A Decision-Theoretic Approach to Subset Selection

by

Jan F. Bjørnstad
Department of Statistics, Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #79-28

December, 1979

(Revised June 1980)

A DECISION-THEORETIC APPROACH TO SUBSET SELECTION

Jan F. Bjørnstad

Purdue University
W. Lafayette, Indiana

ABSTRACT

The multiple decision problem of selecting a random non-empty subset of populations, out of k populations, that are close in some sense to the best population is considered in a decision-theoretic framework. Uniformly optimal procedures for non-negative semi-additive loss are derived. A class of likelihood-ratio type of procedures is shown to be admissible for monotone additive loss.

1. INTRODUCTION

Let $\pi_1,\ldots,\pi_k$ be k populations characterized by $\theta_1,\ldots,\theta_k$ respectively. X_i represents π_i. $X_1,\ldots,X_k$ are assumed to be independent, and X_i has density $f(\cdot,\theta_i)$; $\theta_i \in \Theta \subset \mathbb{R}$, with respect to some σ-finite measure ν. Let $\theta_{[1]} \leq \cdots \leq \theta_{[k]}$ be the ordered θ_i, and let $\pi_{(i)}$ be the population corresponding to $\theta_{[i]}$.

Subset selection is a multiple decision problem concerned with selecting a random non-empty subset of populations that are "close" in some sense to the "best" population, usually defined to be $\pi_{(k)}$. The classical approach, mainly due to Gupta (1965), is to select as few populations as possible while controlling the probability of selecting $\pi_{(k)}$. We shall consider the more general problem of selecting at least t populations, $1 \leq t \leq k-1$, in a decision-theoretic, non-Bayesian framework. Carroll, Gupta and Huang (1975) proposed procedures for this general problem, using the classical approach.

Let $X = (X_1,\ldots,X_k)$, $\theta = (\theta_1,\ldots,\theta_k)$ and $\theta^* = (\theta_{[1]},\ldots,\theta_{[k]})$. The decision-space is denoted by $\mathcal{a}_t = \{a \subset \{1,\ldots,k\}: |a| \geq t\}$. Here $|a|$ is the size of subset a, and decision a is interpreted as selecting the populations π_i for $i \in a$. The loss function is denoted by $\ell(\theta,a)$. A subset selection procedure is a measurable function $\delta: \mathcal{a}_t \times \mathbb{R}^k \to [0,1]$. $\delta(a|x)$ denotes the conditional probability of decision a given that X=x, such that

$$\sum_{a\in\mathcal{a}_t} \delta(a|x) = 1.$$

Let G be the group of permutations g on $\{1,\ldots,k\}$. For $a \in \mathcal{a}_t$, $ga = \{gi: i \in a\}$, and gx, $x \in \mathbb{R}^k$, is defined by $(gx)_i = x_{g^{-1}i}$. It is assumed that ν is permutation-invariant and that the loss function is invariant under G, i.e. $\ell(\theta,a) = \ell(g\theta,ga)$. It follows that the decision problem is invariant under G. δ is by definition a permutation-invariant procedure iff $\delta(ga|gx) = \delta(a|x)$ for all $a \in \mathcal{a}_t$, $x \in \mathbb{R}^k$, $g \in G$. Let $\mathcal{D}_I$ be the class of invariant procedures.

The individual selection probabilities for the procedure δ is given by

$$\Psi_i^\delta(x) = P(\text{selecting } \pi_i|X=x) = \sum_{a\ni i} \delta(a|x).$$

$\Psi^\delta = (\Psi_1^\delta,\ldots,\Psi_k^\delta)$ uniquely determines δ if δ is non-randomized. The risk function of a procedure δ is given by

$$r(\theta|\delta) = \int \sum_{a\in \mathcal{Q}_t} \ell(\theta,a)\delta(a|x)p(x,\theta)d\mu(x)$$

where $p(x,\theta) = \Pi f(x_i,\theta_i)$ and μ is the product-measure of ν on $\mathbb{R}^k$. Since the risk is invariant under G if $\delta \in \mathcal{D}_I$ we have

$$\delta \in \mathcal{D}_I \Rightarrow r(\theta|\delta) = r(\theta^*|\delta) \quad \forall\theta \in \Theta^k. \tag{1.1}$$

We observe that since the decision-space is finite, the conditions of Sion's theorem (see Sion (1958)) are satisfied, and it follows that a minimax procedure always exists.

In Section 2 we consider non-negative, semi-additive loss functions of the form

$$\ell(\theta,a) = \alpha(|a|) \sum_{i\in a} \ell_i(\theta). \tag{1.2}$$

Here $\alpha(|a|) \geq 0$ and $\ell_i(\theta) \geq 0$. It is shown that, under certain conditions, the procedure that selects essentially the t populations corresponding to the t largest X_i's minimizes the risk uniformly in θ for all $\delta \in \mathcal{D}_I$. Since G is a finite group this procedure is also an admissible minimax procedure (see e.g. Ferguson (1967)). This result complements the Bayesian results for the case that t=1 by Deely and Gupta (1968) and Miescke (1979). Miescke (1979) considers the loss (1.2) and shows that under some conditions there exists a Bayes rule that selects exactly one population. Deely and Gupta (1968) considered the case $\ell_i(\theta)=\theta_{[k]}-\theta_i$, and showed a similar result for the Bayes procedure.

The main theorem in Section 2 is proved by employing a certain technique for minimizing $r(\theta|\delta)$, $\delta \in \mathcal{D}_I$, at a fixed point $\theta \in \Theta^k$. This approach is quite general and will now be described for a general loss $\ell(\theta,a)$.

Let $X_{[1]} \leq \cdots \leq X_{[k]}$ be the ordered X_i's and let $Y_i = X_{[i]}$, $i = 1,\ldots,k$. Let $Y = (Y_1,\ldots,Y_k)$, $Y \in \mathcal{Y} = \{y: y_1 \leq \cdots \leq y_k\}$. $G_y = \{g \in G: gy=y\}$, and $m(y)$ is the number of elements in G_y. Define, for fixed θ,

$$h_\theta(a|y) = \sum_{g\in G} \ell(\theta,ga)P_\theta(X=gY|Y=y)/m(y). \tag{1.3}$$

One can regard $h_\theta(a|y)$ as the conditional expected loss of selecting the populations corresponding to Y_i, $i \in a$, given that $Y=y$.

