

ON THE OPTIMALITY OF SOME SUBSET  
SELECTION PROCEDURES\*

by

Lii-Yuh Leu  
National Central University  
Technical Report #84-7

Department of Statistics  
Purdue University

March 1984

\*This research was supported by the Office of Naval Research Contract  
N00014-75-C-0455 at Purdue University. Reproduction in whole or in  
part is permitted for any purpose of the United States Government.

ON THE OPTIMALITY OF SOME SUBSET  
SELECTION PROCEDURES\*

by

Lii-Yuh Leu  
National Central University

ABSTRACT

As measures of goodness of a selection rule, usually two quantities, the probability of a correct selection and the expected size of the selected subset, are considered. Based on these two criteria, Gupta and Huang (1980) proved a theorem to derive a selection procedure with some optimality property. However, the theorem cannot be applied to the unequal sample sizes case. In this paper, we use a different method to generalize this theorem to the unequal sample sizes case. Also a dual problem is investigated. Also, we treat a selection procedure in terms of multiple tests. Based on this approach, we derive an optimality result.

key words: Subset selection, restricted minimax, multiple tests.

\*This research was supported by the Office of Naval Research Contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

ON THE OPTIMALITY OF SOME SUBSET  
SELECTION PROCEDURES\*

by

Lii-Yuh Leu  
National Central University

1. Introduction

Let  $\pi_1, \pi_2, \dots, \pi_k$  be  $k$  independent populations which are characterized by parameters  $\theta_1, \theta_2, \dots, \theta_k$ , respectively. Let  $X_i$  be the observation from population  $\pi_i$ ,  $1 \leq i \leq k$ . Assume  $\underline{X} = (X_1, \dots, X_k)$  is a sufficient statistic for  $\underline{\theta} = (\theta_1, \dots, \theta_k)$ . Suppose that we are interested in selecting a subset of the  $k$  populations containing the largest parameter (or the smallest parameter). For subset selection, let  $\delta = (\delta_1, \dots, \delta_k)$  be a selection procedure, where  $\delta_i(\underline{x})$  is the probability of selecting  $\pi_i$ ,  $1 \leq i \leq k$ , based on the observed vector  $\underline{X} = \underline{x}$ , and  $\sum_{i=1}^k \delta_i(\underline{x}) \geq 1$ , for all  $\underline{x}$  (i.e. at least one population is selected). As measures of goodness of a selection rule, usually two quantities, the probability of a correct selection and the expected size of the selected subset, are considered. Based on these two criteria, optimal selection rules have been considered by Bahadur (1950), Eaton (1967), Lehmann (1961, 1966), Studden (1967), Nagel (1970), Spjøtvoll (1972), Alam (1973), Gupta and Huang (1977, 1980), Berger and Gupta (1980), and Bjørnstad (1981), among others.

Most of the literature on the optimality of selection rules deals with the problem when the sample sizes are all equal and is restricted to the

---

\*This research was supported by the Office of Naval Research Contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

location model. Gupta and Huang (1977, 1980) formulate the problem more generally, including location and scale cases. In this paper we will follow the notation in the papers of Gupta and Huang (1977, 1980) and generalize the theorem in Gupta and Huang (1980) to the unequal sample sizes case.

In Section 3, we use a different approach to deal with the problem of Gupta and Huang (1980) and generalize it to the unequal sample sizes case. Also a dual problem is investigated. In Section 4, we treat a selection problem as a multiple tests problem. Based on this approach, we derive an optimality result.

## 2. Notations and Definitions

Following the notation of Gupta and Huang (1980), let  $\Omega = \{\underline{\theta} = (\theta_1, \dots, \theta_k)\}$  be the parameter space. Let  $\tau_{ij} = \tau_{ij}(\underline{\theta})$  be a measure of separation between  $\pi_i$  and  $\pi_j$ . We assume that there exists a monotonically non-increasing function  $h$  such that  $\tau_{ji} = h(\tau_{ij})$ . Let  $\Omega_i = \{\underline{\theta} | \tau_{ij}(\underline{\theta}) \geq \Delta, \forall j \neq i, 1 \leq j \leq k\}$ ,  $1 \leq i \leq k$ , and  $\Omega_0 = \Omega - \bar{\Omega}$  (an indifference zone), where  $\bar{\Omega} = \bigcup_{i=1}^k \Omega_i$ . For this problem, we assume  $\Delta$  and  $\tau_{ij}$  are known with  $\Delta > \tau_{ij}$  for all  $i$ . Let  $\tau_i = \min_{j \neq i} \tau_{ij}$ ,  $1 \leq i \leq k$ . We define  $\tau^* = \max_{1 \leq i \leq k} \tau_i$ . The population associated with  $\tau^*$  will be called the best population. It should be pointed out that if  $\underline{\theta} \in \Omega_i$ , then  $\tau_i > \tau_j$  for all  $j \neq i$ . Thus if  $\underline{\theta} \in \Omega_i$ , then  $\pi_i$  is the best population. In case of any tie(s) of the populations corresponding to  $\tau^*$ , any one of the tied populations is "tagged" as the best population and selection of any subset containing this population is called a correct selection.

