On the Properties of Subset Selection Procedures*

by

John J. Deely, Purdue University and Sandia Laboratory

and

Shanti S. Gupta, Purdue University

Department of Statistics

Division of Mathematical Statistics

Mimeograph Series No. 49

August 1965

*Research supported by Contract NONR-1100(26) with the Office of Naval Research and by Contract AF 33(657)11737 with the Aerospace Research Laboratories. Reproduction in whole or in part permitted for any purpose of the United States Government.
On the Properties of Subset Selection Procedures*

by

John J. Deely, Purdue University and Sandia Laboratory

and

Shanti S. Gupta, Purdue University

1. Introduction and Summary

Suppose \( \pi_1, \pi_2, \ldots, \pi_k \) are \( k \) normal populations with unknown means \( \mu_1, \mu_2, \ldots, \mu_k \) and a common known variance which we assume to be unity. In many situations the experimenter is interested in selecting a subset of the populations which contains the best population where by best we mean the population with the largest value \( \mu_{[k]} \) of the unknown means, \( \mu_1, \mu_2, \ldots, \mu_k \). For further amplification of this type of selection, reference could be made to Gupta (1965) and to references therein. A selection procedure, denoted by \( R_0 \) for this problem along with some of its properties has been discussed by Gupta (1956, 1965). This selection procedure is a member of the class \( C \) of procedures earlier studied by Seal (1955). The selection rule is such that it selects a non-empty subset of random size and under a specified loss function, the associated risk is always bounded above by \( \alpha = 1 - P^* \) where \( \alpha \) is a small preassigned number \( 0 < \alpha < 1/k \).

In this paper we study some desirable properties of the above selection procedure and make some comparisons with the 'approximate' optimal rule \( D \) of Seal (1955). In particular it is shown that the rule \( R \) is minimax and

*Research supported by Contract NONR-1100(25) with the Office of Naval Research and by Contract AF 33(657)11737 with the Aerospace Research Laboratories. Reproduction in whole or in part permitted for any purposes of the United States Government.
that under the slippage configuration of means, the expected size of the selected subset using $R$ is smaller than that corresponding to $D$; as well as being smallest for a subclass $C'$ of $C$. The probability of a correct selection using $R$ is strictly greater than the probability of a correct selection using $D$.

The subset selection problem is also considered from a Bayesian viewpoint, and under a linear loss function, the Bayes rule for selecting a subset is derived. This latter result is given in a more general form by Deely (1965) than that given by Dunnett (1960).

2. The Class $C$ and $C'$ of rules and the rule $R$

Let $x_1, x_2, \ldots, x_k$ be the observed values of the sample means each based on $n$ independent observations. Let the ordered sample means be denoted by

\[ x_{[1]} \leq x_{[2]} \leq \cdots \leq x_{[k]} \]

(2.1)

Let $c' = (c_1, c_2, \ldots, c_{k-1})$ be a vector whose components are arbitrary real numbers such that $c_i \geq 0$, for all $i$, and $\sum_{i=1}^{k-1} c_i = 1$. Then the class consists of rules

\[ D = D_{c'} = D(c_1 c_2, \ldots, c_{k-1}) \]

with $D$ given as follows:

"Select the population $\pi_{[i]}$ corresponding to the observed sample mean $x_{[1]}$ iff

\[ x_{[i]} \geq c_1 x_{[1]} + \cdots + c_{i-1} x_{[i-1]} + c_i x_{[i+1]} + \cdots + c_{k-1} x_{[k]} - \frac{t(P^*, c')}{\sqrt{n}} \]

(2.2)

where $t(P^*, c')$ is chosen so as to satisfy the basic probability requirement, namely,
\begin{equation}
\inf_{\Omega} P\{CS|D_{\Omega}\} = P^*,
\end{equation}

\(\Omega\) being the parameter space of \(\mu' = (\mu_1, \ldots, \mu_k)\) and \(CS\) standing for correct selection. A correct selection is the selection of any subset which includes the population associated with \(\mu_{[k]}\).

To define the subclass \(\mathcal{C}'\) of \(\mathcal{C}\), we impose the following restrictions on the vector \(c: c_j = 1\) for some \(j = 1, 2, \ldots, k-1\).

An important rule in the class \(\mathcal{C}\) which has been studied in great detail is a member of the class \(\mathcal{C}'\) defined above for \(c' = (0, 0, \ldots, 0, 1)\); this rule denoted in earlier papers by \(R\) is:

Select the population corresponding to the observed sample mean \(x_{[1]}\) iff

\begin{equation}
x_{[1]} \geq x_{[k]} - \frac{d}{\sqrt{n}},
\end{equation}

where \(d = d(k, P^*) = t(P^*, \xi')\) again chosen so as to satisfy the basic probability requirement.

