The Annals of Statistics
1984, Vol. 12, No. 3, 1058–1070

A GENERAL THEORY OF ASYMPTOTIC CONSISTENCY FOR SUBSET SELECTION WITH APPLICATIONS[1]

BY JAN F. BJØRNSTAD

The University of Texas at Austin

The problem of selecting a random nonempty subset from k populations, characterized by $\theta_1, \cdots, \theta_k$ with possible nuisance parameters σ, is considered using a decision-theoretic approach. The concept of asymptotic consistency is defined as the property that the risk of a procedure at $(\boldsymbol{\theta}, \sigma)$ tends to the minimum loss at $(\boldsymbol{\theta}, \sigma)$. Necessary and sufficient conditions for both pointwise and uniform (on compact sets) consistency for permutation-invariant procedures are derived with general loss functions.

Various loss functions when the goal is to select populations with θ_i close to max θ_j are considered. Applications are made to normal populations. It is shown that Gupta's procedure is the only procedure in Seal's class that can be consistent. Other Bayes and admissible procedures are also considered.

1. Introduction. The multiple decision problem of selecting a random nonempty subset from k populations $\pi_1, \cdots, \pi_k$ is considered. $\pi_1, \cdots, \pi_k$ are characterized by $\theta_1, \cdots, \theta_k$ respectively, where $\theta_i \in \Theta \subset \mathbb{R}$ and the parameter-space of $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)$ is $\Omega \subset \Theta^k$. X_i^n is an estimate of θ_i, based on n observations. We shall allow for the presence of nuisance-parameters, denoted by σ with parameter-space Σ. σ is estimated by $S^n \in E$. The joint distribution function of $(\mathbf{X}^n, S^n)$ is denoted by $F_{\boldsymbol{\theta},\sigma}^n$. Lew now G be the group of permutations g on $\{1, \cdots, k\}$. For $\mathbf{x} \in \mathbb{R}^k$, $g\mathbf{x}$ is defined by $(g\mathbf{x})_i = x_{g^{-1}i}$. For any subset A of $\mathbb{R}^k$, $gA = \{g\mathbf{x}: \mathbf{x} \in A\}$. The probability model is assumed to be invariant under G, i.e. (a) if $(\mathbf{X}^n, S^n)$ has cdf $F_{\boldsymbol{\theta},\sigma}^n$ then $(g\mathbf{X}^n, S^n)$ has cdf $F_{g\boldsymbol{\theta},\sigma}^n$, and (b) $g\Omega = \Omega$, $\forall g \in G$.

The decision-space is $\mathscr{A} = \{a \subset \{1, \cdots, k\}\}$, where the decision a is interpreted as selecting the populations π_i, $i \in a$. For $a \in \mathscr{A}$, $ga = \{gi: i \in a\}$. The loss-function $\ell_\sigma(\boldsymbol{\theta}, a)$ is assumed to be permutation-invariant, i.e. $\ell_\sigma(\boldsymbol{\theta}, a) = \ell_\sigma(g\boldsymbol{\theta}, ga)$ for all $g \in G$. It follows that the multiple decision problem is invariant under G. Furthermore, $-\infty < \ell_\sigma(\boldsymbol{\theta}, a) < \infty$, $\forall a \in \mathscr{A}$, $\forall (\boldsymbol{\theta}, \sigma) \in \Omega \times \Sigma$. A subset selection procedure is given by:

$$\delta_n(a \mid \mathbf{x}, s) = \Pr\{\text{decision } a \mid \mathbf{X}^n = \mathbf{x}, S^n = s\}.$$

We shall consider the class of invariant procedures, $\mathscr{D}_I$, where $\delta_n \in \mathscr{D}_I$ if and only if $\delta_n(ga \mid g\mathbf{x}, s) = \delta_n(a \mid \mathbf{x}, s)$ for all $a \in \mathscr{A}$, $\mathbf{x} \in \mathbb{R}^k$, $s \in E$, $g \in G$. The risk-function of δ_n is $r_n(\boldsymbol{\theta}, \sigma \mid \delta_n) = \sum_{a \in \mathscr{A}} \ell_\sigma(\boldsymbol{\theta}, a) E_{\boldsymbol{\theta},\sigma} \delta_n(a \mid \mathbf{X}^n, S^n)$.

Received October 1982; revised March 1984.

[1] This research was supported in part by the National Science Foundation Grant No. MCS-8303620.

AMS 1980 *subject classifications*. Primary 62F07; secondary 62C99.

Key words and phrases. Subset selection, asymptotic theory, decision-theory, invariance, consistency.

The purpose of this paper is to develop a theory of asymptotic consistency for different loss functions in this multiple decision problem. To define the term consistency, let $m_\sigma(\boldsymbol{\theta}) = \min_{a \in \mathcal{A}} \ell_\sigma(\boldsymbol{\theta}, a)$. If $\ell_\sigma(\boldsymbol{\theta}, a_0) = m_\sigma(\boldsymbol{\theta})$ then a_0 is a correct decision when $(\boldsymbol{\theta}, \sigma)$ is true. Obviously, $r_n(\boldsymbol{\theta}, \sigma \mid \delta_n) \geq m_\sigma(\boldsymbol{\theta})$, $\forall(n, \boldsymbol{\theta}, \sigma)$. All limits in this paper are as $n \to \infty$.

DEFINITION 1.1. The sequence of procedures $\{\delta_n\}$ is consistent at $(\boldsymbol{\theta}, \sigma)$ if $r_n(\boldsymbol{\theta}, \sigma \mid \delta_n) \to m_\sigma(\boldsymbol{\theta})$.

We say that δ_n is pointwise consistent on $\Omega \times \Sigma$ if δ_n is consistent at each $(\boldsymbol{\theta}, \sigma) \in \Omega \times \Sigma$.

We shall also consider the concept of uniform consistency. We note that the metric on $\Omega \times \Sigma$ is the usual Euclidean distance.

DEFINITION 1.2. The sequence $\{\delta_n\}$ is uniformly consistent if

$$\sup_{\boldsymbol{\theta} \in K_1, \sigma \in K_2} \{r_n(\boldsymbol{\theta}, \sigma \mid \delta_n) - m_\sigma(\boldsymbol{\theta})\} \to 0 \quad \text{for all compact sets } K_1, K_2 \text{ of } \Omega, \Sigma.$$

Consistency is a desirable property universally in all decision-problems. It simply states that the decision-procedure should take the correct decision as n tends to infinity. The theory for pointwise consistency will require only the following condition:

$$E_{\boldsymbol{\theta},\sigma} |X_i^n - \theta_i| \to 0 \quad \text{for} \quad i = 1, \cdots, k, \quad \forall(\boldsymbol{\theta}, \sigma) \in \Omega \times \Sigma. \tag{1.1}$$

Similarly, the theory for uniformly consistent procedures will require:

$$\sup_{K_1 \times K_2} E_{\boldsymbol{\theta},\sigma} |X_i^n - \theta_i| \to 0 \quad \text{for} \quad i = 1, \cdots, k \tag{1.2}$$

for all compact subsets K_1, K_2 of Ω, Σ.

In Section 2, necessary and sufficient conditions for pointwise and uniform consistency are derived for procedures in $\mathcal{D}_I$ with respect to general loss-functions. Let $\pi_{(i)}$ correspond to $\theta_{(i)}$ where $\theta_{(1)} \leq \cdots \leq \theta_{(k)}$. Section 3 considers different loss functions reflecting the goal to select populations close to $\pi_{(k)}$. These loss functions have been proposed by Chernoff and Yahav (1977), Bickel and Yahav (1977), Goel and Rubin (1977), Gupta and Hsu (1978) and Bjørnstad (1981). It is noted that some of these losses imply that the classical approach, started by Seal (1955) and Gupta (1956), of employing the so-called P^*-condition is not always appropriate.