The density of Y is $\bar{p}(y,\theta)/m(y)$ where

$$\bar{p}(y,\theta) = \sum_{g\in G} p(y,g\theta); \quad y \in \mathcal{Y}.$$

Hence, $P_\theta(X=gY|y) = m(y)p(gy,\theta)/\bar{p}(y,\theta)$ and we have that

$$h_\theta(a|y) = \sum_{g\in G} \ell(\theta,ga)p(gy,\theta)/\bar{p}(y,\theta). \tag{1.4}$$

From (1.4) it follows that $h_\theta(a|y) = h_{\theta*}(a|y)$. Let a_0 minimize $h_\theta(a|y)$, i.e.

$$h_\theta(a_0|y) = \min_{a\in G_t} h_\theta(a|y); \quad y \in \mathcal{Y}.$$

LEMMA 1.1. *Let* $\theta \in \Theta^k$ *be fixed. Then* $r(\theta|\delta)$ *is minimized in* $\mathcal{D}_I$ *by* δ_0 *given by*

$$\delta_0(ga_0|y) = 1/n(y) \quad \forall g \in G_y \quad \forall y \in \mathcal{Y}. \tag{1.5}$$

where $n(y)$ *is the number of different decisions in* $\{ga_0: g \in G_y\}$.

Proof. From (1.1) we may assume $\theta_1 \leq \ldots \leq \theta_k$. $\delta \in \mathcal{D}_I$ implies

$$\delta(a|x) = \sum_{g\in G} \delta(ga|gx)I(gx,y)/m(y) \tag{1.6}$$

where $I(a,b) = 1$ if $a=b$ and 0 otherwise. By conditioning on Y we get from (1.6) that

$$r(\theta|\delta) = E_\theta\{R_\theta(Y|\delta)\}$$

where

$$R_\theta(y|\delta) = \sum_{a\in G_t} \delta(a|y)h_\theta(a|y).$$

Now, $h_\theta(a|y) = h_\theta(ga|y)$ if $y=gy$, and it follows that δ_0, defined by (1.5), minimizes $r(\theta|\delta)$ for $\delta \in \mathcal{D}_I$. For $x \in \mathbb{R}^k$, δ_0 is defined by

$$\delta_0(a|x) = \delta_0(ga|y) \quad \text{iff} \quad gx = y \tag{1.7}$$

δ_0 is well-defined by (1.7) and permutation-invariant, because $\delta_0(a|y) = \delta_0(ga|y)$ if $y=gy$. Q.E.D.

Remarks. 1). From (1.4) we see that δ_0 is also a Bayes procedure with respect to the prior that puts equal weight on each different permutation of θ^*, since the posterior risk in this case is proportional to

$$\sum_{g \in G} \ell(\theta^*, ga) p(gx, \theta^*)/\bar{p}(x, \theta^*)$$

(See also Blackwell and Girshick (1954, Theorem 7.3.1.).)

2). Let $\mathcal{Y}^0 = \{y \in \mathcal{Y}: y_1 < \ldots < y_k\}$. δ_0 is non-randomized on $\mathcal{Y}^0$.

In Section 3 we restrict attention to the case t=1. Recently, several papers have considered this problem from a Bayesian point of view. Some contributions, in addition to those already mentioned, are Goel and Rubin (1977), Chernoff and Yahav (1977), Gupta and Hsu (1977, 1978).

Section 3 deals with additive loss functions of the type

$$\ell(\theta, a) = \sum_{i \in a} \ell_i(\theta) \tag{1.8}$$

where $\ell_i(\theta)$ is now allowed to take negative values. Typically $\ell_i(\theta^*) \geq 0$ for $i \leq k-1$ and $\ell_k(\theta^*) < 0$. This is an example of a two-component loss. The addition of any population in the selected subset increases one component and decreases the other component of the loss. It seems clear that two-component loss functions are more realistic than a non-negative loss. We are mostly concerned with admissibility for such additive, two-component loss. Let the non-randomized subset selection procedure $\delta^{a,b}$ be defined by its individual selection probabilities $\Psi^{a,b} = (\Psi_1^{a,b}, \ldots, \Psi_k^{a,b})$ as follows.

$$\Psi_j^{a,b}(x) = 1 \Leftrightarrow x_j = x_{[k]} \text{ or } C_\ell(a,b) \frac{f(x_j, a+b)}{f(x_j, a)} > \sum_{i=1}^{k} \frac{f(x_i, a+b)}{f(x_i, a)} \tag{1.9}$$

$C_\ell(a,b) = \{\ell_1(\theta_{a,b}) - \ell_k(\theta_{a,b})\}/\ell_1(\theta_{a,b})$, where $\theta_{a,b} = (a, \ldots, a, a+b)$.

Let

$$\mathcal{D}_L = \{\Psi^{a,b}: a, a+b \in \Theta \text{ and } b > 0\}. \tag{1.10}$$

It is shown that under certain conditions $\mathcal{D}_L$ is a class of admissible procedures, provided the probability model is continuous.

2. NON-NEGATIVE SEMI-ADDITIVE LOSS

Assume now that the density f possesses the monotone likelihood ratio property. The loss function considered is (1.2), i.e.

$$\ell(\theta,a) = \alpha(|a|)\sum_{i\in a}\ell_i(\theta)$$

where $\alpha(|a|) \geq 0$ and $\ell_i(\theta) \geq 0$ for $i = 1,\ldots,k$. It is assumed that the ℓ_i's are invariant under G,

$$\ell_i(\theta) = \ell_{gi}(g\theta) \quad \forall(g,i,\theta). \tag{2.1}$$

For $u \in \mathbb{R}$, $[u]$ is the integer value of u.

The main result is the following

THEOREM 2.1. *Assume* (i) $[\frac{r}{t}]\alpha(r) \geq \alpha(t) > 0$ *for* $r \geq t$, *and* (ii) $\theta_i \leq \theta_j \Rightarrow \ell_i(\theta) \geq \ell_j(\theta)$. *Then a uniformly best permutation-invariant procedure is given by*

$$\delta_0(\{k-t+1,\ldots,k\}|y) = 1 \quad \textit{if} \quad y_{k-t} < y_{k-t+1} \tag{2.2}$$

$$\delta_0(a_1 \cup \{k-s+2,\ldots,k\}|y) = \binom{q}{t-s+1}^{-1} \tag{2.3}$$

if $y_{k-s+1-q} < y_{k-s+2-q} = \ldots = y_{k-s+1} < y_{k-s+2}$, *and* $1 \leq s \leq t$, $q \geq t-s+1$. *Here* $|a_1| = t-s+1$ *and* $a_1 \subset \{k-s+2-q,\ldots,k-s+1\}$.

Remarks. (1). Theorem 2.1 implies that δ_0 is minimax and admissible in $\mathcal{D}_I$. Hence δ_0 is minimax and admissible among all subset selection procedures, since G is finite.

(2). Theorem 2.1 also holds if a nuisance parameter σ is present, assuming the density f_σ has the monotone likelihood ratio property, for fixed σ.

(3). If $y \in \mathcal{Y}^0$, δ_0 selects the populations corresponding to $y_k, y_{k-1},\ldots,y_{k-t+1}$.

