SEQUENTIAL SELECTION PROCEDURES
- A DECISION THEORETIC APPROACH*

by

Shanti S. Gupta
Purdue University

Klaus J. Miescke
University of Illinois
at Chicago

Technical Report #82-6

Department of Statistics

March 1982

(Revised July 1983)

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

SEQUENTIAL SELCTION PROCEDURES - A DECISION THEORETIC APPROACH*

by

Shanti S. Gupta
Purdue University

Klaus J. Miescke
University of Illinois
at Chicago

ABSTRACT

Let $\pi_1,\ldots,\pi_k$ be given populations associated with unknown real parameters $\theta_1,\ldots,\theta_k$, which are related to a common underlying exponential family. Permutation invariant sequential procedures of the following type are considered. At stage m, observations are drawn from all eligible populations, i.e. from those which have not been eliminated so far. Then either the procedure stops and makes a final subset selection from the eligible populations, or it selects a subset from the eligible populations and proceeds to stage $m+1$.

Under a general loss structure (favoring large parameters), it is shown that at all stages the finally selected subsets have to be associated with the largest sufficient statistics from the eligible populations. In fact, these natural final decisions are proved to be uniformly optimum in terms of the associated risk. Under the assumption of a strongly unimodal exponential family, several consequences are derived with respect to optimality of natural subset selections at various stages. Especially, in the class of q-stage procedures with fixed predetermined subset sizes at the q stages, the natural procedure is uniformly optimum in terms of the risk. All results are derived using the Bayes approach with respect to permutation symmetric priors. The technique of backward induction is used, and the concept of decrease in transposition (DT), introduced by Hollander, Proschan and Sethuraman (1977), plays a crucial role throughout the paper.

*Research supported by ONR contract N00014-75-C-0455 at Purdue University, West Lafayette, IND, U.S.A.

SEQUENTIAL SELECTION PROCEDURES - A DECISION THEORETIC APPROACH[1,2]

by

Shanti S. Gupta
Purdue University

Klaus J. Miescke
University of Illinois
at Chicago

1. Introduction.

Let $\pi_1,\ldots,\pi_k$ be given populations which are associated with unknown parameters $\theta_1,\ldots,\theta_k \in \Omega$, where $\Omega \subseteq \mathbb{R}$ is an unbounded or bounded interval. Let the goal be to find a subset (of random or fixed size) of populations with large parameters. Sequential procedures will be studied in a general framework which covers the control and non-control, elimination (screening) and non-elimination, truncated as well as open-sequential settings.

Assume that at every Stage $m \in \mathbb{N} = \{1,2,\ldots\}$ samples $\{X_{ijm}\}_{j=1,\ldots,n_m}$ can be drawn from π_i, $i=1,\ldots,k$, where n_m is a fixed common sample size. Let all the observations be real-valued, independent, and have densities with respect to μ, the Lebesgue measure on $\mathcal{X} = \mathbb{R}$ or the counting measure on $\mathcal{X} = \mathbb{Z}$ (or any other lattice on $\mathbb{R}$). Finally, it is assumed that all these densities are members of an exponential family $\mathcal{F} = \{c(\theta)\exp(\theta x)d(x),\ x \in \mathcal{X}\}_{\theta \in \Omega}$, where $\theta = \theta_i$ holds for observations from π_i, $i=1,\ldots,k$. Let $U_{im} = X_{i1m} + \ldots + X_{in_m m}$ be the sufficient statistic for θ_i at Stage m with respect to the samples at Stage m and let $W_{im} = U_{i1} + \ldots + U_{im}$ be the overall (up to Stage m) sufficient statistic for θ_i, $i=1,\ldots,k$, $m \in \mathbb{N}$. For notational convenience, let $\underline{U}_m = (U_{1m},\ldots,U_{km})$, $\underline{V}_m = (\underline{U}_1,\ldots,\underline{U}_m)$, $\underline{W}_m = (W_{1m},\ldots,W_{km}) = \underline{U}_1 + \ldots \underline{U}_m$, $\underline{\theta} = (\theta_1,\ldots,\theta_k)$ and $N_m = n_1 + \ldots + n_m$, $m \in \mathbb{N}$,

[1]Research supported by the Office of Naval Research Contract N00014-75-C-0455.
AMS 1970 subject classifications. Primary 62F07; secondary 62F05, 62F15, 62L99.
Key words and phrases. Multiple decision procedures, sequential procedures, screening procedures, selection procedures, Bayesian analysis.