To save space, the theory in Section 3 is applied only to the selection of means from normally distributed populations in Section 4. It is clear, however, that procedures for binomial, multinomial, multivariate normal and other selection problems can be checked for consistency in a similar way.

It is shown in Section 4, that among all the procedures in the class proposed by Seal (1955), only Gupta's procedure can be consistent. We also consider two classes of admissible procedures, derived by Bjørnstad (1981), and the Bayes procedures derived by Chernoff and Yahav (1977), Goel and Rubin (1977) and Gupta and Hsu (1978) for their respective loss functions and exchangeable normal priors.

2. Consistent invariant procedures. Our first aim is to develop necessary and sufficient conditions for pointwise consistency, with an invariant loss function, for procedures in $\mathscr{D}_I$. For $\mathbf{x} \in \mathbb{R}^k$, $\mathbf{x}^* = (x_{(1)}, \cdots, x_{(k)})$ where $x_{(1)} \leq \cdots \leq x_{(k)}$. Since for any $\delta_n \in \mathscr{D}_I$, $r_n(\boldsymbol{\theta}, \sigma \mid \delta_n) = r_n(\boldsymbol{\theta}^*, \sigma \mid \delta_n)$, we have that δ_n is consistent at $(\boldsymbol{\theta}, \sigma)$ if and only if δ_n is consistent at $(\boldsymbol{\theta}^*, \sigma)$. Let $\mathbf{Y}^n = (Y_1^n, \cdots, Y_k^n) = (X_{(1)}^n, \cdots, X_{(k)}^n)$. We need the following result.

LEMMA 2.1. *Assume* (1.1) *holds, and* $\theta_1 \leq \cdots \leq \theta_k$.
(a) *If* $\theta_i \neq \theta_j$ *then* $P_{\boldsymbol{\theta},\sigma}(X_i^n = Y_j^n) \to 0$.
(b) *If* $g\boldsymbol{\theta} \neq \boldsymbol{\theta}$ *then* $P_{\boldsymbol{\theta},\sigma}(g\mathbf{X}^n = \mathbf{Y}^n) \to 0$.

PROOF. (a): Let first $i > j$ such that $\theta_i > \theta_j$. Then

$$P_{\boldsymbol{\theta},\sigma}(X_i^n = Y_j^n) \leq \sum_{h=1}^{j} P_{\boldsymbol{\theta},\sigma}(X_h^n \geq X_i^n)$$
$$\leq 1/(\theta_i - \theta_j) \sum_{h=1}^{j} \{E_{\boldsymbol{\theta},\sigma}|X_h^n - \theta_h| + E_{\boldsymbol{\theta},\sigma}|X_i^n - \theta_i|\}$$

from Chebyshev's inequality, and the result follows. Let next $i < j$. Then:

$$P_{\boldsymbol{\theta},\sigma}(X_i^n = Y_j^n) \leq 1/(\theta_j - \theta_i) \sum_{h=j}^{k} \{E_{\boldsymbol{\theta},\sigma}|X_i^n - \theta_i| + E_{\boldsymbol{\theta},\sigma}|X_h^n - \theta_h|\} \to 0.$$

Consider next part (b). Let $g\boldsymbol{\theta} = (\theta_{i1}, \cdots, \theta_{ik})$. There exists ij such that $\theta_{ij} \neq \theta_j$. Since $g\mathbf{X}^n = \mathbf{Y}^n$ implies $X_{ij}^n = Y_j^n$, the result follows from (a). □

We can now state and prove the complete solution of pointwise consistency. First, let

$$\mathscr{A}_\sigma(\boldsymbol{\theta}) = \{a \in \mathscr{A}: \ell_\sigma(\boldsymbol{\theta}, a) = m_\sigma(\boldsymbol{\theta})\}. \tag{2.1}$$

THEOREM 2.1. *Assume* (1.1) *holds, and let* $\delta_n \in \mathscr{D}_I$. *Then* (2.2) *and* (2.3) *below are two equivalent, necessary and sufficient conditions for* $\{\delta_n\}$ *to be consistent at* $(\boldsymbol{\theta}, \sigma)$.

$$E_{\boldsymbol{\theta}^*,\sigma}\{\textstyle\sum_{a \in \mathscr{A}_\sigma(\boldsymbol{\theta}^*)} \delta_n(a \mid \mathbf{X}^n, S^n)\} \to 1 \tag{2.2}$$

$$E_{\boldsymbol{\theta}^*,\sigma}\{\textstyle\sum_{a \in \mathscr{A}_\sigma(\boldsymbol{\theta}^*)} \delta_n(a \mid \mathbf{Y}^n, S^n)\} \to 1. \tag{2.3}$$

PROOF. As remarked earlier we may assume $\boldsymbol{\theta} = \boldsymbol{\theta}^*$. Now,

$$r_n(\boldsymbol{\theta}, \sigma \mid \delta_n) = m_\sigma(\boldsymbol{\theta}) + \textstyle\sum_{a \notin \mathscr{A}_\sigma(\boldsymbol{\theta})} \{\ell_\sigma(\boldsymbol{\theta}, a) - m_\sigma(\boldsymbol{\theta})\} E_{\boldsymbol{\theta},\sigma}\delta_n(a \mid \mathbf{X}^n, S^n)$$

and it follows immediately that (2.2) is necessary and sufficient.

It remains to show that (2.2) $\Leftrightarrow$ (2.3). Assume first that (2.3) holds. Let $\mathbf{x} \in \mathbb{R}^k$ and $\mathbf{y} = \mathbf{x}^*$. The function $I(a = b) = 1$ if $a = b$, and 0 otherwise. Then

$$\delta_n(a \mid \mathbf{x}, s) \leq \textstyle\sum_{g \in G} \delta_n(ga \mid \mathbf{y}, s) I(g\mathbf{x} = \mathbf{y}). \tag{2.4}$$

It is therefore enough to show that for all $a \notin \mathscr{A}_\sigma(\boldsymbol{\theta})$, all $g \in G$,

$$E_{\boldsymbol{\theta},\sigma}\delta_n(ga \mid \mathbf{Y}^n, S^n) I(g\mathbf{X}^n = Y^n) \to 0. \tag{2.5}$$

If $ga \notin \mathscr{A}_\sigma(\boldsymbol{\theta})$, (2.5) follows directly from (2.3). If $ga \in \mathscr{A}_\sigma(\boldsymbol{\theta})$, then $\ell_\sigma(\boldsymbol{\theta}, ga) =$

$\ell_\sigma(g^{-1}\boldsymbol{\theta}, a) < \ell_\sigma(\boldsymbol{\theta}, a)$. Hence $g\boldsymbol{\theta} \neq \boldsymbol{\theta}$ and (2.5) follows from Lemma 2.1. The other way follows in the same manner. □

The individual selection functions of a subset selection procedure δ_n are given by

$$\psi_i^n(\mathbf{x}, s) = P(\text{selecting } \pi_i \mid \mathbf{X}^n = \mathbf{x}, S^n = s) = \textstyle\sum_{a \ni i} \delta_n(a \mid \mathbf{x}, s).$$

Let $\psi^n = (\psi_1^n, \cdots, \psi_k^n)$, and let $\psi_{(i)}^n$ correspond to $\theta_{(i)}$. We note that for $\delta_n \in \mathscr{D}_I$,

$$\psi_i^n(\mathbf{x}, s) = \psi_{gi}^n(g\mathbf{x}, s), \quad \forall\{g \in G, \mathbf{x} \in \mathbb{R}^k, s \in E, i \in (1, \cdots, k)\}. \tag{2.6}$$

When convenient, we shall denote the procedure δ_n by its selection functions ψ^n. Immediately from Theorem 2.1 we have the following result.