(4). The result of Theorem 2.1 also holds for $\ell(\theta,a) = \alpha(|a|)\sum_{i\in a}\ell_i(\theta)+\gamma(|a|)v(\theta)$, if $\gamma(r) \geq \gamma(t) > 0$ for $r \geq t$, and $v(\theta) = v(g\theta) \geq 0$.

(5). Theorem 2.1 applies to any loss of the form

$$\ell(\theta,a) = \alpha(|a|) \sum_{i\in a} \ell_i(\theta)+h(\theta) \qquad (2.4)$$

provided $\alpha(r)$ and ℓ_i satisfy the conditions in the theorem.

(6). Broström (1979) considers admissibility of subset selection procedures for two normal populations. A similar result to Theorem 2.1 is given for k=2, although the loss considered by Broström is not quite equivalent to the loss considered here.

For t=1, Theorem 2.1 reduces to

COROLLARY 2.1. *Let* t=1. *Assume* (ii) *of Theorem 2.1 holds and that* $r\alpha(r) \geq \alpha(1) > 0$. *Then a uniformly best permutation-invariance procedure is given by*

$$\delta_0(\{k\}|y) = 1 \quad \text{if} \quad y_{k-1} < y_k,$$

$$\delta_0(\{i\}|y) = 1/q \quad \text{if} \quad y_{k-q} < y_{k-q+1} = \ldots = y_k$$

and $i \geq k-q+1$.

Remark. Bahadur (1950) and Bahadur and Goodman (1952) considered the decision-space $\{a \subset \{1,\ldots,k\}: |a| = 1\}$, and showed a similar result, for a particular loss function. This has been generalized by Lehmann (1966) and Eaton (1967) to the problem of selecting exactly t populations.

Examples of loss functions satisfying the conditions in Theorem 2.1 are:

$$\ell(\theta,a) = [\frac{|a|}{t}]^{-1} \sum_{i\in a} (\theta_{[k-t+1]}-\theta_i)^+$$

$$\ell(\theta,a) = \theta_{[k]} - \frac{1}{|a|} \sum_{i\in a} \theta_i; \; t=1.$$

Let, for t=1,

$$\ell(\theta,a)=|a|+ c\sum_{i\notin a} I(\theta_i,\theta_{[k]}), \quad c > 0. \qquad (2.5)$$

This loss was employed by Gupta and Hsu (1977, 1978), and is of the form (2.4) with $\alpha(r) \equiv 1$, $h(\theta) = c \cdot \#\{i: \theta_i = \theta_{[k]}\}$ and

$$\ell_i(\theta) = \begin{cases} 1 & \text{if } \theta_i < \theta_{[k]} \\ 1-c & \text{if } \theta_i = \theta_{[k]}. \end{cases}$$

Hence Theorem 2.1 applies to (2.5) provided $c \leq 1$. It is interesting to note that Gupta and Hsu (1978) showed that in this case the Bayes procedure is also equal to δ_0.

If θ_i is a scale-parameter then, for t=1, a loss function of the following type may be of interest,

$$\ell(\theta,a) = \frac{1}{|a|} \sum_{i \in a} (\theta_{[k]}/\theta_i). \tag{2.6}$$

(2.6) satisfies the conditions in Theorem 2.1.

Deely and Gupta (1968) considered Bayes procedures for the loss $\ell'(\theta,a) = \sum_{i \in a} \alpha'(a,i)\ell_i(\theta)$ and t=1. However, it is easily seen that if $\ell_1,\ldots,\ell_k$ are linearly independent and (2.1) holds, then for ℓ' to be permutation-invariant it is necessary and sufficient that $\alpha'(a,i) = \alpha(|a|)$ $\forall i \in a$, for some function $\alpha: \{1,\ldots,k\} \to \mathbb{R}$.

In order to prove Theorem 2.1 we shall apply Lemma 1.1. We first need some preliminary results.

Let $h_\theta(a|y)$ be given by (1.3), and let

$$G_{ij} = \{g \in G: i = gj\}. \tag{2.7}$$

Then

$$\begin{aligned} h_\theta(a|y) &= \alpha(|a|) \sum_{g \in G} \sum_{i \in ga} \ell_i(\theta) P_\theta(X=gY|y)/m(y) \\ &= \alpha(|a|) \sum_{j \in a} \sum_{g \in G} \ell_{gj}(\theta) P_\theta(X=gY|y)/m(y) \\ &= \alpha(|a|) \sum_{j \in a} \sum_{i=1}^{k} \ell_i(\theta) \sum_{g \in G_{ij}} P_\theta(X=gY|y)/m(y). \end{aligned}$$

Let now

$$T(i,j) = T_{y,\theta}(i,j) = \sum_{g \in G_{ij}} P_\theta(X=gY|y)/m(y).$$

It follows that

$$T(i,j) = \sum_{g \in G_{ij}} p(gy,\theta)/\bar{p}(y,\theta) = \sum_{g \in G_{ji}} p(y,g\theta)/\bar{p}(y,\theta) \tag{2.8}$$

and

$$h_\theta(a|y) = \alpha(|a|) \sum_{j \in a} \sum_{i=1}^{k} \ell_i(\theta) T(i,j) \tag{2.9}$$

Remark. If there are no ties in y then $T(i,j) = P_\theta(X_i = Y_j | y)$. In case of tied observations it can be shown that

$$P_\theta(X_i = Y_j | y) = \sum_{\{\ell :\ y_\ell = y_j\}} T(i,\ell).$$

Define

$$T^c(j) = \sum_{i=1}^{c} T(i,j). \tag{2.10}$$

We have the following result:

LEMMA 2.1. *Assume* $\theta_1 \leq \ldots \leq \theta_k$. *Let* $\ell < m$. *Then*

$$T^c(\ell) \geq T^c(m), \quad \forall c.$$

The proof goes in the exact same way as the proof of Lemma 3.1 in Bickel and Yahav (1977).

Remark. If $y_1 < \ldots < y_k$, then we can define the antiranks $D = (D_1, \ldots, D_k)$ by $X_{D_i} = Y_i$. Lemma 2.1 says in this case that

$$P_\theta(D_\ell \leq c | Y=y) \geq P_\theta(D_m \leq c | Y=y)$$

provided $\theta_1 \leq \ldots \leq \theta_k$ and $\ell < m$.

Let

$$W(a) = \sum_{j \in a} \sum_{i=1}^{k} \ell_i(\theta) T(i,j)$$

so that $h_\theta(a|y) = \alpha(|a|) W(a)$. By using a technique similar to Bickel and Yahav (1977, Theorem 3.1) the following result is readily shown.

LEMMA 2.2. *Let* $\theta_1 \leq \ldots \leq \theta_k$, *and assume that* $\ell_i(\theta) \geq \ell_j(\theta)$ *if* $\theta_i \leq \theta_j$. *Let for any* $a \in \mathcal{G}_t$, $r = r(a) = |a|$. *Then*

$$W(a) \geq W(k-r+1,\ldots,k) = \sum_{j=k-r+1}^{k} \sum_{i=1}^{k} \ell_i(\theta)T(i,j).$$

Proof of Theorem 2.1. Fix $\theta \in \Theta^k$. We shall show that $r(\theta|\delta_0) = \inf_{\delta \in \mathcal{D}_I} r(\theta|\delta)$, and that $\delta_0 \in \mathcal{D}_I$.