[2]This paper is a revision of an invited paper presented at an international meeting on analysis of sample survey and on sequential analysis at the Hebrew University, Jerusalem, June 14-18, 1982.

in the following. Note that for every $m \in \mathbb{N}$, the density with respect to $\mu_k = \mu\times\ldots\times\mu$ of $\underline{U}_m$ and $\underline{W}_m$ are, respectively,

$$(1) \qquad f_{\underline{\theta}}^{(m)}(\underline{u}) = \prod_{i=1}^{k} c_{n_m}(\theta_i)\exp(\theta_i u_i)d_{n_m}(u_i),\ \underline{u} \in \mathcal{X}^k,\ \underline{\theta} \in \Omega^k, \text{ and}$$

$$g_{\underline{\theta}}^{(m)}(\underline{w}) = \prod_{i=1}^{k} c_{N_m}(\theta_i)\exp(\theta_i w_i)d_{N_m}(w_i),\ \underline{w} \in \mathcal{X}^k,\ \underline{\theta} \in \Omega^k, \text{ where}$$

$c_r(\theta) = c(\theta)^r$, and d_r denotes the r-fold convolution of d w.r.t. μ.

Next, an explicit definition will be given of what is understood to be a (randomized) sequential selection procedure. Briefly, such a procedure can be described as follows: At every stage, it decides <u>either</u> to stop (γ), how many populations to retain (φ), and which populations to select finally (ψ), <u>or</u> not to stop ($1-\gamma$), how many populations to retain ($\tilde{\varphi}$), and which populations to select for further examination at the next stage ($\tilde{\psi}$). There is one restriction, however, which is to be emphasized: Once a certain population has been eliminated at one stage, it may never be selected at subsequent stages.

To make the definition more understandable, let us introduce the notation $S_m = (s_1,\ldots,s_m)$ with $s_m \subseteq s_{m-1} \subseteq \cdots \subseteq s_1$, for the situation at the end of Stage m-1, if populations π_i with $i \in s_{r+1}$ have been selected at the end of Stage r, r=1,...,m-1. Thus, $S_1 = s_1 = \{1,\ldots,k\}$ is the initial situation, $S_2 = (s_1,s_2)$ means that the populations with indices in s_2 have been selected at the end of Stage 1, and so forth. By identifying populations with their indices, selections from $\{1,\ldots,k\}$ are to be understood in a natural way.

<u>Definition 1. (Sequential selection procedure $(\gamma,\varphi,\tilde{\varphi},\psi,\tilde{\psi})$)</u>.

The definition is given by induction with respect to the stage number $m \geq 1$. The starting condition is $S_1 = s_1 = \{1,\ldots,k\}$ and $r_1 = k$.

<u>Stage m</u>: If $S_m = (s_1,\ldots,s_m)$, take additional observations from populations π_i with $i \in s_m$, i.e. observe U_{im} when $i \in s_m$. The decisions at this stage

are based on 5 different decision functions. It will prove to be useful to write them as functions of $\underline{V}_m = (\underline{U}_1,\ldots,\underline{U}_m)$, but it is understood (and clearly indicated by notation) that they depend only on the really observed U_{ip} with $i \in s_p$, $p=1,\ldots,m$. The decisions are made according to the following scheme.

Either, with probability $\gamma_{S_m}(\underline{V}_m)$, the procedure stops, then, with probability $\varphi_{r_{m+1};S_m}(\underline{V}_m)$, it decides that $r_{m+1} \in \{0,1,\ldots,r_m\}$ populations are to be selected from s_m, and finally, with probability $\psi_{s_{m+1};r_{m+1},S_m}(\underline{V}_m)$, it selects $s_{m+1} \subseteq s_m$ with $|s_{m+1}| = r_{m+1}$ (where $|\cdot|$ denotes subset size); or, with probability $1 - \gamma_{S_m}(\underline{V}_m)$, the procedure does not stop, then, with probability $\tilde{\varphi}_{r_{m+1};S_m}(\underline{V}_m)$, it decides that $r_{m+1} \in \{1,2,\ldots,r_m\}$ populations are to be selected from s_m, then, with probability $\tilde{\psi}_{s_{m+1};r_{m+1},S_m}(\underline{V}_m)$, it selects $s_{m+1} \subseteq s_m$ with $|s_{m+1}| = r_{m+1}$, and the procedure continues at Stage m+1.

This process is continued until it is stopped. The procedure is said to be truncated at Stage q if $\gamma_{S_q} = 1$ for all possible S_q.

