THE SUBSET SELECTION PROBLEM.II. ON THE OPTIMALITY OF SOME SUBSET SELECTION PROCEDURES[1]

by

Jan F. Bjørnstad
University of California, Berkeley and Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #78-27

November 1978

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

SUMMARY

The Subset Selection Problem.II. On the Optimality of Some Subset Selection Procedures

by

Jan F. Bjørnstad

University of California, Berkeley and Purdue University

This is Part II of a two-part paper. 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 $\theta_i > \max_{1\le j\le k} \theta_j - \Delta$, and a bad population if $\theta_i \le \max_{1\le j\le k} \theta_j - \Delta$. Here, Δ is a specified positive constant.

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, the optimality considerations suggest three procedures as the main contenders for this problem. Two of these are the "average-type" procedure, and the classical "maximum-type" procedure. The third procedure has not before been considered as a serious contender.

THE SUBSET SELECTION PROBLEM.II. ON THE

OPTIMALITY OF SOME SUBSET SELECTION PROCEDURES[1]

By

Jan F. Bjørnstad

University of California, Berkeley and Purdue University

1. Introduction and Summary

This is part II of a two-part paper dealing with the subset selection approach to the problem of selecting a subset of good populations out of k populations $\Pi_1,\ldots,\Pi_k$. The populations are characterized by $\theta_1,\ldots,\theta_k$ respectively, and X_i is the observation from population Π_i. $\underline{\theta} = (\theta_1,\ldots,\theta_k) \in \Omega$, $\underline{X} = (X_1,\ldots,X_k)$. $\underline{X}$ is a sufficient statistic for $\underline{\theta}$, $X_1,\ldots,X_k$ are assumed to be independent and X_i has density $f(x-\theta_i)$ with respect to Lebesgue measure. The ordered θ_i are denoted by $\theta_{[1]} \leq \cdots \leq \theta_{[k]}$, and $\Pi_{(i)}$, $X_{(i)}$ correspond to $\theta_{[i]}$. The population Π_i is called a best population if $\theta_i = \theta_{[k]}$. The risk criteria we will consider depends only on the individual selection probabilities of the procedure. We can therefore define a subset selection procedure by:

$$\psi(\underline{x}) = [\psi_1(\underline{x}),\ldots,\psi_k(\underline{x})] \tag{1.1}$$

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

$$\sum_{i=1}^{k} \psi_i(\underline{x}) \geq 1, \ \forall \underline{x}. \tag{1.2}$$

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

I.e. we always selected (for non-randomized procedures) at least one population.

In Part I we discussed monotonicity properties of certain risk functions for the class of procedures called Schur-procedures. In Part II we study the problem of finding minimax procedures for the criteria in Part I.

The "subset selection approach" is one of the two basic formulations of the selection problem. The other is the so-called "indifference-zone approach", introduced by Bechhofer (1954), where we wish to select only a single population (or more generally a fixed number of populations), using the natural selection procedure. The problem is to determine how large the sample size from each population has to be, so that the probability of selecting the best population is bounded below by a given constant outside a certain indifference region. The related problem of how to select the best one of several populations was considered in a decision-theoretic framework first by Bahadur (1950). Bahadur showed that for a large class of loss functions and certain families of distributions, the natural selection procedure is the uniformly best permutation-symmetric procedure. The work of Bahadur was generalized and supplemented by Bahadur and Robbins (1950), Bahadur and Goodman (1952), Lehmann (1966) and Eaton (1967).

For the subset selection approach there have been few optimality results.

The first papers on the subset selection problem were Paulson (1949), Seal (1955), Gupta (1956) and Seal (1957). They all dealt with normal means. Two rules were proposed, ψ^a and ψ^m (see also Part I). The procedures are given by

$$\psi_i^a = 1 \quad \text{iff } X_i \geq \frac{1}{k-1} \sum_{j \neq i} X_j - c \tag{1.3}$$

$$\psi_i^m = 1 \quad \text{iff } X_i \geq \max_{1 \leq j \leq k} X_j - d. \tag{1.4}$$

Here c, d are determined such that the classical condition

$$\inf_{\underline{\theta} \in \Omega} P_{\underline{\theta}}(CS|\psi) \geq \gamma \tag{1.5}$$

holds with equality. CS (correct selection) stands for a selection that includes the best population.

ψ^a was proposed by Seal (1955), and ψ^m was proposed by Gupta (1956) and Seal (1957), and also be Paulson (1949) with a slightly different control condition. These two procedures have been the main contenders for this problem.

In 1966, Gupta and Studden gave the first optimality result. Let $S(\underline{\theta},\psi)$ be the expected size of the selected subset. It was shown that if X_i has density $f(x-\theta_i)$ and $f(x-\theta_i)$ has monotone likelihood ratio, MLR, in x then ψ^m minimizes $\sup_\Omega S(\underline{\theta},\psi)$ among all permutation-invariant procedures satisfying the basic requirement (1.5). Since the group of permutations is finite they showed in fact that ψ^m is minimax among all procedures subject to (1.5) (see e.g. Ferguson (1967)). This result is generalized to other families of distributions by Berger (1977), who also considers the risk S', the expected number of non-best populations selected. Berger showed that the minimax risks subject to (1.5) are

$$\begin{aligned} &k\gamma \text{ with respect to } S(\underline{\theta},\psi) \\ \text{and}\quad &(k-1)\gamma \text{ with respect to } S'(\underline{\theta},\psi) \end{aligned}$$

It follows then from a monotonicity result for ψ^m, given by Gupta

(1965), that ψ^m is minimax both with respect to S and S' under the condition (1.5).

Other recent papers that consider optimality problems are Gupta and Miescke (1978), Berger and Gupta (1977), and Gupta and Huang (1977).

We shall here not consider Bayes procedures. Some of the main contributions in this field are Deely and Gupta (1968), Goel and Rubin (1977), Chernoff and Yahav (1977), and Hsu (1977).

This paper deals with minimax theory. The criteria discussed in Part I are considered. For easy reference let us present them here.

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

Here $\psi_{(i)}$ corresponds to $\theta_{[i]}$, for $i = 1,\ldots,k$.

We say that Π_i is a good population if $\theta_i > \theta_{[k]} - \Delta$ and a bad population if $\theta_i \le \theta_{[k]} - \Delta$. Δ is a given positive constant. Then a reasonable criterion is:

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

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

In many cases we want to attach more weight to the worst populations than to those closest to the best. As observed in Part I the following criteria accomplish this and will be considered later.

$$L(\underline{\theta},\psi) = \sum_{i=1}^{k-1} \log[E_{\underline{\theta}}\ \{\psi_{(i)}\}] \tag{1.8}$$

(1.9) $$\ell(\underset{\sim}{\theta},\psi) = \sum_{i \in I_\Delta} \log\{E_{\underset{\sim}{\theta}}(\psi_i)\}.$$

As we have mentioned, a widely used criterion is

(1.10) $$S(\underset{\sim}{\theta},\psi) = \sum_{i=1}^{k} E_{\underset{\sim}{\theta}}(\psi_i).$$

This author feels that $B(\underset{\sim}{\theta},\psi)$ and $S'(\underset{\sim}{\theta},\psi)$ are more appropriate risk functions than $S(\underset{\sim}{\theta},\psi)$. For a discussion of the different criteria we refer to Section 6. For a given criterion we want the procedures to satisfy a certain control condition. As mentioned in Part I several different conditions can be of interest. One that we will consider is

(1.11) $$\inf_{\underset{\sim}{\theta} \in \Omega} R(\underset{\sim}{\theta},\psi) \geq \gamma, \gamma < 1,$$

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

It will turn out that a new procedure ψ^e, which was studied in a different context by Studden (1967), has certain minimax properties. In Section 2 we discuss the control conditions we are interested in, and show how to determine ψ^m, ψ^a and ψ^e so that they satisfy the different conditions.

The minimax problems we consider are presented in Section 3. Section 4 deals with minimax theorems in the general location-model for the risk function $B(\underset{\sim}{\theta},\psi)$, $S'(\underset{\sim}{\theta},\psi)$ and $S(\underset{\sim}{\theta},\psi)$, subject to the condition (1.11). It is also shown that ψ^a is minimax with respect to the criteria $S'(\underset{\sim}{\theta},\psi)$ and $S(\underset{\sim}{\theta},\psi)$ subject to the condition (1.5) if γ is sufficiently large.

In Section 5 normal populations are considered. It is shown that the new procedure ψ^e is minimax over a certain slippage-type subset of Ω for $B(\underset{\sim}{\theta},\psi)$, subject to (1.11). We also prove that ψ^e is minimax among all permutation-invariant procedures with respect to the whole parameter space for $\ell(\underline{\theta},\psi)$, subject to (1.11). We derive some new minimax results for ψ^a, and a new minimax property of ψ^m is proved for the case when k tends to infinity.

2. Control Conditions

The control condition we will most consider is (1.11), i.e.

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

where

$$R(\underline{\theta},\psi) = \sum_{i\in I_\Delta^c} E_{\underline{\theta}}(\psi_i).$$

Let the class of procedures satisfying (1.11) be denoted by $\mathcal{S}(\gamma,\Delta)$. We observe that $R(\underline{\theta},\psi) = P_{\underline{\theta}}(CS|\psi)$ if $\theta_{[k]} - \theta_{[k-1]} \geq \Delta$. For this reason and in its own right, the condition

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

is also of interest. Here

(2.2) $$\Omega(\Delta) = \{\underline{\theta}: \theta_{[k]} - \theta_{[k-1]} \geq \Delta\}.$$

Let $\mathcal{S}'(\gamma,\Delta)$ be the class of procedures satisfying (2.1). Since (1.11) implies (2.1), we have that $\mathcal{S}(\gamma,\Delta) \subset \mathcal{S}'(\gamma,\Delta)$. Let $\mathcal{S}_1(\gamma)$ be the class of procedures satisfying (1.5), the control requirement mostly used in the literature. Note that $\mathcal{S}(\gamma,\Delta)$ and $\mathcal{S}'(\gamma,\Delta)$ are two sequences of sets which as $\Delta \to 0$ tend to $\mathcal{S}_1(\gamma)$. The condition (1.5) means that we require the probability of selecting the best population $\Pi_{(k)}$ to be at least γ, even when $\theta_{[k]}$ is close to the other θ_i. However, in such a case it does not matter much if instead of $\Pi_{(k)}$ we select, say the second-best population $\Pi_{(k-1)}$. As a consequence of this we are really only interested in controlling the probability of selecting the best population when $\theta_{[k]} - \theta_{[k-1]} \geq \Delta$, with Δ not too small. Hence the condition (2.1) seems more appropriate than (1.5). The unsatisfactory feature with (2.1) is that we do not control what happens on $\Omega^c(\Delta)$. This leads then naturally to the control (1.11) as the more appropriate condition of the three we have presented, since it implies (2.1), but also controls what happens on $\Omega^c(\Delta)$.