COROLLARY 2.1. *Assume $\delta_n \in \mathscr{D}_I$ is consistent at $(\boldsymbol{\theta}, \sigma)$, and that* (1.1) *holds. Then $E_{\boldsymbol{\theta}^*,\sigma}\psi_i^n(\mathbf{X}^n, S^n) \to 0$ for all i such that $\{a \in \mathscr{A}: a \ni i\} \cap \mathscr{A}_\sigma(\boldsymbol{\theta}^*) = \varnothing$.*

We now go on to develop necessary and sufficient conditions for uniform consistency in $\mathscr{D}_I$. We shall assume

$$\ell_\sigma(\boldsymbol{\theta}, a) \text{ is continuous in } (\boldsymbol{\theta}, \sigma) \text{ for each } a \in \mathscr{A}. \tag{2.7}$$

Let now $\Omega^* = \{\boldsymbol{\theta} \in \Omega: \theta_1 \leq \cdots \leq \theta_k\}$. Then $\delta_n \in \mathscr{D}_I$ is uniformly consistent if and only if

$$\sup_{\boldsymbol{\theta} \in K_1, \sigma \in K_2}\{r_n(\boldsymbol{\theta}, \sigma \mid \delta_n) - m_\sigma(\boldsymbol{\theta})\} \to 0 \quad \text{for all compact sets } K_1, K_2 \text{ of } \Omega^*, \Sigma. \tag{2.8}$$

Let d be the Euclidean distance in $\mathbb{R}^k$. Define for any compact set K_1, $g \in G$ and $\delta > 0$, $M_{g,\delta} = \{\boldsymbol{\theta} \in K_1: d(g\boldsymbol{\theta}, \boldsymbol{\theta}) \geq \delta\}$.

We need the following modification of Lemma 2.1.

LEMMA 2.2. *Assume* (1.2) *holds. Let K_1, K_2 be compact subsets of Ω^*, Σ respectively. Let $K_{i,j}^\varepsilon = \{\boldsymbol{\theta} \in K_1: |\theta_i - \theta_j| \geq \varepsilon\}$. Then*

(a) $$\sup_{K_{i,j}^\varepsilon \times K_2} P_{\boldsymbol{\theta},\sigma}(X_i^n = Y_j^n) \to 0 \quad \textit{for all} \quad \varepsilon > 0$$

(b) $$\sup_{M_{g,\delta} \times K_2} P_{\boldsymbol{\theta},\sigma}(g\mathbf{X}^n = \mathbf{Y}^n) \to 0, \quad \forall g \in G, \quad \forall \delta > 0.$$

PROOF. (a) We follow the same idea as in the proof of Lemma 2.1. Let first $i > j$. Then

$$\sup_{K_{i,j}^\varepsilon \times K_2} P_{\boldsymbol{\theta},\sigma}(X_i^n = Y_j^n)$$
$$\leq \frac{1}{\varepsilon} \textstyle\sum_{h=1}^{j} \sup_{K_1 \times K_2} E_{\boldsymbol{\theta},\sigma}|X_h^n - \theta_h| + \frac{j}{\varepsilon} \sup_{K_1 \times K_2} E_{\boldsymbol{\theta},\sigma}|X_i^n - \theta_i| \to 0,$$

from (1.2). Similarly for $i < j$,

$$\sup_{K^{\varepsilon}_{i,j}\times K_2} P_{\theta,\sigma}(X_i^n = Y_j^n) \le (1/\varepsilon) \textstyle\sum_{h=j}^{k} \sup_{K_1\times K_2} E_{\theta,\sigma}|X_h^n - \theta_h| + ((k-j+1)/\varepsilon) \sup_{K_1\times K_2} E_{\theta,\sigma}|X_i^n - \theta_i|.$$

(b) For any fixed $(\theta^0, \sigma^0) \in M_{g,\delta} \times K_2$, let $Q = \{i, j: |\theta_i^0 - \theta_j^0| \ge \delta/\sqrt{k}\}$. Then

$$P_{\theta^0,\sigma^0}(g\mathbf{X}^n = \mathbf{Y}^n) \le P_{\theta^0,\sigma^0}\{\cup_Q (X_i^n = Y_j^n)\} \le \textstyle\sum_{i\ne j} \sup_{K^{\delta_0}_{i,j}\times K_2} P_{\theta,\sigma}(X_i^n = Y_j^n),$$

where $\delta_0 = \delta/\sqrt{k}$. The result now follows from part (a). □

The necessary and sufficient conditions for uniform consistency can now be stated. Define for compact sets K_1, K_2 of Ω^*, Σ;

$$K_\varepsilon^a = \{(\theta, \sigma) \in K_1 \times K_2: \ell_\sigma(\theta, a) - m_\sigma(\theta) \ge \varepsilon\}. \tag{2.9}$$

THEOREM 2.2. *Assume* (1.2) *holds, and that* ℓ *satisfies* (2.7). *Let* $\delta_n \in \mathscr{D}_I$. *Then* (2.10) *and* (2.11) *below are two equivalent, necessary and sufficient conditions, for* $\{\delta_n\}$ *to be uniformly consistent.*

$$\sup_{K_\varepsilon^a} E_{\theta,\sigma}\delta_n(a \mid \mathbf{X}^n, S^n) \to 0 \quad as \quad n \to \infty \tag{2.10}$$

$$\sup_{K_\varepsilon^a} E_{\theta,\sigma}\delta_n(a \mid \mathbf{Y}^n, S^n) \to 0 \tag{2.11}$$

for all $a \in \mathscr{A}$, $\varepsilon > 0$, *and all compact sets* K_1 *of* Ω^*, K_2 *of* Σ *such that* $K_\varepsilon^a \ne \varnothing$.

PROOF. Now,

$$r_n(\theta, \sigma \mid \delta_n) - m_\sigma(\theta) = \textstyle\sum_{a\in\mathscr{A}} [\ell_\sigma(\theta, a) - m_\sigma(\theta)] E_{\theta,\sigma}\delta_n(a \mid \mathbf{X}^n, S^n).$$

Using the fact that ℓ and m are bounded on $K_1 \times K_2$ we readily get from (2.10)

$$\limsup \sup_{K_1\times K_2} [\ell_\sigma(\theta, a) - m_\sigma(\theta)] E_{\theta,\sigma}\delta_n(a \mid \mathbf{X}^n, S^n) \le \varepsilon; \quad \forall \varepsilon > 0.$$

Hence (2.8) holds. The other way is obvious.