It may be pointed out that for the zero-one loss function

\begin{equation}
L(S_j, \mu') = \begin{cases} 0 & \text{if } \pi(k) \text{ with } \mu_{[k]} \in S_j, \text{ the selected subset, } j = 1, 2, \ldots, 2^k - 1. \\ 1 & \text{otherwise}, \end{cases}
\end{equation}

the risk is given by

\begin{equation}
\text{Risk} = r(D_{\Omega'}, \mu') = 1 - P\{CS|D_{\Omega}\}.
\end{equation}
Hence, the basic probability requirement (2.3) is equivalent to requiring that

\[(2.7) \quad \sup_{\Omega} r(D_\epsilon, \mu) \leq 1 - P^*\]

The rule \(\overline{D}\) given by Seal (1955) is also a member of \(\mathcal{C}\) and is defined by the vector \(c'\) with \(c_j = \frac{1}{k-1}, j=1,2,\ldots,k-1\). This rule is:

Select the population corresponding to the observed sample mean \(x_i\) iff

\[(2.8) \quad x_i \geq \bar{x} - \frac{t(P^*, c')}{\sqrt{n}} \quad \text{where} \]

\[
\bar{x} = \frac{1}{k-1} \sum_{\substack{j=k \atop j\neq i}}^{j=k} x_j \quad \text{and} \quad t(P^*, c') \quad \text{as stated earlier.}
\]

3. Some Results on Minimaxity and Expected Size

From (2.6) and the definition of a minimax rule we see that \(R\) is minimax in \(\mathcal{C}\) provided

\[(3.1) \quad \min_{\mu \in \Omega} P(CS|R) \geq \min_{\mu \in \Omega} P(CS|D)\]

for every \(D \in \mathcal{C}\). But recalling the definition of the class \(\mathcal{C}\), it can be observed that the quantity \(t(P^*, c')\) is chosen so as to guarantee

\[(3.2) \quad \min_{\mu \in \Omega} P(CS|D) = P^* \quad \text{for any rule} \ D \ \text{in} \ \mathcal{C}.
\]

Hence every rule in \(\mathcal{C}\) and in particular \(R\) is minimax.
It may be pointed out that if we choose \( c' = (1, 0, 0, \ldots, 0) \), the rule \( D_0 \) defined by this choice of \( c' \) selects all populations with probability 1 and

\[
P[CS|D_0] = 1 > P[CS|R].
\]

But the rule \( D_0 \) is quite trivial and not very practical from the experimenters viewpoint for the following reason.

The selection procedures discussed here select a subset whose size is a random variable taking values in the set of integers \( 1, 2, \ldots, k \). Thus in comparing decision procedures of this type it is meaningful to use the expected value of this random variable, to be denoted by \( E[S|D, C] \). For the rule \( R \), \( \Omega \), \( \mu' = (\mu, \mu, \ldots, \mu) \), Gupta (1965) showed that \( E[S|R] \leq kP^* \).

More generally, for any rule \( D, C \in C' \), we have from Seal (1955),

\[
\inf_{\Omega} P[CS|D, C] = \inf_{\mu' = (\mu, \mu, \ldots, \mu)} P[CS|D, C]
\]

and hence the solution for the constants \( t(P^*, C') \), earlier defined by (2.2) and (2.3), can be obtained from equations of the type

\[
P[Y \leq t(P^*, C')] = P^*
\]

where the random variable \( Y \) traverses the real line.

Thus for any rule \( D, C \in C', t(P^*, C') \) will be non-negative provided \( P^* \) is sufficiently large; for example \( P^* > (1/k) \) for the rule \( R \). Hence every rule \( D, C \in C' \) (excluding the rule \( R \) ) will select a subset
consisting of at least 2 populations with probability 1. This implies that for any such $D_\infty$

\[
(3.6) \quad \mathbb{E}[S|D_\infty] \geq 2.
\]

Gupta (1965) showed that in a subset $\Omega(\delta)$ defined by $\mu_{[k]} = \mu + \delta (\delta > 0), \mu_{[i]} \leq \mu, \ i = 1, 2, \ldots, k-1$, $\mathbb{E}[S|R]$ assumes its maximum over $\Omega(\delta)$ when $\mu_{[i]} = \mu, \ i = 1, 2, \ldots, k-1$

\[
(3.7) \quad \frac{1}{\delta} \max_{\Omega(\delta)} \mathbb{E}[S|R] < \int_{-\infty}^{\infty} \phi(z) \frac{1}{\sqrt{2\pi\delta}} \exp\left(-\frac{z^2}{2\delta}\right) dz + \int_{-\infty}^{\infty} (k-1) \phi(z) \frac{1}{\sqrt{2\pi\delta}} \exp\left(-\frac{(z+d)^2}{2\delta}\right) \phi(z) dz.
\]

From (3.6) it follows that there exists a $\delta_0$ such that

\[
(3.8) \quad \max_{\Omega(\delta)} \mathbb{E}[S|R] < 2 \quad \text{for } \delta > \delta_0.
\]

Therefore, if $F^*$ and $\delta$ are sufficiently large, it can be seen from (3.6) and (3.8) that

\[
(3.9) \quad \mathbb{E}[S|R] < \mathbb{E}[S|D_\infty]
\]

for any $\mu' \in \Omega(\delta)$. 

\[
\begin{align*}
\delta & = 0.50 \\
\delta \sqrt{n} & = 2
\end{align*}
\]
So in this case good justification can be made for using the decision procedure $R$. Namely, the $P(CS|R)$ is larger than $P^*$, a preassigned number, and its $E[S|R]$ is smallest among rules in $L'$ under the restrictions in (3.2).