Let the observed sample vector be denoted by  $\underline{x} = (x_1, \dots, x_k)$ , where  $x_i = (x_{i1}, \dots, x_{in_i})$ ,  $1 \leq i \leq k$ ,  $x_{i1}, \dots, x_{in_i}$  are the samples from  $\pi_i$ ,  $1 \leq i \leq k$ . We define

$$(2.1) \quad S(\underline{\theta}, \delta) = P_{\underline{\theta}}(CS|\delta)$$

= the probability of a correct selection using rule  $\delta$ ,

and

$$(2.2) \quad R(\underline{\theta}, \delta) = \sum_{i=1}^k E_{\underline{\theta}} \delta_i$$

= the expected size of the selected subset using rule  $\delta$ .

Let there be a suitably defined statistic  $Z_{ij}$  based on the  $n_i$  and  $n_j$  independent observations from  $\pi_i$  and  $\pi_j$ ,  $1 \leq i, j \leq k$ , respectively, and suppose that for any  $i$ , the statistic  $\underline{Z}_i = \{Z_{ij}; j \neq i, 1 \leq j \leq k\}$  is invariant sufficient under a transformation group  $G$  and let

$\underline{\tau}_i = \{\tau_{ij}; j \neq i, 1 \leq j \leq k\}$  be a maximal invariant under the induced group  $\bar{G}$ . It is well-known that the distribution of  $\underline{Z}_i$  depends only on  $\underline{\tau}_i$ . For example, if the observations from  $\pi_i$  are mutually independently distributed with unknown mean  $\theta_i$ ,  $1 \leq i \leq k$  and known common variance  $\sigma^2$ ,  $Z_{ij}$  might be  $\bar{X}_i - \bar{X}_j$ , where  $\bar{X}_i = \sum_{\ell=1}^{n_i} X_{i\ell}/n_i$  and  $\bar{X}_j = \sum_{\ell=1}^{n_j} X_{j\ell}/n_j$ . For any  $i$ , let the joint density of  $Z_{ij}$ ,  $j \neq i, 1 \leq j \leq k$ , be  $P_{\underline{\theta}}(\underline{z}_i) = P_{\underline{\tau}_i}(\underline{z}_i)$  with respect to some  $\sigma$ -finite measure  $\mu$ . We note that  $E_{\underline{\theta}} \delta_i(\underline{X}) = E_{\underline{\tau}_i} \delta_i(\underline{Z}_i)$ .

The following definitions can be found in the paper of Alam (1973) (see also Gupta and Huang (1981)).

Definition 2.1. A measurable subset  $S$  of the sample space is called monotone non-decreasing if  $\underline{x} \in S$  and  $\underline{y}$  satisfies  $x_i \leq y_i$ ,  $1 \leq i \leq k$ , then  $\underline{y} \in S$ .

Definition 2.2. Let  $P_{\underline{\theta}}(S)$  denote the probability measure of  $S$  under the conditional distribution of  $\underline{X}$ , given  $\underline{\theta}$ . The distribution is said to have stochastically increasing property (SIP) in  $\underline{\theta}$  if  $P_{\underline{\theta}}(S) \leq P_{\underline{\theta}_1}(S)$  for

every monotone non-decreasing set  $S$  and for all  $\theta_i \leq \theta_i'$ ,  $1 \leq i \leq k$ .

Definition 2.3. A function  $\varphi(\underline{x})$  is said to be non-decreasing if  $\varphi(\underline{x}) \leq \varphi(\underline{y})$  for  $x_i \leq y_i$ ,  $1 \leq i \leq k$ .

### 3. Optimal Subset Selection Procedures Based on Criteria $S(\underline{\theta}, \delta)$ and $R(\underline{\theta}, \delta)$

In the following discussion, we will assume that the density function  $P_{\underline{\theta}}(\underline{z}_i) = P_{\tau_i}(\underline{z}_i)$  defined in Section 2 has the SIP in  $\tau_i$ . Let  $P_{\underline{\theta}}(\underline{z}_i)$  be denoted by  $P_0(\underline{z}_i)$  when  $\tau_{ij} = \tau_{ji} = \text{constant}$ ,  $j \neq i$ ,  $1 \leq j \leq k$ , and by  $P_j(\underline{z}_i)$  when  $\tau_{ij} = \Delta$ ,  $j \neq i$ ,  $1 \leq j \leq k$ . Our goal is to generalize the theorem of Gupta and Huang (1980). It should be pointed out that the proof is different from that of Gupta and Huang (1980).

First we quote two lemmas from Alam (1973) and Lehmann (1961) for completeness.

#### Lemma 3.1. (Alam (1973))

Let  $\{P_{\underline{\theta}}\}$  be a family of distributions which has SIP in  $\underline{\theta}$ . Then  $E_{\underline{\theta}}\varphi(\underline{X}) \leq E_{\underline{\theta}'}\varphi(\underline{X})$  for all non-decreasing integrable function  $\varphi(\underline{x})$  and  $\theta_i \leq \theta_i'$ ,  $1 \leq i \leq k$ . (Thus, if  $P_{\underline{\theta}}$  has SIP in  $\underline{\theta}$ , and if  $\varphi(\underline{x})$  is non-decreasing in  $x_j$ , then  $E_{\underline{\theta}}\varphi(\underline{X})$  is non-decreasing in  $\theta_j$ .)