To show that (2.11) ⇒ (2.10), it is enough from (2.4) to show

$$\sup_{K_\varepsilon^a} E_{\theta,\sigma}\delta_n(ga \mid \mathbf{Y}^n, S^n) I(g\mathbf{X}^n = \mathbf{Y}^n) \to 0, \quad \forall g \in G. \tag{2.12}$$

Let $g \in G$ be arbitrary and define $\Delta_\sigma(\theta, a) = \ell_\sigma(\theta, a) - m_\sigma(\theta)$. Since the loss-function is continuous, ∃ a $\delta > 0$ such that: $d(g\theta, \theta) < \delta$ and $\theta \in K_1 \Rightarrow |\Delta_\sigma(g\theta, ga) - \Delta_\sigma(\theta, ga)| \le \varepsilon/2$; $\forall \sigma \in K_2$. Then, for $(\theta, \sigma) \in K_1 \times K_2$, $\Delta_\sigma(\theta, a) \ge \varepsilon \Leftrightarrow \Delta_\sigma(g\theta, ga) \ge \varepsilon \Rightarrow d(g\theta, \theta) \ge \delta$ or $\Delta_\sigma(\theta, ga) \ge \varepsilon/2$. Hence, $K_\varepsilon^a \subset K_{\varepsilon/2}^{ga} \cup M_{g,\delta} \times K_2$ and (2.12) follows from (2.11) and Lemma 2.2. (2.10) ⇒ (2.11) in a similar way by showing that

$$\sup_{K_\varepsilon^a} E_{\theta,\sigma}\delta_n(ga \mid \mathbf{X}^n, S^n) I(\mathbf{X}^n = g\mathbf{Y}^n) \to 0, \quad \forall g \in G. \quad □$$

3. Some specific loss functions for selecting π_i close to $\pi_{(k)}$. In this section we shall apply the theory in the previous section to loss functions that more or less reflect the desire to have $\pi_{(k)}$ in the selected subset a, while keeping the size $|a|$ of the subset small. The invariant loss-functions to be considered are:

$$\ell_1(\boldsymbol{\theta}, a) = \theta_{(k)} - (1/|a|)\textstyle\sum_{j\in a}\theta_j + r(\theta_{(k)} - \max_{i\in a}\theta_i); \quad r > 0 \tag{3.1}$$

$$\ell_2(\boldsymbol{\theta}, a) = \theta_{(k)} - (1/|a|)\textstyle\sum_{j\in a}\theta_j + LI(\max_{i\in a}\theta_i < \theta_{(k)}); \quad L > 0. \tag{3.2}$$

Here $I(a < b) = 1$ if $a < b$ and 0 otherwise.

$$\ell_3(\boldsymbol{\theta}, a) = c|a| + \theta_{(k)} - \max_{i\in a}\theta_i; \qquad c > 0 \tag{3.3}$$

$$\ell_4(\boldsymbol{\theta}, a) = c_1 I(\max_{i\in a}\theta_i < \theta_k) + c_2|a|; \qquad c_1, c_2 > 0 \tag{3.4}$$

$$\ell_5(\boldsymbol{\theta}, a) = |a| + c\textstyle\sum_{i\notin a} I(\theta_i = \theta_{(k)}); \qquad c > 0 \tag{3.5}$$

$$\ell_6(\boldsymbol{\theta}, a) = \textstyle\sum_{i\in a}(\theta_{(k)} - \theta_i) + \alpha\sum_{i\notin a} I(\theta_i = \theta_{(k)}); \qquad \alpha > 0. \tag{3.6}$$

ℓ_1 was considered by Chernoff and Yahav (1977). They derived a Bayes procedure for normal populations. We show in Section 4 that this Bayes procedure is uniformly consistent for ℓ_1.

ℓ_2 was proposed by Bickel and Yahav (1977). This loss is not continuous in $\boldsymbol{\theta}$ for given a, so for ℓ_2 only pointwise consistency will be discussed. The loss ℓ_3 has been used by Goel-Rubin (1977), who derived a Bayes procedure. In the case of normal populations we show in Section 4 that the Bayes procedure is uniformly consistent for ℓ_3. Gupta and Hsu (1978) employed $\ell_4(\boldsymbol{\theta}, a)$. ℓ_5 and ℓ_6 are members of the class of additive loss-functions considered by Bjørnstad (1981). We note that ℓ_4, ℓ_5, ℓ_6 are not continuous in $\boldsymbol{\theta}$ for fixed a. Since all the loss-functions are independent of σ, we will use the notation $m(\boldsymbol{\theta})$ and $\mathscr{A}(\boldsymbol{\theta})$ (see (2.1)).

$$E_\varepsilon^a = \{\boldsymbol{\theta} \in K_1 : \ell(\boldsymbol{\theta}, a) - m(\boldsymbol{\theta}) \ge \varepsilon\} \quad \text{for any compact set} \quad K_1 \text{ of } \Omega^*$$

such that, from (2.9), $K_\varepsilon^a = E_\varepsilon^a \times K_2$. Define for any compact set K_1 of Ω^* and $\varepsilon > 0$,

$$K_i^\varepsilon = \{\boldsymbol{\theta} \in K_1 : \theta_k - \theta_i \ge \varepsilon\}. \tag{3.7}$$

THEOREM 3.1. *Let the loss be* ℓ_1, *given by* (3.1), *and let* $\psi^n \in \mathscr{D}_I$.

(a) *Assume* (1.1) *holds. Then* ψ^n *is consistent at* $(\boldsymbol{\theta}, \sigma)$ *with* $\theta_{(p-1)} < \theta_{(p)} = \theta_{(k)}$ *if and only if*

$$E_{\boldsymbol{\theta}^*,\sigma}\psi_i^n(\mathbf{X}^n, S^n) \to 0 \quad \text{for} \quad i \le p - 1. \tag{3.8}$$

(b) *Assume* (1.2) *holds. Then* ψ^n *is uniformly consistent if and only if*

$$\sup_{K_i^\varepsilon \times K_2} E_{\boldsymbol{\theta},\sigma}\psi_i^n(\mathbf{X}^n, S^n) \to 0 \quad \text{for} \quad i = 1, \cdots, k-1, \tag{3.9}$$

for all compact sets K_1, K_2 *of* Ω^*, Σ *and all* $\varepsilon > 0$ *such that* $K_i^\varepsilon \times K_2 \ne \varnothing$.

PROOF. (a) Using Theorem 2.1, the result follows from Corollary 2.1 and the fact that $\delta_n(a \mid \mathbf{x}, s) \le \psi_i^n(\mathbf{x}, s)$ if $a \ni i$.

(b) Since $K_i^\varepsilon \subset E_{\varepsilon/k}^a$, for $a \ni i$, (3.9) follows from (2.10). Assume now (3.9). Let $E_\varepsilon^a \times K_2 \ne \varnothing$, and $a = \{i_1, \cdots, i_q\}$ where $i_1 < \cdots < i_q$. Then $\delta_n(a \mid \mathbf{x}, s) \le$

$\psi^n_{i_1}(\mathbf{x}, s)$. Furthermore, $E^a_\varepsilon \subset K^\delta_{i_1}$ where $\delta = \varepsilon/(1 + r)$, and

$$\sup_{E^a_\varepsilon \times K_2} E_{\theta,\sigma}\delta_n(a \mid \mathbf{X}^n, S^n) \le \sup_{K^\delta_{i_1} \times K_2} E_{\theta,\sigma}\psi^n_{i_1}(\mathbf{X}^n, S^n) \to 0$$

from (3.9). □

REMARK. $m(\theta)$ and $\mathscr{A}(\theta)$ are the same for ℓ_2, given by (3.2), as for ℓ_1. Hence Theorem 3.1 (a) is valid also for ℓ_2.

THEOREM 3.2. *Let the loss be* ℓ_3, *given by* (3.3), *and assume* $\psi^n \in \mathscr{D}_I$.

(a) *Assume* (1.1) *holds. Then* ψ^n *is consistent at* (θ, σ) *with* $\theta_{(p-1)} < \theta_{(p)} = \theta_{(k)}$ *if and only if*

(3.10) $E_{\theta^*,\sigma}\{\sum_{i=1}^k \psi^n_i(\mathbf{X}^n, S^n)\} \to 1$ *and* $E_{\theta^*,\sigma}\, \psi^n_i(\mathbf{X}^n, S^n) \to 0$ *for* $i \le p - 1$.