From (1.1) it follows that we may assume $\theta_1 \leq \ldots \leq \theta_k$. From Lemma 1.1 it is enough to minimize $h_\theta(a|y)$ given by (2.9). Let $a \in \mathcal{G}_t$, $r = |a|$, then from Lemma 2.2 we see that

$$h_\theta(a|y) \geq \alpha(r)W(k-r+1,\ldots,k).$$

Let $m=[r/t]$, such that $mt \leq r < (m+1)t$.

Now,

$$W(k-r+1,\ldots,k) = \sum_{j=1}^{r} [\sum_{i=1}^{k-1} T^i(k-r+j)(\ell_i(\theta) - \ell_{i+1}(\theta)) + \ell_k(\theta)].$$

It follows that

$$W(k-r+1,\ldots,k) \geq \sum_{q=1}^{m} \sum_{j=1}^{t} [\sum_{i=1}^{k-1} T^i(k-r+(q-1)t+j)(\ell_i(\theta)-\ell_{i+1}(\theta))+\ell_k(\theta)]$$

$k-r+(q-1)t+j \leq k-t+j$. Hence from Lemma 2.1,

$$T^i(k-r+(q-1)t+j) \geq T^i(k-t+j)$$

and therefore

$$W(k-r+1,\ldots,k) \geq [r/t]W(k-t+1,\ldots,k).$$

Hence

$$h_\theta(a|y) \geq [\tfrac{r}{t}]\alpha(r)W(k-t+1,\ldots,k) \geq \alpha(t)W(k-t+1,\ldots,k) = h_\theta(a_0|y)$$

where $a_0=\{k-t+1,\ldots,k\}$. From Lemma 1.1 a $\mathcal{D}_I$-best procedure at θ is given by

$$\delta_0(ga_0|y) = [n(y)]^{-1} \quad \forall g \in G_y = \{g \in G: gy=y\}$$

and $n(y)$ = number of different decisions in $\{ga_0: g \in G_y\}$. Clearly, if $y_{k-t} < y_{k-t+1}$ then $n(y)=1$ and (2.2) follows. If $y_{k-s+1-q} < y_{k-s+2-q} = \ldots = y_{k-s+1} < y_{k-s+2}$, $1 \leq s \leq t$, $q \geq t-s+1$, then $n(y) = \binom{q}{t-s+1}$, and for any $g \in G_y$, $ga_0 = a_1 \cup \{k-s+2,\ldots,k\}$ for $a_1 \subset \{k-s+2-q,\ldots,k-s+1\}$ and $|a_1| = t-s+1$. Hence (2.3) follows. Q.E.D.

3. ADMISSIBILITY FOR ADDITIVE LOSS

3.1. *Introduction and preliminary results*

It is now assumed that t=1, such that the decision-space is $\mathcal{G} = \mathcal{G}_1 = \{a \subset \{1,\ldots,k\}: |a| \geq 1\}$. We consider additive loss of the type (1.8):

$$\ell(\theta,a) = \sum_{i \in a} \ell_i(\theta).$$

(2.1) is assumed to hold such that $\ell(\theta,a)$ is permutation-invariant. We are now mostly interested in the case where $\ell_k(\theta^*) < 0$. The admissibility of the class $\mathcal{D}_L$ given by (1.9) and (1.10), will be studied.

We shall also be concerned with properties like orderedness and unbiasedness. Let us therefore briefly describe these concepts and some related basic facts.

Definition 3.1. Let $\mathcal{G}_0 = \{a \in \mathcal{G}: a=(r,r+1,\ldots,k);\ r=1,\ldots,k\}$, and let

$$\mathcal{G}_{o,y} = \bigcup_{g \in G_y} g\,\mathcal{G}_o, \quad \text{for } y \in \mathcal{Y},$$

where $G_y = \{g: gy=y\}$. Then δ is said to be an ordered procedure if

$$\sum_{a \in g_1 \mathcal{G}_{o,y}} \delta(a|x)=1, \text{ when } x = g_1 y;\ y \in \mathcal{Y}.$$

Let $\mathcal{D}_0$ denote the class of ordered procedures. Consider the statement (A): $x_i < x_j \Rightarrow \Psi_i^\delta(x) \leq \Psi_j^\delta(x)$. It is readily shown that if $\delta \in \mathcal{D}_0$ then (A) holds, but (A) does not necessarily imply that $\delta \in \mathcal{D}_0$. To see this, let k = 2 and δ be defined as follows. If $x_1 < x_2$, $\delta(\{1\}|x) = .4$ and $\delta(\{2\}|x) = .6$. If $x_1 > x_2$, $\delta(\{1\}|x)=.6$ and $\delta(\{2\}|x) = .4$.

However, if δ is non-randomized we have that δ is ordered if and only if (A) holds.

We are mostly concerned with the class $\mathcal{D}_I$ of permutation-invariant procedures. The following result follows directly from Definition 3.1.

LEMMA 3.1. *Let* $\delta \in \mathcal{D}_I$. *Then*

$$\delta \in \mathcal{D}_0 \Leftrightarrow \sum_{a \in \mathcal{A}_{0,y}} \delta(a|y)=1 \quad \forall y \in \mathcal{Y}.$$

Definition 3.2. δ is said to be an unbiased procedure if $\theta_j \leq \theta_i$ implies that $E_\theta(\Psi_j^\delta) \leq E_\theta(\Psi_i^\delta)$.

Definition 3.3. δ is a monotone procedure if

$x_j \leq x_j'$, and $x_i \geq x_i'$ for $\forall i \neq j \Rightarrow \Psi_j^\delta(x) \leq \Psi_j^\delta(x')$.

The properties of orderedness and monotonicity of procedures are not really related, since orderedness (for nonrandomized procedures) concerns different Ψ_i's for the same x, while monotonicity relates to the same ψ_i for different sets of observations. So a procedure can be monotone and not ordered or vice versa. E.g. let δ be given by

$$\Psi_i^\delta = 1 \quad \text{iff} \quad x_i \geq \min(x_{[k]}, x_{[k]}/\bar{x}), \quad \bar{x} = \Sigma x_i/k.$$

Then δ is ordered but not monotone. However, if δ is nonrandomized, permutation-invariant and monotone, then δ is also ordered. (See Bjørnstad (1979) for proof.)

From Nagel (1970) we have the following result.