A table giving the values of the expected proportion of the populations in the selected subset for the slippage configuration is given below.

**Table**

This table gives the expected proportion of $k$ normal populations retained in the selected subset by the suggested procedure $R$ under the configuration that one of the populations has its mean greater than all the others by $8$ times the common known standard deviation of the sample means (selected values of $k$, $P^*$ and $\delta^*$ are considered).

<table>
<thead>
<tr>
<th>$P^*$</th>
<th>$k$</th>
<th>$6_{kn}$</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>25</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.00</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
</tr>
<tr>
<td>.10</td>
<td>.7495</td>
<td>.7494</td>
<td>.7495</td>
<td>.7497</td>
<td>.7498</td>
<td>.7497</td>
<td>.7497</td>
<td>.7497</td>
</tr>
<tr>
<td>.25</td>
<td>.7467</td>
<td>.7463</td>
<td>.7469</td>
<td>.7479</td>
<td>.7489</td>
<td>.7492</td>
<td>.7492</td>
<td>.7492</td>
</tr>
<tr>
<td>.50</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
</tr>
<tr>
<td>1.00</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
</tr>
<tr>
<td>2.00</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
</tr>
<tr>
<td>3.00</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
<td>.7500</td>
</tr>
<tr>
<td></td>
<td>.00</td>
<td>.9000</td>
<td>.9000</td>
<td>.9000</td>
<td>.9000</td>
<td>.9000</td>
<td>.9000</td>
<td>.9000</td>
</tr>
<tr>
<td>.10</td>
<td>.8994</td>
<td>.8995</td>
<td>.8996</td>
<td>.8998</td>
<td>.8998</td>
<td>.8998</td>
<td>.8998</td>
<td>.8998</td>
</tr>
<tr>
<td>.25</td>
<td>.8965</td>
<td>.8968</td>
<td>.8976</td>
<td>.8985</td>
<td>.8992</td>
<td>.8995</td>
<td>.8995</td>
<td>.8995</td>
</tr>
<tr>
<td>.50</td>
<td>.8861</td>
<td>.8872</td>
<td>.8891</td>
<td>.8937</td>
<td>.8968</td>
<td>.8981</td>
<td>.8981</td>
<td>.8981</td>
</tr>
<tr>
<td>1.00</td>
<td>.8569</td>
<td>.8480</td>
<td>.8574</td>
<td>.8710</td>
<td>.8840</td>
<td>.8901</td>
<td>.8901</td>
<td>.8901</td>
</tr>
<tr>
<td>2.00</td>
<td>.7219</td>
<td>.7023</td>
<td>.7151</td>
<td>.7527</td>
<td>.8015</td>
<td>.8302</td>
<td>.8302</td>
<td>.8302</td>
</tr>
<tr>
<td>3.00</td>
<td>.6001</td>
<td>.5271</td>
<td>.5065</td>
<td>.5370</td>
<td>.6090</td>
<td>.6643</td>
<td>.6643</td>
<td>.6643</td>
</tr>
<tr>
<td></td>
<td>.00</td>
<td>.9500</td>
<td>.9500</td>
<td>.9500</td>
<td>.9500</td>
<td>.9500</td>
<td>.9500</td>
<td>.9500</td>
</tr>
<tr>
<td>.10</td>
<td>.9496</td>
<td>.9496</td>
<td>.9497</td>
<td>.9499</td>
<td>.9500</td>
<td>.9500</td>
<td>.9500</td>
<td>.9500</td>
</tr>
<tr>
<td>.25</td>
<td>.9474</td>
<td>.9478</td>
<td>.9484</td>
<td>.9490</td>
<td>.9496</td>
<td>.9498</td>
<td>.9498</td>
<td>.9498</td>
</tr>
<tr>
<td>.50</td>
<td>.9394</td>
<td>.9409</td>
<td>.9432</td>
<td>.9459</td>
<td>.9480</td>
<td>.9489</td>
<td>.9489</td>
<td>.9489</td>
</tr>
<tr>
<td>1.00</td>
<td>.9082</td>
<td>.9120</td>
<td>.9202</td>
<td>.9305</td>
<td>.9396</td>
<td>.9437</td>
<td>.9437</td>
<td>.9437</td>
</tr>
<tr>
<td>2.00</td>
<td>.7950</td>
<td>.7902</td>
<td>.8079</td>
<td>.8114</td>
<td>.8801</td>
<td>.9014</td>
<td>.9014</td>
<td>.9014</td>
</tr>
<tr>
<td>3.00</td>
<td>.6584</td>
<td>.6114</td>
<td>.6095</td>
<td>.6490</td>
<td>.7185</td>
<td>.7670</td>
<td>.7670</td>
<td>.7670</td>
</tr>
</tbody>
</table>
4. Comparison of $E[S|D]$ and $E[S|IR]$ 