(b) *Assume* (1.2) *holds. Then* ψ^n *is uniformly consistent if and only if*

(3.11) $$\sup_{K_1 \times K_2} E_{\theta,\sigma}\{\textstyle\sum_{i=1}^k \psi^n_i(\mathbf{X}^n, S^n)\} \to 1$$

and

(3.12) $$\sup_{K^\varepsilon_i \times K_2} E_{\theta,\sigma}\psi^n_i(\mathbf{X}^n, S^n) \to 0 \quad \text{for} \quad i \le k - 1$$

and for all compact sets K_1, K_2 *of* Ω^*, Σ *and all* $\varepsilon > 0$ *such that* $K^\varepsilon_i \times K_2 \ne \varnothing$.

PROOF. (a) Let $\theta_{(p-1)} < \theta_{(p)} = \theta_{(k)}$ and $\theta = \theta^*$. Then $\mathscr{A}(\theta) = \{(p), \cdots, (k)\}$. Now using the property that $\delta_n(a \mid \mathbf{x}, s) \le \psi^n_i(\mathbf{x}, s)$ if $a \ni i$ and the equation

$$\textstyle\sum_{i=1}^k \psi^n_i(\mathbf{x}, s) = 1 + \sum_{q=2}^k (q - 1) \sum_{\{a: |a| = q\}} \delta_n(a \mid \mathbf{x}, s),$$

the result follows immediately from Theorem 2.1.

(b) Assume (2.10) holds. Let $E^a_\varepsilon = \{\theta \in K_1: \ell_3(\theta, a) \ge c + \varepsilon\}$. Then $K^a_\varepsilon = E^a_\varepsilon \times K_2$. For $|a| \ge 2$ and $\varepsilon < c$, $E^a_\varepsilon = K_1$, and (3.11) follows. Also, (3.12) follows from the fact that $E^{\{i\}}_\varepsilon = K^\varepsilon_i$.

Now, let us assume that (3.11) and (3.12) hold. Clearly for $|a| \ge 2$, (3.11) ⇒ (2.10). For $a = \{i\}$, $i \le k - 1$: $\delta_n(\{i\} \mid \mathbf{X}^n, S^n) \le \psi^n_i(\mathbf{X}^n, S^n)$ and (2.10) follows from the fact that $E^{\{i\}}_\varepsilon = K^\varepsilon_i$. □

REMARK. $\mathscr{A}(\theta)$ is the same for ℓ_4 as for ℓ_3. Hence Theorem 3.2 (a) is valid also for ℓ_4.

Most of the research on subset selection has assumed that the procedures satisfy a certain control condition. The most common is the so-called P^*-condition, due primarily to Gupta (1956, 1965) and Seal (1955). Let a subset that includes $\pi_{(k)}$ be called a correct selection, CS. The P^*-condition is:

(3.13) $$\inf_{\Omega \times \Sigma} P^n_{\theta,\sigma}\{\text{CS} \mid \delta_n\} = \inf_{\Omega \times \Sigma} E_{\theta,\sigma}\{\psi^n_{(k)}\} \ge P^*; \quad 1/k < P^* < 1.$$

Suppose $\Omega \supset \Omega_0 = \{\theta: \theta_1 = \cdots = \theta_k\}$. If $\psi^n \in \mathscr{D}_I$ is pointwise consistent for ℓ_3 and ℓ_4 on $\Omega \times \Sigma$, it follows from Theorem 3.2 (a) that for $\theta \in \Omega_0$, $E_{\theta,\sigma}\psi^n_1 = \cdots =$

$E_{\theta,\sigma}\psi_k^n \to 1/k$, and therefore any pointwise consistent invariant procedure δ_n must have $\lim\sup_n\{\inf_{\Omega\times\Sigma} P_{\theta,\sigma}^n(\text{CS} \mid \delta_n)\} \le 1/k$, and cannot satisfy (3.13) as $n \to \infty$. Here we used that fact, derived from (2.6), that for any $\psi^n \in \mathscr{D}_I$, $E_\theta\psi_i^n = E_\theta\psi_j^n$ if $\theta_i = \theta_j$. In a similar way we see that if $P^* > 1/2$, no procedure satisfying (3.13) for all n can be consistent at any $\boldsymbol{\theta}$ where $\theta_{(k-1)} = \theta_{(k)}$.

Let $\Omega_1 = \{\boldsymbol{\theta} \in \Omega: \theta_{(k-1)} < \theta_{(k)}\}$. Procedures that are consistent on Ω_1 can of course satisfy the P^*-condition. Now, for any compact set K_1 in Ω_1^* there exists $\varepsilon > 0$ such that $K_1 \subset \{\boldsymbol{\theta} \in \Omega_1^*: \theta_k - \theta_{k-1} \ge \varepsilon\}$. From Theorem 3.2(b), we readily see that if (1.2) holds and $\psi^n \in \mathscr{D}_I$, then ψ^n is uniformly consistent on $\Omega_1 \times \Sigma$ for ℓ_3 iff (3.12) holds.

Let us now consider ℓ_5 with $c > 1$ and ℓ_6.

Theorem 3.3. *Let the loss be ℓ_5 with $c > 1$ or ℓ_6, given by* (3.5) *and* (3.6). *Assume* (1.1) *holds, and $\psi^n \in \mathscr{D}_I$. Then ψ^n is consistent at $(\boldsymbol{\theta}, \sigma)$ with $\theta_{(p-1)} < \theta_{(p)} = \theta_{(k)}$ if and only if*

$$E_{\theta^*,\sigma}\psi_i^n(\mathbf{X}^n, S^n) \to \begin{cases} 0 & \text{for } i \le p-1 \\ 1 & \text{for } i \ge p. \end{cases} \tag{3.14}$$

Remark. Comparing (3.14) with (3.8) and (3.10), we see that ℓ_5, ℓ_6 requires one to select all populations π_i with $\theta_i = \theta_{(k)}$ and excluding all others, while ℓ_1, ℓ_2, ℓ_3, ℓ_4 essentially requires one to exclude all π_i with $\theta_i < \theta_{(k)}$ and including only at least one π_i with $\theta_i = \theta_{(k)}$.

Proof. Let $\theta_{(p-1)} < \theta_{(p)} = \theta_{(k)}$ and $\boldsymbol{\theta} = \boldsymbol{\theta}^*$. Let $a_0 = \{p, \cdots, k\}$. By expressing ℓ_5 as $\ell_5(\boldsymbol{\theta}, a) = \#\{i \in a: \theta_i < \theta_k\} + (1-c)\#\{i \in a: \theta_i = \theta_k\} + c\#\{i: \theta_i = \theta_k\}$ we see that $\mathscr{A}(\boldsymbol{\theta}) = a_0$ for both ℓ_5 and ℓ_6, since $1 - c < 0$. From Theorem 2.1 it remains to show that (3.14) is equivalent to

$$E_{\theta^*\sigma}\delta_n(a_0 \mid \mathbf{X}^n, S^n) \to 1. \tag{3.15}$$

Obviously, (3.15) $\Rightarrow$ (3.14). Assume now that (3.15) does not hold. Then there exists $a_1 \ne a_0$ such that $\lim\sup E_{\theta,\sigma}\delta_n(a_1 \mid \mathbf{X}^n, S^n) = \beta > 0$. If there is an $i \in a_1$, $i \le p - 1$, then $\lim\sup E_{\theta,\sigma}\psi_i^n \ge \beta$, violating (3.14). If $\{i \in a_1 \Rightarrow i \ge p\}$ there must exist $j \ge p$, $j \notin a_1$ and therefore $\lim\inf E_{\theta,\sigma}\psi_j^n \le 1 - \beta$, implying again that (3.14) does not hold. □

ψ^n is said to be a just procedure if $x_i \ge x_i'$ and $x_j \le x_j'$ for $j \ne i$ implies that $\psi_i^n(\mathbf{x}) \ge \psi_i^n(\mathbf{x}')$.