LEMMA 3.2. *Assume* f *has the monotone likelihood ratio property. Let* δ *be a permutation-invariant and monotone procedure. Then* δ *is unbiased.*

We say that δ_0 is a $\mathcal{D}_I$-best procedure if there exists $\theta \in \Theta^k$ such that $r(\theta|\delta_0)$ minimizes $r(\theta|\delta)$, $\forall \delta \in \mathcal{D}_I$. As mentioned in Section 1, a $\mathcal{D}_I$-best procedure δ_0 is also a Bayes procedure, so that δ_0 is admissible in $\mathcal{D}_I$ if it is unique Bayes in $\mathcal{D}_I$. This is the technique we will apply to show admissibility of a procedure. We state this result as a lemma to be able to refer to it later.

LEMMA 3.3. *Let* δ_0 *be a best procedure in* $\mathcal{D}_I$ *at* θ_0. Assume $\delta \in \mathcal{D}_I$ *and that* $r(\theta_0|\delta) = r(\theta_0|\delta_0)$ *implies* $r(\theta|\delta) = r(\theta|\delta_0)$, $\forall \theta \in \Theta^k$. *Then* δ_0 *is admissible in* $\mathcal{D}_I$ *and hence among all procedures.*

3.2. *General admissibility-theory for $\mathcal{D}_I$-best procedures*

We shall first consider how to minimize $r(\theta|\delta)$, $\delta \in \mathcal{D}_I$ at a fixed point $\theta \in \Theta^k$. The approach given by Lemma 1.1 can still be used. However, since we are now dealing with additive loss it is more convenient to employ a technique similar to the one used by Studden (1967). Let $p(x,\theta) = \Pi f(x_i,\theta_i)$, and let

$$P_{ji}(x,\theta) = \sum_{g \in G_{ji}} p(x,g\theta).$$

We see that for $y \in \mathcal{Y}$, $P_{ji}(y,\theta) = \bar{p}(y,\theta)T(i,j)$. G_{ji} is given by (2.7). Let

$$T_j(x,\theta) = \sum_{i=1}^{k} \ell_i(\theta)P_{ji}(x,\theta).$$

Then T_j is permutation-invariant in the following way.

$$T_j(x,\theta)=T_{gj}(gx,\theta) \text{ and } T_j(x,\theta)=T_j(x,g\theta),\ \forall g \in G. \tag{3.1}$$

For the additive loss we consider, $r(\theta|\delta) = \sum \ell_i(\theta)E_\theta(\Psi_i^\delta)$. Hence two procedures are equivalent if their individual selection probabilities are identical. Therefore, from now on a procedure δ will only be defined by its selection probabilities $\Psi=(\Psi_1,\ldots,\Psi_k)$. Let $T_{[1]} \leq \cdots \leq T_{[k]}$ be the ordered T_j's and let $I = \{i: T_i=T_{[1]}\}$. Then the $\mathcal{D}_I$-best procedure at θ is given in the following result.

THEOREM 3.1. *Let $\theta \in \Theta^k$ be fixed. Define Ψ^0 by:*

$$\Psi_j^0 = \begin{cases} 1 & \text{if } T_j < 0 \\ q^{-1} & \text{if } T_{[1]} \geq 0 \text{ and } j \in I; \quad q = |I| \\ 0 & \text{if } T_j \geq 0 \text{ and } T_j > T_{[1]} \end{cases}$$

Then Ψ^0 is a best procedure in $\mathcal{D}_I$ at θ.

Proof. From (3.1) it follows that $\Psi^0 \in \mathcal{D}_I$. Let now $\Psi \in \mathcal{D}_I$. Then it is readily seen that

$$k!r(\theta|\Psi) = \sum_{j=1}^{k} \int \Psi_j(x)T_j(x,\theta)d\mu(x).$$

It is therefore enough to show

$$\sum_{j=1}^{k}[\Psi_j(x)-\Psi_j^0(x)]T_j(x,\theta) \geq 0 \quad \text{for all } x. \tag{3.2}$$

If $T_{[1]} < 0$ then (3.2) is obvious. Assume now that $T_{[1]} \geq 0$. Then $\Sigma\Psi_j^0 T_j = T_{[1]}$, while $\Sigma\Psi_j T_j \geq T_{[1]}\Sigma\Psi_j \geq T_{[1]}$ and (3.2) follows.

Q.E.D.

Remark. For the decision-space G_t, $t > 1$, a similar result holds. Let Ψ^t be defined by:

(i) If $T_{[t]} < 0$ then $\Psi_j^t = 1$ if $T_j < 0$ and $\Psi_j^t = 0$ if $T_j \geq 0$.

(ii) If $T_{[t]} \geq 0$ and $T_{[s-1]} < T_{[s]} = \ldots = T_{[s+q-1]} < T_{[s+q]}$, $s \leq t \leq s+q-1$, then $\Psi_j^t=1$ if $T_j \leq T_{[s-1]}$, $\Psi_j^t=(t-s+1)/q$ if $T_j=T_{[s]}$, and $\Psi_j^t=0$ if $T_j \geq T_{[s+q]}$.

Then it can be shown that Ψ^t is a best procedure in $\mathcal{D}_I$ at θ, with respect to G_t.

In order to apply Lemma 3.3 in admissibility-considerations, we must show that Ψ^0, given by Theorem 3.1, is essentially unique. Let P_θ be the probability-measure defined by the joint density $p(x,\theta)$. A sufficient condition for uniqueness is given in the following result.

THEOREM 3.2. *Let $\theta \in \Theta^k$ be fixed. Assume*

$$P_\theta(T_j=0)=0 \quad \textit{for} \quad j=1,\ldots,k. \tag{3.3}$$

Then Ψ^0 given by Theorem 3.1 *is a.s. (P_θ)-unique in $\mathcal{D}_I$, i.e. $\Psi \in \mathcal{D}_I$ and* $r(\theta|\Psi)=r(\theta|\Psi^0)$ *implies* $\Psi=\Psi^0$ *a.s.* (P_θ).

Proof. Let $\Psi \in \mathcal{D}_I$ and assume $r(\theta|\Psi)=r(\theta|\Psi^0)$. Let $X_- = \{x: T_{[1]}(x,\theta) < 0 \text{ and } T_j(x,\theta) \neq 0 \text{ for } j=1,\ldots,k\}$, and let $X_+ = \{x: T_{[1]}(x,\theta) > 0\}$. It then follows from (3.3) that

$$\int_{X_-}\sum_{j=1}^{k}(\Psi_j-\Psi_j^0)T_j d\mu + \int_{X_+}\sum_{j=1}^{k}(\Psi_j-\Psi_j^0)T_j d\mu = 0. \tag{3.4}$$

From (3.2), this implies that $(\Psi_j-\Psi_j^o)T_j = 0$ a.e. (μ) on X_- for $j = 1,\ldots,k$ since $\{\Psi_j-\Psi_j^o\}T_j \geq 0$ on X_- for all j. Hence $\Psi = \Psi^o$ a.e. (μ) on X_-. It remains to show that $\Psi = \Psi^o$ a.e. (μ) on X_+. From (3.4) and (3.2) it follows that $\Psi_j = 0$ if $T_j > T_{[1]}$ a.e. (μ) on X_+, and also that

$$\sum_{\{j:T_j=T_{[1]}\}} \Psi_j=1 \qquad \text{a.e. } (\mu) \text{ on } X_+.$$

Since $\Psi \in \mathcal{D}_I$ this shows that $\Psi_j = \Psi_j^o$ if $T_j = T_{[1]}$ a.e. (μ) on X_+.