From the above remarks it follows that we are mostly interested in the classes $\mathcal{S}(\gamma,\Delta)$ and $\mathcal{S}'(\gamma,\Delta)$. We will, however, also present new minimax results for the class $\mathcal{S}_1(\gamma)$.

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

(2.3) $$\sup_{\underline{\theta}\in\Omega} S(\underline{\theta},\psi) \leq \beta.$$

The class of procedures satisfying (2.3) is denoted by $\mathcal{S}_2(\beta)$. A condition like (2.3) may be imposed for example because of practical restrictions on how many populations we can select.

Let us now consider the procedures ψ^a and ψ^m given by (1.3) and (1.4). The critical constants c and d are determined according to what control condition we want the procedures to satisfy. It is easily seen (see e.g. Lemma 3.1 in Part I, and Gupta (1965)) that

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

and

$$\inf_{\underline{\theta} \in \Omega(\Delta)} P_{\underline{\theta}}(CS|\psi) \text{ occurs at } \theta_{[1]} = \ldots = \theta_{[k-1]} = \theta_{[k]} - \Delta$$

for ψ equal to ψ^a or ψ^m. Let $d_1(\gamma)$, $c_1(\gamma)$ be the values of d, c such that ψ^a and ψ^m satisfy (1.5) with equality, and let $d'(\gamma)$, $c'(\gamma)$ be the values of d, c such that (2.1) is satisfied with equality. Then

$$(2.4) \qquad \begin{aligned} d' &= d_1 - \Delta \\ c' &= c_1 - \Delta \end{aligned}$$

We observe that (1.2) is satisfied, i.e. at least one population is always selected if and only if $d \geq 0$ and $c \geq 0$. Hence if $\psi^a(\psi^m) \in \mathcal{D}'(\gamma,\Delta)$ then (1.2) holds if $\Delta \leq c_1$ $(\Delta \leq d_1)$.

Let us now turn to our main control condition (1.11), and procedure ψ^m. Let $d = d'$ so that $\psi^m \in \mathcal{D}'(\gamma,\Delta)$, and assume $d' \geq 0$. Then it is readily seen, by using a method similar to the one used by Gupta (1965) to find $\sup_{\Omega} S(\underline{\theta},\psi^m)$, that ψ^m satisfies (1.11) with equality, if the density $f(x-\theta)$ has MLR in x. From Gupta (1965) we also see that $\psi^m \in \mathcal{D}_2(\beta)$ if $f(x-\theta)$ has MLR in x, $\beta \geq 1$ and $d = d_1(\beta/k)$. This follows from the fact that

$$\sup_{\theta \in \Omega} S(\underline{\theta},\psi^m) \text{ occurs when } \theta_1 = \ldots = \theta_k$$

if $f(x-\theta)$ has MLR in x and $d \geq 0$.

For the remaining part of this section we will consider the case of normal populations with known variance, i.e. $X_1,\ldots,X_k$ are independent, normally distributed with variance 1 and $E_{\underline{\theta}}(X_i) = \theta_i$ for $i = 1,\ldots,k$. Let the $\mathcal{N}(0,1)$-density be denoted by $\phi(x)$. Then $f(x-\theta) = \phi(x-\theta)$ in this case. Let Φ be the $\mathcal{N}(0,1)$-distribution function, and let $z(\gamma)$ be defined by $\Phi(z(\gamma)) = \gamma$. Then

$$(2.5)\qquad \begin{aligned} c_1(\gamma) &= \sqrt{\frac{k}{k-1}}\, z(\gamma) \\ c'(\gamma) &= \sqrt{\frac{k}{k-1}}\, z(\gamma) - \Delta. \end{aligned}$$

$d_1(\gamma)$ is tabulated by Gupta (1956) and Gupta (1965).

Let us now consider the problem of determining c such that ψ^a satisfies (1.11). It turns out that usually $c = c'$, given by (2.5). It will be shown in Sections 4 and 5 and ψ^a has optimality properties only if γ is sufficiently large. In particular for the class $\mathcal{S}(\gamma,\Delta)$, ψ^a has a certain minimax property if $\gamma > (k-1)/k$. Let

$$(2.6)\qquad \Delta_a(\gamma) = \sqrt{\frac{k}{k-1}}\ \min\{[\tfrac{1}{2}z(\gamma) + \tfrac{1}{2}z(\tfrac{k-1}{k})],\ [z(\gamma) - z(\tfrac{k-1}{k})]\}.$$

It is straightforward to show that if $\gamma > (k-1)/k$, $\Delta \le \Delta_a(\gamma)$ and $c = c'$, then ψ^a satisfies (1.11) with equality, and hence $\psi^a \in \mathcal{S}(\gamma,\Delta)$. We note that $\Delta_a(\gamma) = \sqrt{\frac{k}{k-1}}\{z(\gamma) - z(\frac{k-1}{k})\}$ if and only if $z(\gamma) \le 3z(\frac{k-1}{k})$. As is seen above both for ψ^a and ψ^m, if we want to satisfy condition (1.11) we first determine c and d such that (2.1) is satisfied, and then show that (2.1) implies (1.11) if Δ is not too large.

Let for any c,

$$\gamma(c) = P(Y_1 \ge \frac{1}{k-1} \sum_{j=2}^{k} Y_j - c)$$

where $Y_1,\ldots,Y_k$ are independent, $\mathcal{N}(0,1)$ random variables. It is shown in Section 4 (see Theorem 4.3) that

$$\sup_{\underline{\theta}\in\Omega} S(\underline{\theta},\psi^a) \text{ is attained at } \theta_1 = \ldots = \theta_k$$
$$\text{if and only if } \gamma(c) \ge \frac{k-1}{k}.$$

It follows that ψ^a satisfies condition (2.3) with equality if

$$c = c_1(\beta/k) = \sqrt{\frac{k}{k-1}}\, z(\beta/k) \text{ and } \beta \ge k-1.$$

Berger (1977) and Gupta and Studden (1966) showed that ψ^m has certain minimax properties in the class $\mathcal{D}_1(\gamma)$ if $f(x-\theta)$ has MLR in x. We show in Section 4 that ψ^a has the same minimax properties under some additional assumptions. In Section 5 it is shown that in the case of normal populations ψ^a also has some minimax properties in the classes $\mathcal{D}'(\gamma,\Delta)$, $\mathcal{D}(\gamma,\Delta)$ and $\mathcal{D}_2(\beta)$.

As will turn out in Section 5, in addition to ψ^a and ψ^m, a third procedure, denoted by ψ^e, has certain minimax properties in the classes $\mathcal{D}(\gamma,\Delta)$ and $\mathcal{D}'(\gamma,\Delta)$ when we have normal populations. ψ^e is given by

$$\psi_i^e = 1 \text{ iff } Ce^{\Delta X_i} \geq \sum_{j\neq i} e^{\Delta X_j}. \tag{2.7}$$

Remark. In the general case, when X_i is the sample mean of n independent normal variables with known variance σ^2, then ψ^e is given by

$$\psi_i^e = 1 \text{ iff } Ce^{\Delta_0\sqrt{n}X_i/\sigma} \geq \sum_{j\neq i} e^{\Delta_0\sqrt{n}X_j/\sigma} \tag{2.8}$$

where

$$\Delta_0 = \sqrt{n}\,\Delta/\sigma. \tag{2.9}$$

C is determined such that (2.1) is satisfied with equality. From Part I, Lemma 3.1 we have that

$$\inf_{\underline{\theta}\in\Omega(\Delta)} P_{\underline{\theta}}(CS|\psi^e) \text{ occurs when } \theta_{[1]} = \cdots = \theta_{[k-1]} = \theta_{[k]} - \Delta.$$

This implies that C is determined by

$$\gamma = P(Ce^{\Delta Y_k+\Delta^2} \geq \sum_{j=1}^{k-1} e^{\Delta Y_j}) \tag{2.10}$$

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 (2.1) or (1.11). However, we see from (2.10) that ψ^e satisfies (1.5) if we instead of C use $e^{\Delta^2}C$ as the critical constant.

We next discuss the problem of determining C such that $\psi^e \in \mathcal{S}(\gamma,\Delta)$, i.e.

$$\inf_{\underline{\theta}\in\Omega} R(\underline{\theta},\psi^e) = \gamma.$$

As seen in Theorem 2.1 below, usually C will be given by (2.10).

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

<u>Theorem 2.1</u>. *Let* $X_1,\ldots,X_k$ *be independent;* X_i *is* $\mathcal{N}(\theta_i,1)$ *for* $i = 1,\ldots,k$.

