A CLASS OF SCHUR-PROCEDURES AND MINIMAX THEORY FOR SUBSET SELECTION[1]

by

Jan F. Bjørnstad
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #80-17

July 1980

FOOTNOTES

[1]This research was supported by the Norwegian Research Council for Science and the Humanities, and also in part by the National Science Foundation Grant No. MCS-7903780. American Mathematical Society 1970 subject classifications. Primary 62F07; Secondary 62C99, 26A51.

Key words and phrases. Subset selection, location-model, Schur-concave functions, minimax procedures.

SUMMARY

A CLASS OF SCHUR-PROCEDURES AND MINIMAX

THEORY FOR SUBSET SELECTION

Abbreviated Title: MINIMAX SUBSET SELECTION

by

Jan F. Bjørnstad

University of California, Berkeley and Purdue University

The problem of selecting a random subset of good populations out of k populations is considered. The populations $\Pi_1,\ldots,\Pi_k$ are characterized by the location parameters $\theta_1,\ldots,\theta_k$ and Π_i is said to be a good population if $\theta_i > \max_{1\leq j\leq k} \theta_j-\Delta$, and a bad population if $\theta_i \leq \max_{1\leq j\leq k} \theta_j-\Delta$. Here, Δ is a specified positive constant.

A theory for a special class of procedures, called Schur-procedures, is developed, and applied to certain minimax problems. Subject to controlling the minimum expected number of good populations selected or the probability that the best population is in the selected subset, procedures are derived which minimize the expected number of bad populations selected or some similar criterion.

For normal populations it is known that the classical "maximum-type" procedure has certain minimax properties. In this paper, two other procedures are shown to have several minimax properties. One is the "average-type" procedure. The other procedure has not before been considered as a serious contender.

A CLASS OF SCHUR-PROCEDURES AND MINIMAX

THEORY FOR SUBSET SELECTION[1]

Abbreviated Title: MINIMAX SUBSET SELECTION

by

Jan F. Bjørnstad

University of California, Berkeley and Purdue University

1. Introduction, Basic Concepts and Notation

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

The k populations $\pi_1,\ldots,\pi_k$ are 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 statistic that is used to represent π_i . $\underline{X} = (X_1,\ldots,X_k)$. $X_1,\ldots,X_k$ are assumed to be independent and X_i has density $f(x-\theta_i)$ with respect to Lebesgue measure. We will assume that $f(x-\theta)$ has monotone likelihood ratio (MLR) in x. Let $p(\underline{x}-\underline{\theta}) = \Pi f(x_i-\theta_i)$. Normal populations with X_i being the sample mean from π_i is of course such a MLR-location model. Another example is the case of double exponential populations, letting X_i be the sample medians of say 2m+1 observations. Gupta and Leong (1979) have shown that the location-density of X_i has MLR in x.

The ordered θ_i are denoted by $\theta_{[1]} \le \cdots \le \theta_{[k]}$, and $\pi_{(i)}$, $X_{(i)}$ correspond to $\theta_{[i]}$. Π_i is called a best population if $\theta_i = \theta_{[k]}$. Following an idea similar to Lehmann (1961), π_i is said to be a good population if $\theta_i > \theta_{[k]}-\Delta$ and a bad population if $\theta_i \le \theta_{[k]}-\Delta$. Here Δ is a given positive constant.

For the risk criteria we shall consider, two subset selection procedures are equivalent if their individual selection probabilities are the same. Therefore we can define a subset selection procedure by $\psi(\underline{x}) = [\psi_1(\underline{x}),\ldots,\psi_k(\underline{x})]$, where $\psi_i(\underline{x}) = P(\text{selecting } \Pi_i | \underline{X} = \underline{x})$, for $i = 1,\ldots,k$.

We will usually require that at least one population is selected which implies that

$$(1.1) \qquad \sum_{i=1}^{k} \psi_i(\underline{x}) \geq 1, \ \forall \ \underline{x}.$$

A correct selection (CS) is defined to be a selection that includes the best population $\Pi_{(k)}$. The usual basic condition has been to require that

$$(1.2) \qquad \inf_{\underline{\theta}\in\Omega} P_{\underline{\theta}}(CS|\psi) = \inf_{\underline{\theta}\in\Omega} E_{\underline{\theta}}\{\psi_{(k)}\} \geq \gamma, \ \gamma < 1.$$

Here $\psi_{(i)}$ corresponds to $\theta_{[i]}$. The control condition we will mostly consider is to require that ψ satisfies

$$(1.3) \qquad \inf_{\underline{\theta}\in\Omega} R(\underline{\theta},\psi) \geq \gamma,$$

where $R(\underline{\theta},\psi)$ is the expected number of good populations selected. Let the class of procedures satisfying (1.3) be denoted by $\mathcal{D}(\gamma,\Delta)$. The condition

$$(1.4) \qquad \inf_{\underline{\theta}\in\Omega(\Delta)} P_{\underline{\theta}}(CS|\psi) \geq \gamma$$

is also of interest. (Also used by Ryan and Antle (1976).) Here $\Omega(\Delta) = \{\underline{\theta}: \theta_{[k]}-\theta_{[k-1]} \geq \Delta\}$. Let $\mathcal{D}'(\gamma,\Delta)$ be the class of procedures satisfying (1.4), so that $\mathcal{D}'(\gamma) = \mathcal{D}'(\gamma, 0)$ is the class of procedures satisfying (1.2). We see that $\mathcal{D}(\gamma,\Delta) \subset \mathcal{D}'(\gamma,\Delta)$.

It can be argued that it is unnecessary to require that we select $\Pi_{(k)}$ if $\theta_{[k]}$ is close to the other θ_i. Therefore (1.4) seems more appropriate than (1.2). However (1.4) does not control what happens on $\Omega^c(\Delta)$, leading then to control (1.3).

A fourth control condition, different in nature from the three mentioned above, is

$$\sup_{\underline{\theta}\in\Omega} S(\underline{\theta},\psi) \le \beta. \tag{1.5}$$

Here, $S(\underline{\theta},\psi) = \sum E_{\underline{\theta}}(\psi_i)$, i.e. S is the expected size of the selected subset. The class of procedures satisfying (1.5) is denoted by $\mathcal{D}_1(\beta)$.

Subject to a chosen control condition a procedure should exclude the bad or non-best populations. One criterion for excluding non-best populations is

$$S'(\underline{\theta},\psi) = \sum_{i=1}^{k-1} E_{\underline{\theta}}\{\psi_{(i)}\}. \tag{1.6}$$

$S'(\underline{\theta},\psi)$ is the expected number of non-best populations selected when $\theta_{[k-1]} < \theta_{[k]}$. Let $I_\Delta = I_\Delta(\underline{\theta}) = \{i: \theta_i \le \theta_{[k]}-\Delta\}$. The related criterion for excluding only the bad populations is

$$B(\underline{\theta},\psi) = \sum_{i\in I_\Delta} E_{\underline{\theta}}(\psi_i). \tag{1.7}$$

$B(\underline{\theta},\psi)$ is the expected number of bad populations selected. A widely used criterion in the literature has been $S(\underline{\theta},\psi)$. The author feels that $B(\underline{\theta},\psi)$ and $S'(\underline{\theta},\psi)$ are more appropriate risk functions than $S(\underline{\theta},\psi)$, since $S(\underline{\theta},\psi)$ includes the probability of selecting the best population.

The first papers on the subset selection problem dealt with normal means. Two rules were proposed, ψ^a and ψ^m (see Seal (1955, 1957) and Gupta (1956)). The procedures are given by

$$\psi_i^a = 1 \text{ iff } X_i \ge \frac{1}{k-1}\sum_{j\ne i} X_j - c \tag{1.8}$$

$$\psi_i^m = 1 \text{ iff } X_i \geq \max_{1 \leq j \leq k} X_j - d. \qquad (1.9)$$

Here c,d are determined such that the classical condition (1.2) holds with equality.

Our main concern is minimax theory. The lack of monotonicity results for the risk criteria considered has been the main drawback for deriving minimax results. Sections 2,3,4 deal with this problem. In Section 2 a special class, the class of Schur-procedures, is defined. Section 3 presents a convexity result that is used in Section 4 to show that the class of Schur-procedures has nice monotonicity-properties for the risk-criteria $S'(\underline{\theta},\psi)$ and $B(\underline{\theta},\psi)$.