Our main interest is on permutation invariant sequential selection procedures which treat all k populations symmetrically. More precisely, they are defined as follows.

Definition 2. (Permutation invariant procedures).

A procedure $(\gamma, \varphi, \tilde{\varphi}, \psi, \tilde{\psi})$ is called permutation invariant if for every $m \geq 1$, the 5 decision functions at Stage m are permutation invariant in the following sense. Let $S_m = (s_1,\ldots,s_m)$, $\underline{V}_m = \underline{v}_m$ and permutation σ of $(1,\ldots,k)$ be fixed. For notational convenience, let $\sigma(S_m)=(\sigma(s_1),\ldots,\sigma(s_m))$, where $\sigma(s_r) = \{\sigma(i) | i \in s_r\}$, $r=1,\ldots,m$, and let $\sigma(\underline{v}_m) = (\sigma(\underline{u}_1),\ldots,\sigma(\underline{u}_m))$, where $\sigma(\underline{u}_r) = (u_{\sigma(1)r},\ldots,u_{\sigma(k)r})$, $r=1,\ldots,m$, $\underline{v}_m = (\underline{u}_1,\ldots,\underline{u}_m)$. Let $\sigma(s_{m+1})$ have an analogous meaning. Then

(2)
$$\gamma_{\sigma(S_m)}(\underline{v}_m) = \gamma_{S_m}(\sigma(\underline{v}_m)),$$
$$\varphi_{r_{m+1};\sigma(S_m)}(\underline{v}_m) = \varphi_{r_{m+1};S_m}(\sigma(\underline{v}_m)),$$
$$\psi_{\sigma(s_{m+1});r_{m+1},\sigma(S_m)}(\underline{v}_m) = \psi_{s_{m+1};r_{m+1},S_m}(\sigma(\underline{v}_m)),$$

and the conditions for $\tilde{\varphi}$ and $\tilde{\psi}$ are the same as for φ and ψ, respectively.

Remark 1. The symbol σ is being used simultaneously for a permutation of $(1,\ldots,k)$ as well as for several other operations. There should, however, be no confusion in the sequel, since the argument of $\sigma(\cdot)$ always will indicate in a natural way which operation is meant in the context.

Many procedures of the above type have been proposed in the literature. A few examples for the non-control case will be given in Section 3. An example for the control case is considered in Gupta and Miescke (1982a) where 2-stage procedures are studied. Further references and examples can be found in Bechhofer, Kiefer and Sobel (1968) and Gupta and Panchapakesan (1979). In contrast to non-control settings, in control problems (where $\theta_1,\ldots,\theta_k$ are to be compared with a control value θ_0), the empty set may be selected finally with the interpretation that "no population is better than the control".

A first step towards reasonable procedures is to find appropriate candidates for decision functions ψ and $\tilde{\psi}$. The present paper is focussing on that point. It will be shown that two natural versions ψ^* and $\tilde{\psi}^*$ (cf. Definition 3) are optimum under fairly general assumptions on the loss functions. Since cost of sampling has no influence on these results, no assumptions in this respect are made explicitly in the following. It should be pointed out, however, that in subsequent steps, where γ, φ and $\tilde{\varphi}$ are considered, cost of sampling would play a crucial role for finding optimum procedures.

Assumption (L1) (Loss structure).

For $m \geq 1$, let $L_m(\underline{\theta}, S_{m+1})$ be a real-valued loss which occurs at $\underline{\theta} \in \Omega^k$, if at Stage m the procedure stops at the subset-configuration S_{m+1}. Let L_m be permutation invariant and favoring parameters with large values. More precisely, let

(3a) $$L_m(\underline{\theta}, \sigma(S_{m+1})) = L_m(\sigma(\underline{\theta}), S_{m+1}),$$

for every permutation σ of $(1,\ldots,k)$, and

(3b) $$L_m(\underline{\theta}, \tilde{S}_{m+1}) \leq L_m(\underline{\theta}, S_{m+1}),$$

if the following holds for one pair (i,j) with $\theta_i \leq \theta_j$: For every $q \in \{1,\ldots,m+1\}$ with $i \in s_q$ and $j \notin s_q$, $\tilde{s}_q = (s_q \setminus \{i\}) \cup \{j\}$, and $\tilde{s}_q = s_q$, otherwise.

Assumption (3b) states that it is worthwhile in terms of loss to exchange the roles of two populations in a sequence of selected subsets $S_{m+1} = (s_1,\ldots,s_{m+1})$ if the better of the two populations is eliminated at an earlier stage than the worse one.