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

$$\inf_{\underline{\theta}\in\Omega(\Delta)} P_{\underline{\theta}}(CS|\psi^e) = \gamma \Rightarrow \inf_{\underline{\theta}\in\Omega} R(\underline{\theta},\psi^e) = \gamma. \tag{2.11}$$

(ii) *Let* $k \geq 4$ *and* $C \geq k-1$. *Then (2.11) holds if*

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

where

$$b_k(\gamma) = \begin{cases} \gamma/9 & \text{if } k = 4 \\ \gamma/7 & \text{if } k = 5 \\ (11/75)\gamma & \text{if } k \geq 6 \end{cases} \tag{2.12}$$

and

$$\underline{\theta}^\Delta = (0,\Delta,\ldots,\Delta).$$

Remark. Let Δ_γ be defined by

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

$E_{\underline{\theta}^\Delta}(\psi_1^e)$ is tabulated in Table 2. We see that $E_{\underline{\theta}^\Delta}(\psi_1^e)$ typically increases in Δ to begin with and decreases from a certain point. It is clear from Table 2 that if $\Delta \leq \Delta_\gamma$, then $\psi^e \in \mathcal{S}(\gamma,\Delta)$. (For general $\sigma/\sqrt{n}$ we require $\Delta_0 \leq \Delta_\gamma$). Using Table 2 we can find approximate values of Δ_γ. For example if $k = 10$ and $\gamma = .95$ then $\Delta_\gamma \approx 1.5$.

To prove Theorem 2.1 we need the following two lemmas.

Lemma 2.1. *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 $\theta_1 \leq \dots \leq \theta_k$ and consider $r(\underline{\theta}) = E_{\underline{\theta}}(\psi_i^e + \psi_j^e)$. Now,

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

We get:

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

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

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

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

$$(2.14) \qquad [1 + \sum_{\ell\neq 1} e^{\Delta(y_\ell - y_i)}] \cdot \exp\{(\theta_j-\theta_i)\left(y_i - \frac{1}{\Delta}\log[\frac{1}{C}\sum_{\ell=1}^{k-1} e^{\Delta y_\ell}]\right)\} \geq 1.$$

Let $y_{max} = \max(y_1,\ldots,y_{k-1})$. Then

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

Hence the left side of (2.14) is greater than or equal to

$$\exp\{(\theta_j-\theta_i)\left(y_{max} - \frac{1}{\Delta}\log[\frac{1}{C}\sum_{\ell=1}^{k-1} e^{\Delta y_\ell}]\right) \geq 1,$$

since $C \geq k-1$ implies that

$$y_{max} \geq \frac{1}{\Delta}\log[\frac{1}{C}\sum_{\ell=1}^{k-1} e^{\Delta y_\ell}]. \qquad \text{Q.E.D.}$$

Remark. If $C < k-1$, then Lemma 2.1 is not necessarily true as seen by the following example.

Let $k = 2$. If Lemma 2.1 holds then

$$E_{\theta_1=\theta_2}(\psi_1^e+\psi_2^c) \geq E_{\theta_1=\theta_2-\Delta}(\psi_1^e+\psi_2^e) \geq \gamma$$

since

$$E_{\theta_1=\theta_2-\Delta}(\psi_2^e) = \gamma.$$

However:

$$E_{\theta_1=\theta_2}(\psi_1^e+\psi_2^e) = 2\Phi(\frac{\log C}{\Delta\sqrt{2}}) < \gamma$$

if

$$C < e^{\Delta\sqrt{2}\, z(\gamma/2)} < 1 = k-1.$$

Lemma 2.2. *Let* $k \geq 4$ *and* $C \geq k-1$, *and let* p *be such that* $1 \leq p \leq k-3$. *Assume that*

$$R(\underline{\theta},\psi^e) = \sum_{i=p+1}^{k} E_{\underline{\theta}}(\psi_{(i)}^e),$$

i.e. $\theta_{[p+1]} > \theta_{[k]} - \Delta$ *and* $\theta_{[p]} \leq \theta_{[k]} - \Delta$. *If*

$$\inf_{\underline{\theta}\in\Omega(\Delta)} P_{\underline{\theta}}(CS|\psi^e) = \gamma$$

then

$$R(\underline{\theta},\psi^e) \geq \frac{k-p}{2^{k-1-p}}\gamma + (k-p)\{1 - (\tfrac{1}{2})^{k-1-p}\}E_{\underline{\theta}^{(p+1)}}(\psi_1^e). \tag{2.15}$$

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

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

Proof. Let $\psi = \psi^e$. We may assume that $\theta_i = \theta_{[i]}$ for $i = 1,\ldots,k$. Then

$$R(\underline{\theta},\psi) = \sum_{i=p+1}^{k} E_{\underline{\theta}}(\psi_i).$$

Each $E_{\underline{\theta}}(\psi_i)$, $i \geq p+1$, is nonincreasing in each of $\theta_1,\ldots,\theta_p$, so we can let $\theta_1 = \cdots = \theta_p = \theta_k - \Delta$. Define $\underline{\theta}^q$ by:

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

Let $i \ge p+3$. Then

$$E_{\underline{\theta}}(\psi_i) \ge \frac{k-p}{4(k-p-2)} E_{\underline{\theta}}(\psi_i) + \frac{2(k-p)-4}{4(k-p-2)} E_{\underline{\theta}}(\psi_{p+1}) + \frac{(k-p)-4}{4(k-p-2)} E_{\underline{\theta}}(\psi_{p+2}).$$

It follows that

$$(2.16) \qquad \sum_{i=p+1}^{k} E_{\underline{\theta}}(\psi_i) \ge \frac{k-p}{4} E_{\underline{\theta}}(\psi_{p+1}+\psi_{p+2}) + \frac{k-p}{4(k-p-2)} \sum_{i=p+3}^{k} E_{\underline{\theta}}(\psi_{p+1}+\psi_i).$$

Let q be an integer, $1 \le q \le k-2$. Then for $i \ge q+2$:

$$E_{\underline{\theta}}(\psi_i) \ge \frac{k-q}{2(k-q-1)} E_{\underline{\theta}}(\psi_i) + \frac{k-q-2}{2(k-q-1)} E_{\underline{\theta}}(\psi_{q+1})$$

This implies that

$$(2.17) \qquad \sum_{i=q+1}^{k} E_{\underline{\theta}}(\psi_q+\psi_i) \ge \frac{k-q}{2} E_{\underline{\theta}}(\psi_q+\psi_{q+1}) + \frac{k-q}{2(k-q-1)} \sum_{i=q+2}^{k} E_{\underline{\theta}}(\psi_q+\psi_i).$$

The idea in the proof is to apply Lemma 2.1 to decrease first θ_{p+1}, then θ_{p+2} etc. to $\theta_k-\Delta$ in $E_{\underline{\theta}}(\psi_k)$ and then use the fact that $\gamma = E_{\underline{\theta}^{k-1}}(\psi_k)$. Let $\underline{\theta}_0 = \underline{\theta}^{(p+1)}$. From (2.16) and (2.17) it now follows from induction that

$$(2.18) \qquad \sum_{i=p+1}^{k} E_{\underline{\theta}}(\psi_i) \ge (k-p)\left(\sum_{i=1}^{m-1} 2^{-i}\right)E_{\underline{\theta}_0}(\psi_1)$$

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

for $m = 2,\ldots,k-p-1$.

By letting $m = k-p-1$ in (2.18) we get:

$$\sum_{i=p+1}^{k} E_{\underline{\theta}}(\psi_i) \geq (k-p)\left(\sum_{i=1}^{k-p-2} 2^{-i}\right)E_{\underline{\theta}_0}(\psi_1) + \frac{k-p}{2^{k-p-1}} E_{\underline{\theta}^{k-1}}(\psi_{k-1}+\psi_k)$$

$$\geq (k-p)\left(\sum_{i=1}^{k-p-1} 2^{-i}\right)E_{\underline{\theta}_0}(\psi_1) + \frac{k-p}{2^{k-1-p}}\gamma. \qquad \text{Q.E.D.}$$

Proof of Theorem 2.1. The result for $k = 2$ follows directly since for $\underline{\theta} \in \Omega^c(\Delta)$, $R(\underline{\theta},\psi^e) = E_{\underline{\theta}}(\psi_1^e+\psi_2^e) \geq 1$. (2.11) for $k = 3$ follows immediately from Lemma 2.1. For $k \geq 4$, we have from Lemma 2.2 that (2.11) holds if

$$\frac{k-p}{2^{k-p-1}}\gamma + (k-p)\left(1 - \left(\tfrac{1}{2}\right)^{k-p-1}\right)E_{\underline{\theta}^{p+1}}(\psi_1^e) \geq \gamma$$

for all p such that $1 \leq p \leq k-3$. Since $E_{\underline{0}^{p+1}}(\psi_1^e) \geq E_{\underline{\theta}^{\Delta}}(\psi_1^e)$ for all $p \geq 1$, (2.11) holds if

(2.19) $$E_{\underline{\theta}^{\Delta}}(\psi_1^e) \geq \gamma \cdot \max_{3\leq m\leq k-1} g(m).$$

where

$$g(m) = \frac{1-m(1/2)^{m-1}}{m-m(1/2)^{m-1}} \quad ; \quad 3 \leq m \leq k-1.$$

It is readily seen that

$$\max_{3\leq m\leq k-1} g(m) = \begin{cases} g(3) = 1/9 & \text{if } k = 4 \\ g(4) = 1/7 & \text{if } k = 5 \\ g(5) = 11/75 & \text{if } k \geq 6. \end{cases} \qquad \text{Q.E.D.}$$

Remarks. (a). Let $k = 2$. Then it is readily seen that

$$\inf_{\underline{\theta}\in\Omega} R(\underline{\theta},\psi^e) < \gamma$$

if

$$\Delta > \sqrt{2}\{z(\gamma) + z(1 - \tfrac{\gamma}{2})\} .$$

Hence (2.11) is not necessarily true if Δ becomes too large.