Section 5 deals with minimax theorems in the general location-model for the risk functions $B(\underline{\theta},\psi)$, $S'(\underline{\theta},\psi)$ and $S(\underline{\theta},\psi)$, in the class $\mathcal{D}'(\gamma,\Delta)$. Gupta and Studden (1966) and Berger (1979) showed by using a monotonicity result for ψ^m given by Gupta (1965), that ψ^m is minimax with respect to the criteria $S'(\underline{\theta},\psi)$ and $S(\underline{\theta},\psi)$ in the class $\mathcal{D}'(\gamma)$ provided $f(x-\theta)$ has MLR in x. We show that ψ^a has the same minimax properties if γ is sufficiently large.

In Section 6 we consider normal populations. It is shown that a new procedure ψ^e, which was studied in a different context by Studden (1967), has certain minimax properties in the class $\mathcal{D}(\gamma,\Delta)$ for the criteria $B(\underline{\theta},\psi)$ and $S'(\underline{\theta},\psi)$. If $f(x)$ is the standard normal density, then ψ^e is defined by

$$\psi_i^e = 1 \text{ iff } Ce^{\Delta X_i} \geq \sum_{j \neq i} e^{\Delta X_j}. \qquad (1.10)$$

Here C is determined such that (1.4) is satisfied with equality. In Section 6 we also present some new minimax results for ψ^a. In Section 7 the minimax properties of ψ^e are proved. Section 8 proves the minimax results for ψ^a.

2. The Class of Schur-Procedures

Let $\underline{s}, \underline{t} \in \mathbb{R}^k$ with ordered components $s_{[1]} \leq \cdots \leq s_{[k]}$, $t_{[1]} \leq \cdots \leq t_{[k]}$. $\underline{s}$ is majorized by $\underline{t}$, $\underline{s} \underset{m}{<} \underline{t}$, if $\sum s_i = \sum t_i$ and

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

Let g be a real-valued function from $\mathbb{R}^k$. Then g is Schur-concave if $\underline{s} \underset{m}{<} \underline{t} \Rightarrow g(\underline{s}) \geq g(\underline{t})$. A subset A of $\mathbb{R}^k$ is a Schur-concave set if the indicator function $I_A(\underline{u})$, is Schur-concave, i.e., if $\underline{u}' \underset{m}{<} \underline{u}$, $\underline{u} \in A \Rightarrow \underline{u}' \in A \Rightarrow \underline{u}' \in A$.

If g (or log g) is concave and permutation-symmetric then g is Schur-concave. Also, a Schur-concave function achieves maximum at a point where the coordinates are equal.

Applying results from Marshall and Olkin (1974), Mudholkar (1969) and Lehmann (1959), p. 330 we see that the MLR-assumption is equivalent with the assumption that $p(\underline{x})$ is Schur-concave.

A procedure 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 $i = 1,\ldots,k$. Let $\mathcal{N}$ be the class of just, permutation-invariant and translation-invariant procedures. Nagel (1970) showed that if ψ is just and permutation-invariant then $E_{\underline{\theta}}\{\psi_{(i)}\}$ is nondecreasing in i. For $\underline{u} \in \mathbb{R}^k$ and $i \in \{1,\ldots,k\}$ let

$$\underline{u}_i^* = (u_1-u_i,\cdots, u_{i-1}-u_i, u_{i+1}-u_i,\cdots,u_k-u_i).$$

It is readily shown (see Berger and Gupta (1980)) that $\psi \in \mathcal{N}$ if and only if there exists a permutation-symmetric function $\psi': \mathbb{R}^{k-1} \to \mathbb{R}$ which is non-increasing in each component, such that for every i

$$\psi_i(\underline{x}) = \psi'(\underline{x}_i^*).$$

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

Definition 2.1. A subset selection procedure $\psi = (\psi_1,\ldots,\psi_k)$ is said to be a Schur-procedure if $\psi \in \mathcal{N}$ and ψ' is a Schur-concave function.

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

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

Then ψ is a Schur-procedure if $\psi_i(\underline{x}) = I_B(\underline{x}_i^*)$ for some Schur-concave set $B \subset \mathbb{R}^{k-1}$ and B is a monotone decreasing set, i.e. if $\underline{u} \in B$ and $v_j \le u_j$, $j = 1,\ldots,k-1$, then $\underline{v} \in B$.

$\underline{X}_i^*$ has a location-density with parameter $\underline{\theta}_i^*$. If ψ is translation-invariant, ψ_i is a function only of $\underline{X}_i^*$. From well-known properties of a location-family of distributions (See Lehmann (1955) and Alam (1973)) we have the following result.

Lemma 2.1. *Let* ψ *be a just and translation-invariant procedure. If for some* $i \in \{1,\ldots,k\}$, $\theta_j'-\theta_i' \le \theta_j-\theta_i$ *for all* $j \ne i$, *then* $E_{\underline{\theta}'}(\psi_i) \ge E_{\underline{\theta}}(\psi_i)$. *In particular,* $\inf_{\underline{\theta}\in\Omega} P_{\underline{\theta}}(CS|\psi)$ *occurs when* $\theta_1 = \ldots = \theta_k$.

Remark. If we want ψ to satisfy the basic condition (1.2) with equality we must have

$$E_{\underline{0}}(\psi_i) = \gamma, \text{ for } i = 1,\ldots,k; \qquad \underline{0} = (0,\ldots,0). \tag{2.1}$$

As the following observation indicates, many reasonable procedures are Schur-procedures. Consider the class $\mathcal{C}$ discussed by Seal (1955), which can be described as follows. Let $X'_{[1]} \le \cdots \le X'_{[k-1]}$ be the ordered $\{X_j : j \ne i\}$. Let $\underline{c} = (c_1,\ldots,c_{k-1}) \in \mathbb{R}^{k-1}$, $c_i \ge 0$ and $\sum c_i = 1$. The procedure $\psi^{\underline{c}}$ is defined by

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

Here $d(\underline{c})$ is determined such that (1.2) holds with equality, using Lemma 2.1 and (2.1). All such procedures belong to $\mathcal{N}$. Note that condition (1.1) implies that $d(\underline{c}) \geq 0$ which in turn implies that $\gamma \geq 1/k$. Seal's class is given by

$$\mathcal{C} = \{\psi^{\underline{c}} : \sum_{i=1}^{k} c_i = 1 \text{ and } \gamma \geq 1/k\}.$$

The following result is proved in Bjørnstad (1978).

Lemma 2.2. *Let* $\mathcal{C}_0 = \{\psi^{\underline{c}} \in \mathcal{C}: c_1 \leq \ldots \leq c_{k-1}\}$.
Assume $\psi^{\underline{c}} \in \mathcal{C}$.

Then $\psi^{\underline{c}} \in \mathcal{C}_0 \Leftrightarrow \psi^{\underline{c}}$ *is a Schur-procedure.*

Remark. The two procedures ψ^a, ψ^m defined by (1.8) and (1.9) are both in $\mathcal{C}_0$, and are therefore Schur-procedures.

To show that the procedure ψ^e, defined by (1.10), is a Schur-procedure, we can use the following result from Ostrowski (1952). A permutation-symmetric and differentiable function, $h: \mathbb{R}^m \to \mathbb{R}$, is Schur-concave if and only if $(\partial h(\underline{x})/\partial x_i - \partial h(\underline{x})/\partial x_j)(x_i - x_j) \leq 0$ $\forall i \neq j$ and $\forall (x_1, \ldots, x_m)$. Using this result it is readily seen that ψ^e is a Schur-procedure.

3. A Fundamental Convexity Lemma

Definition 3.1. For $\underline{u}, \underline{v} \in \mathbb{R}^n$, $\underline{v} \prec \underline{u}$ iff for some $a \geq 0$

$$(v_1 + a, \ldots, v_n + a) \underset{m}{<} (u_1, \ldots, u_n).$$

The following result is then easily proved.