The main purpose of this paper is to show that under fairly general conditions the natural candidates for ψ and $\tilde{\psi}$, ψ^* and $\tilde{\psi}^*$, say, to be defined below, are optimal with respect to the risk (expected loss) or the Bayes risk.

Definition 3. (ψ^* and $\tilde{\psi}^*$).

For every fixed $m \geq 1$, S_m, $r_{m+1} \leq |s_m|$ and $\underline{v}_m = (\underline{u}_1,\ldots,\underline{u}_m) \in \mathbb{R}^{km}$, let $\psi^*_{s_{m+1};r_{m+1},S_m}(\underline{v}_m)$ be equal to a positive constant for all $s_{m+1} \subseteq s_m$ with $|s_{m+1}| = r_{m+1}$, which satisfy $\max\{w_{im} | i \in s_m \setminus s_{m+1}\} \leq \min\{w_{jm} | j \in s_{m+1}\}$, and let it be equal to 0 otherwise. Thus, $s_{m+1} \subseteq s_m$ with $|s_{m+1}| = r_{m+1}$ is selected if it is associated with the r_{m+1} largest values of $w_{im} = u_{i1} + \ldots + u_{im}$, $i \in s_m$, where ties are broken at random. Let $\tilde{\psi}^* = \psi^*$ be similar.

2. Auxiliary Results.

Since by assumption (3a) the loss is permutation invariant, every permutation invariant sequential selection procedure $\wp = (\gamma, \varphi, \tilde{\varphi}, \psi, \tilde{\psi})$ has a risk (expected loss) $R(\underline{\theta}, \wp)$, say, at $\underline{\theta} \in \Omega^k$, which is likewise permutation invariant, i.e.

(4) $\quad R(\underline{\theta}, \wp) = R(\sigma(\underline{\theta}), \wp)$, for every permutation σ of $(1,\ldots,k)$.

Here and in the sequel we assume that the risk always exists, a condition which is met at least for every truncated procedure, where the action space is finite. (4) can be rewritten as

(5) $$R(\underline{\theta}, \wp) = \sum_{\sigma} R(\sigma(\underline{\theta}), \wp)\,(k!)^{-1}, \quad \underline{\theta} \in \Omega^k,$$

where the sum is taken over all permutations σ of $(1,\ldots,k)$. Thus for every fixed $\underline{\theta} \in \Omega^k$, $R(\underline{\theta}, \wp)$ can also be interpreted as the Bayes risk for that prior which gives equal mass $1/k!$ to every point $\sigma(\underline{\theta})$, σ permutation of $(1,\ldots,k)$. It will prove useful and interesting on its own to study the form of Bayes procedures with respect to any permutation invariant prior τ, say, which is defined on $\mathcal{B}(\Omega^k)$, the Borel sets of Ω^k. By doing this in the following, we will assume tacitly that the loss as a function of $\underline{\theta} \in \Omega^k$ always is measurable and integrable properly. In the Bayes approach, the parameter vector is viewed to be random, denoted by $\underline{\Theta} = (\Theta_1, \ldots, \Theta_k)$ in the sequel, which has the probability distribution τ. The Bayes risk of a procedure $\wp$ under a (permutation invariant) prior then is given by

(6) $$r(\tau, \wp) = E[R(\underline{\Theta}, \wp)] = \int_{\Omega^k} R(\underline{\theta}, \wp)\,d\tau(\underline{\theta}).$$

When studying the form of Bayes rules, typically posterior expectations and the technique of backward induction will be applied. To simplify the derivation of the main results, some useful facts will now be presented and proved separately for convenience.

To begin with, let us consider a fundamental property of multivariate distributions which was called "property M" by Eaton (1967) and, more recently, "decreasing in transposition property (DT)" by Hollander, Proschan and Sethuraman (1977). Let $\mathcal{B}(\cdot)$ stand for "Borel sets of" in the following.

Definition 4. (Decreasing in transposition property (DT)).
Let $A,B \in \mathcal{B}(\mathbb{R})$. A function $h: A^k \times B^k \to \mathbb{R}$ is said to be decreasing in transposition (DT), if for every fixed $\underline{a} \in A^k$, $\underline{b} \in B^k$,

(7a) $h(\underline{a},\underline{b}) = h(\sigma(\underline{a}),\sigma(\underline{b}))$, for every permutation σ of $(1,\ldots,k)$, and

(7b) $h(\underline{a},\underline{b}) \leq h(\underline{a},\tilde{\sigma}(\underline{b}))$, if for some permutation $\tilde{\sigma}$ and $i, j \in \{1,\ldots,k\}$, $(a_i-a_j)(b_i-b_j) \leq 0$, $\tilde{\sigma}(i) = j$, $\tilde{\sigma}(j) = i$, and $\tilde{\sigma}(r) = r$ for $r \neq i,j$.