(b) By applying the geometric-arithmetic mean inequality, it is readily seen that

$$C \geq (k-1)\exp\{\Delta\sqrt{\frac{k}{k-1}}\, z(\gamma) - \Delta^2\}$$

and therefore

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

We conclude this section with a few comments on the case of unknown variance σ^2, i.e. X_i is the sample mean of n normal random variables with unknown variance σ^2, such that $\mathrm{Var}(X_i) = \sigma^2/n$. Let S^2 be the usual unbiased estimator of σ^2 such that $\nu S^2/\sigma^2$ has a chi-square distribution with $\nu = k(n-1)$ degrees of freedom. Then ψ^a, ψ^m is modified by replacing d with $d_\nu S/\sqrt{n}$ and c with $c_\nu \cdot S/\sqrt{n}$ (see Gupta (1965) and Seal (1955)). It is easily seen by conditioning on S that

$$\inf_{\underline{\theta},\sigma\in\Omega} P_{\underline{\theta},\sigma}(CS|\psi) \text{ occurs at } \theta_1 = \ldots = \theta_k, \text{ for } \psi = \psi^a, \psi^m.$$

Selected values of d_ν such that $\psi^m \in \mathcal{S}_1(\gamma)$ is tabulated in Gupta and Sobel (1957). The values of c_ν such that $\psi^a \in \mathcal{S}_1(\gamma)$ can be obtained from the t-distribution.

The obvious modification of ψ^e, given by (2.8) is to replace σ by S. Hence the modified rule ψ^e_m is given by

$$\psi^e_{m,i} = 1 \text{ iff } C_m e^{\Delta_0\sqrt{n}X_i/S} \geq \sum_{j\neq i} e^{\Delta_0\sqrt{n}X_j/S} .$$

By conditioning on S we easily see that

$$\inf_{\substack{\theta_{[k]}-\theta_{[k-1]} \\ \geq \sigma\Delta_0/\sqrt{n}}} P_{\underline{\theta}}(CS|\psi^e) \text{ occurs at } \theta_{[1]} = \cdots = \theta_{[k-1]} = \theta_{[k]} - \sigma\Delta_0/\sqrt{n}.$$

However no tabulated values of C_m exist.

3. Presentation of the Optimality Problems

The notion of good and bad populations is the main concept. Our approach is in principal similar to the one considered by Lehmann (1961) for the problem of comparing populations with a standard or control.

The risk functions we are mostly interested in are $B(\underline{\theta},\psi)$ and $\ell(\underline{\theta},\psi)$, given by (1.7) and (1.9). Specifically we consider the problem of finding subset selection procedures, which subject to (1.11)

$$\text{minimize} \sup_{\underline{\theta}\in\Omega_1} B(\underline{\theta},\psi) \tag{3.1}$$

and

$$\text{minimize} \sup_{\underline{\theta}\in\Omega_1} \ell(\underline{\theta},\psi). \tag{3.2}$$

Ω_1 (to be specified later) is the set of parameter-values where we want the selection-procedure to have good performance. In Section 4 a solution to the goal (3.1) is found for Ω_1 equal to a slippage-set of the type given in Part I, (5.6). In Section 5 we find a solution to the goal (3.2) for normal populations with $\Omega_1 = \Omega$, the whole parameter-space.

The general minimax theorems in Section 4 are given for the class $\mathcal{S}'(\gamma,\Delta)$ of procedures satisfying the control condition (2.1). For applications, we then have to show that the optimal procedure also lies in

$\mathcal{D}(\gamma,\Delta)$, so that the procedure is also optimal in the class $\mathcal{D}(\gamma,\Delta)$. As is seen in Section 5 for normal populations, this will usually be the case if Δ is not too large. (See also Theorem 2.1.)

Other risk functions of interest for the class $\mathcal{D}(\gamma,\Delta)$ are $S'(\underline{\theta},\psi)$ and $L(\underline{\theta},\psi)$ given by (1.6) and (1.8). As is shown later, the optimal procedures for the goals to (3.1) and (3.2) will also have certain minimax properties for S' and L. For the class of procedures $\mathcal{D}_1(\gamma)$ satisfying the classical control condition (1.5), Berger (1977) has shown that the minimax risks are $(k-1)\gamma$ and $k\gamma$ for respectively $S'(\underline{\theta},\psi)$ and $S(\underline{\theta},\psi)$. Using this result Berger showed that ψ^m is minimax for S and S' in the class $\mathcal{D}_1(\gamma)$. In Section 4 it is shown that ψ^a is also minimax if γ is large enough.

For the class of procedures $\mathcal{D}_2(\gamma)$ satisfying (2.3), the goal is to

(3.3) $$\text{maximize } \inf_{\underline{\theta}\in\Omega} R(\underline{\theta},\psi) \text{ and } \inf_{\underline{\theta}\in\Omega(\Delta)} P_{\underline{\theta}}(CS|\psi).$$

Also the dual problem of

(3.4) $$\text{minimizing } \inf_{\underline{\theta}\in\Omega} S(\underline{\theta},\psi) \quad \text{for } \psi \in \mathcal{D}(\gamma,\Delta)$$

is considered.

4. Some General Minimax Theorems in the Location-Model

The model: $X_1,\ldots,X_k$ are independent. X_i has density $f(x-\theta_i)$ with respect to Lebesgue-measure. The joint density is

$$p(\underline{x}-\underline{\theta}) = \sum_{i=1}^{k} f(x_i-\theta_i).$$

Let $\underline{\delta}_i$ be the vector in $\mathbb{R}^k$ where the i^{th} coordinate is equal to 1 and the rest is equal to zero. Define $\underline{\theta}_i^\Delta$ by

(4.1) $$\underline{\theta}_i^\Delta = \Delta\underline{\delta}_i^\Delta.$$

Let $p_i(\underset{\sim}{x}) = p(\underset{\sim}{x}-\underset{\sim}{\theta}_i^{\Delta})$, and define the statistic

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

ψ^0 is the subset selection procedure given by:

(4.3) $$\psi_i^0(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } T_i < C \\ 0 & \text{if } T_i > C \end{cases}$$

where C is determined by

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

Let us recall from Part I (see Definitions 3.3 and 3.4), that a procedure ψ is said to be just if $\psi_i(\underset{\sim}{x})$ is non-decreasing in x_i and non-increasing in x_j, $j \neq i$, for $i = 1,\ldots,k$, and ψ is said to be a <u>Schur-procedure</u> if (i) ψ is just and translation-invariant and (ii) ψ_i is the same Schur-concave function of $\underset{\sim}{x}_i^* = \{x_j - x_i : j\neq i\}$, for all i. (For a definition of a Schur-concave function we refer to Part I, Section 3.) Let now

(4.5) $$\Omega_p(\Delta) = \{\underset{\sim}{\theta} \in \Omega:\ \theta_{[k]}-\theta_{[p]}<\Delta \text{ and } \theta_{[k]}-\theta_{[p-1]}\geq\Delta + \sum_{i=1}^{p-2}(\theta_{[p-1]}-\theta_{[i]})\},$$

for $p = 1,\ldots,k$ and let

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

Let us review the properties of Schur-procedures, derived in Part I, we need to prove our first result. (See Part I, Theorems 5.3 and 5.4.)

If $f(x-\theta)$ has MLR in x, and ψ is a Schur-procedure, then

$$\text{(4.7)}\qquad \sup_{\underline{\theta}\in\Omega_1} B(\underline{\theta},\psi) = B(\underline{\theta}_i^{\Delta},\psi) \text{ and } \sup_{\underline{\theta}\in\Omega_k(\Delta)} S'(\underline{\theta},\psi) = S(\underline{\theta}_i^{\Delta},\psi)$$

for $i = 1,\ldots,k$.

The following result concerns $B(\underline{\theta},\psi)$ and $S'(\underline{\theta},\psi)$ for $\psi \in \mathcal{D}'(\gamma,\Delta)$.

Theorem 4.1. *Assume* $f(x-\theta)$ *has MLR in* x, *and that* ψ^0, *defined by* (4.3), *is a Schur-procedure. Let* Ω_1 *be given by* (4.6). *Then* ψ^0 *minimizes for all* $c \in \mathcal{D}'(\gamma,\Delta)$

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

Proof. We prove the theorem only for $B(\underline{\theta},\psi)$. The proof for $S'(\underline{\theta},\psi)$ is exactly the same. It is readily seen from (4.4) that $\psi^0 \in \mathcal{D}'(\gamma,\Delta)$. (See e.g. Part I, Lemma 3.1.) From (4.7) we have that

$$\sup_{\underline{\theta}\in\Omega_1} B(\underline{\theta},\psi^0) = B(\underline{\theta}_i^{\Delta},\psi^0) \quad \text{for } i=1,\ldots,k.$$

This implies that for any $\psi \in \mathcal{D}'(\gamma,\Delta)$

$$\sup_{\underline{\theta}\in\Omega_1} B(\underline{\theta},\psi) - \sup_{\underline{\theta}\in\Omega_1} B(\underline{\theta},\psi^0)$$

$$\geq \frac{1}{k}\sum_{i=1}^{k} B(\underline{\theta}_i^{\Delta},\psi) - \frac{1}{k}\sum_{i=1}^{k} B(\underline{\theta}_i^{\Delta},\psi^0)$$

$$= \frac{1}{k}\sum_{j=1}^{k}\sum_{i\neq j}\int (\psi_j-\psi_j^0)p_i\,d\nu$$

$$\geq \frac{1}{k}\sum_{j=1}^{k}\int(\psi_j-\psi_j^0)(\sum_{i\neq j} p_j)d\nu + \frac{C}{k}\sum_{j=1}^{k}\{\int\psi_j^0 p_j d\nu - \int\psi_j p_j d\nu\}$$

since

$$\int\psi_j^0 p_j d\nu = \gamma \quad \text{and } \int\psi_j p_j d\nu \geq \gamma \text{ for } j = 1,\ldots,k.$$

It follows that

$$\sup_{\underset{\sim}{\theta}\in\Omega_1} B(\underset{\sim}{\theta},\psi) - \sup_{\underset{\sim}{\theta}\in\Omega_1} B(\underset{\sim}{\theta},\psi^0) \geq \frac{1}{k}\sum_{j=1}^{k}\int(\psi_j-\psi_j^0)(\sum_{i\neq j} p_i - Cp_j)d\nu \geq 0.$$

Q.E.D.