Q.E.D.

Definition 3.4. A family $\mathcal{P}$ of probability-measures is called *homogeneous* if $P(A) = 0$ implies $P'(A) = 0$ for all $P, P' \in \mathcal{P}$.

COROLLARY 3.1. *Let* $\theta_o \in \Theta^k$ *be fixed. Assume that* $\{P_\theta: \theta\in\Theta^k\}$ *is homogeneous and that* (3.3) *holds. Let* Ψ^o *be the best procedure in* $\mathcal{D}_I$ *at* θ_o. *Then* Ψ^o *is admissible.*

3.3. *Monotone additive loss*

It is now assumed that f possesses the strict monotone likelihood ratio property, i.e. $f(u,\theta_i')/f(u,\theta_i)$ is strict increasing in $u \in \mathbb{R}$ for all $\theta_i < \theta_i'$. The additive loss is now monotone, in the sense that if $\theta_i \leq \theta_j$ then $\ell_i(\theta) \geq \ell_j(\theta)$. We shall also assume that $\ell_1(\theta^*) > \ell_k(\theta^*)$ if $\theta_{[1]} < \theta_{[k]}$.

From (3.3) we see that $T_j(y,\theta)$, $j=1,\ldots,k$ and $y \in \mathcal{Y}$ uniquely determines $T_j(x,\theta)$ for all j, x. It is therefore enough to consider $T_j(y,\theta)$ for $y \in \mathcal{Y}$. Let now

$$U_j = \sum_{i=1}^{k} \ell_i(\theta)T(i,j),$$

so $h_\theta(a|y) = \alpha(|a|)\sum_{j\in a} U_j$, and $\alpha(|a|)U_j$ can be regarded as the conditional expected loss of selecting the population corresponding to Y_j given that Y=y. It now follows from (2.8) that if $\bar{p}(y,\theta) > 0$ then

$$T_j = \bar{p}\, U_j; \quad j = 1,\ldots,k. \tag{3.5}$$

Let

$$M_\theta = \{x\colon 0 < f(x_i,\theta_j) < \infty \text{ for all } (i,j)\}$$

Then we can strengthen Lemma 2.1 to

LEMMA 3.4. Assume $\theta_1 \leq \ldots \leq \theta_k$ and $\theta_1 < \theta_k$. Let $\ell < m$, and $y \in M_\theta \cap \mathcal{Y}$. Then, provided $y_\ell < y_m$,

$$\sum_{i=1}^{c} T(i,\ell) > \sum_{i=1}^{c} T(i,m) \qquad c \leq k-1.$$

LEMMA 3.5. *Assume* $\theta_{[1]} < \theta_{[k]}$, $y \in M_\theta$ *and* $y_j < y_{j+1}$ for $j = 1,2,\ldots,k-1$. *Then* $T_j(y,\theta) < T_{j-1}(y,\theta)$ for $j = 2,3,\ldots,k$.

Proof. From (3.1) it follows that $T_j(y,\theta) = T_j(y,\theta^*)$, $\forall j$. Hence we may assume $\theta_1 \leq \ldots \leq \theta_k$. From (3.5) it is enough to show that $U_j < U_{j-1}$. Now

$$U_j = \sum_{i=1}^{k-1} T^i(j)\{\ell_i(\theta)-\ell_{i+1}(\theta)\} + \ell_k(\theta). \tag{3.6}$$

Result now follows from Lemma 3.4 and the fact that $\ell_i(\theta)<\ell_{i+1}(\theta)$ for at least one i. Q.E.D.

We can apply Lemma 3.5 to show that the $\mathcal{D}_I$-best procedures are ordered.

LEMMA 3.6. *Let* $\theta \in \Theta^k$ *be fixed. Assume* $P_\theta(Y \in M_\theta) = 1$ *and* $\theta_{[1]} < \theta_{[k]}$. *Then there exists a version of the* $\mathcal{D}_I$*-best procedure at* θ*, given by Theorem 3.1., that is ordered.*

Proof. We may assume that $y \in M_\theta$, $y_1 \leq \ldots \leq y_k$. Let first $T_k(y) < 0$. Then the $\mathcal{D}_I$-best procedure δ_0 is given by $\delta_0(a_1|y) = 1$, where $a_1 = \{r,r+1,\ldots,k\}$ and $T_r(y) < 0$, $T_{r-1}(y) \geq 0$.

Now $a_1 \in \mathcal{G}_0 \subset \mathcal{G}_{0,y}$. So Lemma 3.1 holds. Consider next the case $T_k(y) \geq 0$ and $y_k = \ldots = y_p > y_{p-1}$. Then $T_k(y) = \ldots = T_p(y)$, from (3.1). Also $T_{p-1}(y) > T_p(y)$ from Lemma 3.5. Consider the following version of ψ^ρ:

$$\delta_o(\{j\}|y) = 1/(k-p+1) \text{ for } j=p, p+1,\ldots,k.$$

Then $\{j\} \in G_{o,y}$ for $j \geq p$, and it follows that

$$\sum_{a\in G_{o,y}} \delta_o(a|y) = 1.$$

From Lemma 3.1, δ_o is ordered. Q.E.D.

We note that $P_\theta(\bar{p}(Y,\theta) < \infty) = 1$. Hence, $P_\theta(Y \in M_\theta) = 1$ if and only if $P_\theta(p(Y,g\theta) > 0) = 1$ for all $g \in G$. If $\{P_\theta: \theta \in \Theta^k\}$ is homogeneous then $P_{g\theta}(p(X,\theta) = 0)$, for all $g \in G$ and therefore $P_\theta(p(Y,g\theta) = 0) = 0$, for all $g \in G$. It follows that $P_\theta(Y \in M_\theta) = 1$ if $\{P_\theta: \theta \in \Theta^k\}$ is homogeneous.