It can be seen that the expected size of the selected subset using any rule $D$ is given by:

\[ E[S|D] = \sum_{i=1}^{k} P_i \text{selecting the population with mean } \mu_{i,j} | D \]

So we shall first derive the expression for the probability of selecting the population $\pi_{(j)}$ with mean $\mu_{[j]}$ for the rule $D$. (Recall $\mu_{[1]} \leq \mu_{[2]} \leq \cdots \leq \mu_{[k]}$ are the ranked values of the mean vector $\mu = (\mu_1, \mu_2, \ldots, \mu_k)$.) Let $X_{(j)}$ be the sample mean (unknown) which is associated with the population $\pi_{(j)}$. Then

\[
P_{\pi_{(j)}} \text{ is selected | } D = \frac{1}{k-1} \sum_{i=1}^{k-1} X_{(j)} - \frac{t(P^*, \xi')}{\sqrt{n}} \]

where $\xi' = (\frac{1}{k-1}, \ldots, \frac{1}{k-1})$ for the rule $D$. Equation (4.8) reduces to

\[
P_{\pi_{(j)}} \text{ is selected | } D = \Phi\left( \frac{t(P^*, \xi')}{\sqrt{n}} \right) - \frac{1}{k-1} \sum_{i=1}^{k-1} \mu_{[i]} \]

where $\Phi$ and $\phi$ are the standard normal density and cumulative distribution functions respectively. In particular from (4.8) we obtain

\[
P_{\pi_{(k)}} \text{ is selected | } D = \Phi\left( \frac{t(P^*, \xi')}{\sqrt{n}} - \frac{1}{k-1} \sum_{i=1}^{k} \mu_{[i]} \right)
\]
From (2.3) we see that the constant \( t(P^*, \alpha') \) to carry out the procedure can be obtained by using (4.6) and solving

\[
\frac{t(P^*, \alpha')}{(\frac{k}{k-1})^{1/2}} = P^* \quad \text{for} \quad t(P^*, \alpha')
\]

which gives

\[
t(P^*, \alpha') = (\frac{k}{k-1})^{1/2} \gamma; \quad \text{letting} \quad \gamma = \phi^{-1}(P^*).
\]

Then using (4.5) in (4.3) we obtain

\[
P(CS|\overline{D}) = \phi(\gamma + (\frac{n}{k(k-1)})^{1/2} \sum_{i=1}^{k-1} (\mu_{[k]i} - \mu_{[i]}))
\]

From Gupta (1965) we have

\[
P(CS|R) = \int_{-\infty}^{\infty} \prod_{j=1}^{k-1} \phi(z + d + \sqrt{n} (\mu_{[k]j} - \mu_{[j]})) \phi(z)dz.
\]

Let \( \Omega' \subset \Omega \) be the space of mean vectors \( \mu' \) such that \( \mu' = (\mu, \mu, \ldots, \mu, \mu + \delta) \) where \( -\infty < \mu < \infty \) and \( \delta \geq 0 \). Then for \( \mu' \in \Omega' \) we have

\[
P(CS|\overline{D}) = \phi(\gamma + (\frac{k-1}{k})^{1/2} \sqrt{n} \delta)
\]

and

\[
P(CS|R) = \int_{-\infty}^{\infty} \phi^{k-1}(z + d + \sqrt{n} \delta) \phi(z)dz.
\]
Then using (4.1), (4.3), and (4.6) we have

\[ E[S|D] = \sum_{j=1}^{k} \tilde{\phi}(y + (n \frac{\mu_{[j]} - \mu_{[i]}}{k(k-1)})^{1/2} \sum_{i=1}^{i=k} (\mu_{[j]} - \mu_{[i]})), \]

Now from Gupta (1965) we have the expression for the expected size of the selected subset using the rule R

\[ E[S|R] = \sum_{j=1}^{k} \int_{-\infty}^{\infty} \tilde{\phi}(z+d+(\mu_{[j]} - \mu_{[i]})/\sqrt{n}) \varphi(z) dz. \]

Let \( \Omega' \subset \Omega \) be the space of all mean vectors \( \mu' \) such that \( \mu' = (\mu, \mu, \ldots, \mu, \mu + \delta) \) where \( -\infty < \mu < \infty \) and \( \delta \geq 0 \). Then for \( \mu' \in \Omega' \), we can write

\[ E[S|R] = \sum_{j=1}^{k-1} \tilde{\phi}(y - (n \frac{\mu_{[j]} - \mu_{[i]}}{k(k-1)})^{1/2} \delta/\sqrt{n}) + \tilde{\phi}(y + (k-1)^{1/2} \delta/\sqrt{n}) \]

and

\[ E[S|R] = (k-1) \tilde{\phi}(y - (n \frac{1}{k(k-1)})^{1/2} \delta) + \tilde{\phi}(y + (k-1)^{1/2} \delta/\sqrt{n}) \]