Remarks. 1) In Section 5, Theorem 4.1 is applied to normal populations. It is shown that ψ^0 is the procedure ψ^e, defined by (2.7). It follows from Theorem 2.1 that in this case $\psi^0 \in \mathcal{D}(\gamma,\Delta)$ if $\Delta \leq \Delta_\gamma$; Δ_γ is defined by (2.13). 2) As mentioned in Part I, one way to interpret Ω, is to say that Ω_1 consists of the cases where the good populations have "slipped" from the bad populations.

In Section 5 it will be shown that for normal populations, ψ^0 is minimax also with respect to $\ell(\underset{\sim}{\theta},\psi)$ and $L(\underset{\sim}{\theta},\psi)$. More precisely, in the normal case we have that ψ^0 also minimizes $\sup_{\underset{\sim}{\theta}\in\Omega} \ell(\underset{\sim}{\theta},\psi)$ and $\sup_{\underset{\sim}{\theta}\in\Omega(\Delta)} L(\underset{\sim}{\theta},\psi)$ for all permutation-invariant procedures in $\mathcal{D}'(\gamma,\Delta)$.

Next we consider the problem of finding solutions to the dual goals (3.3) and (3.4). First we will consider the two simpler problems of

(4.8) maximizing $\inf_{\underset{\sim}{\theta}\in\Omega(\Delta)} P_{\underset{\sim}{\theta}}(CS|\psi)$ for $\psi \in \mathcal{D}_2(\beta)$

and

(4.9) minimizing $\sup_{\underset{\sim}{\theta}\in\Omega} S(\underset{\sim}{\theta},\psi)$ for $\psi \in \mathcal{D}'(\gamma,\Delta)$.

By the Hunt-Stein theorem we can restrict attention to translation-invariant subset selection procedures, i.e. we can assume that ψ_i is a function of $\underset{\sim}{x}_i^* = \{x_j - x_i : j\neq i\}$. Let $\underset{\sim}{\theta}_i^* = \{\theta_j-\theta_i : j\neq i\}$. Then $\underset{\sim}{x}_i^*$ has location-density $g(\underset{\sim}{y}-\underset{\sim}{\theta}_i^*)$, $\underset{\sim}{y} \in R^{k-1}$, where g is the density of $(U_1-U_k,\ldots,U_{k-1}-U_k)$, and $U_1,\dots,U_k$ are i.i.d. with density $f(u)$.

The solutions of problems (4.8) and (4.9) are given in the following result.

Theorem 4.2. *Let* $X_1,\dots,X_k$ *be independent;* X_i *has density* $f(x-\theta_i)$ *for* $i = 1,\dots,k$. *Define* ψ^* *by*

$$\psi_i^*(\underline{x}_i^*) = \begin{cases} 1 & \text{if } g(\underline{x}_i^*+\underline{\Delta}) > cg(\underline{x}_i^*) \\ 0 & \text{if } g(\underline{x}_i^*+\underline{\Delta}) < cg(\underline{x}_i^*) \end{cases} \tag{4.10}$$

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

$$\sup_{\underline{\theta}\in\Omega} S(\underline{\theta},\psi^*) \quad \text{occurs at } \theta_1 = \dots = \theta_k. \tag{4.11}$$

If c *is determined by*

$$\int\psi_i^*(\underline{y})g(\underline{y})d\nu(\underline{y}) = \beta/k \quad \text{for } i = 1,\dots,k \tag{4.12}$$

then ψ^* *maximizes for all* $\psi \in \mathcal{D}_2(\beta)$

$$\inf_{\underline{\theta}\in\Omega(\Delta)} P_{\underline{\theta}}(CS|\psi). \tag{4.13}$$

If c *is determined by*

$$\int\psi_i^*(\underline{y})g(\underline{y}+\underline{\Delta})d\nu(\underline{y}) = \gamma \quad \text{for } i = 1,\dots,k \tag{4.14}$$

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

$$\sup_{\underline{\theta}\in\Omega} S(\underline{\theta},\psi).$$

Proof. Let $\psi \in \mathcal{D}_2(\beta)$. We can assume that ψ is translation-invariant. Since ψ^* is just,

$\inf_{\underline{\theta}\in\Omega(\Delta)} P_{\underline{\theta}}(CS|\psi^*)$ occurs at the k points $\underline{\theta}_i^* = -\underline{\Delta}$; i = 1,...,k.

Hence

$$\inf_{\underline{\theta}\in\Omega(\Delta)} P_{\underline{\theta}}(CS|\psi^*) - \inf_{\underline{\theta}\in\Omega(\Delta)} P_{\underline{\theta}}(CS|\psi)$$

$$\geq \frac{1}{k}\sum_{i=1}^{k}\int[\psi_i^*(\underline{y}) - \psi_i(\underline{y})]g(\underline{y}+\underline{\Delta})d\nu(\underline{y})$$

$$\geq \frac{1}{k}\sum_{i=1}^{k}\int[\psi_i^*(\underline{y}) - \psi_i(\underline{y})]g(\underline{y}+\underline{\Delta})d\nu(\underline{y}) - \frac{c}{k}\{S(\underline{0},\psi^*) - S(\underline{0},\psi)\}$$

$$= \frac{1}{k}\sum_{i=1}^{k}\int[\psi_i^*(\underline{y}) - \psi_i(\underline{y})][g(\underline{y}+\underline{\Delta}) - cg(\underline{y})]d\nu(\underline{y}) \geq 0.$$

The proof of the second result goes in a similar way and is omitted.

Q.E.D.

Remarks. 1) Assume

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

Then if (4.12) is satisfied, ψ^* maximizes, for all $\psi \in \mathcal{D}_2(\beta)$

$$\inf_{\underline{\theta}\in\Omega} R(\underline{\theta},\psi).$$

If (4.14) is satisfied then ψ^* minimizes for all $\psi \in \mathcal{D}(\gamma,\Delta)$

$$\sup_{\underline{\theta}\in\Omega} S(\underline{\theta},\psi).$$

In Section 5 it is seen that for normal populations, (4.15) is true if Δ is not too large.

2) By employing an idea used by Spjøtvoll (1972) for multiple comparison problems, Gupta and Huang (1977) prove a result similar to

(4.13). However, they do not assume that $\psi \in \mathcal{S}_2(\beta)$, only that $S(\underline{\theta},\psi) \leq \beta$ for $\underline{\theta} \in \Omega_0 = \{\underline{\theta}: \theta_1 = \dots = \theta_k\}$. We will see that for normal populations ψ^* in fact satisfies (4.11) if Δ is not too large and β is large enough.

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 $\mathcal{S}_1(\gamma)$. From Berger (1977) we have that

$$\text{(4.16)} \qquad \begin{aligned} &\inf_{\mathcal{S}_1(\gamma)} \sup_\Omega S(\underline{\theta},\psi) = k\gamma \\ &\inf_{\mathcal{S}_1(\gamma)} \sup_\Omega S'(\underline{\theta},\psi) = (k-1)\gamma. \end{aligned}$$

As mentioned in Section 1, ψ^m is minimax in $\mathcal{S}_1(\gamma)$ for S and S' if $f(x-\theta)$ has MLR in x. Berger (1977) observed that if $\gamma < (k-1)/k$ then ψ^a is not minimax for S, and if $\gamma < (k-2)/(k-1)$ then ψ^a is not minimax for S'. We shall show that if $\gamma \geq (k-1)/k$ $(\gamma \geq (k-2)/(k-1))$, then ψ^a is in fact minimax for S (S'). Hence, Berger's condition is not only sufficient but also necessary. More precisely the following result holds.

<u>Theorem 4.3.</u> $X_1,\dots,X_k$ *are independent.* X_i *has density* $f(x-\theta_i)$ *for* $i = 1,\dots,k$. *Assume that* $f(x-\theta)$ *has MLR in* x *and that* $f(x) = f(-x)$ *for all* x. *Then*

$$\text{(4.17)} \qquad \sup_{\underline{\theta}\in\Omega} S'(\underline{\theta},\psi^a) = S'(\underline{0},\psi^a) = (k-1)\gamma \textit{ if and only if } \gamma \geq \frac{k-2}{k-1}.$$

$$\text{(4.18)} \qquad \sup_{\underline{\theta}\in\Omega} S(\underline{\theta},\psi^a) = S(\underline{0},\psi^a) = k\gamma \textit{ if and only if } \gamma \geq \frac{k-1}{k}.$$

From (4.16) we then have the following corollary.

Corollary 4.1. *Assume the conditions in Theorem 4.3 hold. Then* ψ^a *is minimax in* $\mathcal{S}_1(\gamma)$ *for* S *if and only if* $\gamma \geq (k-1)/k$, *and for* S' *if and only if* $\gamma \geq (k-2)/(k-1)$.

Remark. Berger (1977) showed that if ψ is minimax in $\mathcal{S}_1(\gamma)$ for the risk S, then ψ is minimax also for the risk S'. Corollary 4.1 implies that in the other direction the result is not true, since ψ^a is minimax for S', but not for S if

$$\frac{k-2}{k-1} \leq \gamma < \frac{k-1}{k}.$$

We prove Theorem 4.3 only for S. The proof of (4.17) 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 a strongly unimodal density. 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 4.1. *Assume the conditions in Theorem 4.3 hold. Let*

$$V_i = \frac{1}{k-1} \sum_{j \neq i} (X_j - X_i)$$

and

$$\mu_i = \frac{1}{k-1} \sum_{j \neq i} (\theta_j - \theta_i), \text{ for } 1,\dots,k.$$

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 $\theta_1 \le \cdots \le \theta_k$. Let $t_i = (\theta_{i+1} - \theta_i)/(k-1)$. Then

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

where $\underline{t} = (t_1,\ldots,t_{k-1})$ and $t_i \ge 0$ for all i.

The "if" part is the important and difficult part. The next lemma considers this case for "large" $\underline{t}$.

Lemma 4.2. *Assume the conditions in Theorem 4.3 hold. Let* $\gamma \ge (k-1)/k$ *and* $k \ge 3$. *Then*

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

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

$$E_{\underline{\theta}}(\psi_1^a) \le G(-c-t_1) = 1 - G(c+t_1),$$

and result follows. Q.E.D.