Lemma 3.1. $g: \mathbb{R}^n \to \mathbb{R}$ *is nonincreasing with respect to* "$\prec$" *iff*

(*i*) g *is Schur-concave*

and (*ii*) $g(\underline{x}) \geq g(x_1+a,\ldots,x_n+a)$ $\forall \underline{x}$, $\forall\, a \geq 0$.

i.e. g *is non-increasing for simultaneous shifts in all components.*

The monotonicity-results for the different risk criteria are based on the following lemma that deals with the convexity-property for a certain sum of functions.

Lemma 3.2. *Let* g *be a real-valued, Schur-concave function from* $\mathbb{R}^{k-1}$, *non-increasing for simultaneous shifts in all components. Define* $G: \mathbb{R}^k \to \mathbb{R}$ by:

$$G(\underline{u}) = \sum_{i=1}^{k-1} g(\underline{u}^*_{[i]})$$

where for $i = 1,\ldots,k$ *and* $u_{[1]} \leq \cdots \leq u_{[k]}$,

$$(3.1) \quad \underline{u}^*_{[i]} = (u_{[1]} - u_{[i]},\ldots,u_{[i-1]} - u_{[i]},\ u_{[i+1]} - u_{[i]},\ldots,u_{[k]} - u_{[i]}).$$

Let $\underline{v} \underset{m}{\leq} \underline{u}$ *and* $v_{[i]} \geq u_{[i]}$ *for* $i = 1,\ldots,k-1$. *Then* $G(\underline{u}) \leq G(\underline{v})$.

Proof. Let $\underline{v} \underset{m}{\leq} \underline{u}$. For every $i \in \{1,\ldots,k-1\}$, $\underline{u}^*_{[i]} \underset{m}{\geq} \underline{v}^*_{[i]} + \frac{k}{k-1}(v_{[i]}-u_{[i]})\underline{1}$ where $\underline{1} = (1,\ldots,1)$. Hence $\underline{v}^*_{[i]} \prec \underline{u}^*_{[i]}$, and from Lemma 3.1, $g(\underline{u}^*_{[i]}) \leq g(\underline{v}^*_{[i]})$. Since this is true for every i, $G(\underline{u}) \leq G(\underline{v})$. Q.E.D.

Remark. A natural question is whether $G(\underline{u})$ is in fact Schur-concave. This is not in general true, as can be seen by letting $k = 3$ and $g(x_1,x_2) = \Phi(-x_1-x_2)$ where Φ is the distribution function of the $\mathcal{N}(0,1)$-distribution. It can be shown that if g, in addition to satisfying the conditions in Lemma 3.2, is also concave then G is Schur-concave. A proof for the case where g admits partial derivatives can be found in Bjørnstad (1978).

4. Some Properties of Schur-Procedures

Applying Proposition 5.1 and Theorem 2.1 from Marshall and Olkin (1974) we obtain the first interesting result about Schur-procedures.

Theorem 4.1. *Let ψ be a Schur-procedure. Then $E_{\underline{\theta}}(\psi_i)$ is Schur-concave in $\underline{\theta}_i^*$.*

Consider now the risk function $S'(\underline{\theta},\psi)$ defined by (1.6). One of the main results for Schur-procedures is a monotonicity-result for $S'(\underline{\theta},\psi)$.

Theorem 4.2. *Let ψ be a Schur-procedure and let $\underline{\theta}' \underset{m}{<} \underline{\theta}$ and $\theta'_{[i]} \geq \theta_{[i]}$ for $i = 1,\ldots,k-1$. Then $S'(\underline{\theta},\psi) \leq S'(\underline{\theta}',\psi)$.*

Proof. Let g be defined by $g(\underline{\theta}^*_{[i]}) = E_{\underline{\theta}}\{\psi_{(i)}\}$, where $\underline{\theta}^*_{[i]}$ is defined in (3.1). (Since ψ is a Schur-procedure g does not depend on i.) Then

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

From Theorem 4.1 and Lemma 2.1 we see that the assumptions in Lemma 3.2 are satisfied and result follows. Q.E.D.

As mentioned in Section 2 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 4.2 we can show a similar result for $S'(\underline{\theta},\psi)$ over certain subsets of Ω.

Theorem 4.3. *Let ψ be a Schur-procedure. Let $\delta \geq 0$ and define the slippage-set*

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

Then

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

where

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

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 $\theta_1 \leq \ldots \leq \theta_k$. From Lemma 2.1 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 $\underline{\theta}^0 = (\theta_{k-1},\ldots,\theta_{k-1},\theta_{k-1}+\delta)$. Then $\underline{\theta}^0 \underset{m}{\leq} \underline{\theta}$ and $\theta^0_{[i]} \geq \theta_{[i]}$ for $i \leq k-1$. From Theorem 4.2 it follows that $S'(\underline{\theta},\psi) \leq S'(\underline{\theta}^0,\psi) = S'(\underline{\theta}^{\delta},\psi)$. Q.E.D.

Remark. Gupta (1965) showed that $\sup_{\Omega(\delta)} S'(\underline{\theta},\psi^m) = S'(\underline{\theta}^{\delta},\psi^m)$, where $\Omega(\delta) = \{\underline{\theta}: \theta_{[k]}-\theta_{[k-1]} \geq \delta\}$. This is not true in general for Schur-procedures, as can be seen by considering ψ^a, and the case $\delta = 0$, $\gamma < (k-2)/(k-1)$.

We shall next consider the corresponding problem for the risk $B(\underline{\theta},\psi)$ given by (1.7).

Theorem 4.4. *Let ψ be a Schur-procedure. Let $\Omega_1(\Delta) = \{\underline{\theta} \in \Omega: \theta_{[k]}-\theta_{[1]} < \Delta\}$, and let for* $p = 2,\ldots,k-1$

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

Let

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

Then

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

where $\underline{\theta}^{\Delta}$ is defined by (4.1) *for* $\delta = \Delta$.

Proof. We can assume $\theta_1 \leq \ldots \leq \theta_k$. Let $\underline{\theta} \in \Omega_1$. Then $\underline{\theta} \in \Omega_p(\Delta)$ for some $p \in \{1,\ldots,k\}$. Let $\underline{\theta}'$ be defined by

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

Clearly, from Lemma 2.1

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

Since $\underline{\theta}' \in \Omega_k(\Delta)$, the result now follows from Theorem 4.3. Q.E.D.

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

In the next section Theorems 4.3 and 4.4 are applied to derive a certain optimal procedure which will be minimax with respect to slippage sets of the type Ω_1 and $\Omega_k(\Delta)$.

5. Some General Minimax Theorems in the Location-Model

Let $\underline{\delta}_i$ be the vector in $\mathbb{R}^k$ where the i^{th} coordinate is equal to 1 and the rest are equal to zero. Let $\underline{\theta}_i^{\Delta} = \Delta\underline{\delta}_i$, and let $p_i(\underline{x}) = p(\underline{x} - \underline{\theta}_i^{\Delta})$. Define the statistic

$$T_i(\underline{x}) = \frac{1}{p_i(\underline{x})} \sum_{j \neq i} p_j(\underline{x}), \quad \text{for } i = 1,\ldots,k.$$

ψ^0 is the subset selection procedure given by:

$$(5.1) \qquad \psi_i^0(\underline{x}) = \begin{cases} 1 & \text{if } T_i < C \\ 0 & \text{if } T_i > C \end{cases}$$

where C is determined by

(5.2) $$E_{\underline{\theta}_i^{\Delta}}(\psi_i^0) = \gamma, \text{for } i = 1,\ldots,k.$$

Theorem 5.1. *Assume that* ψ^0, *defined by* (5.1), *is a Schur-procedure. Let* Ω_1 *be given by* (4.3). *Then* ψ^0 *minimizes for all* $\psi \in \mathscr{D}'(\gamma,\Delta)$

$$\sup_{\underline{\theta}\in\Omega_1} B(\underline{\theta},\psi) \text{ and } \sup_{\underline{\theta}\in\Omega_k(\Delta)} S'(\underline{\theta},\psi).$$