#### Lemma 3.2. (Lehmann (1961))

Let  $\mu$  and  $\lambda$  be two probability distributions on  $\omega_0$  and  $\omega_1$  (subsets of  $\Omega$ ), respectively. Let  $A$  and  $B$  be two positive constants and let  $\delta^0$  maximize the integral

$$(3.1) \quad B \int_{\omega_0} S(\underline{\theta}, \delta) d\mu(\underline{\theta}) - A \int_{\omega_1} R(\underline{\theta}, \delta) d\lambda(\underline{\theta})$$

where  $S(\underline{\theta}, \delta)$  and  $R(\underline{\theta}, \delta)$  are defined by (2.1) and (2.2), respectively.

Then

(i)  $\delta^0$  minimizes  $\sup_{\underline{\theta} \in \omega_1} R(\underline{\theta}, \delta)$  subject to  $\inf_{\underline{\theta} \in \omega_0} S(\underline{\theta}, \delta) \geq \gamma$  provided

$$(3.2) \quad \int_{\omega_1} R(\underline{\theta}, \delta^0) d\lambda(\underline{\theta}) = \sup_{\underline{\theta} \in \omega_1} R(\underline{\theta}, \delta^0)$$

and

$$(3.3) \quad \int_{\omega_0} S(\underline{\theta}, \delta^0) d\mu(\underline{\theta}) = \inf_{\underline{\theta} \in \omega_0} S(\underline{\theta}, \delta^0) = \gamma.$$

(ii)  $\delta^0$  maximizes  $\inf_{\underline{\theta} \in \omega_0} S(\underline{\theta}, \delta)$  subject to  $\sup_{\underline{\theta} \in \omega_1} R(\underline{\theta}, \delta) \leq \gamma'$

provided

$$(3.4) \quad \int_{\omega_1} R(\underline{\theta}, \delta^0) d\lambda(\underline{\theta}) = \sup_{\underline{\theta} \in \omega_1} R(\underline{\theta}, \delta^0) = \gamma'$$

and

$$(3.5) \quad \int_{\omega_0} S(\underline{\theta}, \delta^0) d\mu(\underline{\theta}) = \inf_{\underline{\theta} \in \omega_0} S(\underline{\theta}, \delta^0).$$

The following theorem is a generalization of Gupta and Huang (1980).

Theorem 3.3. Suppose that for any  $i$ ,  $p_i(z_i)/p_0(z_i)$  is non-decreasing in  $z_i$  and that  $P_{\underline{\theta}}(z_i)$  has the SIP. If  $R(\underline{\theta}, \delta^0)$  is maximized at  $\tau_{ij} = \tau_{ji} =$  constant, for all  $i, j$ , where  $\delta^0$  is given by

$$(3.6) \quad \delta_i^0(z_i) = \begin{cases} 1 & \text{if } P_i(z_i) > c_i P_0(z_i) \\ \lambda_i & \text{if } P_i(z_i) = c_i P_0(z_i) \\ 0 & \text{if } P_i(z_i) < c_i P_0(z_i), \end{cases}$$

$c_i (> 0)$  and  $\lambda_i$  are determined by  $\int \delta_i^0 P_i = \gamma$ ,  $1 \leq i \leq k$ . Then  $\delta^0 = (\delta_1^0, \dots, \delta_k^0)$

minimizes  $\sup_{\underline{\theta} \in \Omega} R(\underline{\theta}, \delta)$  subject to  $\inf_{\underline{\theta} \in \Omega} S(\underline{\theta}, \delta) \geq \gamma$ .

Proof. Let  $\mu$  be the probability distribution which assigns probability  $a_i$  (will be determined later) to the set  $\omega_i = \{\underline{\theta} | \tau_{ij} = \Delta\} \subset \Omega_i$ ,  $1 \leq i \leq k$ , and  $\lambda$  be the probability distribution which assigns probability one to the set  $\omega_0 = \{\underline{\theta} | \tau_{ij} = \tau_{ij} = \text{constant}\}$ .

Let  $A$  and  $B$  be two positive constants (will be determined later). Then

$$\begin{aligned} & B \int_{\bar{\Omega}} S(\underline{\theta}, \delta) d\mu(\underline{\theta}) - A \int_{\Omega} R(\underline{\theta}, \delta) d\lambda(\underline{\theta}) \\ &= B \sum_{i=1}^k \int_{\omega_i} E_{\underline{\theta}} \delta_i d\mu(\underline{\theta}) - A \int_{\omega_0} \left( \sum_{i=1}^k E_{\underline{\theta}} \delta_i \right) d\lambda(\underline{\theta}) \\ &= \int \sum_{i=1}^k \delta_i (Ba_i P_i - AP_0) \end{aligned}$$