We desire values of \( \delta \) for which

\[ E[S|R] - E[S|D] \leq 0. \]
To do this we will find the common values of $\delta$ for which

\[(4.20)\quad \int_{-\infty}^{\infty} \frac{k-1}{\sqrt{n}} \varphi(z)dz \leq \delta \left( \gamma + \frac{k-1}{k} \frac{1}{\sqrt{n}} \right) \leq 0
\]

and

\[(4.21)\quad \int_{-\infty}^{\infty} \frac{k-2}{\sqrt{n}} \varphi(z)dz \leq \delta \left( \gamma - \frac{n}{k(k-1)} \right) \leq 0.
\]

Now

\[(4.22)\quad \int_{-\infty}^{\infty} \frac{k-1}{\sqrt{n}} \varphi(z)dz \leq \delta \left( \gamma + \frac{k-1}{k} \right) \frac{1}{\sqrt{n}}
\]

\[\leq \int_{-\infty}^{\infty} \frac{\varphi(z)dz}{\sqrt{n}} \left( \gamma + \frac{k-1}{k} \right) \frac{1}{\sqrt{n}}
\]

\[= \delta \left( \frac{\varphi(z)dz}{\sqrt{n}} \right) \left( \gamma + \frac{k-1}{k} \right) \frac{1}{\sqrt{n}} < 0
\]

provided

\[(4.23)\quad \gamma + \frac{k-1}{k} \frac{1}{\sqrt{n}} > \frac{d - 8\sqrt{n}}{\sqrt{2}}
\]

which yields

\[(4.24)\quad \sqrt{n} \delta > \frac{d - \gamma}{\sqrt{2} \left( \frac{k-1}{k} \right)^{1/2} - \frac{1}{\sqrt{2}}}
\]

Thus \((4.20)\) is satisfied for all $\delta$ as in \((4.24)\).

For the inequality \((4.21)\) we observe that...
\[
\int_0^\infty \frac{k-2(z+d)}{\sqrt{n} \delta} \varphi(z) \, dz - \Phi(\gamma - \left(\frac{n}{k(k-1)}\right)^{1/2} \delta)
\]

\leq \int_0^\infty \frac{1}{\sqrt{n} \delta} \varphi(z) \, dz - \Phi(\gamma - \left(\frac{n}{k(k-1)}\right)^{1/2} \delta)

= \Phi\left(\frac{d - \delta \sqrt{n}}{\sqrt{2}}\right) - \Phi(\gamma - \left(\frac{n}{k(k-1)}\right)^{1/2} \delta) < 0

\]

provided

\[
\frac{d - \delta \sqrt{n}}{\sqrt{2}} < \gamma - \left(\frac{n}{k(k-1)}\right)^{1/2} \delta,
\]

which yields

\[
\frac{1}{\sqrt{n} \delta} > \frac{\frac{d}{\sqrt{2}} - \gamma}{\frac{1}{\sqrt{2}} - \left(\frac{1}{k(k-1)}\right)^{1/2}}, \quad k \geq 3.
\]

Thus (4.31) is satisfied for all \( \delta \) as in (4.37). For \( k \geq 3 \) we have

\[
\frac{1}{\sqrt{n} \delta} < \frac{\frac{k-1}{k} \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} - \left(\frac{1}{k(k-1)}\right)^{1/2}}{\left(\frac{k-1}{k}\right)^{1/2} - \frac{1}{\sqrt{2}}}
\]

which gives the following result.

\[
E[S|R] < E[S|D], \quad \text{when} \quad \sqrt{n} \delta > \frac{\frac{d}{\sqrt{2}} - \gamma}{\left(\frac{k-1}{k}\right)^{1/2} - \frac{1}{\sqrt{2}}},
\]

Thus we see that the rule \( R \) yields a smaller subset size on the average than the rule \( D \) for all mean vectors \( \mu^* \) in \( \Omega \).
To show the existence of such a $d$, it remains to show that the right hand side of \((4.14)\) is a positive number. The denominator is positive if $k \geq 3$. Thus we need only show that

\[
\frac{2c}{(4.14)} \quad d - \sqrt{2} \gamma > 0 \quad \text{for} \quad k \geq 3.
\]

Recall that "\(d\)" is chosen so as to make

\[
(4.12) \quad \inf_{\Omega} P[CS|R] = \int_{-\infty}^{\infty} \frac{\phi^{-1}(z+d)}{\phi(z)} \varphi(z) \, dz = P^*\]

and "\(\gamma\)" is such that $\bar{\gamma}(\gamma) = P^*$. Now

\[
P^* = \int_{-\infty}^{\infty} \frac{\phi(z+d)}{\varphi(z)} \, dz = \frac{\bar{\gamma}(d)}{\sqrt{2}}
\]

which implies $d = \sqrt{2} \gamma$ when $k = 2$. But for $k \geq 3$, $\phi^{-1}(z+d) < \bar{\gamma}(z+d)$ for every $z$; hence if $d \leq \sqrt{2} \gamma$, then

\[
\int_{-\infty}^{\infty} \phi^{-1}(z+d) \varphi(z) \, dz < \int_{-\infty}^{\infty} \bar{\gamma}(z+d) \varphi(z) \, dz \leq P^*,
\]

a contradiction. Therefore \((4.12)\) is true for $k \geq 3$. (Go to top of p. 14)
Thus we see that the rule $R$ yields a smaller subset size on the average than the rule $D$ for all mean vectors $\mu'$ assuming the mean vector $\Omega'$.

configuration) except when $\delta$ is relatively small.