A family $\left\{P_{\underline{b}}\right\}_{\underline{b} \in B^k}$ of probability measures on $\mathcal{B}(A^k)$ is said to be decreasing in transposition (DT) if for every $\underline{b} \in B^k$, $P_{\underline{b}}$ has a density $h_{\underline{b}}(\underline{a})$, $\underline{a} \in A^k$, with respect to a permutation invariant sigma-finite measure ν on $\mathcal{B}(A^k)$, such that h is DT .

Lemma 1. Let $A, B \in \mathcal{B}(\mathbb{R})$. If a family $\left\{P_{\underline{b}}\right\}_{\underline{b} \in B^k}$ of probability measures on $\mathcal{B}(A^k)$ is DT, then the posterior family with respect to every permutation invariant prior on $\mathcal{B}(B^k)$ also is DT.

Proof: According to Definition 4, let $P_{\underline{b}}$ have a density $h_{\underline{b}}$ with respect to ν, $\underline{b} \in B^k$, and let ρ be a permutation invariant prior on $\mathcal{B}(B^k)$. Then, at $\underline{a} \in A^k$, the posterior distribution has a density $g_{\underline{a}}$ with respect to ρ, which at $\underline{b} \in B^k$ is given by

(8) $$g_{\underline{a}}(\underline{b}) = h_{\underline{b}}(\underline{a})q(\underline{a}), \text{ where } q(\underline{a}) = 1/\int_{B^k} h_{\underline{e}}(\underline{a})d\rho(\underline{e}).$$

Since $q(\underline{a}) = q(\sigma(\underline{a}))$ for every permutation σ of $(1,\ldots,k)$, it is easy to see that g is DT . Thus, the proof is completed.

The next fact is more closely related to the setting of the sequential selection problem under consideration.

Lemma 2. Let $m \geq 1$ be fixed and let τ be a permutation invariant prior of $\underline{\theta}$ on $\mathcal{B}(\Omega^k)$. Based on the joint distribution of $(\underline{\theta}, \underline{W}_m, \underline{W}_{m+1})$, let $P^{\tau}_{\underline{w}}$ denote the conditional distribution of $\underline{W}_{m+1}$, given $\underline{W}_m = \underline{w}$, $\underline{w} \in \mathcal{X}^k$. If the function $d(x)$, $x \in \mathcal{X}$, is log-concave, i.e. if the basic underlying exponential family $\mathfrak{F}$ is strongly unimodal, then the family $\left\{P^{\tau}_{\underline{w}}\right\}_{\underline{w} \in \mathcal{X}^k}$ is DT.

Proof: Let d be log-concave. Clearly, this holds true if and only if in the family $\mathfrak{F}$, every density $c(\theta)\exp(\theta x)d(x)$, $x \in \mathcal{X}$, is log-concave, $\theta \in \Omega$. Since in the discrete as well as in the continuous case, log-concavity of densities with respect to μ is being preserved under convolutions (cf. Barndorff-Nielsen (1978)), the function $d_r(x)$, $x \in \mathcal{X}$, is log-concave for every $r \in \mathbb{N}$.

Let $m \geq 1$ be fixed and let τ be a permutation invariant prior on $\mathcal{B}(\Omega^k)$. The joint (marginal) distribution of $(\underline{U}_{m+1}, \underline{W}_m)$ in view of (1) has the following density with respect to μ_k.

$$(9) \qquad \delta^{(m+1)}(\underline{u},\underline{w}) = \int_{\Omega^k} f_{\underline{\theta}}^{(m+1)}(\underline{u}) g_{\underline{\theta}}^{(m)}(\underline{w}) d\tau(\underline{\theta}), \ \underline{u}, \ \underline{w} \in \mathcal{X}^k.$$

Therefore, the conditional distribution of $\underline{W}_{m+1}$, given $\underline{W}_m = \underline{w}$, has the density with respect to μ_k,

$$(10) \qquad \xi^{(m+1)}(\underline{z}|\underline{w}) = \delta^{(m+1)}(\underline{z}-\underline{w},\underline{w}) / \int_{\Omega^k} g_{\underline{\theta}}^{(m)}(\underline{w}) d\tau(\underline{\theta}), \ \underline{z} \in \mathcal{X}^k.$$