Proof. We prove the theorem only for $B(\underline{\theta},\psi)$. The proof for $S'(\underline{\theta},\psi)$ is exactly the same. From (5.2), $\psi^0 \in \mathscr{D}'(\gamma,\Delta)$. From Theorem 4.4 $\sup_{\underline{\theta}\in\Omega_1} B(\underline{\theta},\psi^0) = B(\underline{\theta}_i^{\Delta},\psi^0)$ for $i = 1,\ldots,k$. For any $\psi \in \mathscr{D}'(\gamma,\Delta)$

$$\sup_{\underline{\theta}\in\Omega_1} B(\underline{\theta},\psi) \geq \frac{1}{k}\sum_{i=1}^{k} B(\underline{\theta}_i^{\Delta},\psi) = \frac{1}{k}\sum_{j=1}^{k}\int\psi_j\Big(\sum_{i\neq j} p_i\Big)d\nu$$

$$\geq \frac{1}{k}\sum_{j=1}^{k}\int\psi_j^0\Big(\sum_{i\neq j} p_i\Big)d\nu = \frac{1}{k}\sum_{i=1}^{k} B(\underline{\theta}_i^{\Delta},\psi^0) = \sup_{\underline{\theta}\in\Omega_1} B(\underline{\theta},\psi^0).$$

The second inequality follows from the Neyman-Pearson Lemma since ψ_j^0 minimizes $\int\psi_j(\sum_{i\neq j} p_i)d\nu$ subject to $\int\psi_j p_j d\nu \geq \gamma$. Q.E.D.

Next we consider the problem of finding solutions to the dual goals

(5.3) maximizing $\inf_{\underline{\theta}\in\Omega(\Delta)} P_{\underline{\theta}}(CS|\psi)$ for $\psi \in \mathscr{D}_1(\beta)$

and

(5.4) minimizing $\sup_{\underline{\theta}\in\Omega} S(\underline{\theta},\psi)$ for $\psi \in \mathscr{D}'(\gamma,\Delta)$.

It is shown by the Hunt-Stein theorem that we can restrict attention to translation-invariant subset selection procedures, i.e. we can assume that ψ_i is a function of $\underline{x}_i^*$, and $\underline{X}_i^*$ has location-density $g(\underline{y}-\underline{\theta}_i^*)$, $\underline{y} \in \mathbb{R}^{k-1}$, where g is the density of $(U_1-U_k,\ldots,U_{k-1}-U_k)$, and $U_1,\ldots,U_k$ are i.i.d. with density $f(u)$.

Theorem 5.2. *Assume Ω is translation-invariant. Define ψ^* by*

(5.5) $$\psi_i^*(\underline{x}_i^*) = \begin{cases} 1 \text{ if } g(\underline{x}_i^* + \underline{\Delta}) > cg(\underline{x}_i^*) \\ 0 \text{ if } g(\underline{x}_i^* + \underline{\Delta}) < cg(\underline{x}_i^*). \end{cases}$$

Here $\underline{\Delta} = (\Delta,\ldots,\Delta)$. *Assume ψ^* is a just procedure and that*

(5.6) $$\sup_{\underline{\theta}\in\Omega} S(\underline{\theta},\psi^*) \text{ occurs at } \theta_1 = \ldots = \theta_k.$$

If c *is determined by*

(5.7) $$\int \psi_i^*(\underline{y}) g(\underline{y}) d\nu(\underline{y}) = \beta/k \text{ for } i = 1,\ldots,k$$

then ψ^* maximizes $\inf_{\underline{\theta}\in\Omega(\Delta)} P_{\underline{\theta}}(CS|\psi)$ *for all* $\psi \in \mathcal{D}_1(\beta)$.

If c *is determined by*

(5.8) $$\int \psi_i^*(\underline{y}) g(\underline{y} + \underline{\Delta}) d\nu(\underline{y}) = \gamma \text{ for } i = 1,\ldots,k$$

then ψ^* minimizes $\sup_{\underline{\theta}\in\Omega} S(\underline{\theta},\psi)$ *for all* $\psi \in \mathcal{D}'(\gamma,\Delta)$.

Proof. From the Hunt-Stein Theorem and Example 7, p.337 in Lehmann (1959) it can be shown that for any procedure ψ there exists a translation-invariant ψ^I such that for $j = 1,\ldots,k$, $E_{\underline{\theta}}(\psi_j^I) = \lim_{i\to\infty} \int_G E_{g\underline{\theta}}(\psi_j) d\nu_{n_i}(g)$, where G is the group of translations and $\{\nu_{n_i}\}$ is a subsequence of the uniform probability distributions ν_n on the interval $I(-n,n) = \{g: -n \le g \le n\}$. This implies that $\inf_{\Omega(\Delta)} P(CS|\psi^I) \ge \inf_{\Omega(\Delta)} P(CS|\psi)$. Also, since $S(\underline{\theta},\psi^I) = \lim_{i\to\infty} \int_G S(g\underline{\theta},\psi) d\nu_{n_i}(g)$ $\le \sup_{\Omega} S(\underline{\theta},\psi)$, $\sup_{\Omega} S(\underline{\theta},\psi^I) \le \sup_{\Omega} S(\underline{\theta},\psi)$. Let $\psi \in \mathcal{D}_1(\beta)$. It follows that we may assume that ψ is translation-invariant. The first result now follows easily from a result of Gupta and Huang (1977), and the proof of the second result goes in a similar way. Q.E.D.

Remark. Assume

$$(5.9) \qquad \inf_{\underline{\theta}\in\Omega} R(\underline{\theta},\psi^*) = \inf_{\underline{\theta}\in\Omega(\Delta)} P_{\underline{\theta}}(CS|\psi^*).$$

Then if (5.7) is satisfied, ψ^* maximizes, for all $\psi \in \mathcal{D}_1(\beta)$, $\inf_{\underline{\theta}\in\Omega} R(\underline{\theta},\psi)$.

If (5.8) is satisfied then ψ^* minimizes for all $\psi \in \mathcal{D}(\gamma,\Delta)$, $\sup_{\underline{\theta}\in\Omega} S(\underline{\theta},\psi)$.

At last in this section we consider the classical problem of minimizing $\sup_{\Omega} S(\underline{\theta},\psi)$ and $\sup_{\Omega} S'(\underline{\theta},\psi)$ in the class $\mathcal{D}'(\gamma)$. From Berger (1979) we have that $\inf_{\mathcal{D}'(\gamma)} \sup_{\Omega} S(\underline{\theta},\psi)(S'(\underline{\theta},\psi)) = k\gamma((k-1)\gamma)$.

As mentioned in Section 1, ψ^m is minimax in $\mathcal{D}'(\gamma)$ for S and S' if $f(x-\theta)$ has MLR in x. We shall next show that ψ^a has the same minimax property if γ is large enough. For the remainder of this section it is assumed that $f(x) = f(-x)$ for all x.

Theorem 5.3.

$$\sup_{\underline{\theta}\in\Omega} S'(\underline{\theta},\psi^a) = S'(\underline{0},\psi^a) = (k-1)\gamma \text{ if and only if } \gamma \geq \frac{k-2}{k-1}.$$

$$\sup_{\underline{\theta}\in\Omega} S(\underline{\theta},\psi^a) = S(\underline{0},\psi^a) = k\gamma \text{ if and only if } \gamma \geq \frac{k-1}{k}.$$

We then have the following corollary.

Corollary 5.1. *ψ^a is minimax in $\mathcal{D}'(\gamma)$ for S if and only if $\gamma \geq (k-1)/k$, and for S' if and only if $\gamma \geq (k-2)/(k-1)$.*

We prove Theorem 5.3 only for S. The proof for S' is completely analogous. The proof goes by a series of lemmas.

A location-density has MLR in x if and only if it is strongly unimodal. Using the result from Ibragimov (1956), that the convolution of two strongly unimodal densities is again strongly unimodal, we readily get the following result.

Lemma 5.1. *Let for* $1,\ldots,k$, $V_i = (k-1)^{-1}\sum_{j\neq i}(X_j-X_i)$ *and* $\mu_i = (k-1)^{-1}\sum_{j\neq i}(\theta_j-\theta_i)$. *Then* V_i *has density* $g(v-\mu_i)$, *where* g *is symmetric around zero, i.e.* $g(v) = g(-v)$ *and* $g(v-\mu)$ *has* MLR *in* v.