which is maximized by putting  $\delta_i = 1$  or  $0$  as  $Ba_i P_i >$  or  $<$   $AP_0$ . Let  $A$ ,  $B$ ,  $a_i$ ,  $1 \leq i \leq k$  be satisfied the conditions  $a_i > 0$ ,  $\sum_{i=1}^k a_i = 1$ , and  $c_i = A/a_i B$ , then  $\delta^0$  defined by (3.6) maximizes  $B \int_{\bar{\Omega}} S(\underline{\theta}, \delta) d\mu(\underline{\theta}) - A \int_{\Omega} R(\underline{\theta}, \delta) d\lambda(\underline{\theta})$ . Now, by assumption  $P_i(z_i)/P_0(z_i)$  is non-decreasing in  $z_i$ , then  $\delta_i^0$  is non-decreasing in  $z_i$  and by Lemma 3.1, for any  $\underline{\theta} \in \bar{\Omega}$ , we have

$$S(\underline{\theta}, \delta^0) = E_{\underline{z}_i} \delta_i^0 \geq E_{\underline{\Delta}} \delta_i^0 = \int \delta_i^0 P_i = \gamma.$$

Hence  $\inf_{\underline{\theta} \in \bar{\Omega}} S(\underline{\theta}, \delta^0) \geq \gamma$ .

On the other hand,

$$\int_{\bar{\Omega}} S(\underline{\theta}, \delta^0) d\mu(\underline{\theta}) = \sum_{i=1}^k a_i \int \delta_i^0 P_i = \gamma.$$

Therefore  $\int_{\bar{\Omega}} S(\underline{\theta}, \delta^0) d\mu(\underline{\theta}) = \inf_{\underline{\theta} \in \bar{\Omega}} S(\underline{\theta}, \delta^0) = \gamma$ .

Next, we have

$$\int_{\Omega} R(\underline{\theta}, \delta^0) d\lambda(\underline{\theta}) = \int_{\omega_0} R(\underline{\theta}, \delta^0) d\lambda(\underline{\theta}) = \sup_{\underline{\theta} \in \Omega} R(\underline{\theta}, \delta^0), \text{ by assumption.}$$

The theorem follows by applying Lemma 3.2. (i).



Remark: In the theorem of Gupta and Huang (1980),  $c_i = c$ ,  $1 \leq i \leq k$  which is a special case of Theorem 3.3. We note that if the sample sizes are not equal, in order to satisfy the condition  $\int \delta_i^0 p_i = \gamma$ ,  $1 \leq i \leq k$ ,  $c_i$  should be different.

Furthermore, we have the following theorem, which is a dual of Theorem 3.3.

Theorem 3.4. Suppose that for any  $i$ ,  $\bar{p}_i(z_i)/p_0(z_i)$  is non-decreasing in  $z_i$  and that  $P_{\underline{\theta}}(z_i)$  has the SIP. If  $R(\underline{\theta}, \delta^0)$  is maximized at  $\tau_{ij} = \tau_{ji} = \text{constant}$ , for all  $i, j$ , where  $\delta^0$  is given by (3.6) and  $c_i (> 0)$  and  $\lambda_i$  are determined by

$$\sum_{i=1}^k \int \delta_i^0 p_0 = \gamma' \text{ and } \int \delta_i^0 p_i \text{ is independent of } i. \text{ Then } \delta^0 = (\delta_1^0, \dots, \delta_k^0)$$

maximizes  $\inf_{\underline{\theta} \in \Omega} S(\underline{\theta}, \delta)$  subject to  $\sup_{\underline{\theta} \in \Omega} R(\underline{\theta}, \delta) \leq \gamma'$ .

Proof. By the same argument as the proof of Theorem 3.3, we have  $\delta^0$  defined by (3.6) maximizes

$$B \int_{\bar{\Omega}} S(\underline{\theta}, \delta) d\mu(\underline{\theta}) - A \int_{\Omega} R(\underline{\theta}, \delta) d\lambda(\underline{\theta}).$$

Now, if  $\underline{\theta} \in \omega_0$ , we have

$$R(\underline{\theta}, \delta^0) = \sum_{i=1}^k \int \delta_i^0 p_0 = \gamma', \text{ by assumption.}$$

Hence

$$\int_{\Omega} R(\underline{\theta}, \delta^0) d\lambda(\underline{\theta}) = \sup_{\underline{\theta} \in \Omega} R(\underline{\theta}, \delta^0) = \gamma'.$$

Furthermore,

$$\int_{\bar{\Omega}} S(\underline{\theta}, \delta^0) d\mu(\underline{\theta}) = \sum_{i=1}^k a_i \int \delta_i^0 p_i = \int \delta_i^0 p_i, \text{ since } \int \delta_i^0 p_i \text{ is independent of } i,$$

and

$$\inf_{\underline{\theta} \in \bar{\Omega}} S(\underline{\theta}, \delta^0) \geq \min_{1 \leq i \leq k} \int \delta_i^0 p_i = \int \delta_i^0 p_i.$$

Hence

$$\int_{\bar{\Omega}} S(\underline{\theta}, \delta^0) d\mu(\underline{\theta}) = \inf_{\underline{\theta} \in \bar{\Omega}} S(\underline{\theta}, \delta^0).$$

The theorem follows by applying Lemma 3.2 (ii).