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

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

Lemma 4.3. *Assume the conditions in Theorem 4.3 hold. Let* $k \ge 3$ *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* $t_m \le t_m^0$ *for* $1 \le m \le k-1$.

Proof. Let first $m \ge 2$, and let $\underline{v} = (v_1,\ldots,v_m,0,\ldots,0) = (t_1^0,\ldots,t_{m-1}^0, t_m,0,\ldots,0)$, $t_m \le t_m^0$. Then $\underline{v} \in A(\gamma)$. Let $h(v_m) = H(\underline{v})$. We shall show that the derivative

$$h'(v_m) \leq 0 \quad \text{for} \quad v_m \leq t_m^0.$$

It is easily seen that

(4.19) $$h'(v_m) \leq 0 \leftrightarrow m \leq \sum_{i=1}^{m} r_i(\underline{v}).$$

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 $y = \sum_{j=i}^{m} jv_j + \sum_{j=i}^{m} (k-j)v_j$. Then $r_i(\underline{v}) = g(a-y)/g(a)$, and for $i \geq 3$ we have that $y \leq 2a$. From Lemma 4.1 it follows that $r_i \geq 1$ for $i \geq 3$. To show (4.19) it remains to show that $r_1 + r_2 \geq 2$. Since $(r_1+r_2)/2 \geq \sqrt{r_1 r_2}$ it is enough to show that

(4.20) $$r_1 r_2 \geq 1. \; .$$

From the MLR-property of g, and the fact that $(k-2)v_1 + a \leq 2c$ we get

$$r_1 r_2 \geq \frac{g(c+v_1)g(c-(k-1)v_1)}{g^2(c+v_1+b)}.$$

where $b = \sum_{j=2}^{m} jv_j$. Now, $g(c+v_1) \geq g(c+v_1+b)$ and since $kv_1 < 2(c+v_1)$

(4.21) $$g(c-(k-1)v_1) \geq g(c+v_1).$$

Hence (4.20) is proved, and the Lemma is proved for $m \geq 2$. Now let $m = 1$. Then

$$h'(v_1) \leq 0 \Leftrightarrow 1 \leq g(c-(k-1)v_1)/g(c+v_1)$$

which follows from (4.21). Q.E.D.

Proof of Theorem 4.3. The "only if" part is seen by letting $t_1 \to \infty$. 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 4.3 we get that

$$\underline{t} \in \Lambda(\gamma) \Rightarrow H(\underline{t}) \leq H(\underline{0}) = k\gamma$$

Together with Lemma 4.2 this completes the proof.

For later use we will also consider the case where $\psi^a \in \mathcal{D}'(\gamma,\Delta)$. i.e. c is determined by (2.4). Then $c = c(\gamma) - \Delta$. In the same way as we proved Theorem 4.3, the following result can be proved.

Theorem 4.4. *Assume the conditions in Theorem 4.3 hold. Let* $\psi^a \in \mathcal{D}'(\gamma,\Delta)$. *Then*

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

5. Optimal Subset Selection Procedures for Normal Populations

Let X_{ij} = (i = 1,...,k; j = 1,...,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

$$\theta_i > \theta_{[k]} - \Delta_0 \frac{\sigma}{\sqrt{n}}, \quad \Delta_0 > 0. \tag{5.1}$$

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 indpependent, and X_i is $\mathcal{N}(\theta_i,1)$.

Let us first consider the problem of minimizing

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

for all $\psi \in \mathcal{D}'(\gamma,\Delta)$. Here Ω_1 is given by (4.6) and $\Omega_k(\Delta)$ is given by (4.5).

From Theorem 4.1, the optimal procedure ψ^0 is given by (4.3). We find that

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

Hence the optimal procedure is ψ^e given by (2.7) and C is determined by (2.10). Note that for a general $\sigma/\sqrt{n}$, ψ^e has the form in (2.8).

From Ostrowski (1952) we have that a permutation-symmetric and differentiable function, h: $R^m \to R$, is Schur-concave if and only if

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

Using this result it is readily seen that ψ^e is a Schur-procedure. It follows that ψ^e minimizes

$$\sup_{\underline{\theta}\in\Omega_1} B(\underline{\theta},\psi) \text{ and } \sup_{\underline{\theta}\in\Omega_k(\Delta)} S'(\underline{\theta},\psi) \tag{5.2}$$

for all $\psi \in \mathcal{D}'(\gamma,\Delta)$.

From Theorem 2.1 we have that $\psi^e \in \mathcal{D}(\gamma,\Delta)$ if $\Delta \leq \Delta_\gamma$, where Δ_γ is defined by (2.13). Hence if $\Delta \leq \Delta_\gamma$, then ψ^e minimizes (5.2) for all $\psi \in \mathcal{D}(\gamma,\Delta)$.

<u>Remark</u>. 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 $\theta_{[1]} = \cdots = \theta_{[k-1]} = \theta_{[k]} - \Delta$.

Let us next consider the problem of minimizing

$$\sup_{\underline{\theta}\in\Omega} \ell(\underline{\theta},\psi) \text{ and } \sup_{\underline{\theta}\in\Omega(\Delta)} L(\underline{\theta},\psi)$$

for all permutation-invariant procedures $\psi \in \mathcal{D}'(\gamma,\Delta)$. We shall see that ψ^e is also optimal for this problem. We should remark that this does not necessarily mean that ψ^e is minimax for all $\psi \in \mathcal{D}'(\gamma,\Delta)$, since the criteria ℓ and L do not correspond to any loss functions.

Let $\mathcal{D}_I$ be the class of permutation-invariant procedures. In order to describe $\mathcal{D}_I$, let π be a permutation of $(1,\ldots,k)$ such that πi is the new position of element i, i.e.

$$\pi(1,\ldots,k) = (\pi^{-1}1,\ldots,\pi^{-1}k).$$

Then the permutation $\pi\underline{x}$ of $\underline{x} \in \mathbb{R}^k$ is defined by

$$(\pi x)_i = x_{\pi^{-1}i}, \text{ for } i = 1,\ldots,k.$$

<u>Definition 5.1.</u> $\psi = (\psi_1,\ldots,\psi_k)$ is permutation-invariant if

$$\psi_i(\underline{x}) = \psi_{\pi i}(\pi\underline{x})$$

for $i = 1,\ldots,k$ and all $(\pi,\underline{x})$.

It is readily seen that if $\psi \in \mathcal{D}_I$ then

$$\int\psi_j p_i = \int\psi_{j'} p_{i'}, \quad \text{for all } (i,j) \neq (i',j') \tag{5.3}$$

Here

$$p_i(\underline{x}) = p(\underline{x}-\underline{\theta}_i^{\Delta}) = (2\pi)^{-k/2} \cdot \exp\{-\frac{1}{2}\sum_{\ell=1}^{k} x_\ell^2 - \frac{1}{2}\Delta^2 + x_i\Delta\}.$$

We recall from Part I that a procedure $\psi \in K$, the class of non-randomized convex procedures if and only if ψ is just, translation-invariant and satisfies

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

where A is a permutation-symmetric, monotone decreasing, convex set. It was shown in Part I (see Corollary 5.1), that

$$(5.4) \qquad \sup_{\underline{\theta}\in\Omega} \ell(\underline{\theta},\psi) = \sup_{\underline{\theta}\in\Omega(\Delta)} L(\underline{\theta},\psi) = \ell(\underline{\theta}_i^{\Delta},\psi) \text{ for } i = 1,\ldots,k$$

if $\psi \in K$.

The main result for the criteria ℓ and L now follows.

Theorem 5.1. *Assume* $X_1,\ldots,X_k$ *are independent and* X_i *is* $\mathcal{N}(\theta_i,1)$. *Then* ψ^e *minimizes*

$$\sup_{\underline{\theta}\in\Omega} \ell(\underline{\theta},\psi) \quad \textit{and} \quad \sup_{\underline{\theta}\in\Omega(\Delta)} L(\underline{\theta},\psi)$$

for all $\psi \in \mathcal{D}'(\gamma,\Delta) \cap \mathcal{D}_I$.

Proof. ψ^e is given by:

$$\psi_i^e(\underline{x}) = I_A(\underline{x}_i^*)$$

where

$$A = \{\underline{y} \in \mathbb{R}^{k-1} : \sum_{j=1}^{k-1} e^{\Delta y_j} \le C\}.$$

A is a convex set. Hence ψ^e is a member of the class K, and from (5.4) we have that

$$\sup_{\underline{\theta}\in\Omega} \ell(\underline{\theta},\psi^e) = \sup_{\underline{\theta}\in\Omega(\Delta)} L(\underline{\theta},\psi^e) = \ell(\underline{\theta}_i^\Delta,\psi^e) \quad \text{for } i = 1,\ldots,k.$$

This implies that it is enough to consider $L(\underline{\theta},\psi)$. Let $\psi \in \mathcal{D}'(\gamma,\Delta) \cap \mathcal{D}_I$. Then

$$\begin{aligned}
&\sup_{\underline{\theta}\in\Omega(\Delta)} L(\underline{\theta},\psi) - \sup_{\underline{\theta}\in\Omega(\Delta)} L(\underline{\theta},\psi^e) \\
&\geq \log \prod_{i=2}^{k} \int \psi_i p_1 - \log \prod_{i=2}^{k} \int \psi_i^e p_1 \\
&= \log(\int \psi_2 p_1)^{k-1} - \log(\int \psi_2^e p_1)^{k-1} = \log(\int \psi_2 p_1 / \int \psi_2^e p_1)^{k-1}
\end{aligned}$$

from (5.3).

From the proof of Theorem 4.1 it follows that for any $\psi \in \mathcal{D}'(\gamma,\Delta)$ we have

$$\sum_{j=1}^{k} \sum_{i\neq j} \int \psi_j p_i \geq \sum_{j=1}^{k} \sum_{i\neq j} \int \psi_j^e p_i = k(k-1)\int \psi_2^e p_1. \tag{5.5}$$

From (5.3) the left side of (5.5) is equal to $k(k-1)\int \psi_2 p_1$ since $\psi \in \mathcal{D}_I$. Hence $\int \psi_2 p_1 \geq \int \psi_2^e p_1$. Q.E.D.