Let $G(v-\mu_i)$ be the distribution function of V_i, and let $c(\gamma)$ be the γ-quantile in G, i.e. $G(c(\gamma)) = \gamma$. Then the critical constant c in ψ^a is equal to $c(\gamma)$. $S(\underline{\theta},\psi^a)$ is permutation-symmetric in $(\theta_1,\ldots,\theta_k)$, so we can assume $\theta_1 \leq\ldots\leq \theta_k$. Let $t_i = (\theta_{i+1}-\theta_i)/(k-1)$. Then

$$(5.10)\qquad S(\underline{\theta},\psi^a) = H(\underline{t}) = \sum_{i=1}^{k} G\{c(\gamma) + \sum_{j=1}^{i-1} jt_j - \sum_{j=i}^{k-1}(k-j)t_j\}$$

where $\underline{t} = (t_1,\ldots,t_{k-1})$ and $t_i \geq 0$ for all i. The next lemma considers the "if" part for "large" $\underline{t}$.

Lemma 5.2. *Let* $\gamma \geq (k-1)/k$ *and* $k \geq 3$. *Then*

$$(k-2)t_1 + \sum_{j=2}^{k-1}(k-j)t_j \geq 2c(\gamma) \Rightarrow H(\underline{t}) \leq k\gamma.$$

Proof. It is enough to show that $E_{\underline{\theta}}(\psi_1^a + \psi_2^a) \leq 1$. Let $c = c(\gamma)$. Now,

$$E_{\underline{\theta}}(\psi_1^a) \leq G(-c-t_1) = 1-G(c+t_1), \text{ and } E_{\underline{\theta}}(\psi_2^a) \leq G(c+t_1). \qquad \text{Q.E.D.}$$

It remains to consider $H(\underline{t})$ for $\underline{t} \in A(\gamma)$, where

$$A(\gamma) = \{\underline{t}\colon (k-2)t_1 + \sum_{j=2}^{k-1}(k-j)t_j < 2c(\gamma)\}.$$

Lemma 5.3. *Let* $k \geq 3$ *and let* $\underline{t}^0 = (t_1^0,\ldots,t_{k-1}^0) \in A(\gamma)$. *Then* $H(t_1^0,\ldots,t_{m-1}^0,t_m,0,\ldots,0)$ *is nonincreasing in* t_m *for* $t_m \leq t_m^0$ *for* $1 \leq m \leq k-1$.

Proof. Let $m \geq 1$, and let $\underline{v} = (v_1,\ldots,v_m,0,\ldots,0) = (t_1^0,\ldots,t_{m-1}^0, t_m,0,\ldots,0)$, $t_m \leq t_m^0$. Then $\underline{v} \in A(\gamma)$. Let $h(v_m) = H(\underline{v})$. We shall show that the derivative $h'(v_m) \leq 0$ for $v_m \leq t_m$.

It is easily seen that

(5.11) $$h'(v_m) \leq 0 \Leftrightarrow m \leq \sum_{i=1}^{m} r_i(\underline{v}),$$

where

$$r_i(\underline{v}) = g(c+\sum_{j=1}^{i-1} jv_j - \sum_{j=i}^{m}(k-j)v_j)/g(c+\sum_{j=1}^{m} jv_j).$$

Let $a = c+\sum_{j=1}^{m} jv_j$ and $y_i = \sum_{j=i}^{m} jv_j + \sum_{j=i}^{m}(k-j)v_j$. Then $r_i(\underline{v}) = g(a-y_i)/g(a)$, and $y_i \leq 2a$, for all i. From Lemma 5.1 it follows that $r_i \geq 1$ for $i = 1,\ldots,m$, and (5.11) follows. Q.E.D.

Proof of Theorem 5.3. As shown by Berger (1977), the "only if" part follows by letting $t_1 \to \infty$ in (5.10). Now assume $\gamma \geq (k-1)/k$. Consider first the case $k = 2$. It is readily seen that $H'(t_1) \leq 0$ since $c(\gamma) \geq 0$. Let now $k \geq 3$. From Lemma 5.3 we get that $t \in A(\gamma) \Rightarrow H(\underline{t}) \leq H(\underline{0}) = k\gamma$. Together with Lemma 5.2 this completes the proof.

For later use we will also consider the case where $\psi^a \in \mathcal{D}'(\gamma,\Delta)$. It is seen from Lemma 2.1 that for any just and translation-invariant procedure ψ, $\inf_{\underline{\theta}\in\Omega(\Delta)} P_{\underline{\theta}}(CS|\psi)$ occurs at $\theta_{[1]} = \ldots = \theta_{[k-1]} = \theta_{[k]}-\Delta$. Hence ψ^a satisfies (1.4) with equality if in (1.8) $c = c(\gamma)-\Delta$. In the same way as we proved Theorem 5.3, the following result can be proved.

Theorem 5.4. *Let* $c = c(\gamma)-\Delta$ *such that* $\psi^a \in \mathcal{D}'(\gamma,\Delta)$. *Then*

$$\sup_{\underline{\theta}\in\Omega} S(\underline{\theta},\psi^a) = S(\underline{0},\psi^a) \text{ if and only if } \Delta \leq c(\gamma)-c(\tfrac{k-1}{k}).$$

6. Optimal Subset Selection Procedures for Normal Populations

Let X_{ij} $(i = 1,\ldots,k;\ j = 1,\ldots,n)$ be independent and normally distributed. X_{ij} is $\mathcal{N}(\theta_i,\sigma^2)$ where σ^2 is known. A sufficient statistic is $\underline{X} = (X_1,\ldots,X_k)$ where $X_i = (n^{-1})\sum_j X_{ij}$. Let $\Delta_0 = \sqrt{n}\Delta/\sigma$, so that Π_i is a good population if

$\theta_i > \theta_{[k]} - \Delta_0 \frac{\sigma}{\sqrt{n}}$, $\Delta_0 > 0$. In this general case ψ^e is given by

$$\psi_i^e = 1 \text{ iff } Ce^{\Delta_0 \sqrt{n} X_i/\sigma} \geq \sum_{j \neq i} e^{\Delta_0 \sqrt{n} X_j/\sigma} .$$

C is determined such that (1.4) holds with equality. Hence

$$\text{(6.1)} \qquad \gamma = P(Ce^{\Delta_0 Y_k + \Delta_0^2} \geq \sum_{j=1}^{k-1} e^{\Delta_0 Y_j})$$

where $Y_1,\ldots,Y_k$ are independent, $\mathcal{N}(0,1)$ random variables. The critical constant C is tabulated in Table 1, for $k \leq 10$. For ψ^e we are really only interested in satisfying (1.4) or (1.3). However, we see from (6.1) that ψ^e satisfies (1.2) if we instead of C use $e^{\Delta_0^2}C$ as the critical constant.

Since σ is known, we may just as well assume $\sigma/\sqrt{n} = 1$, and denote Δ_0 by Δ. Hence we assume that $X_1,\ldots,X_k$ are independent, and X_i is $\mathcal{N}(\theta_i,1)$. Let $\underline{\theta}_0^\Delta = (0,0,\Delta,\ldots,\Delta)$ and let

$$\text{(6.2)} \qquad b_k(\gamma) = \begin{cases} \gamma/9 & \text{if } k = 4 \\ \gamma/7 & \text{if } k = 5 \\ (11/75)\gamma & \text{if } k \geq 6. \end{cases}$$

Note that we always select at least one population with ψ^e, if and only if $C \geq k-1$.

The main minimax properties of ψ^e are given in the following theorem (proved in Section 7).