Remark: For equal sample sizes case,  $p_i(z_i)/p_0(z_i)$  is independent of  $i$ . If we choose  $c(> 0)$  and  $\lambda_0$  such that  $\int \delta_i^0 p_0 = \gamma'/k$ , then

$$\delta_i^0(z_i) = \begin{cases} 1 & \text{if } p_i(z_i) > cp_0(z_i) \\ \lambda_0 & \text{if } p_i(z_i) = cp_0(z_i) \\ 0 & \text{if } p_i(z_i) < cp_0(z_i) \end{cases}$$

maximizes  $\inf_{\underline{\theta} \in \bar{\Omega}} S(\underline{\theta}, \delta)$  subject to  $\sup_{\underline{\theta} \in \bar{\Omega}} R(\underline{\theta}, \delta) \leq \gamma'$ .

Example:

Let  $X_{i1}, \dots, X_{in_i}$  be a random sample from  $N(\theta_i, \sigma^2)$ ,  $1 \leq i \leq k$ , where  $\sigma^2$  is known. Then  $(\bar{X}_1, \dots, \bar{X}_k)$  is a sufficient statistic for  $\underline{\theta}' = (\theta_1, \dots, \theta_k)$ ,

where  $\bar{X}_i = \sum_{\ell=1}^{n_i} X_{i\ell}/n_i \sim N(\theta_i, \sigma^2/n_i)$ . Consider the transformations

$g_c(\bar{x}_1, \dots, \bar{x}_k) = (\bar{x}_1+c, \dots, \bar{x}_k+c)$ , then  $Z_i' = \{\bar{X}_i - \bar{X}_j; j \neq i, 1 \leq j \leq k\}$  is a maximal invariant. The induced group  $\bar{G} = \{\bar{g}_c | \bar{g}_c(\theta_1, \dots, \theta_k) = (\theta_1+c, \dots, \theta_k+c)\}$  has maximal invariant  $\tau_i' = \{\theta_i - \theta_j; j \neq i, 1 \leq j \leq k\}$  and the distribution of  $Z_i$  depends only on  $\tau_i$ . For any  $i$ , the joint density of  $Z_i$  is given by

$$P_{\underline{\theta}}(Z_i) = P_{\tau_i}(Z_i) = (2\pi\sigma^2)^{-(k-1)/2} |\Sigma_i|^{-1/2} \exp\{- (Z_i - \tau_i)' \Sigma_i^{-1} (Z_i - \tau_i) / 2\sigma^2\}$$

where

$$\Sigma_i = \begin{pmatrix} \frac{1}{n_i} + \frac{1}{n_1} & & & \\ & \ddots & & \\ & & \frac{1}{n_i} & \\ & & & \ddots \\ \frac{1}{n_i} & & & & \frac{1}{n_i} + \frac{1}{n_k} \end{pmatrix} \quad (k-1) \times (k-1)$$

" $\leftrightarrow$ " (" $\updownarrow$ ") means that the  $i$ th row ( $i$ th column) is deleted.

$P_{\underline{z}_i}(z_i)$  has SIP in  $\underline{z}_i$ ,  $1 \leq i \leq k$  and

$$P_i(z_i)/P_0(z_i) = \exp\{\Delta n_i (\sum_{j \neq i} z_{ij} n_j) / N \sigma^2 - \Delta^2 (n_i - \frac{n_i^2}{N}) / 2 \sigma^2\}$$

which is non-decreasing in  $\underline{z}_i$  ( $N = \sum_{i=1}^k n_i$ ). Furthermore,

$$P_i(z_i)/P_0(z_i) > c_i \quad \text{iff} \quad \bar{X}_i > \sum_{j \neq i} n_j \bar{X}_j / \sum_{j \neq i} n_j + d_i.$$

Thus

$$(3.7) \quad \delta_i^0 = \begin{cases} 1 & \text{if } \bar{X}_i \geq \sum_{j \neq i} n_j \bar{X}_j / \sum_{j \neq i} n_j + d_i \\ 0 & \text{if } \bar{X}_i < \sum_{j \neq i} n_j \bar{X}_j / \sum_{j \neq i} n_j + d_i, \end{cases}$$

if  $R(\underline{\theta}, \delta^0)$  is maximized at  $\theta_1 = \theta_2 = \dots = \theta_k$ , then we have

(i) if  $d_i = \Delta - \Phi^{-1}(\gamma) \sigma \sqrt{\frac{1}{n_i} + \frac{1}{\sum_{j \neq i} n_j}}$ , then  $\delta^0$  defined by (3.7) minimizes

$\sup_{\underline{\theta} \in \Omega} R(\underline{\theta}, \delta)$  subject to  $\inf_{\underline{\theta} \in \bar{\Omega}} S(\underline{\theta}, \delta) \geq \gamma$ .

(ii) If  $\sum_{i=1}^k \Phi\left(\frac{-d_i}{\sigma \sqrt{\frac{1}{n_i} + \frac{1}{\sum_{j \neq i} n_j}}}\right) = \gamma'$  and  $\Phi\left(\frac{\Delta - d_i}{\sigma \sqrt{\frac{1}{n_i} + \frac{1}{\sum_{j \neq i} n_j}}}\right) = \text{constant}$ ,

then  $\delta^0$  defined by (3.7) maximizes  $\inf_{\underline{\theta} \in \bar{\Omega}} S(\underline{\theta}, \delta)$  subject to  $\sup_{\underline{\theta} \in \Omega} R(\underline{\theta}, \delta) \leq \gamma'$ .