Corollary 3.1. *Assume ψ^n is just, invariant and pointwise consistent for ℓ_5 with $c > 1$ and ℓ_6 on $\Omega \times \Sigma$, $\Omega \supset \Omega_0 = \{\boldsymbol{\theta}: \theta_1 = \cdots = \theta_k\}$. Then* $\inf_{\Omega\times\Sigma} P_{\theta,\sigma}\{CS \mid \psi^n\} \to 1$.

Proof. Nagel (1970) showed that for any just procedure, $\inf_{\Omega\times\Sigma} P_{\theta,\sigma}\{CS \mid \psi^n\}$ occurs at some $\boldsymbol{\theta} \in \Omega_0$. From (3.14) we have that for $\boldsymbol{\theta} \in \Omega_0$,

$$E_{\theta,\sigma}\psi_1^n = \cdots = E_{\theta,\sigma}\psi_k^n \to 1. \quad \square$$

Corollary 3.1 means that no just, invariant procedure satisfying (3.13) with equality can be pointwise consistent for ℓ_5 or ℓ_6, if P^* is chosen independent of n. Hence, if these loss functions reflect the true losses involved in the selection problem the P^*-condition is not appropriate. It seems clear that it is the term $\sum_{i \notin a} I(\theta_i = \theta_{(k)})$ that makes (3.13) inappropriate.

Finally, consider ℓ_5 with $c \le 1$. The following result is needed.

LEMMA 3.1. $\mathbf{Y}^n = (\mathbf{X}^n)^*$. (a) (1.1) $\Rightarrow E_{\theta^*,\sigma} | Y_i^n - \theta_i^* | \to 0$ *for* $i = 1, \cdots, k$.
(b) (1.2) $\Rightarrow \sup_{K_1 \times K_2} E_{\theta,\sigma} | Y_i^n - \theta_i | \to 0$ *for* $i = 1, \cdots, k$ *and all compact sets* K_1, K_2 *in* Ω^*, Σ.

PROOF. Let $\theta = \theta^*$. Then $| Y_i^n - \theta_i | \le \sum_{j=1}^k | X_j^n - \theta_j | + \sum_{j \ne i} | \theta_j - \theta_i | \cdot I(X_j^n = Y_i^n)$. Hence

$$E_{\theta,\sigma} | Y_i^n - \theta_i | \le \sum_{j=1}^k E_{\theta,\sigma} | X_j^n - \theta_j | + \sum_{j \ne i} | \theta_j - \theta_i | P_{\theta,\sigma}(X_j^n = Y_i^n).$$

Then (a) follows directly from (1.1) and Lemma 2.1, and (b) follows from (1.2) and Lemma 2.2. □

For this particular loss we shall assume that $X_1^n, \cdots, X_k^n$ are independent, each X_i^n has density $f_\sigma^n(\cdot, \theta_i)$ with respect to a σ-finite measure, and $\Omega = \Theta^k$. It is assumed that for fixed (n, σ), f_σ^n has the monotone likelihood-ratio property.

Bjørnstad (1981) showed that there is a uniformly minimum risk procedure in $\mathscr{D}_I$ for ℓ_5 when $c \le 1$. It is given by:

$$\delta_0(\{i\} | \mathbf{y}) = 1/q \quad \text{for} \quad i \ge k - q + 1,$$

when

$$y_{k-q} < y_{k-q+1} = \cdots = y_k; \quad \forall \mathbf{y} \in \mathscr{Y} = \{\mathbf{x} \in \mathbb{R}^k : x_1 \le \cdots \le x_k\}.$$

Obviously, δ_0 is the only interesting procedure in $\mathscr{D}_I$ for this loss. Even though ℓ_5 is not continuous in θ we can say something about uniform consistency of δ_0 as the next result shows.

THEOREM 3.4. *The loss is* $\ell_5(\theta, a) = | a | + c \sum_{i \notin a} I(\theta_i = \theta_{(k)})$ *with* $0 < c \le 1$.

(a) *Assume* (1.1) *holds. Then* δ_0 *is pointwise consistent on* $\Omega \times \Sigma$.

(b) *Assume* (1.2) *holds and that* $f_\sigma^n(x, \theta)$ *is a continuous function of* (θ, σ). *Then* δ_0 *is uniformly consistent on* $\Omega_1 \times \Sigma$, *where* $\Omega_1 = \{\theta \in \Omega: \theta_{(k)} > \theta_{(k-1)}\}$.

PROOF. (a) Let $\theta = \theta^*$ and assume $\theta_{p-1} < \theta_p = \theta_k$. We see that when $c \le 1$, $\mathscr{A}(\theta) = \{\{p\}, \cdots, \{k\}\}$. From (2.3) of Theorem 2.1 we need to show that $E_{\theta,\sigma} \sum_{i=p}^k \delta_0(\{i\} | \mathbf{Y}^n) \to 1$. $\sum_{i=1}^k \delta_0(\{i\} | \mathbf{Y}^n) \equiv 1$ so we must show that $E_{\theta,\sigma}\delta_0(\{i\} | \mathbf{Y}^n) \to 0$ for $i \le p - 1$. Now, $E_{\theta,\sigma}\delta_0(\{i\} | \mathbf{Y}^n) \le P_{\theta,\sigma}(Y_i^n = Y_k^n) \to 0$ since $Y_k^n - Y_i^n \to_P \theta_k - \theta_i > 0$, from Lemma 3.1.

(b) On Ω_1, $m(\theta) = 1$ so we must show that $\sup_{K_1 \times K_2} r_n(\theta, \sigma | \delta_0) \to 1$ for all compact sets K_1, K_2 in Ω_1^*, Σ. We readily derive that

$$r_n(\boldsymbol{\theta}, \sigma \mid \delta_0) \leq P_{\boldsymbol{\theta},\sigma}(X_k^n > \max_{1\leq j\leq k-1} X_j^n) + (1+c)P_{\boldsymbol{\theta},\sigma}(X_k^n \leq \max_{1\leq j\leq k-1} X_j^n).$$

It follows that it is sufficient to show $\inf_{K_1\times K_2} P_{\boldsymbol{\theta},\sigma}(X_k^n > X_j^n) \to 1$ for $j \leq k-1$.