Let us now summarize the minimax properties of ψ^e.

<u>Theorem 5.2</u>. $X_1,\ldots,X_k$ *are independent.* X_i *is* $\mathcal{N}(\theta_i,1)$ *for* $i = 1,\ldots,k$. *Assume* $C \geq k-1$, *i.e.* $\sum_{i=1}^{k} \psi_i^e \geq 1$, Δ_γ *is given by (2.13). If* $k \geq 4$ *and* $\Delta \leq \Delta_\gamma$ *or* $k \leq 3$ *then* ψ^e *minimizes*

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

for all $\psi \in \mathcal{D}(\gamma,\Delta)$, *and*

(ii) $\sup_{\underline{\theta}\in\Omega} \ell(\underline{\theta},\psi)$ *and* $\sup_{\underline{\theta}\in\Omega(\Delta)} L(\underline{\theta},\psi)$

for all $\psi \in \mathcal{D}(\gamma,\Delta) \cap \mathcal{D}_I$.

Remark. As is seen from Table 1 and Table 2 usually $\Delta \leq \Delta_\gamma$ implies that $C \geq k-1$.

Next we consider the problem of finding solutions to the goals (3.3) and (3.4), by applying Theorem 4.2. We shall therefore first consider the two simpler problems (4.8) and (4.9). Secondly we will show when the optimal procedures satisfy (4.15), and hence are solutions to (3.3) and (3.4).

The density g in (4.10) 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

$$\frac{g(\underline{y}+\underline{\Delta})}{g(\underline{y})} = e^{-\frac{\Delta}{k}\sum_{i=1}^{k-1} y_i - \frac{k-1}{2k}\Delta^2}.$$

It follows from Theorem 4.2 that the optimal procedure ψ^* for the problems (4.8) and (4.9) is given by

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

i.e. $\psi^* = \psi^a$. For given c, let $\psi^a = \psi^a(c)$. For problem (4.8) c is given by

$$c = c_1 = \sqrt{\frac{k}{k-1}}\; z(\beta/k)$$

(see Section 2). For problem (4.9), c is given by (see (2.5))

$$c = c' = \sqrt{\frac{k}{k-1}}\ z(\gamma) - \Delta.$$

From Theorem 4.3 and Theorem 4.4 we see that ψ^a is the solution of problem (4.8) if $\beta \geq k-1$ and of problem (4.9) if

$$\Delta \leq \sqrt{\frac{k}{k-1}}\ \{z(\gamma) - z(\tfrac{k-1}{k})\}. \tag{5.6}$$

From Section 2, $\psi^a(c') \in \mathcal{D}(\gamma,\Delta)$ if $\Delta \leq \Delta_a(\gamma)$, where $\Delta_a(\gamma)$ is given by (2.6). For $\psi^a(c_1)$ it can be similarily shown that (4.15) holds if $k \geq 4$ and $\Delta \leq \Delta_2(\beta)$ or $k \leq 3$. Here

$$\Delta_2(\beta) = \sqrt{\frac{k}{k-1}}\ \{z(\tfrac{\beta}{k}) + z(\tfrac{k-1}{k})\}..$$

We now summarize the minimax properties of ψ^a. (See also Corollary 4.1.)

Theorem 5.3. $X_1,\ldots,X_k$ *are independent.* X_i *is* $\mathcal{N}(\theta_i,1)$ *for* $i = 1,\ldots,k$.

(a) *Let* $\psi^a \in \mathcal{D}_1(\gamma)$, *i.e.* $c = \sqrt{\frac{k}{k-1}}\ z(\gamma)$. *Then* ψ^a *minimizes for all* $\psi \in \mathcal{D}_1(\gamma)$

$$\sup_{\underline{\theta}\in\Omega} S'(\underline{\theta},\psi) \quad (\sup_{\underline{\theta}\in\Omega} S(\underline{\theta},\psi))$$

if and only if $\gamma \geq (k-2)/(k-1)\ ((k-1)/k)$.

(b) *Let* $\psi^a \in \mathcal{D}_2(\beta)$, *i.e.* $c = \sqrt{\frac{k}{k-1}}\ z(\beta/k)$. *Let* $\beta \geq k-1$ *and assume* $k \geq 4$ *and* $\Delta \leq \Delta_2(\beta)$ *or* $k \leq 3$. *Then* ψ^a *maximizes for all* $\psi \in \mathcal{D}_2(\beta)$

$$\inf_{\underline{\theta}\in\Omega} R(\underline{\theta},\psi)$$

(c) *Let* $\psi^a \in \mathcal{D}(\gamma,\Delta)$, *i.e.* $c = \sqrt{\frac{k}{k-1}}\, z(\gamma) - \Delta$. *Assume* $\Delta \leq \Delta_a(\gamma)$. *Then* ψ^a *minimizes for all* $\psi \in \mathcal{D}(\gamma,\Delta)$

$$\sup_{\underline{\theta}\in\Omega} S(\underline{\theta},\psi).$$

Remark. The general normal case where $\mathrm{Var}(X_i) = \sigma^2/n$, will give exactly the same procedures using $\Delta_0 = \Delta\sqrt{n}/\sigma$ instead of Δ.

We will now discuss the case where k is large, and same normal model. Consider first ψ^e, and let $C_k = C$ be determined by (2.10). Let $k \to \infty$. Then it is easily seen that $C_0 = \lim_{k\to\infty} (C_k/k-1)$ exists and is given by

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

Therefore, as a supplement to (2.20) we have

$$\lim_{k\to\infty} (C_k/k-1) \geq 1 \qquad \Delta \leq 2z(\gamma).$$

Consider next the upper bound $\Delta_\gamma(k) = \Delta_\gamma$, given by (2.13), that insures $\psi^e \in \mathcal{D}(\gamma,\Delta)$. For given $\Delta > 0$ we have

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

Hence

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

from Theorem 2.1.

Let us now compare the procedures ψ^m, ψ^a, ψ^e for the case $k \to \infty$. We use the standardized risk $B_k(\underline{\theta},\psi) = B(\underline{\theta},\psi)/k$, and assume that the procedures lie in the class $\mathcal{S}'(\gamma,\Delta)$. It is readily seen that

$$\sup_{\underline{\theta}\in\Omega_1} B_k(\underline{\theta},\psi^e) \to \Phi(z(\gamma)-\Delta) \text{ as } k\to\infty$$

and

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

It follows that ψ^a is asymptotically minimax in the class $\mathcal{S}'(\gamma,\Delta)$ with respect to the slippage set Ω_1.

Let us next consider the procedure ψ^m, and let $d_k = d'$, given by (2.4). I.e. d_k is determined by

$$\gamma = P(Y_1 + \Delta \geq \max_{2\leq i\leq k} Y_i - d_k).$$

where $Y_1,\ldots,Y_k$ are i.i.d. $\mathcal{N}(0,1)$.

We will assume that Δ is such that $d_k > 0$. Then, as mentioned in Section 2, $\psi^m \in \mathcal{S}(\gamma,\Delta)$.

From Gnedenko (1943) (see also David (1970), p. 214) we have that

$$\text{(5.7)}\qquad \max_{2\leq i\leq k} Y_i - \sqrt{2}\sqrt{\log(k-1)} \xrightarrow{p} 0 \text{ as } k\to\infty.$$

Let $D_k = d_k - \sqrt{2}\sqrt{\log(k-1)}$. From (5.7) it follows that $D_0 = \lim_{k\to\infty} D_k$ exists and is given by

$$D_0 = z(\gamma) - \Delta.$$

It also follows that

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

Hence

$$\sup_{\underline{\theta} \in \Omega} B_k(\underline{\theta},\psi^m) \to \Phi(z(\gamma)-\Delta) \text{ as } k \to \infty.$$

We have now proved the following result.

Theorem 5.4. *Let* $\psi \in \mathcal{D}(\gamma,\Delta)$ *or* $\mathcal{D}'(\gamma,\Delta)$. *Then*

$$\lim_{k\to\infty} \inf\{\sup_{\underline{\theta} \in \Omega} B_k(\underline{\theta},\psi)\} \geq \lim_{k\to\infty} \{\sup_{\underline{\theta} \in \Omega} B_k(\underline{\theta},\psi^m)\}.$$

I.e. ψ^m *is asymptotically minimax in* $\mathcal{D}'(\gamma,\Delta)$ *and* $\mathcal{D}(\gamma,\Delta)$ *with respect to the whole parameter-space.*

Remark. Bickel and Yahav (1977) considered the case $k \to \infty$ in a decision-theoretic framework, for the identification problem where $(\theta_{[1]},\ldots,\theta_{[k]})$ are known.

Δ is typically a small number. It is therefore of interest to see what happens to $\psi_\Delta^e = \psi^e$ as we let $\Delta \to 0$. Of course, for $\Delta = 0$ the situation reduces to the case of minimizing $\sup_{\underline{\theta} \in \Omega} S'(\underline{\theta},\psi)$ for $\psi \in \mathcal{D}_1(\gamma)$. Let now $\psi^a \in \mathcal{D}_1(\gamma)$. Then from Studden (1967) we have

$$\psi_\Delta^e(\underline{x}) \overset{a.s.}{\to} \psi^a(\underline{x}) \quad \text{as } \Delta \to 0.$$

It follows that

$$B(\underline{\theta},\psi_\Delta^e) \to S'(\underline{\theta},\psi^a) \text{ as } \Delta \to 0 \text{ for, } \forall\underline{\theta}.$$

This result together with Corollary 4.1 indicates that ψ^e, for small Δ, is minimax for the risk B with respect to the whole parameter-space Ω if

$\gamma \geq (k-2)/(k-1)$.

