helpful discussions.

References

[1] Alam, K. (1973). On a multiple decision rule. Ann. Statist. 1, 750-755.

[2] Berger, R. L. and Gupta, S. S. (1977). Minimax subset selection with applications to unequal variance problems. Mimeograph Series #495, Dept. of Statist., Purdue University.

[3] Gupta, S. S. (1956). On a decision rule for a problem in ranking means. Mimeograph Series #150, Inst. of Statist., University of North Carolina at Chapel Hill.

[4] Gupta, S. S. (1965). On some multiple decision (selection and ranking) rules. Technometrics 7, 225-245.

[5] Hájek, J. and Šidák, Z. (1967). Theory of Rank Tests. Academic Press, New York.

[6] Hardy, G. H., Littlewood, J. E. and Pólya, G. (1952). Inequalities. 2nd ed. Cambridge University Press, Cambridge.

[7] Hollander, M., Proschan, F. and Sethuraman, J. (1977). Functions decreasing in transposition and their applications in ranking problems. Ann. Statist. 5, 722-734.

[8] Ibragimov, J. A. (1956). On the composition of unimodal distributions. (Russian). Teoriya veroyatnostey 1, 283-288.

[9] Kanter, M. (1977). Unimodality and dominace for symmetric random vectors. Trans. Amer. Math. Soc. 229, 65-85.

[10] Lehmann, E. L. (1955). Ordered families of distributions. Ann. Math. Statist. 26, 399-419.

[11] Lehmann, E. L. (1959). Testing Statistical Hypotheses. Wiley, New York.

[12] Marshall, A. W. and Olkin, I. (1974). Majorization in multivariate distributions. Ann. Statist. 2, 1189-1200.

[13] Nagel, K. (1970). On subset selection rules with certain optimality properties. Mimeograph Series #222, Dept. of Statist., Purdue University.

[14] Nevius, S. E., Proschan, F. and Sethuraman, J. (1977). Schur functions in statistics. II. Stochastic majorization. Ann. Statist. 5, 263-274.

[15] Ostrowski, A. (1952). Sur quelques applications des functions convexes et concaves au sens de I. Schur. J. Math. Pure Appl. 31, 255-292.

[16] Paulson, E. (1949). A multiple decision procedure for certain problems in the analysis of variance. Ann. Math. Statist. 20, 95-98.

[17] Proschan, F. and Sethuraman, J. (1977). Schur functions in statistics. I. The preservation theorem. Ann. Statist. 5, 256-263.

[18] Seal, K. C. (1955). On a class of decision procedures for ranking means of normal populations. Ann. Math. Statist. 26, 387-398.

[19] Seal, K. C. (1957). An optimum decision rule for ranking means of normal populations. Calcutta Statist. Assoc. Bull. 7, 131-150.

[20] Sverdrup, E. (1953). Similarity, unbiasedness, minimaxibility and admissibility of statistical test procedures. Skand. Aktuar. Tidskrift 36, 64-86.