After inserting the exponential families (1) into (9) and (10), one gets

$$\xi^{(m+1)}(\underline{z}|\underline{w}) = \alpha_{N_{m+1}}(\underline{z})\beta_{\underline{w}}(\underline{z})\alpha_{N_m}(\underline{w})^{-1}, \ \underline{z}\in\mathcal{X}^k, \tag{11}$$

where $\alpha_r(\underline{x}) = \int_{\Omega^k} \prod_{i=1}^{k} c_r(\theta_i)\exp(\theta_i x_i)d\tau(\underline{\theta}), \ \underline{x}\in\mathcal{X}^k, \ r\in\mathbb{N}$,

and $\beta_{\underline{w}}(\underline{z}) = \prod_{i=1}^{k} d_{n_{m+1}}(z_i-w_i), \ \underline{z}\in\mathcal{X}^k$.

Obviously, the functions α_r are permutation invariant. Moreover, standard arguments show that β is DT if and only if $d_{n_{m+1}}$ is log-concave. Since the latter is given, $\xi^{(m+1)}$ is DT and the proof is completed.

In the remainder of this section, it will be shown that the loss structure, which is described in Assumption (L1), is preserved under certain operations. First a slightly more general and closely related structure will be introduced for convenience.

<u>Definition 5. (Property $\mathcal{D}$ (m,A)).</u>

Let $m \geq 1$ and $A\in \mathcal{B}(\mathbb{R})$ be fixed. For every m+1 disjoint subsets $t_1,\ldots,t_{m+1} \subseteq \{1,\ldots,k\}$ with $t_1 \cup \ldots \cup t_{m+1} = \{1,\ldots,k\}$, let $T_{m+1} = (t_1,\ldots,t_{m+1})$, and let $\mathcal{L}_m(\underline{a},T_{m+1})$, $\underline{a}\in A^k$, be a real valued measurable function of $\underline{a}$. $\mathcal{L}_m$ is said to have Property $\mathcal{D}$ (m,A), if for every $\underline{a}\in A^k$ and T_{m+1} the following two conditions are satisfied.

$$\mathcal{L}_m(\underline{a},\sigma(T_{m+1})) = \mathcal{L}_m(\sigma(\underline{a}),T_{m+1}), \tag{12a}$$

$\sigma(T_{m+1}) = (\sigma(t_1),\ldots,\sigma(t_{m+1}))$, for every permutation σ of $(1,\ldots,k)$, and

$$\mathcal{L}_m(\underline{a},\tilde{T}_{m+1}) \leq \mathcal{L}_m(\underline{a},T_{m+1}), \tag{12b}$$

if the following holds for one pair (i,j) with $a_i \leq a_j$: There exist integers $\alpha < \beta \leq m + 1$, such that $i\in t_\beta$, $j\in t_\alpha$, $\tilde{t}_\alpha = (t_\alpha\setminus\{j\})\cup\{i\}$, $\tilde{t}_\beta = (t_\beta\setminus\{i\})\cup\{j\}$, and $\tilde{t}_q = t_q$ for $q \neq \alpha,\beta$.

Remark 2. The relationship between the assumed loss structure (L1) and functions which have Property $\mathscr{D}(m,\Omega)$, $m \geq 1$, is of a fairly natural type. Let $m \geq 1$ be fixed and let $L_m(\underline{\theta},S_{m+1})$ be the loss at $\underline{\theta} \in \Omega^k$ for $S_{m+1} = (s_1,\ldots,s_{m+1})$, $s_1 \supseteq \ldots \supseteq s_{m+1}$, at Stage m. Let $\mathscr{T}_m(S_{m+1}) = (s_1 \backslash s_2, s_2 \backslash s_3, \ldots, s_m \backslash s_{m+1}, s_{m+1})$ $= (t_1,\ldots,t_{m+1}) = T_{m+1}$, say. Then, $t_1,\ldots,t_m$ are the populations which have been eliminated at Stages $1,\ldots,m$, and t_{m+1} are the populations which are selected at the end of Stage m. Now, let $\mathscr{L}_m(\underline{\theta},T_{m+1}) = L_m(\underline{\theta},S_{m+1})$, $\underline{\theta} \in \Omega^k$. Then it is easy to see that L_m satisfies the loss assumptions (3) if and only if $\mathscr{L}_m$ has Property $\mathscr{D}(m,\Omega)$.