The table at the end of Section 3 gives the expected size of the selected subset for the rule $R$ at various values of $k$ and $\delta$ assuming the mean vector $\Omega'$ to be in $\Omega'$.

5. **Bayesian Approach**

A. Introduction and main theorem

In this section we will make the further assumption that each population mean is itself a random variable with a distribution $G_i$, $i=1,2,\ldots,k$.

The distribution $G_i$ is called an a priori distribution and $G = \prod_{i=1}^{k} G_i$ is called an a priori distribution on the parameter space $\Omega$. This is the so-called Bayesian approach to the multiple decision problem. And whereas the zero-one loss function (2.5) was used in the previous work we will assume here that the loss function is the so-called linear loss:

\[
L(S_j,\mu') = \sum_{q \in S_j} c_{jq}(\mu[k] - \mu_q)
\]

where $j=1,2,\ldots,2^{k-1}$ and $\alpha_{jq} \geq 0$, $S_j$ being a subset of the $k$ populations. This loss function is analogous to the loss function considered by Bahadur and Robbins (1950), Dunnett (1960), and Bland (1961) for the problem of selecting only one population. We then define the Bayes risk (overall expected loss, or average risk) of a decision procedure $D$ with respect to the a priori distribution $G$ as:

\[
B(D,G) = \int_{\Omega} \left\{ \int_{E^k} L(D,\mu') f(x|\mu') dx \right\} dG(\mu),
\]
where \( x \in \mathbb{R}^k \), Euclidean \( k \)-space, and \( f(x|\mu') \) is the product of the \( k \)-independent normal densities; with the density corresponding to \( \pi_1 \) having mean \( \mu_1 \), variance 1.

By Fubini's theorem we can write

\[
B(D,G) = \int_{\mathbb{R}^k} \left\{ \int_\Omega L(D,\mu') f(x|\mu') dG(\mu') \right\} dx
\]

and then let

\[
\psi_G(D,x) = \int_\Omega L(D,\mu') f(x|\mu') dG(\mu').
\]

An optimal rule is the so-called Bayes procedure and is defined as any procedure \( D^* \) such that

\[
B(D^*,G) \leq B(D,G) \quad \text{for any } D.
\]

It can be seen that \( D^* \) is Bayes with respect to \( G \) if

\[
\psi_G(D^*,x) \leq \psi_G(D,x)
\]

for every \( D \) at each \( x \in \mathbb{R}^k \). It is noted that condition \((5.3)'\) is sufficient but not necessary. Now each rule \( D \) after observing a vector \( x \) must select one of the \( 2^k - 1 \) subsets. Hence if at each \( x \) we compute the \( 2^k - 1 \) numbers

\[
\psi_G(S_j,x) = \int_\Omega L(S_j,\mu') f(x|\mu') dG(\mu); j=1,2,...,2^k-1,
\]
and select the smallest of these (or anyone of those that may equal the smallest), we then have a Bayes decision procedure with respect to the a priori $G$. That is, a Bayes decision procedure $D^*$ is defined by:

$$(5.5) \quad D^* = D^*(x) = S_j \quad \text{where } j \text{ is any positive integer } 1,2,\ldots,2^{k-1} \text{ such that }$$

$$\psi_G(S_j;x) = \min \psi_G(S_i;x): 1 \leq i \leq 2^{k-1}.$$ 

From looking at the loss function (5.1) one might intuitively feel that the Bayes procedure would select a subset consisting of only one population since only the one element subsets can make the loss function zero. We now state and prove this result as a theorem. Note that the theorem is stated in a more general framework than for the normal means problem. After proving the theorem and two corollaries, the exact Bayes procedure relative to some specific a priori distributions will be given for the normal means problem.

For convenience we adopt the following notation:

$S_j = \{\text{one element, population } \pi_j \}$ for $j = 1,2,\ldots,k$ and no explicit knowledge about $S_j$ for $j = k+1,k+2,\ldots,2^k-1$; $f(x|\mu^i)$ will be the product of the $k$-independent densities from which an observation vector (or matrix) $X$ is taken; each density being conditional upon a parameter $\mu_i$, $i=1,2,\ldots,k$.

$$(5.6) \textbf{Theorem:} \text{ In the loss function (5.1) let } \alpha_{jq} = \alpha > 0 \text{ for } j=1,2,\ldots,k. \text{ Let}$$
\[ a_q = \int_{\Omega} (\mu[k]-\mu_q) f(x|\mu) d\mu, \]

\[ a_{[1]} = \min\{a_q : 1 \leq q \leq k\}, \quad \text{and} \]

\[ b_q = a_q - a_{[1]}. \]

If \( a_{[1]} \neq 0 \), then a necessary and sufficient condition that

\[ \min\{\psi_q(S_j, \bar{x}) : 1 \leq j \leq 2^k-1\} = \min\{\psi_G(S_j, \bar{x}) : 1 \leq j \leq k\} \]

is that

\[ \sum_{q \in S_j} a_{q,j} \geq c^{-1/a_{[1]}} (\sum_{q \in S_j} b_q) \]

for every \( j=1,2,\ldots,2^k-1 \). If \( a_{[1]} = 0 \), then (5.6.1) is true.