<u>Theorem 6.1</u>. *Assume* $C \geq k-1$. *If* $k \geq 4$ *and* Δ *is such that* $E_{\underline{\theta}_0^\Delta}(\psi_1^e) \geq b_k(\gamma)$ *or* $k \leq 3$ *then* $\psi^e \in \mathcal{D}(\gamma,\Delta)$ *and minimizes for all* $\psi \in \mathcal{D}(\gamma,\Delta)$,

$$\sup_{\underline{\theta}\in\Omega_1} B(\underline{\theta},\psi) \text{ and } \sup_{\underline{\theta}\in\Omega_k(\Delta)} S'(\underline{\theta},\psi).$$

Remarks. 1) Let Δ_γ be defined by

$$E_{\underline{\theta}_0^{\Delta_\gamma}}(\psi_1^e) = b_k(\gamma). \tag{6.3}$$

Calculations have shown that $E_{\underline{\theta}_0^{\Delta}}(\psi_1^e)$ seems to be decreasing in Δ. If so, $\Delta \leq \Delta_\gamma$ implies $\psi^e \in \mathcal{D}(\gamma,\Delta)$. Table 2 gives approximate values of Δ_γ for k = 4, 10 and the limiting value ($k \to \infty$) (see (6.6) below). Δ_γ does not seem to vary much for different values of k. For general $\sigma/\sqrt{n}$ we require $\Delta_0 \leq \Delta_\gamma$.

2) Studden (1967) considered the identification problem, i.e. the case where $\theta_{[1]},\ldots,\theta_{[k]}$ are known. It was shown that ψ^e is the best permutation-invariant procedure in $\mathcal{D}'(\gamma,\Delta)$ for the risk $S(\underline{\theta},\psi)$, when $\theta_{[1]} = \ldots = \theta_{[k-1]} = \theta_{[k]} - \Delta$.

3) By applying the geometric-arithmetic mean inequality, we find

$$\Delta \leq \sqrt{\frac{k}{k-1}}\, z(\gamma) \Rightarrow C \geq k-1. \tag{6.4}$$

4) It can be shown that ψ^e has similar minimax properties also for normal models like the two-way layout without interaction.

We shall next consider the procedure $\psi^a = \psi^a(c)$, defined by (1.8). Let

$$\Delta_a(\gamma) = \sqrt{\frac{k}{k-1}} \min\{[\tfrac{1}{2} z(\gamma) + \tfrac{1}{2} z(\tfrac{k-1}{k})],\ [z(\gamma) - z(\tfrac{k-1}{k})]\},$$

and let $\Delta_1(\beta) = \sqrt{\frac{k}{k-1}} \{z(\frac{\beta}{k}) + z(\frac{k-1}{k})\}$. Here $z(\gamma)$ is defined by $\Phi\{z(\gamma)\} = \gamma$, where Φ is the $\mathcal{N}(0,1)$-distribution function. ψ^a has the following minimax properties. (Proof is given in Section 8.)

Theorem 6.2. a) *Let* $c = \sqrt{\frac{k}{k-1}}\, z(\gamma)$. *Then* $\psi^a \in \mathcal{D}'(\gamma)$ *and minimizes* $\sup_{\underline{\theta}\in\Omega} S'(\underline{\theta},\psi)(\sup_{\underline{\theta}\in\Omega} S(\underline{\theta},\psi))$ *for all* $\psi \in \mathcal{D}'(\gamma)$ *if and only if* $\gamma \geq (k-2)/(k-1)$ $((k-1)/k)$.

b) *Let* $\beta \geq k-1$ *and* $c = \sqrt{\frac{k}{k-1}}\, z(\beta/k)$. *Then* $\psi^a \in \mathcal{D}_1(\beta)$. *Assume further that* $k \geq 4$ *and* $\Delta \leq \Delta_1(\beta)$ *or* $k \leq 3$. *Then* ψ^a *maximizes* $\inf_{\underline{\theta}\in\Omega} R(\underline{\theta},\psi)$ *for all* $\psi \in \mathcal{D}_1(\beta)$.

c) *Assume* $\Delta \leq \Delta_a(\gamma)$, *and let* $c = \sqrt{\frac{k}{k-1}}\, z(\gamma)-\Delta$. *Then* $\psi^a \in \mathcal{D}(\gamma,\Delta)$, *and minimizes* $\sup_{\underline{\theta}\in\Omega} S(\underline{\theta},\psi)$ *for all* $\psi \in \mathcal{D}(\gamma,\Delta)$.

Remark. $\Delta_a(\gamma) = \sqrt{\frac{k}{k-1}}\, \{z(\gamma)-z(\frac{k-1}{k})\}$ if and only if $z(\gamma) \leq 3z(\frac{k-1}{k})$.

Let now $k \to \infty$, and let $C_k = C$ be determined by (6.1). Then from the strong law of large numbers it is easily seen that $C_0 = \lim_{k\to\infty}(C_k/k-1)$ exists and is given by

$$\log C_0 = \Delta z(\gamma) - \frac{1}{2}\Delta^2. \tag{6.5}$$

This can be used to find approximate values of C_k for large and moderate k. From Table 1 we see that the asymptotic value C_0 is a good approximation to $C_k/k-1$ already for $k = 10$ in case of smaller Δ values. Also, from (6.5), we have as a supplement to (6.4) that $\lim_{k\to\infty} (C_k/k-1) \geq 1$ iff $\Delta \leq 2z(\gamma)$.

Consider next the upper bound $\Delta_\gamma(k) = \Delta_\gamma$, given by (6.3), that insures $\psi^e \in \mathcal{D}(\gamma,\Delta)$. For given $\Delta > 0$ we have

$$E_{\underline{\theta}_0^\Delta}(\psi_1^e) \to \Phi(z(\gamma)-2\Delta) \text{ as } k \to \infty.$$

Hence, from (6.2) and (6.3)

$$\lim_{k\to\infty} \Delta_\gamma(k) = \frac{1}{2}\{z(\gamma) + z(1 - \frac{11}{75}\gamma)\}. \tag{6.6}$$

7. Proof of Theorem 6.1

The proof consists of two theorems. The first proves the result for the class $\mathcal{D}'(\gamma,\Delta)$, and the second shows that $\psi^e \in \mathcal{D}(\gamma,\Delta)$.

Theorem 7.1. *ψ^e minimizes, for all $\psi \in \mathcal{D}'(\gamma,\Delta)$,*

$$\sup_{\underline{\theta}\in\Omega_1} B(\underline{\theta},\psi) \quad \text{and} \quad \sup_{\underline{\theta}\in\Omega_k(\Delta)} S'(\underline{\theta},\psi).$$

Proof. From Theorem 5.1, the optimal procedure ψ^0 is given by (5.1). We find that

$$T_i = \sum_{j\neq i} e^{\Delta(X_j-X_i)}.$$

Hence the minimax procedure is ψ^e. Q.E.D.

The following result completes the proof of Theorem 6.1.

Theorem 7.2. *(i) Let $k = 2,3$ and $C \geq k-1$. Then*

$$\text{(7.1)} \qquad \inf_{\underline{\theta}\in\Omega(\Delta)} P_{\underline{\theta}}(CS|\psi^e) = \gamma \Rightarrow \inf_{\underline{\theta}\in\Omega} R(\underline{\theta},\psi^e) = \gamma.$$

(ii) Let $k \geq 4$ and $C \geq k-1$. Then (7.1) holds if

$$E_{\underline{\theta}_0^\Delta}(\psi_1^e) \geq b_k(\gamma)$$

where $\underline{\theta}_0^\Delta = (0,0,\Delta,\ldots,\Delta)$ and $b_k(\gamma)$ is given in (6.2).

To prove Theorem 7.2 we need the following two lemmas.