In particular, if  $n_1 = n_2 = \dots = n_k = n$ , then

$$(3.8) \quad \delta_i^0 = 1 \quad \text{iff} \quad \bar{X}_i \geq \frac{1}{k-1} \sum_{j \neq i} \bar{X}_j + d.$$

We know that  $R(\underline{\theta}, \delta^0)$  is maximized at  $\theta_1 = \dots = \theta_k$  iff  $\inf_{\underline{\theta} \in \bar{\Omega}} S(\underline{\theta}, \delta^0) \geq \frac{k-1}{k}$

(see Bjørnstad (1981)). Therefore, if  $\Phi\left(\frac{(k-1)/k}{\frac{1}{2} n^{\frac{1}{2}} d / \sigma}\right) \leq \frac{1}{k}$  and

$d = \Delta^{-\sigma} \Phi^{-1}(\gamma) n^{-\frac{1}{2}} \left(\frac{k-1}{k}\right)^{-\frac{1}{2}}$ , then  $\delta^0$  defined by (3.8) minimizes  $\sup_{\underline{\theta} \in \Omega} R(\underline{\theta}, \delta)$  subject to  $\inf_{\underline{\theta} \in \bar{\Omega}} S(\underline{\theta}, \delta) \geq \gamma$ . If  $\Phi\left(\left(\frac{k-1}{k}\right)^{\frac{1}{2}} n^{\frac{1}{2}} d / \sigma\right) \leq \frac{1}{k}$  and  $d = -\Phi^{-1}\left(\frac{\gamma'}{k}\right) \sigma n^{-\frac{1}{2}} \left(\frac{k-1}{k}\right)^{-\frac{1}{2}}$ , then  $\delta^0$  defined by (3.8) maximizes  $\inf_{\underline{\theta} \in \bar{\Omega}} S(\underline{\theta}, \delta)$  subject to  $\sup_{\underline{\theta} \in \Omega} R(\underline{\theta}, \delta) \leq \gamma'$ .

#### 4. Optimal Selection Rules in Relation to Multiple Tests

Let  $\underline{X} = (X_1, \dots, X_k)$  be a random vector with probability distribution depending on a parameter vector  $\underline{\theta} = (\theta_1, \dots, \theta_k) \in \Omega$ . Consider a family of hypothesis testing problems

$$(4.1) \quad H_i: \underline{\theta} \in \Omega_{0i} \quad \text{against} \quad K_i: \underline{\theta} \in \Omega_i$$

where  $\Omega_{0i} = \Omega - \Omega_i$ ,  $1 \leq i \leq k$ , and  $\Omega_i$ ,  $1 \leq i \leq k$ , are defined as in Section 2.

We know that  $\Omega_0 = \bigcap_{i=1}^k \Omega_{0i}$ . A test of the hypotheses (4.1) will be defined to be a vector  $(\delta_1(\underline{x}), \dots, \delta_k(\underline{x}))$ , where the elements of the vector are ordinary test functions. When  $\underline{x}$  is observed, we reject  $H_i$  with probability  $\delta_i(\underline{x})$ ,  $1 \leq i \leq k$ . The power function of a test  $(\delta_1, \dots, \delta_k)$  is defined to be the vector  $(\beta_1(\underline{\theta}), \dots, \beta_k(\underline{\theta}))$ , where  $\beta_i(\underline{\theta}) = E_{\underline{\theta}} \delta_i$ ,  $1 \leq i \leq k$ . For  $\underline{\theta} \in \Omega_i$ , we know that  $\beta_i(\underline{\theta})$  is the probability of a correct selection and  $\delta_i(\underline{x})$  is the probability of selecting the best population  $\pi_i$ .

Let  $S_{\gamma_1}$  be the class of all tests  $(\delta_1, \dots, \delta_k)$  such that

$$(4.2) \quad \sup_{\underline{\theta} \in \Omega_0} \sum_{i=1}^k E_{\underline{\theta}} \delta_i \leq \gamma_1.$$

Hence the expected subset size for the selection rule  $\delta$  over  $\Omega_0$  in  $S_{\gamma_1}$  is less than or equal to  $\gamma_1$ . For each  $i$ ,  $1 \leq i \leq k$ , we would, subject to (4.2), like to have  $\beta_i(\underline{\theta})$  large when  $\underline{\theta} \in \Omega_i$ . For  $\underline{\theta} \in \Omega_i$ , if we make  $\beta_i(\underline{\theta})$  large, then  $\beta_j(\underline{\theta})$  will often have to be small for  $j \neq i$ , if (4.2) is to be satisfied.