As mentioned earlier, there exists $\varepsilon > 0$ such that $\boldsymbol{\theta} \in K_1 \Rightarrow \theta_k - \theta_{k-1} \geq \varepsilon$. Since $f_\sigma^n(\cdot, \boldsymbol{\theta})$ is continuous in $(\boldsymbol{\theta}, \sigma)$, $P_{\boldsymbol{\theta},\sigma}(X_k^n > X_j^n)$ is a continuous function of $(\boldsymbol{\theta}, \sigma)$. Hence infimum occurs at some $(\boldsymbol{\theta}^n, \sigma^n) \in K_1 \times K_2$. From (1.2), $X_i^n - \theta_i^n \to_P 0$ under $(\boldsymbol{\theta}^n, \sigma^n)$, and

$$P_{\boldsymbol{\theta}^n,\sigma^n}(X_k^n - X_j^n > 0) \geq P_{\boldsymbol{\theta}^n,\sigma^n}\{(X_k^n - \theta_k^n) - (X_j^n - \theta_j^n) > -\varepsilon\} \to 1. \quad \square$$

4. Selection of means from normal populations. The k populations are now assumed to be normally distributed, and X_i^n is the sample mean of size n from π_i. Hence $X_1^n, \cdots, X_k^n$ are independent and X_i^n is $N(\theta_i, \sigma^2/n)$, where σ is unknown, $\sigma \in \langle 0, \infty\rangle$. Moreover, let $S^n = S_n^2$, the usual U.M.V.U. estimate of σ^2. Then $S_n^2 \to_P \sigma^2$. In this section $\Omega = \mathbb{R}^k$ and (1.2) clearly holds. We shall apply the theory in the previous section for the loss functions $\ell_1 - \ell_6$ on some subset selection procedures that have been studied in the literature. It is now assumed that $c > 1$ in ℓ_5.

Consider the class $\mathscr{S}$, proposed by Seal (1955). $\mathscr{S} = \{\psi^{\mathbf{c},n}\colon \sum_{i=1}^{k-1} c_i = 1, c_i \geq 0; \forall i\}$, where

$$\psi_i^{\mathbf{c},n} = 1 \Leftrightarrow X_i^n \geq \sum_{j=1}^{k-1} c_j X_{(j)}^{1,n} - S_n D_n(\mathbf{c}); \quad D_n(\mathbf{c}) \geq 0. \tag{4.1}$$

Here $X_{(1)}^{1,n} \leq \cdots \leq X_{(k-1)}^{1,n}$ are the ordered X_j^n, $j \neq i$. Seal assumed $D_n(\mathbf{c})$ is determined such that the P^*-condition (3.13) holds with equality. We shall, however, consider $\psi^{\mathbf{c},n}$ for any sequence $\{D_n(\mathbf{c})\}$. If we want (3.13) to be satisfied, it is readily seen, since infimum of $P(\mathrm{CS} \mid \psi^{\mathbf{c},n})$ occurs when $\theta_1 = \cdots = \theta_k$, that $\sqrt{n}D_n(\mathbf{c}) \to t(\mathbf{c})$ where

$$P\{\textstyle\sum_{j=1}^{k-1} c_j Z_{(j)} - Z_k \leq t(\mathbf{c})\} = P^*. \tag{4.2}$$

Here $Z_{(1)} < \cdots < Z_{(k-1)}$ are the ordered $Z_1, \cdots, Z_{k-1}$, and $Z_1, \cdots, Z_k$ are i.i.d. $N(0, 1)$.

One procedure in $\mathscr{S}$ has received special attention in the literature by many authors. Gupta (1956, 1965) suggested the use of $c_{k-1} = 1$. Let us call this procedure $\psi^{G,n}$, and denote $D_n(\mathbf{c})$ by d_n, such that

$$\psi_i^{G,n} = 1 \Leftrightarrow X_i^n \geq X_{(k)}^n - S_n d_n. \tag{4.3}$$

Applying Lemma 2.1, the following two results can be readily shown, using Theorems 3.1–3.3.

THEOREM 4.1. *Let the loss be one of* $\ell_1 - \ell_6$, *given by* (3.1)–(3.6), *where* $c > 1$ *for* ℓ_5. *Assume* $c_{k-1} < 1$. *Then* $\psi^{\mathbf{c},n}$, *given by* (4.1), *is not pointwise consistent on* $\mathbb{R}^k \times \langle 0, \infty\rangle$ *for any sequence* $\{D_n(\mathbf{c})\}$.

This result shows that no procedure in $\mathscr{S}$, except $\psi^{G,n}$, has a chance of being consistent for the losses $\ell_1 - \ell_6$.

The cases of $\{d_n\}$ when $\psi^{G,n}$ is consistent for the different losses are specified in the next result.

THEOREM 4.2. *Let $\psi^{G,n}$ be given by* (4.3).
(a) *$\psi^{G,n}$ is uniformly (pointwise) consistent on $\mathbb{R}^k \times \langle 0, \infty\rangle$ for $\ell_1(\ell_2)$, given by* (3.1), ((3.2)), *if and only if $d_n \to 0$.*
(b) *$\psi^{G,n}$ is uniformly (pointwise) consistent on $\mathbb{R}^k \times \langle 0, \infty\rangle$ for $\ell_3(\ell_4)$, given by* (3.3), ((3.4)), *if and only if $\sqrt{n}d_n \to 0$.*
(c) *$\psi^{G,n}$ is pointwise consistent on $\mathbb{R}^k \times \langle 0, \infty\rangle$ for ℓ_5 with $c > 1$ and ℓ_6 if and only if $d_n \to 0$ and $\sqrt{n}d_n \to \infty$.*

REMARK. If d_n is determined such that $\psi^{G,n}$ satisfies (3.13) with equality, then, from (4.2), $\psi^{G,n}$ is uniformly (pointwise) consistent for $\ell_1(\ell_2)$, but not consistent for any of the other losses.

For the rest of this section we assume σ is known. Two classes of invariant, admissible procedures for ℓ_6 and ℓ_5 with $c > 1$ are given below.

(4.4) $\quad \psi_i^{1,n} = 1 \Leftrightarrow c\exp(b_n X_i^n) \geq \sum_{j=1}^k \exp(b_n X_j^n) \quad \text{or} \quad X_i^n = X_{(k)}^n.$

(4.5) $\quad \psi_i^{2,n} = 1 \Leftrightarrow (1 + (\alpha/b_n))\exp(b_n X_i^n) \geq \sum_{j=1}^k \exp(b_n X_j^n) \quad \text{or} \quad X_i^n = X_{(k)}^n.$

Bjørnstad (1981) showed that $\psi^{1,n}$ is admissible for ℓ_5, and $\psi^{2,n}$ is admissible for ℓ_6, for all $b_n > 0$. From Theorem 3.3, the following result is easily shown.

THEOREM 4.3. (a) *Let $\psi^{1,n}$ be given by* (4.4), *and assume $c > k$ in ℓ_5. Then $\psi^{1,n}$ is pointwise consistent on $\mathbb{R}^k$ for ℓ_5, ℓ_6 if and only if $b_n \to \infty$ and $b_n/\sqrt{n} \to 0$.*
(b) *Let $\psi^{2,n}$ be given by* (4.5). *Then $\psi^{2,n}$ is pointwise consistent on $\Omega_1 = \{\theta \in \mathbb{R}^k, \theta_{(k-1)} < \theta_{(k)}\}$ for ℓ_5, ℓ_6 if and only if* $\liminf b_n \geq \alpha$.

REMARK 1. It is readily seen that $\liminf b_n \geq \alpha$ implies

$$E_{\theta^*}\psi_k^{2,n'} \to 1/(k - p + 1),$$

for some subsequence, for all $p \leq k - 1$. By (b) of Theorem 4.3 this implies, from Theorem 3.3, that $\psi^{2,n}$ is not pointwise consistent on $\mathbb{R}^k$ for any $\{b_n\}$.

REMARK 2. It can be shown that if $c \leq k$, then $\psi^{1,n}$ is not pointwise consistent on $\mathbb{R}^k$ for any $\{b_n\}$.