Lemma 7.1. *Let $i < j$. Assume $C \geq k-1$ and $\theta_{[j]}-\theta_{[i]} \leq \Delta$. Then $E_{\underline{\theta}}(\psi_{(i)}^e) + E_{\underline{\theta}}(\psi_{(j)}^e)$ is nondecreasing in $\theta_{[i]}$.*

Proof. We can without loss of generality assume $\theta_1 \leq\ldots\leq \theta_k$ and consider $r(\underline{\theta}) = E_{\underline{\theta}}(\psi_i^e + \psi_j^e)$. Now,

$$E_{\underline{\theta}}(\psi_i^e) = \int\Phi(\theta_i - \frac{1}{\Delta}\log[\frac{1}{C}\sum_{\ell=1}^{i-1} e^{\Delta(y_\ell+\theta_\ell)} + \frac{1}{C}\sum_{\ell=i}^{k-1} e^{\Delta(y_\ell+\theta_{\ell+1})}])\prod_{\ell=1}^{k-1}\phi(y_\ell)d\nu(\underline{y}).$$

We get:

$$\frac{\partial r}{\partial \theta_i} = \int\{\phi(\theta_i - \frac{1}{\Delta}\log[\frac{1}{C}\sum_{\ell=1}^{k-1} e^{\Delta y_\ell}])\phi(y_i-\theta_j)$$

$$- \phi(\theta_j - \frac{1}{\Delta}\log[\frac{1}{C}\sum_{\ell=1}^{k-1} e^{\Delta y_\ell}])\phi(y_i-\theta_i)\cdot(\frac{e^{\Delta y_i}}{\sum_{\ell=1}^{k-1} e^{\Delta y_\ell}})\}$$

$$\cdot \prod_{\ell=1}^{i-1} \phi(y_\ell-\theta_\ell) \prod_{\ell=i+1}^{j-1} \phi(y_\ell-\theta_\ell) \prod_{\ell=j}^{k-1} \phi(y_\ell-\theta_{\ell+1})d\nu(\underline{y}).$$

From this expression we find that $\partial r/\partial\theta_i \geq 0$ if

$$(7.2)\quad [1+\sum_{\ell\neq i} e^{\Delta(y_\ell-y_i)}]\cdot\exp\{(\theta_j-\theta_i)(y_i - \frac{1}{\Delta}\log[\frac{1}{C}\sum_{\ell=1}^{k-1} e^{\Delta y_\ell}])\} \geq 1.$$

Let $y_{max} = \max(y_1,\ldots,y_{k-1})$. Then, since $\Delta \geq \theta_j-\theta_i$,

$$[1+\sum_{\ell\neq i} e^{\Delta(y_\ell-y_i)}]e^{(\theta_j-\theta_i)y_i} \geq e^{(\theta_j-\theta_i)y_{max}}$$

and (7.2) follows. Q.E.D.

Remark. If $C < k-1$, then Lemma 7.1 is not necessarily true as seen by the following example. Let $k = 2$. Then $C < 1$ iff $\Delta > \sqrt{2}z(\gamma)$. If Lemma 7.1 holds then $R((\theta,\theta),\psi^e) \geq E_{\theta-\Delta,\theta}(\psi_1^e+\psi_2^e) \geq \gamma$. Now, $R((\theta,\theta),\psi^e) = 2\Phi(\frac{\log C}{\Delta\sqrt{2}})$. Let $\Lambda > \sqrt{2}\ \{z(\gamma)-z(\gamma/2)\}$. Then $C < 1$ and $R((\theta,\theta),\psi^e) < \gamma$.

Lemma 7.2. *Let* $C \geq k-1$ *and let* p *be such that* $1 \leq p \leq k-1$. *Assume* $\theta_{[p+1]} > \theta_{[k]}-\Delta \geq \theta_{[p]}$. *If* $\inf_{\underline{\theta}\in\Omega(\Delta)} P_{\underline{\theta}}(CS|\psi^e) = \gamma$ *then*

$$(7.3)\qquad R(\underline{\theta},\psi^e) \geq \frac{k-p}{2^{k-1-p}}\gamma + (k-p)\{1-(\frac{1}{2})^{k-1-p}\}E_{\underline{\theta}^{(p+1)}}(\psi_1^e).$$

Here $\underline{\theta}^{(p+1)}$ *is given by*

$$\theta_i^{(p+1)} = \begin{cases} \theta_{[k]}-\Delta & \text{for } i \leq p+1 \\ \theta_{[k]} & \text{for } i \geq p+2. \end{cases}$$

Proof. Let $\psi = \psi^e$. We may assume $\theta_1 \leq\cdots\leq \theta_k$. Then $R(\underline{\theta},\psi) = \sum_{i=p+1}^{k} E_{\underline{\theta}}(\psi_i)$.

(7.3) is true for $p = k-1$ since the right hand side is γ and $R(\underline{\theta},\psi) = P_{\underline{\theta}}(CS|\psi) \geq \gamma$. We now prove (7.3) for general p, $1 \leq p \leq k-2$ by assuming (7.3) is true for p+1 and proving it for p. Each $E_{\underline{\theta}}(\psi_i)$, $i \geq p+1$, is nonincreasing in each $\theta_1,\ldots,\theta_p$, so we can let $\theta_1 = \ldots = \theta_p = \theta_k - \Delta$. Define $\underline{\theta}^q$ by:

$$\theta_i^q = \begin{cases} \theta_k - \Delta \text{ for } i \leq q \\ \theta_i \text{ for } i \geq q+1. \end{cases}$$

$E_{\underline{\theta}}(\psi_i)$ is nondecreasing in i, hence for $i \geq p+2$:

$$E_{\underline{\theta}}(\psi_i) \geq \frac{k-p}{2(k-p-1)} E_{\underline{\theta}}(\psi_i) + \frac{k-p-2}{2(k-p-1)} E_{\underline{\theta}}(\psi_{p+1}).$$

This implies that

$$(7.4) \quad R(\underline{\theta},\psi) = \sum_{i=p+1}^{k} E_{\underline{\theta}}(\psi_i) \geq \frac{k-p}{2(k-p-1)} \sum_{i=p+2}^{k} E_{\underline{\theta}}(\psi_{p+1}+\psi_i)$$

$$\geq \frac{k-p}{2(k-p-1)} \sum_{i=p+2}^{k} E_{\underline{\theta}^{p+1}}(\psi_{p+1}+\psi_i),$$

the second inequality being a result of Lemma 7.1. Now $E_{\underline{\theta}^{p+1}}(\psi_{p+1}) = E_{\underline{\theta}^{p+1}}(\psi_1) \geq E_{\underline{\theta}^{(p+1)}}(\psi_1)$. By the induction hypothesis, $R(\underline{\theta}^{p+1},\psi) = \sum_{i=p+2}^{k} E_{\underline{\theta}^{p+1}}(\psi_i)$ can be bounded by the right side of (7.3) (substituting p+1 for p). Making these two substitutions and $E_{\underline{\theta}^{(p+2)}}(\psi_1) \geq E_{\underline{\theta}^{(p+1)}}(\psi_1)$ in (7.4) yields

$$R(\underline{\theta},\psi) \geq \frac{k-p}{2} E_{\underline{\theta}^{(p+1)}}(\psi_1) + \frac{k-p}{2} \left[\frac{1}{2^{k-p-2}} \gamma + \{1-(\tfrac{1}{2})^{k-p-2}\} E_{\underline{\theta}^{(p+1)}}(\psi_1)\right]$$

$$= \frac{k-p}{2^{k-1-p}} \gamma + (k-p)\{1-(\tfrac{1}{2})^{k-1-p}\} E_{\underline{\theta}^{(p+1)}}(\psi_1). \qquad \text{Q.E.D.}$$

<u>Proof of Theorem 7.2</u>. The result for k = 2 follows directly since for $\underline{\theta} \in \Omega^c(\Delta)$, $R(\underline{\theta},\psi^e) = E_{\underline{\theta}}(\psi_1^e+\psi_2^e) \geq 1$. For $k \geq 3$, we have from Lemma 7.2 that (7.1) holds if

$$E_{\underset{\sim}{\theta}_0^{\Delta}}(\psi_1^e) \geq \gamma \cdot \max_{2 \leq m \leq k-1} g(m).$$

where

$$g(m) = \frac{1-m(1/2)^{m-1}}{m-m(1/2)^{m-1}}; \quad 2 \leq m \leq k-1.$$

It is readily seen that

$$\gamma \cdot \max_{2 \leq m \leq k-1} g(m) = \begin{cases} 0 & \text{if } k = 3 \\ b_k(\gamma) & \text{if } k \geq 4. \end{cases}$$

Q.E.D.

Remark. Let k = 2. From remark after Lemma 7.1 we see that (7.1) is not necessarily true if Δ gets too large.

8. Proof of Theorem 6.2

Part (a) follows directly from Corollary 5.1. We next show that ψ^a is the solution to the problems (5.3) and (5.4) for the normal case.