Therefore, we will restrict attention to tests which

(1) maximize the minimum average power over  $\Omega_i$ ,  $1 \leq i \leq k$ , i.e. maximize

$$\sum_{i=1}^k \inf_{\underline{\theta} \in \Omega_i} E_{\underline{\theta}}^{\delta_i}(X) \text{ among tests } \delta \text{ in } S_{\gamma_1}, \text{ or}$$

(2) maximize the minimum power over  $\Omega_i$ ,  $1 \leq i \leq k$ , i.e. maximize

$$\min_{1 \leq i \leq k} \inf_{\underline{\theta} \in \Omega_i} E_{\underline{\theta}}^{\delta_i}(X) \text{ among tests } \delta \text{ in } S_{\gamma_1} \text{ (see Spjøtvoll (1972) and}$$

Gupta and Huang (1977)).

As discussed in Section 2, we will assume that, for any  $i$ , the statistic  $\underline{Z}_i = \{Z_{ij}; j \neq i, 1 \leq j \leq k\}$  is sufficient invariant under a transformation group and has joint distribution which depends only on  $\underline{\tau}_i$ , say  $P_{\underline{\tau}_i}(\underline{z}_i)$ , with SIP in  $\underline{\tau}_i$ . Let  $P_{\underline{\Delta}}(\underline{z}_i) = p_i(\underline{z}_i)$ ,  $1 \leq i \leq k$ , and  $P_{\underline{\tau}_i}(\underline{z}_i) = P_0(\underline{z}_i)$  when  $\tau_{ij} = \tau_{ji} = \text{constant}$ ,  $j \neq i$ ,  $1 \leq j \leq k$ .

Gupta and Huang (1977) have considered the first problem. In this section, we will consider the second problem.

Theorem 4.1. Suppose that for any  $i$ ,  $p_i(\underline{z}_i)/p_0(\underline{z}_i)$  is non-decreasing in  $\underline{z}_i$ . If  $\delta^0$  is given by

$$(4.3) \quad \delta_i^0(\underline{z}_i) = \begin{cases} 1 & \text{if } c_i p_i(\underline{z}_i) > p_0(\underline{z}_i) \\ \lambda_i & \text{if } c_i p_i(\underline{z}_i) = p_0(\underline{z}_i) \\ 0 & \text{if } c_i p_i(\underline{z}_i) < p_0(\underline{z}_i) \end{cases}$$

where  $c_i$  ( $> 0$ ) and  $\lambda_i$  are determined by  $\sum_{i=1}^k \int \delta_i^0 p_0 = \gamma_1$  and  $\int \delta_i^0 p_i$  is independent of  $i$ . If  $\sup_{\underline{\theta} \in \Omega_0} \sum_{i=1}^k E_{\underline{\theta}}^{\delta_i^0}$  occurs at  $\tau_{ij} = \tau_{ji} = \text{constant}$ , then  $\delta^0$  maximizes

$$\min_{1 \leq i \leq k} \inf_{\underline{\theta} \in \Omega_i} E_{\underline{\theta}}^{\delta_i} \text{ among all rules } \delta \text{ in } S_{\gamma_1}.$$

Proof. Since  $p_i(z_i)/p_0(z_i)$  is non-decreasing in  $z_i$ , then  $\delta_i^0(z_i)$  is non-decreasing in  $z_i$ . By Lemma 3.1, for any  $\underline{\theta} \in \Omega_i$ ,

$$E_{\underline{\theta}} \delta_i^0 \geq E_{\underline{\Delta}} \delta_i^0 = \int \delta_i^0 p_i$$

which is independent of  $i$ , by assumption.

Hence

$$(4.4) \quad \inf_{\underline{\theta} \in \Omega_i} E_{\underline{\theta}} \delta_i^0 = \int \delta_i^0 p_i.$$

Furthermore,

$$\begin{aligned} & \left( \sum_{i=1}^k c_i \right) \left( \min_{1 \leq i \leq k} E_{\underline{\Delta}} \delta_i^0 - \min_{1 \leq i \leq k} E_{\underline{\Delta}} \delta_i \right) \\ &= \sum_{i=1}^k c_i \left( \min_{1 \leq i \leq k} E_{\underline{\Delta}} \delta_i^0 - \min_{1 \leq i \leq k} E_{\underline{\Delta}} \delta_i \right) \\ &\geq \sum_{i=1}^k c_i \left( \int \delta_i^0 p_i - \int \delta_i p_i \right) \\ &= \int \sum_{i=1}^k (\delta_i^0 - \delta_i) (c_i p_i - p_0) + \int \sum_{i=1}^k (\delta_i^0 - \delta_i) p_0 \\ &\geq 0, \text{ by definition of } \delta_i^0 \text{ and the fact} \\ & \sum_{i=1}^k \int \delta_i p_0 \leq \sup_{\underline{\theta} \in \Omega_0} \sum_{i=1}^k E_{\underline{\theta}} \delta_i \leq \gamma_1 = \sum_{i=1}^k \int \delta_i^0 p_0. \end{aligned}$$

Hence

$$(4.5) \quad \min_{1 \leq i \leq k} E_{\underline{\Delta}} \delta_i^0 \geq \min_{1 \leq i \leq k} E_{\underline{\Delta}} \delta_i.$$

From (4.4) and (4.5), we have

$$\min_{1 \leq i \leq k} \inf_{\underline{\theta} \in \Omega_i} E_{\underline{\theta}} \delta_i^0 = \min_{1 \leq i \leq k} E_{\underline{\Delta}} \delta_i^0 \geq \min_{1 \leq i \leq k} E_{\underline{\Delta}} \delta_i \geq \min_{1 \leq i \leq k} \inf_{\underline{\theta} \in \Omega_i} E_{\underline{\theta}} \delta_i.$$

This completes the proof of the theorem.