Lemma 3. Let $m \geq 1$ and $A, B \in \mathscr{B}(\mathbb{R})$ be fixed, and let $\mathscr{L}_m$ have Property $\mathscr{D}(m,A)$. Let $\{h_{\underline{b}}\}_{\underline{b} \in B}$ be a family of densities with respect to a permutation invariant sigma-finite measure ν on $\mathscr{B}(A^k)$, where h is DT . For every T_{m+1}, let

$$(13) \qquad \tilde{\mathscr{L}}_m(\underline{b},T_{m+1}) = \int_{A^k} \mathscr{L}_m(\underline{a},T_{m+1}) h_{\underline{b}}(\underline{a}) d\nu(\underline{a}), \ \underline{b} \in B^k.$$

Then $\tilde{\mathscr{L}}_m$ has Property $\mathscr{D}(m,B)$.

Proof: Let $k_0 = 0$ and $k_1,\ldots,k_{m+1} \in \{1,\ldots,k\}$ with $k_1 + \ldots + k_{m+1} = k$ be fixed. For every $T_{m+1} = (t_1,\ldots,t_{m+1})$ with $|t_r| = k_r$, $r=1,\ldots,m+1$, and every $\underline{a} \in A^k$, let $K_r = k_0 + \ldots + k_r$, $r = 0,1,\ldots,m+1$, and

$$(14) \qquad \mathscr{K}_{\underline{k}}(\underline{a},(\sigma(1),\ldots,\sigma(k))) = -\mathscr{L}_m(\underline{a},T_{m+1}), \ \underline{k} = (k_1,\ldots,k_{m+1}),$$

for all permutations σ of $(1,\ldots,k)$ with $t_r = \{\sigma(K_{r-1}+1),\ldots,\sigma(K_r)\}$, $r = 1, \ldots, m+1$. Let $E = \{1,\ldots,k\}$ and take the following auxiliary function $\mathscr{K}_{\underline{k}}: A^k \times E^k \to \mathbb{R}$, where for every $\underline{a} \in A^k$, $\mathscr{K}_{\underline{k}}(\underline{a},\underline{e})$ is defined by (14) if $\underline{e} \in E^k$ is a permutation of $(1,\ldots,k)$, and where $\mathscr{K}_{\underline{k}}(\underline{a},\underline{e}) = 0$, otherwise. Let $\tilde{\mathscr{K}}_{\underline{k}}$ be defined analogously with respect to $\tilde{\mathscr{L}}_m$. Then an equation analogous to (13) holds for $\tilde{\mathscr{K}}_{\underline{k}}$ and $\mathscr{K}_{\underline{k}}$.

Now let $\mathcal{L}_m$ have Property $\mathcal{D}(m,A)$. Then, apparently, $\mathcal{K}_{\underline{k}}$ is DT. Thus if h is DT, by Theorem 3.3 of Hollander, Proschan and Sethuraman (1977), $\tilde{\mathcal{K}}_{\underline{k}}$ also is DT. Therefore, $\tilde{\mathcal{L}}_m$ has the properties (12a) and (12b) for all T_{m+1} with $|t_r| = k_r$, $r=1,\ldots,m+1$, and all $\underline{b} \in B$. Since this holds true for every $\underline{k}$ as specified at the beginning of the proof, it follows that $\tilde{\mathcal{L}}_m$ has Property $\mathcal{D}(m,B)$. Thus the proof is completed.

Remark 3. Eaton (1967) considered 1-Stage procedures that select (in the present notation) the k_{m+1} best, k_m second best,...,k_1 worst populations, where $k_1,\ldots,k_{m+1}$ are fixed and predetermined with $k_1 + \ldots + k_{m+1} = k$. His loss assumptions are analogous to Property $\mathcal{D}(m,\Omega)$, where (12b), however, is assumed to hold only for $\alpha = \beta - 1$. Eaton's (1967) main result states that the natural rule is uniformly best in terms of risk, and it may be interesting to note that his proof is essentially a combination of Lemma 1 and Lemma 3 of the present paper. Further details are given in Remark 5.