**Proof:** First observe that using the notation of the theorem, we have from (5.4) that

\[ \psi_G(S_j, \bar{x}) = c a_{j} ; \quad j=1,2,\ldots,k \]

\[ \psi_G(S_j, \bar{x}) = \sum_{q \in S_j} c_{q,j} a_{q} ; \quad j+k+1,\ldots,2^k-1. \]

Next it is always true that,

\[ \min\{\psi_G(S_j, \bar{x}) : 1 \leq j \leq 2^k-1\} \leq \min\{\psi_G(S_j, \bar{x}) : 1 \leq j \leq k\} . \]

Hence (5.6.1) is true iff
\[(5.6.5) \quad \min\{q_i(s_j, x) : 1 \leq i \leq k\} \leq q_i(s_j, x)\]

for every \(j = 1, 2, \ldots, 2^k - 1\). Now from \((5.6.3)\) and definition of \(a_{[1]}\) we have

\[(5.6.6) \quad \min\{q_i(s_j, x) : 1 \leq i \leq l\} = c_{a_{[1]}}.\]

Thus \((5.6.1)\) is true iff

\[(5.6.7) \quad c_{a_{[1]}} \leq \sum_{q \in S_j} c_{j_q} a_{q}, \quad (j = 1, 2, \ldots, 2^k - 1).\]

Note that if \(a_{[1]} = 0\) then \((5.6.7)\) is clearly true and hence \((5.6.1)\) holds.

If \(a_{[1]} \neq 0\), we have condition \((5.6.7)\) iff

\[(5.6.8) \quad c_{a_{[1]}} \leq \sum_{q \in S_j} c_{j_q} (a_{[1]} + b_q)\]

which holds iff

\[(5.6.9) \quad (c - \sum_{q \in S_j} c_{j_q}) \leq (1/a_{[1]}) \sum_{q \in S_j} c_{j_q} b_q\]

iff

\[(5.6.10) \quad \sum_{q \in S_j} c_{j_q} \geq c \cdot (1/a_{[1]}) \sum_{q \in S_j} c_{j_q} b_q\]

for \(j = 1, 2, \ldots, 2^k - 1\), which is condition \((5.6.2)\) and thus completes the proof.
Remark: In general the quantity $a_q$ is difficult to compute and hence $b_q$ is not available. Thus the necessary and sufficient condition (5.6.2) is not too useful from a practical viewpoint. For this reason the following corollary is significant.

(5.7) Corollary Under the conditions of Theorem (5.6),

\[(5.7.1) \quad \min \{ \psi_G(s_j, \mathbf{X}) : 1 \leq j \leq 2^k - 1 \} = \min \{ \psi_G(s_j, \mathbf{X}) : 1 \leq j \leq k \}\]

if

\[(5.7.2) \quad \sum_{q \in S_j} \alpha_j q \geq \alpha \quad \text{for every } j=1, 2, \ldots, 2^k - 1.\]

Proof: In the proof of the theorem it was shown that (5.6.7) is necessary and sufficient for (5.7.1). But

\[\sum_{q \in S_j} c_{jq} a_q \geq \sum_{q \in S_j} \alpha_j a_{q[l]} = a_{q[l]} \sum_{q \in S_j} \alpha_j q \geq \alpha a_{q[l]}\]

using (5.7.2), which completes the proof.

The following corollary is the main result:

(5.8) Corollary: Let (5.1) be the loss function and let $\sum_{q \in S_j} \alpha_j q \geq \alpha$ for every $j=1, 2, \ldots, 2^k - 1$. Then the Bayes procedure with respect to an a priori distribution $G$ for selecting a subset containing the best of $k$-populations is given by:

$D^* = D^*(\mathbf{x}) = S_j$ where $j$ is any positive integer $1, 2, \ldots, k$ such that
\[ \psi_G(S_j, \bar{x}) = \min\{\psi_G(S_i, \bar{x}): 1 \leq i \leq k\}. \]

**Proof:** From (5.5) we see that the Bayes procedure is obtained from the \(\min\{\psi_G(S_i, \bar{x}): 1 \leq i \leq 2^k - 1\}\). But from Corollary (5.7), this minimum occurs among the first \(k\)-one element subsets; i.e.,

\[ \min\{\psi_G(S_i, \bar{x}): 1 \leq i \leq 2^k - 1\} = \min\{\psi_G(S_i, \bar{x}): 1 \leq i \leq k\}. \]

Thus the Bayes procedure is to select the one element subset which minimizes the quantity \(\psi_G(S_j, \bar{x})\) among the \(k\)-such numbers. This completes the proof.