Remarks:

(1) Theorem 4.1 implies Theorem 3.4. Since  $\min_{1 \leq i \leq k} \inf_{\underline{\theta} \in \Omega_i} E_{\underline{\theta}}^{\delta_i} = \inf_{\underline{\theta} \in \bar{\Omega}} S(\underline{\theta}, \delta)$

and if  $\sup_{\underline{\theta} \in \bar{\Omega}} R(\underline{\theta}, \delta) \leq \gamma_1$ , then  $\delta \in S_{\gamma_1}$ .

(2) Theorem 4.1 can also be proved by using Lemma 3.2 and following the same arguments as in the proof of Theorem 3.4.

(3) If  $c_i = c$ ,  $1 \leq i \leq k$ , then Theorem 4.1 follows from the theorem of Gupta and Huang (1977). Since

$$\begin{aligned} & \min_{1 \leq i \leq k} \inf_{\underline{\theta} \in \Omega_i} E_{\underline{\theta}}^{\delta_i^0} - \min_{1 \leq i \leq k} \inf_{\underline{\theta} \in \Omega_i} E_{\underline{\theta}}^{\delta_i} \\ &= \frac{1}{k} \sum_{i=1}^k \inf_{\underline{\theta} \in \Omega_i} E_{\underline{\theta}}^{\delta_i^0} - \min_{1 \leq i \leq k} \inf_{\underline{\theta} \in \Omega_i} E_{\underline{\theta}}^{\delta_i} \\ &\geq \frac{1}{k} \sum_{i=1}^k \inf_{\underline{\theta} \in \Omega_i} E_{\underline{\theta}}^{\delta_i^0} - \frac{1}{k} \sum_{i=1}^k \inf_{\underline{\theta} \in \Omega_i} E_{\underline{\theta}}^{\delta_i} \\ &\geq 0. \end{aligned}$$

## REFERENCES

1. Alam, K. (1973). On a multiple decision rule. Ann. Statist. 1, 750-755.
2. Bahadur, R. R. (1950). On a problem in the theory of k populations. Ann. Math. Statist. 21, 362-375.
3. Berger, R. L. and Gupta, S. S. (1980). Minimax subset selection rules with applications to unequal variance (unequal sample size) problems. Scand. J. of Statist. 7, 21-26.
4. Bjørnstad, J. F. (1981). A class of Schur-Procedures and minimax theory for subset selection. Ann. Statist. 9, 777-791.
5. Eaton, M. L. (1967). Some optimum properties of ranking procedures. Ann. Math. Statist. 38, 124-137.
6. Gupta, S. S. and Huang, D. Y. (1977). Some multiple decision problems in analysis of variance. Commun. Statist. A6(11), 1035-1054.
7. Gupta, S. S. and Huang, D. Y. (1980). A note on optimal subset selection procedures. Ann. Statist. 8, 1164-1167.
8. Gupta, S. S. and Huang, D. Y. (1981). Multiple statistical decision theory: recent developments. Lecture Notes in Statistics 6, Springer-Verlag, New York.
9. Lehmann, E. L. (1961). Some model I problems of selection. Ann. Math. Statist. 32, 990-1012.
10. Lehmann, E. L. (1966). On a theorem of Bahadur and Goodman. Ann. Math. Statist. 37, 1-6.
11. Nagel, K. (1970). On subset selection rules with certain optimizing properties. Ph.D. Thesis (Mimeo. Ser. No. 222), Dept. of Statist., Purdue University, West Lafayette, IN.
12. Spjøtvoll, E. (1972). On the optimality of some multiple comparison procedures. Ann. Math. Statist. 43, 398-411.
13. Studden, W. J. (1967). On selecting a set of k populations containing the best. Ann. Math. Statist. 38, 1072-1078.



REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report #84-7	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ON THE OPTIMALITY OF SOME SUBSET SELECTION PROCEDURES		5. TYPE OF REPORT & PERIOD COVERED Technical
7. AUTHOR(s) Lii-Yuh Leu		6. PERFORMING ORG. REPORT NUMBER Technical Report #84-7
9. PERFORMING ORGANIZATION NAME AND ADDRESS Purdue University Department of Statistics West Lafayette, IN 47907		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0455
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington, DC		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE March 1984
		13. NUMBER OF PAGES 14
		15. SECURITY CLASS. (of this report) Unclassified
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release, distribution unlimited.		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Subset selection, restricted minimax, multiple tests.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) As measures of goodness of a selection rule, usually two quantities, the probability of a correct selection and the expected size of the selected subset, are considered. Based on these two criteria, Gupta and Huang (1980) proved a theorem to derive a selection procedure with some optimality property. However, the theorem cannot be applied to the unequal sample sizes case. In this paper, we use a different method to generalize this theorem to the unequal sample sizes case. Also a dual problem is investigated. Also, we treat a selection procedure in terms of multiple tests. Based on this approach, we derive an optimality result.		