Define the quantity

$$H_j^i(y,\theta) = 1 + \sum_{b=i+1}^{k} \left[\sum_{a=1}^{i} \frac{f(y_j,\theta_a)}{f(y_j,\theta_b)} \left\{ \frac{\sum_{g\in G_{ja}} \prod_{p\neq j} f(y_p,\theta_{g^{-1}p})}{\sum_{g\in G_{jb}} \prod_{q\neq j} f(y_q,\theta_{g^{-1}q})} \right\} \right]^{-1}. \tag{3.7}$$

We have that

$$T^i(j) = \sum_{a=1}^{i} T(a,j) = \frac{\sum_{a=1}^{i} \sum_{g\in G_{ja}} p(y,g\theta)}{\sum_{b=1}^{k} \sum_{g\in G_{jb}} p(y,g\theta)}$$

$$= \left[1 + \sum_{b=i+1}^{k} \left\{ \sum_{a=1}^{i} \frac{\sum_{g\in G_{ja}} p(y,g\theta)}{\sum_{g\in G_{jb}} p(y,g\theta)} \right\}^{-1} \right]^{-1}.$$

Hence,

$$T^i(j) = [H_j^i(y,\theta)]^{-1}. \tag{3.8}$$

Now, if $\theta_1 \leq \ldots \leq \theta_k$ and $\theta_1 < \theta_k$, then $f(y_j,\theta_a)/f(y_j,\theta_b)$ is non-increasing in y_j $\forall (a \leq i,\ b \geq i+1)$ and for a=1, b=k we get that $f(y_j,\theta_1)/f(y_j,\theta_k)$ is strict decreasing in y_j. It follows from (3.8) that $T^i(j)$ is strict decreasing in y_j, if $y \in M_\theta$ and from

(3.6) we then have that U_j is strict decreasing in y_j for fixed $\{y_\ell: \ell \neq j\}$, if $y \in M_\theta$.

This enables us to give necessary conditions for (3.3) to hold.

LEMMA 3.7. *Let* ν *be the Lebesgue-measure, and assume* $\theta_{[1]} < \theta_{[k]}$. *If* $P_\theta(Y \in M_\theta) = 1$, *then* $P_\theta(T_j=0) = 0$, $\forall j$.

Proof. We can assume $\theta_1 \leq \ldots \leq \theta_k$. Since $P_\theta\{\bar{p}(Y,\theta)=0\} = 0$, it is enough to show that $P_\theta(U_j=0) = 0$. We only need to consider $y \in M_\theta$. Let $y^{(j)} = (y_1,\ldots,y_{j-1},y_{j+1},\ldots,y_k)$. For given $y^{(j)}$ there exists at most one value of y_j for which $U_j = 0$. Let $\Delta(y^{(j)})$ be this value. Define $\Delta(x)$, $\forall x \in \mathbb{R}^{k-1}$, be letting $\Delta(y^{(j)}) = 0$ if no such value exists, and $\Delta(x) = 0$ for all $x \notin \{v: v_1 \leq \ldots \leq v_{k-1}\}$. Then

$$P_\theta(U_j=0) \leq \sum_{i=1}^{k} \sum_{g \in G_i} P_\theta\{X_i=\Delta(gX^{(i)})\},$$

where G_i is the group of permutations on $X^{(i)}$.
Since ν is Lebesgue-measure, X_i is a continuous random variable. Also we have that X_i and $\Delta(gX^{(i)})$ are independent. It follows that $P_\theta\{X_i = \Delta(gX^{(i)})\} = 0$; $\forall i$, $\forall g \in G_i$. Q.E.D.

We can now state the main admissibility result.

THEOREM 3.3. *Assumptions:* A1) $\{P_\theta: \theta \in \Theta^k\}$ *is homogeneous.* A2) ν *is Lebesgue-measure.* A3) $\theta \in \Theta^k$ *such that* $\theta_{[1]} < \theta_{[k]}$. *Let* Ψ^0 *be the best procedure in* $\mathcal{D}_I$ *at* θ. *Then* Ψ^0 *is admissible.*

Proof. We shall apply Corollary 3.1. Hence, we only have to show that (3.3) holds. This follows from Lemma 3.7, since A1) implies that $P_\theta(Y \in M_\theta) = 1$. Q.E.D.

Remarks. 1) When ν is the Lebesgue-measure Ψ^0 can be chosen non-randomized as follows. If $T_{[1]}(x,\theta) < 0$, $\Psi_j^0(x) = 1$ iff $T_j(x,\theta) < 0$. If $T_{[1]}(x,\theta) \geq 0$, $\Psi_j^0(x) = 1$ iff $x_j = x_{[k]}$.

2) If $\ell_k(\theta^*) \geq 0$, the loss is a special case of the loss considered in Section 2. If $\ell_1(\theta^*) < 0$, then $T_j(x) < 0$, $\forall j$, and $\Psi_j^0(x) = 1$, $\forall j$. Therefore the interesting case is when $\ell_1(\theta^*) \geq 0$ and $\ell_k(\theta^*) < 0$. Examples of such loss functions are considered in Section 3.4.

To summarize the results in this section, we have shown that if A1), A2), A3) hold then the $\mathcal{D}_I$-best procedures are admissible and if A1), A3) hold then the $\mathcal{D}_I$-best procedures are ordered.

3.4. *A class of admissible procedures*

We make the same assumptions on f and ℓ as in Section 3.3. We shall now consider the class $\mathcal{D}_L$ defined by (1.9) and (1.10).

THEOREM 3.4. *Assume* A1), A2) *hold. Then* $\mathcal{D}_L$ *is a class of admissible procedures.*

Proof. Assume $\theta = (a,\dots,a, a+b)$, $b > 0$. Then $\ell_1(\theta) = \dots = \ell_{k-1}(\theta)$, and

$$T_j(x,\theta) < 0 \Leftrightarrow P_{jk}(x,\theta)/\bar{p}(x,\theta) > C_\ell^{-1}(a,b).$$

In this slippage-configuration

$$P_{jk}(x,\theta)/\bar{p}(x,\theta) = \left\{\frac{f(x_j,a+b)}{f(x_j,a)}\right\} / \sum_{\ell=1}^{k} \frac{f(x_\ell,a+b)}{f(x_\ell,a)}$$

From Theorem 3.1 it follows that $\Psi^{a,b}$, given by (1.9), is the $\mathcal{D}_I$-best procedure at θ, and hence from Theorem 3.3, $\Psi^{a,b}$ is admissible. Q.E.D.

Since f has the monotone likelihood ratio property we see that the procedures in $\mathcal{D}_L$ are monotone and hence, from Lemma 3.2, unbiased, i.e.: $\theta_j \leq \theta_i \Rightarrow E_\theta(\Psi_j^{a,b}) \leq E_\theta(\Psi_i^{a,b})$.

Applications.

Let ν the Lebesgue measure and consider the exponential family of distributions given by the density

$$f(x_i,\theta_i) = A(\theta_i)h(x_i)e^{\theta_i x_i}. \tag{3.9}$$

This family is homogeneous, such that the assumptions of Theorem 3.4 hold. The procedures $\Psi^{a,b}$ in $\mathcal{D}_L$ are now given by

$$\Psi_j^{a,b}=1 \text{ iff } C_\ell(a,b)e^{bX_j} > \sum_{i=1}^{k} e^{bX_i} \text{ or } X_j=X_{[k]}. \tag{3.10}$$

Therefore the exponential procedures defined by (3.10) are admissible, and also ordered and unbiased. Examples of selection problems with continuous exponential distributions of the form (3.9) are (i) selecting the largest normal mean with known variance, and (ii) selecting the smallest scale parameter for gamma distributed populations.