Theorem 8.1. *(i) Let* $\beta \geq k-1$ *and* $c = \sqrt{k/(k-1)}\, z(\beta/k)$. *Then* $\psi^a \in \mathcal{D}_1(\beta)$, *and* ψ^a *maximizes* $\inf_{\Omega(\Delta)} P_{\underset{\sim}{\theta}}(CS|\psi)$, *for all* $\psi \in \mathcal{D}_1(\beta)$.

(ii) Let $c = \sqrt{k/(k-1)}\, z(\gamma)-\Delta$ *such that* $\psi^a \in \mathcal{D}'(\gamma,\Delta)$. *Assume* $\Delta \leq \sqrt{k/(k-1)} \cdot \{z(\gamma)-z((k-1)/k)\}$. *Then* ψ^a *minimizes* $\sup_{\Omega} S(\underset{\sim}{\theta}, \psi)$ *for all* $\psi \in \mathcal{D}'(\gamma,\Delta)$.

Proof. We shall apply Theorem 5.2. The density g in (5.5) is the $\mathcal{N}_{k-1}(0,\Sigma)$-density, where $\Sigma = (\sigma_{ij})$ and $\sigma_{ii} = 2$; $\sigma_{ij} = 1$ for $i \neq j$. This implies that

$$g(\underset{\sim}{y}+\underset{\sim}{\Delta})/g(\underset{\sim}{y}) = \exp\{-\frac{\Delta}{k}\sum_{i=1}^{k-1} y_i - \frac{k-1}{2k}\Delta^2\}.$$

It follows from Theorem 5.2 that the optimal procedure ψ^* is given by

$$\psi_i^* = 1 \text{ iff } \frac{1}{k-1}\sum_{j \neq i}(X_j - X_i) \leq c.$$

i.e. $\psi^* = \psi^a$. From Theorem 5.3 and Theorem 5.4 we see that the conditions in (i), (ii) insure that (5.6) holds. The results now follow from Theorem 5.2. Q.E.D.

Let $c_1 = \sqrt{\frac{k}{k-1}}\, z(\beta/k)$. Part (b) of Theorem 6.2 now follows from the fact that $\psi^a(c_1)$ satisfies (5.9) if $k \geq 4$ and $\Delta \leq \Delta_1(\beta)$ or $k \leq 3$. This can be seen as follows. If $\theta_{[p]} > \theta_{[k]}-\Delta$, $\theta_{[p-1]} \leq \theta_{[k]}-\Delta$ for some $2 \leq p \leq k-1$, then $R(\underline{\theta},\psi^a(c_1)) \geq (k-p)\Phi\{z(\beta/k)-(k-p)\Delta/\sqrt{k(k-1)}\} + \Phi\{z(\beta/k)+(p-1)\Delta/\sqrt{k(k-1)}\}$. If $\Delta \leq \Delta_1(\beta)$, then this lower bound on R is greater than or equal to one. For k = 3 this bound is at least $\inf_{\Omega(\Delta)} P_{\underline{\theta}}(CS|\psi^a(c_1)) = \Phi\{z(\beta/3) + 2\Delta/\sqrt{6}\}$. Let now $c' = \sqrt{k/(k-1)}\, z(\gamma)-\Delta$. If $\Delta \leq \Delta_a(\gamma)$, then $\gamma > (k-1)/k$ and $\psi^a(c')$ satisfies (1.3) with equality, and hence $\psi^a(c') \in \mathcal{D}(\gamma,\Delta)$. This is seen by observing that if there are (p-1) bad populations, then

$$R(\underline{\theta},\psi^a(c')) \geq (k-p)\Phi\{z(\gamma)- \frac{2k-p-1}{\sqrt{k(k-1)}}\Delta\} + \Phi\{z(\gamma)- \frac{(k-p)\Delta}{\sqrt{k(k-1)}}\}.$$

$\Delta \leq \Delta_a(\gamma)$ implies that the first term on the right side of the inequality is at least (k-p)/k and the second term is bounded below by (k-1)/k. Hence $R(\underline{\theta},\psi^a(c')) \geq 1$ if $\Delta \leq \Delta_a(\gamma)$. This, together with Theorem 8.1 (ii), proves part (c) of Theorem 6.2.

Acknowledgment.

This paper is based on a part of the author's Ph.D. dissertation. I would like to thank Professor Erich L. Lehmann for his valuable advice and suggestions. I am grateful to two referees and an associate editor for an extremely thorough review that lead to a substantial improvement in the presentation and more straightforward proofs of several results. Thanks are also due to Peter Guttorp and Regina Becker who programmed the Monte Carlo experiments that were used to construct the tables.

TABLE 1

The critical constant C for procedure ψ^e, given by (6.1).

k \ $\sqrt{n}\Delta/\sigma$	.10	.25	.50	1.0	1.5	2.0	3.0
				$\gamma = .95$			
3	2.44	3.20	4.69	7.37	7.51	4.87	.46
4	3.63	4.72	6.83	10.83	12.11	8.45	1.05
5	4.80	6.17	8.82	14.13	15.83	11.39	1.50
6	5.98	7.66	11.00	17.67	20.57	15.26	2.22
7	7.14	9.07	12.83	20.76	23.18	17.83	2.67
8	8.33	10.59	14.94	23.56	26.75	20.81	3.31
9	9.50	12.03	16.80	27.07	30.56	23.12	3.76
10	10.67	13.50	18.83	30.12	34.02	25.88	4.29
				$\gamma = .99$			
3	2.64	3.90	6.93	16.55	25.19	24.66	5.96
4	3.93	5.74	10.10	25.76	43.62	46.78	13.71
5	5.19	7.51	12.98	33.29	59.04	69.01	21.93
6	6.46	9.25	15.98	38.65	71.15	87.79	31.11
7	7.71	11.02	19.06	45.29	81.13	95.94	34.82
8	8.99	12.77	21.63	50.93	87.22	109.38	40.11
9	10.26	14.60	24.54	55.25	96.25	115.48	45.54
10	11.48	16.22	27.13	59.52	101.30	126.29	52.35

For k = 2, C is given by

$$C = \exp\{\sqrt{2n}\,\Delta z(\gamma)/\sigma - n\Delta^2/\sigma^2\}.$$

TABLE 2

Values of Δ_γ, given by (6.3) and (6.6).

γ \ k	4	10	∞
.95	2.0	1.5	1.37
.99	2.4	1.9	1.69

REFERENCES

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

[2] Berger, R. L. (1977). Minimax, admissible and gamma-minimax multiple decision rules. *Mimeograph Series* #489, Dept. of Statistics, Purdue University.

[3] Berger, R. L. (1979). Minimax subset selection for loss measured by subset size. *Ann. Statist.* 7, 1333-1338.

[4] Berger, R. L. and Gupta, S. S. (1980). Minimax subset selection rules with applications to unequal variance (unequal sample size) problems. *Scand. J. Statist.* 7, 21-26.

[5] Bjørnstad, J. F. (1978). The subset selection problem. I. The class of Schur-procedures and its properties. *Mimeograph Series* #78-24, Dept. of Statistics, Purdue University.

[6] 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.

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

[8] Gupta, S. S. and Huang, D. Y. (1977). Some multiple decision problems in analysis of variance. *Comm. Statist. - Theor. Meth.* A6(11), 1035-1054.

[9] Gupta, S. S. and Leong, Y. K. (1979). Some results on subset selection procedures for double exponential populations. In *Decision Information* (ed. by C. P. Tsokos and R. M. Thrall), Academic Press, New York, 277-305.

[10] Gupta, S. S. and Studden, W. J. (1966). Some aspects of selection and ranking procedures with applications. *Mimeograph Series* #81, Dept. of Statistics, Purdue University.

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

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

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

[14] Lehmann, E. L. (1961). Some model 1 problems of selection. *Ann. Math. Statist.* 32, 990-1012.

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

[16] Mudholkar, G. S. (1969). A generalized monotone character of d.f.'s and moments of statistics from some well-known populations. *Ann. Inst. Statist. Math.* 21, 277-285.

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

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

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

[20] Ryan, T. A., Jr. and Antle, C. E. (1976). A note on Gupta's selection procedure. *J. Amer. Statist. Assoc.* 71, 140-142.

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

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

[23] Studden, W. J. (1967). On selecting a set of k populations containing the best. *Ann. Math. Statist.* 38, 1072-1078.