At last in this section we consider the Bayes-procedures derived for normal exchangeable priors for ℓ_1, ℓ_3, ℓ_4 by Chernoff and Yahav (1977), Goel and Rubin (1977) and Gupta and Hsu (1978) respectively. The prior is: $\theta' \sim N_k(m\mathbf{e}, rI + tU)$, where $\mathbf{e} = (1, \cdots, 1)'$ and $U = \mathbf{ee}'$, $r > 0$, $t \geq 0$. As shown by the authors mentioned above, the risks of the Bayes-procedures do not depend on m, t so we may let $m = 0$, $t = 0$. If so, $(\theta \mid \mathbf{X}^n = \mathbf{x})$ is

$$N\left(\frac{r}{q_n + r}\mathbf{x}, \frac{rq_n}{q_n + r}I\right), \quad \text{where} \quad q_n = \frac{\sigma^2}{n} \quad \text{and} \quad \hat{\theta}^n = \frac{r}{q_n + r}\mathbf{X}^n$$

is the usual squared error loss Bayes estimate.

Let us consider first ℓ_1 and its Bayes procedure δ_1^B. Consider $T_n(a \mid \mathbf{X}^n) = E\{\ell_1(\boldsymbol{\theta}, a) \mid \mathbf{X}^n\} = (1 + r)E\{\theta_{(k)} \mid \mathbf{X}^n\} - (1/|a|) \sum_{j \in a} \hat{\theta}_j^n - rE\{\max_{i \in a} \theta_i \mid \mathbf{X}^n\}$. Then, $\delta_1^B(a \mid \mathbf{X}^n) = 1$ iff a minimizes $T_n(a' \mid \mathbf{X}^n)$ for all $a' \in \mathscr{A}$. Clearly, δ_1^B is permutation-invariant. Lemma 3.1(b) implies that $T_n(a \mid \mathbf{Y}^n) - \ell_1(\boldsymbol{\theta}_n, a) \to_P 0$ for any sequence $\{\boldsymbol{\theta}_n\} \in K$, and compact set K of Ω^*. We shall show that δ_1^B is uniformly consistent for ℓ_1 on $\mathbb{R}^k$. In order to do so it is enough, from Theorem 2.2, to show that $\sup_{E_\varepsilon^a} E_{\boldsymbol{\theta}} \delta_1^B(a \mid \mathbf{Y}^n) \to 0$, where $E_\varepsilon^a = \{\boldsymbol{\theta} \in K_1 : \ell_1(\boldsymbol{\theta}, a) \geq \varepsilon\}$. Now,

$$E_{\boldsymbol{\theta}} \delta_1^B(a \mid \mathbf{Y}^n) \leq P_{\boldsymbol{\theta}}\{T_n(a \mid \mathbf{Y}^n) \leq T_n(\{k\} \mid \mathbf{Y}^n)\},$$

which is a continuous function in $\boldsymbol{\theta}$. Hence, for some $\boldsymbol{\theta}^n \in E_\varepsilon^a$,

$$\sup_{E_\varepsilon^a} E_{\boldsymbol{\theta}} \delta_1^B(a \mid \mathbf{Y}^n) \leq P_{\boldsymbol{\theta}^n}\{T_n(a \mid \mathbf{Y}^n) \leq T_n(\{k\} \mid \mathbf{Y})\}$$
$$\leq P_{\boldsymbol{\theta}^n}\{T_n(a \mid \mathbf{Y}^n) - \ell_1(\boldsymbol{\theta}^n, a) - T_n(\{k\} \mid \mathbf{Y}^n) \leq -\varepsilon\} \to 0.$$

Next, we consider the Bayes-procedure δ_3^B for ℓ_3. Let $\gamma_n^2 = rq_n/(q_n + r) = 1/(r^{-1} + n\sigma^{-2})$. According to Corollary 5 of Goel and Rubin (1977), $\delta_3^B(\{k\} \mid \mathbf{y}) = 1$, for $\mathbf{y} = \mathbf{x}^*$, $\forall \mathbf{x} \in \mathbb{R}^k$, provided $c/\gamma_n \geq 1/\sqrt{\pi}$. Since $\gamma_n \to 0$, it now follows immediately from (2.11) of Theorem 2.2 that δ_3^B is uniformly consistent for ℓ_3 on $\mathbb{R}^k$, since $K_\varepsilon^a \neq \varnothing$ implies $a \neq \{k\}$.

Finally, let δ_4^B be the Bayes-procedure for ℓ_4, given by (3.4). Gupta and Hsu (1978) showed that δ_4^B is given by

$$\psi_{4,i}^B = 1 \Leftrightarrow X_i^n \geq \max_{j \neq i} X_j^n \quad \text{or} \quad P\{\theta_i = \theta_{(k)} \mid \mathbf{X}^n\} \geq c_2/c_1.$$

Clearly ψ_4^B is permutation-invariant. Since ℓ_4 is not continuous in $\boldsymbol{\theta}$ we shall discuss only the pointwise consistency properties of ψ_4^B. Let now $\theta_{(p-1)} < \theta_{(p)} = \theta_{(k)}$, and $\boldsymbol{\theta} = \boldsymbol{\theta}^*$. Then for $i \leq p - 1$,

$$P(\theta_i = \theta_{(k)} \mid \mathbf{X}^n) \leq \Phi((\hat{\theta}_i - \hat{\theta}_k)/\gamma_n\sqrt{2}) \to_P 0,$$

since $\hat{\theta}_i - \hat{\theta}_k \to_P \theta_i - \theta_k$ and $\gamma_n \to 0$. Therefore $E_{\boldsymbol{\theta}} \psi_{4,i}^B \to 0$ for $i \leq p - 1$. This implies that (3.10) holds if $p = k$, and from the remark after Theorem 3.2, we have shown that ψ_4^B is pointwise consistent on $\Omega_1 = \{\boldsymbol{\theta} \in \mathbb{R}^k : \theta_{(k)} > \theta_{(k-1)}\}$. If $c_2/c_1 \geq 1/2$, then $P(\theta_k = \theta_{(k)} \mid \mathbf{X}^n) \geq c_2/c_1 \Rightarrow X_k^n \geq \max_{j \leq k-1} X_j^n$ and $E_{\boldsymbol{\theta}} \psi_{4,k}^B \to 1/(k - p + 1)$. Therefore, from (3.10), ψ_4^B is pointwise consistent on $\mathbb{R}^k$, if $c_2/c_1 \geq 1/2$. However, if $c_2/c_1 < 1/2$ it is straightforward to show that ψ_4^B is not consistent on all points in $\mathbb{R}^k - \Omega_1$.

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

REFERENCES

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

BJØRNSTAD, J. F. (1981). A decision-theoretic approach to subset selection. *Comm. Statist.-Theor. Meth.* **A10** (23) 2411–2433.

CHERNOFF, H. and YAHAV, J. A. (1977). A subset selection problem employing a new criterion. *Statistical Decision Theory and Related Topics II* 93–119. (Ed. S. S. Gupta and D. S. Moore). Academic, 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. (1956). On a decision rule for a problem in ranking means. Mimeo Series #150, Inst. Statist. Univ. North Carolina, Chapel Hill.

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

GUPTA, S. S. and HSU, J. C. (1978). On the performance of some subset selection procedures. *Comm. Statist.-Sim. Comp.* **B7** (6) 561–591.

NAGEL, K. (1970). On subset selection rules with certain optimality properties. Mimeo Series 222, Dept. of Statistics, Purdue University, W. Lafayette, Ind.

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

UNIVERSITY OF TROMSÖ
INSTITUTE OF MATHEMATICAL AND PHYSICAL SCIENCES
P.O. BOX 953
N-9001 TROMSÖ
NORWAY