Bjørnstad (1978) showed that the exponential class (3.10) has certain minimax properties for the normal case, with C_ℓ and b now determined to satisfy a certain control condition.

If V_i has the gamma density $\{\Gamma(p)\eta_i\}^{-1}(v_i/\eta_i)^{p-1}e^{-v_i/\eta_i}$, then we let $\theta_i=1/\eta_i$ and $X_i = -V_i$ in (3.9). It follows from (3.10) that

$$\Psi_j^{a,b}=1 \text{ iff } e^{bV_j} < C_\ell(a,b)(\sum_{i=1}^{k} e^{-bV_i})^{-1} \text{ or } V_j=V_{[1]}.$$

It is clear that the admissibility result, Theorem 3.3, holds for any loss function of the form

$$\ell(\theta,a) = \sum_{i\in a} \ell_i'(\theta) + h(\theta) \tag{3.11}$$

where the ℓ_i''s satisfy the conditions in the theorem. We shall now consider some loss functions that are equivalent to monotone additive loss functions, i.e. of the from (3.11).

$$\ell_1(\theta,a) = \sum_{i\in a}(\theta_{[k]}-\theta_i)-\varepsilon|a|,\ \varepsilon > 0.$$

This loss is also considered by Miescke (1979). Here, $C_\ell(a,b) = b/(b-\varepsilon)$.

$$\ell_2(\theta,a) = \sum_{i\in a}(\theta_{[k]}-\theta_i) +\alpha\sum_{i\notin a} I(\theta_i,\theta_{[k]});\ \alpha > 0.$$

Here, $I(u,v) = 1$ if $u=v$ and 0 otherwise. We see that ℓ_2 is of the form (3.11) with $h(\theta) = \alpha\cdot\#\{i: \theta_i = \theta_{[k]}\}$, and

$$\ell_i'(\theta) = \begin{cases} \theta_{[k]}-\theta_i & \text{if } \theta_i < \theta_{[k]} \\ -\alpha & \text{if } \theta_i = \theta_{[k]}. \end{cases}$$

ℓ_2 is similar to the loss considered by Bickel and Yahav (1977), and implies that $C_\ell(a,b) = (b+\alpha)/b$.

$$\ell_3(\theta,a) = \sum_{i\in a} \ell'(\theta_i-\theta_{[k]}+\Delta) + \sum_{i\notin a} \ell''(\theta_i-\theta_{[k]}+\Delta),$$

where ℓ' is non-increasing, ℓ'' is non-decreasing, $\ell'(u) = 0$ if $u > 0$ and $\ell''(u) = 0$ if $u \leq 0$. ℓ_3 is of the form (3.11) with $\ell_i'(\theta) = \ell'(\theta_i-\theta_{[k]}+\Delta) - \ell''(\theta_i-\theta_{[k]}+\Delta)$ and $h(\theta) = \sum_i \ell''(\theta_i-\theta_{[k]}+\Delta)$. ℓ_3 is an appropriate loss when it is of interest to select populations with $\theta_i > \theta_{[k]}-\Delta$, for some fixed $\Delta > 0$. Here, $C_\ell(a,b) = \{\ell'(\Delta-a)-\ell''(\Delta-a)+\ell''(\Delta)\}/\{\ell'(\Delta-a)-\ell''(\Delta-a)\}$. Kim (1979) considered this loss.

The loss given by (2.5) is also equivalent with a monotone additive loss. This loss gives $C_\ell(a,b) = c$. Deely and Gupta (1968), Berger (1979) and Gupta and Miescke (1980) deal with aspects of the problem of minimizing the expected subset size subject to controlling the probability of selecting $\pi_{(k)}$. In particular, Gupta and Miescke (1980) consider the class of procedures defined by Seal (1955) and show that Gupta's maximum-type procedure is optimal in Seal's class for three normal populations, when $\theta_{[k]}-\theta_{[1]}$ is sufficiently large.

From Theorem 3.1 it follows that when the two components $|a|$ and $\sum_{i\notin a} I(\theta_i,\theta_{[k]})$ are considered simultaneously, then for any θ with $\theta_{[1]} < \theta_{[k]}$, no procedure in Seal's class is a $\mathcal{D}_I$-best procedure.

ACKNOWLEDGEMENT

This research was supported by the National Science Foundation Grant No. MCS-7903780.

The author would like to thank two referees for helpful comments.

BIBLIOGRAPHY

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

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

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

Bickel, P. J. & Yahav, J. A. (1977). On selecting a set of good populations. *Statistical Decision Theory and Related Topics II.* (Ed. S. S. Gupta and D. S. Moore.) Academic Press, New York, 37-55.

Bjørnstad, J. F. (1978). The subset selection problem. II. On the optimality of some subset selection procedures. *Mimeograph Series #78-27,* Dept. of Statistics, Purdue University.

Bjørnstad, J. F. (1979). On the probability of correct selection in the subset selection problem. *Mimeograph Series #79-32,* Dept. of Statistics, Purdue University.

Blackwell, D. and Girshick, M. A. (1954). *Theory of Games and Statistical Decisions.* Wiley, New York.

Broström, G. (1979). On the admissibility of subset selection procedures. *Technical report,* Institute of Mathematical Statistics, University of Umea, Sweden.

Carroll, R. J., Gupta, S. S. and Huang, D. Y. (1975). On selection procedures for the t best populations and some related problems. *Comm. Statist.* 4, 987-1008.

Chernoff, H. and Yahav, J. A. (1977). A subset selection problem employing a new criterion. *Statistical Decision Theory and Related Topics II.* (Ed. S. S. Gupta and D. S. Moore.) Academic Press, New York, 93-119.

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

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

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

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

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

Gupta, S. S. and Hsu, J. C. (1977). On the monotonicity of Bayes subset selection procedures. *Proceedings of the 41st session of the I.S.T.*, New Delhi.

Gupta, S. S. and Hsu, J. C. (1978). On the performance of some subset selection procedures. *Comm. Statist.* B7(6), 561-591.

Gupta, S. S. and Miescke, K. J. (1980). Optimality of subset selection procedures for ranking means of three normal populations. (To appear *Sankhyā*.)

Kim, W. C. (1979). On Bayes and gamma-minimax subset selection rules. *Mimeograph Series #79-7*, Dept. of Statistics, Purdue University.

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

Miescke, K. J. (1979). Bayesian subset selection for additive and linear loss functions. *Comm. Statist.* A8(12), 1205-1226.

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

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

Sion, M. (1958). On general minimax theorems. *Pacific J. Math.* 8, 171-176.

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