The minimax properties of ψ^e hold also for other normal models. Assume we have a two-way layout without interaction, i.e. the model is: X_{ij}; $j = 1,\ldots,n$, $i = 1,\ldots,k$ are independent. X_{ij} is $\mathcal{N}(\theta_{ij},\sigma^2)$, where σ^2 is known and

$$\theta_{ij} = \mu + \alpha_i + \beta_j;\ \sum_i \alpha_i = \sum_j \beta_j = 0.$$

Let $\underline{\underline{\theta}} = \{\theta_{ij}\}$, $\underline{\underline{X}} = \{X_{ij}\}$, $\underline{\alpha} = \{\alpha_i\}$, $\underline{\beta} = \{\beta_j\}$. β_j corresponds to the block effects, so the main effects are

$$\theta_i = \mu + \alpha_i; \quad \text{for } i = 1,\ldots,k.$$

Let $\underline{\beta}$ be arbitrary and let $p_i(\underline{\underline{x}}|\underline{\beta})$ be the density corresponding to $\theta_i = \Delta$, $\theta_j = 0$ for $j \neq i$. Then

$$T_i(\underline{\underline{x}}) = (p_i(\underline{\underline{x}}|\underline{\beta}))^{-1} \sum_{j\neq i} p_j(\underline{\underline{x}}|\underline{\beta}) = \sum_{j\neq i} e^{\Delta_0 \frac{\sqrt{n}}{\sigma}(x_j - x_i)}$$

where as before

$$\Delta_0 = \sqrt{n}\Delta/\sigma \text{ and } x_i = (n)^{-1} \sum_{j=1}^{n} x_{ij}.$$

Hence the procedure ψ^0 defined by (4.3) and (4.4) is again ψ^e. Since ψ^0 is independent of $\underline{\beta}$ and is a Schur-procedure in $\underline{X} = (X_1,\ldots,X_k)$ we see from the proof that Theorem 4.1 still applies. It follows that for the two-way layout without interaction ψ^c minimizes

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

for all $\psi \in \mathcal{B}(\gamma,\Delta)$ if $\Delta_0 \leq \Delta_\gamma$. Here $\Omega_1' = \{\underline{\underline{\theta}}:\ \underline{\theta} \in \Omega_1\}$ and $\Omega_k'(\Delta) = \{\underline{\underline{\theta}}:\ \underline{\theta} \in \Omega_k(\Delta)\}$.

6. Concluding Comments

Usually in a given subset selection problem the first thing to decide is what kind of control condition we want the procedures to satisfy. From the discussion in Section 2 we see that the three control conditions (1.5), (1.11), (2.1) seem to be of special interest. In Section 2 we found (1.11) as the most appropriate condition of these.

Using these control conditions we have derived minimax procedures with respect to different risk functions. However, it is clear that not all the risk criteria we have considered are equally appropriate or meaningful. A criterion for comparing procedures should measure how well a procedure excludes populations that are inferior to the best population. Now $S(\underline{\theta},\psi)$ includes the probability of selecting the best population, so clearly $S'(\underline{\theta},\psi)$ is a more appropriate measure of performance than $S(\underline{\theta},\psi)$. Analyzing this point of view further, we do not need to protect against populations, that are "close" to the best population, i.e. populations Π_i where $\theta_i > \theta_{[k]} - \Delta$. Hence $B(\underline{\theta},\psi)$ seems to be a natural criterion. So of the criteria S, S', B we regard B as the most meaningful. The criteria L, ℓ do not have such a nice intuitive appeal as S' or B, but as shown in Part I they do have the nice feature of placing more weights on the worst populations that on those closest to the best.

As a conclusion we regard $B(\underline{\theta},\psi)$ and $\ell(\underline{\theta},\psi)$ as the most appropriate of the criteria considered in this paper.

From Section 5, we see that for normal populations, the procedures ψ^e, ψ^m and ψ^a have certain minimax properties. One of them, ψ^e, has not been considered before as a serious contender, but as seen in Theorem 5.2, ψ^e seems to have the most desirable minimax properties of the three procedures.

Acknowledgment. This paper is 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 Professor Kjell Doksum for helpful discussions, and to Peter Guttorp 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 (2.8).

k \ $\sqrt{n}\Delta/\sigma$	.10	.25	.50	1.0	1.5	2.0	3.0
			γ = .75				
3	2.16	2.35	2.52	2.15	1.16	.39	.01
4	3.23	3.51	3.78	3.35	1.95	.74	.03
5	4.29	4.67	5.04	4.52	2.74	1.09	.04
6	5.36	5.83	6.27	5.76	3.56	1.47	.06
7	6.42	6.97	7.52	6.84	4.38	1.84	.08
8	7.50	8.13	8.79	8.14	5.22	2.30	.11
9	8.56	9.30	10.09	9.49	6.29	2.77	.14
10	9.63	10.46	11.29	10.60	7.11	3.17	.17
			γ = .90				
3	2.34	2.87	3.76	4.89	4.11	2.15	.13
4	3.47	4.20	5.45	7.05	6.29	3.48	.27
5	4.60	5.54	7.12	9.28	8.17	4.75	.39
6	5.73	6.88	8.82	11.61	10.54	6.38	.56
7	6.84	8.15	10.33	13.56	12.68	8.07	.77
8	7.97	9.48	12.03	15.59	14.74	9.23	.94
9	9.09	10.81	13.65	17.70	16.89	10.73	1.13
10	10.23	12.14	15.31	19.60	18.74	12.07	1.29
			γ = .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

TABLE 1 (continued)

k \ $\sqrt{n}\Delta/\sigma$	.10	.25	.50	1.0	1.5	2.0	3.0
				γ = .975			
3	2.54	3.53	5.66	11.04	13.95	11.15	1.66
4	3.77	5.18	8.17	16.36	21.11	17.72	3.27
5	4.99	6.77	10.68	20.89	28.59	25.29	4.94
6	6.22	8.40	13.03	25.27	33.55	29.73	6.24
7	7.45	10.04	15.71	30.37	41.12	36.78	8.06
8	8.68	11.72	18.00	35.32	48.33	43.66	9.84
9	9.86	13.21	20.56	39.85	54.33	52.07	12.57
10	11.09	14.86	22.89	44.51	66.18	65.44	15.24
				γ = .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

The probability of selecting population Π_1, using ψ^e, when $\theta_2=\ldots=\theta_k=\theta_1+\Delta$

k \ $\sqrt{n}\Delta/\sigma$	.5	1.0	1.5	2.0	3.0
			γ = .90		
4	.36	.33	.14	.03	.00
5	.34	.30	.11	.02	.00
6	.33	.30	.10	.01	.00
7	.33	.29	.09	.01	.00
8	.32	.28	.09	.01	.00
9	.32	.27	.08	.01	.00
10	.32	.26	.07	.01	.00
			γ = .95		
4	.41	.50	.20	.05	.00
5	.39	.48	.18	.04	.00
6	.38	.47	.17	.03	.00
7	.36	.44	.14	.03	.00
8	.37	.43	.14	.03	.00
9	.37	.42	.14	.03	.00
10	.36	.41	.13	.02	.00
			γ = .975		
4	.45	.66	.30	.10	.00
5	.43	.63	.29	.09	.00
6	.42	.62	.26	.07	.00
7	.41	.62	.25	.07	.00
8	.41	.60	.25	.06	.00
9	.39	.60	.24	.06	.00
10	.39	.58	.25	.06	.00

TABLE 2 (continued)

k \ $\sqrt{n}\Delta/\sigma$	.5	1.0	1.5	2.0	3.0
		$\gamma = .99$			
4	.48	.77	.48	.19	.01
5	.46	.75	.46	.17	.01
6	.46	.73	.43	.16	.01
7	.44	.71	.39	.14	.01
8	.43	.71	.37	.13	.01
9	.44	.71	.36	.12	.01
10	.43	.67	.35	.12	.01

References

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

[2] Bahadur, R. R. and Goodman, L. A. (1952). Impartial decision rules and sufficient statistics. Ann. Math. Statist. 23, 553-562.

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

[4] 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.

[5] Berger, R. L. (1977). Minimax subset selection for loss measured by subset size. Mimeograph Series #489, Dept. of Statistics, Purdue University.

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

[7] Bickel, P. J. and Yahav, J. A. (1977). On selecting a set of good populations. In Statistical Decision Theory and Related Topics II. (ed. by S. S. Gupta and D. S. Moore), Academic Press, New York, pp. 37-55.

[8] Chernoff, H. and Yahav, J. A. (1977). A subset selection problem employing a new criterion. In Statistical Decision Theory and Related Topics II. (ed. by S. S. Gupta and D. S. Moore), Academic Press, New York, pp. 93-119.

[9] David, H. A. (1970). Order Statistics. Wiley, New York.

[10] Deely, J. J. and Gupta, S. S. (1968). On properties of subset selection procedures. Sankhyā 30, 37-50.

[11] Eaton, M. L. (1967). Some optimum properties of ranking procedures. Ann. Math. Statist. 38, 124-137.

[12] Ferguson, T. S. (1967). Mathematical Statistics. A Decision-Theoretic Approach. Academic Press, New York.

[13] Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. Ann. Mathematics 44, 423-453.

[14] Goel, P. K. and Rubin, H. (1977). On selecting a subset containing the best population - A Bayesian approach. Ann. Statist. 5, 969-984.

[15] 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.

[16] Gupta, S. S. (1963). Probability integrals of the multivariate normal and multivariate t. Ann. Math. Statist. 34, 792-828.

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

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

[19] Gupta, S. S. and Miescke, K. J. (1978). Optimality of subset selection procedures for ranking means of three normal populations, Mimeograph Series #78-19, Dept. of Statistics, Purdue University

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

[21] 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.

[22] Hsu, J. C. (1977). On some decision-theoretic contributions to the problem of subset selection. Mimeograph Series #491, Dept. of Statistics, Purdue University.

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

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

[25] Lehmann, E. L. (1966). On a theorem of Bahadur and Goodman. Ann. Math. Statist. 37, 1-6.

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

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

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

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

[30] Spjøtvoll, E. (1972). On the optimality of some multiple comparison procedures. Ann. Math. Statist. 43, 398-411.

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