**Remark:** The following are three examples of loss functions for which Corollary (5.8) is true. Note: \(|S_j| = \text{number of populations in the subset } S_j.\)

(i) \(L(S_j, \mu') = \sum_{q \in S_j} (\mu_{[k]} - \mu_q), \text{ sum of losses.}\)

(ii) \(L(S_j, \mu') = \frac{1}{|S_j|} \sum_{q \in S_j} (\mu_{[k]} - \mu_q), \text{ average loss.}\)

(iii) \(L(S_j, \mu') = (k+1-|S_j|) \sum_{q \in S_j} (\mu_{[k]} - \mu_q).\)

**B. Normal means problem with specific a priori distributions.**

We will consider the normal means problem and give the Bayes procedure for selecting a subset when the a priori distribution is: (1) normal and (2) uniform. These procedures are the same as those for the problem of selecting the best when the loss function is:

\[ (5.9) \quad L(S_j, \mu') = \mu_{[k]} - \mu_j; \quad j = 1, 2, \ldots, k. \]
and hence the derivations of these procedures will be given as part of a later publication. A loss function of the form (5.1) satisfying Corollary (5.8) is assumed.

(1) $g_i$ is normal with mean $\lambda_i$ and variance $\beta_i^2$.

The Bayes procedure $D^*_k$ given by Corollary (5.8) says:

Select $\pi_j$ such that

$$\frac{n\beta_i^2 x_i + \lambda_i}{1 + n\beta_i^2} = \max_{1 \leq i \leq k} \left\{ \frac{n\beta_i^2 x_i + \lambda_i}{1 + n\beta_i^2} \right\},$$

where $x_i$ is the sample mean of $n$ observations.

(2) $g_i$ is uniform on $(\lambda_j - d_j, \lambda_j + d_j)$.

The Bayes procedure $D^*_k$ given by Corollary (5.8) says:

Select $\pi_j$ such that

$$\frac{\varphi(\beta_j) - \varphi(\alpha_j)}{\varphi(\alpha_j) - \varphi(\beta_j)} + \sqrt{n} x_j = \max_{1 \leq i \leq k} \left\{ \frac{\varphi(\beta_j) - \varphi(\alpha_j)}{\varphi(\alpha_j) - \varphi(\beta_j)} + \sqrt{n} x_j \right\},$$

where

$$\alpha_j = \sqrt{n} (\lambda_j - d_j - x_j),$$

$$\beta_j = \sqrt{n} (\lambda_j + d_j - x_j),$$

and

$x_j$ being the sample mean based on $n$ observations.
References


4. Deely, J.J. (1965). Multiple decision procedures from an empirical Bayes approach. Mimeograph Series No. 45, Department of Statistics, Purdue University.


On the Properties of Subset Selection Procedures

Technical Report, August 1965

Deely, John J. and Gupta, Shanti S.

August 1965

NONR-1100(26)

Mimeograph Series No. 49

Releasable without limitation on dissemination

Also supported by the Aerospace Research Laboratories

Logistics and Mathematical Sciences Branch
Office of Naval Research
Washington, D.C. 20360

In this paper we study some desirable properties of a selection procedure which selects the normal population $\pi_i$ with mean $\mu_i$ and variance unity ($i=1,2,\ldots,k$) iff the observed sample mean $x_i$ from $\pi_i \in [x_{\max} - d, x_{\max}]$. This rule earlier studied by Gupta (1956, 1965) is compared with the "approximate" optimal rule $\overline{D}$ of Seal (1955). It is shown that the rule $R$ is minimax. It is also shown that under the slippage configuration of means given by $(\mu_1, \mu_2, \ldots, \mu_k)$ the expected size of the selected subset using $R$ is smaller than that corresponding to $\overline{D}$ and that the probability of a correct selection using $R$ is strictly greater than that of $\overline{D}$, provided $\delta$ satisfies some inequalities. Under a more general linear loss function, the Bayes rule for selecting a subset is also derived.
### Key Words

- Selection Ranking
- Multiple Decision
- Subset Selection
- Best Population
- Bayes Rule

### Instructions

1. **Originating Activity:** Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (corporate author) issuing the report.

2a. **Report Security Classification:** Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. **Group:** Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, where applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. **Report Title:** Complete the report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. **Descriptive Notes:** If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. **Author(s):** Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. **Report Date:** Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. **Total Number of Pages:** The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. **Number of References:** Enter the total number of references cited in the report.

8a. **Contract or Grant Number:** If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. **Project Number:** Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. **Originator's Report Number(s):** Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. **Other Report Number(s):** If the report has been assigned any other report numbers (either by the originator or by the sponsor), also enter this number(s).

10. **Availability/Limitation Notices:** Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

   1. "Qualified requesters may obtain copies of this report from DDC."
   2. "Foreign announcement and dissemination of this report by DDC is not authorized."
   3. "U.S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through "
   4. "U.S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through "
   5. "All distribution of this report is controlled. Qualified DDC users shall request through "

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. **Supplementary Notes:** Use for additional explanatory notes.

12. **Sponsoring Military Activity:** Enter the name of the departmental project office or laboratory sponsoring (paying for) the research and development. Include address.

13. **Abstract:** Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

   It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (TS). (S). (C). or (U).

   There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. **Key Words:** Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.