Lemma 4. Let $m \geq 1$ and $A \in \mathcal{B}(\mathbb{R})$ be fixed. Let $\mathcal{L}_m$ have Property $\mathcal{D}(m,A)$. For every disjoint $t_1,\ldots,t_m \subseteq \{1,\ldots,k\}$ with $t_1 \cup \ldots \cup t_m = \{1,\ldots,k\}$, let $T_m = (t_1,\ldots,t_m)$ and

$$(15)\qquad \mathcal{L}_{m-1}(\underline{a},T_m) = \min\{\mathcal{L}_m(\underline{a},(t_1,\ldots,t_{m-1},\hat{t}_m,\hat{t}_{m+1}))\mid \hat{t}_m \cup \hat{t}_{m+1} = t_m,\ \hat{t}_m \cap \hat{t}_{m+1} = \emptyset\},\ \underline{a} \in A^k.$$

Then $\mathcal{L}_{m-1}$ has Property $\mathcal{D}(m-1,A)$.

Proof: Let $\underline{a} \in A^k$ and T_m, as specified in Lemma 4, be fixed. Then for every permutation σ of $(1,\ldots,k)$,

$$\begin{aligned}
&\mathcal{L}_{m-1}(\sigma(\underline{a}),T_m)\\
&\quad= \min\{\mathcal{L}_m(\underline{a},\sigma(t_1,\dots,t_{m-1},\hat{t}_m,\hat{t}_{m+1}))\,|\,\hat{t}_m\cup\hat{t}_{m+1}=t_m,\hat{t}_m\cap\hat{t}_{m+1}=\emptyset\}\\
&\quad= \min\{\mathcal{L}_m(\underline{a},(\sigma(t_1),\dots,\sigma(t_{m-1}),t^*_m,t^*_{m+1}))\,|\,t^*_m\cup t^*_{m+1}=\sigma(t_m),t^*_m\cap t^*_{m+1}=\emptyset\}\\
&\quad= \mathcal{L}_{m-1}(\underline{a},\sigma(T_m)),
\end{aligned}$$

where the first equality follows from the invariance property (12a) of $\mathcal{L}_m$. Thus, $\mathcal{L}_{m-1}$ has the analogous invariance property.

Additionally, let a pair (i,j) be fixed with $a_i \le a_j$, for which there exist $\alpha < \beta \le m$ with $i \in t_\beta$ and $j \in t_\alpha$. Let $\tilde{T} = (\tilde{t}_1,\dots,\tilde{t}_m)$ with $\tilde{t}_\alpha = (t_\alpha\backslash\{j\})\cup\{i\}$, $\tilde{t}_\beta = (t_\beta\backslash\{i\})\cup\{j\}$, and $\tilde{t}_q = t_q$ for $q \ne \alpha,\beta$. Two cases are considered separately.

Case 1: $\beta \le m-1$. Since in this case $t_m = \tilde{t}_m$, it follows that for all disjoint $\hat{t}_m,\hat{t}_{m+1}$ with $\hat{t}_m\cup\hat{t}_{m+1} = t_m$,

$$\mathcal{L}_m(\underline{a},(\tilde{t}_1,\dots,\tilde{t}_{m-1},\hat{t}_m,\hat{t}_{m+1})) \le \mathcal{L}_m(\underline{a},(t_1,\dots,t_{m-1},\hat{t}_m,\hat{t}_{m+1}))$$

holds, and therefore $\mathcal{L}_{m-1}(\underline{a},\tilde{T}_m) \le \mathcal{L}_{m-1}(\underline{a},T_m)$.

Case 2: $\beta = m$. Let $\hat{t}_m,\hat{t}_{m+1}$ be disjoint with $\hat{t}_m\cup\hat{t}_{m+1} = t_m$. If $i\in\hat{t}_m$, let $\bar{t}_m = (\hat{t}_m\backslash\{i\})\cup\{j\}$ and $\bar{t}_{m+1} = \hat{t}_{m+1}$, and if $i\in\hat{t}_{m+1}$, let $\bar{t}_{m+1} = (\hat{t}_{m+1}\backslash\{i\})\cup\{j\}$ and $\bar{t}_m = \hat{t}_m$. Then in either case,

$$\mathcal{L}_m(\underline{a},(\tilde{t}_1,\dots,\tilde{t}_{m-1},\bar{t}_m,\bar{t}_{m+1})) \le \mathcal{L}_m(\underline{a},(t_1,\dots,t_{m-1},\hat{t}_m,\hat{t}_{m+1}))$$

where $\bar{t}_m,\bar{t}_{m+1}$ are disjoint with $\bar{t}_m\cup\bar{t}_{m+1} = \tilde{t}_m$. This implies $\mathcal{L}_{m-1}(\underline{a},\tilde{T}_m) \le \mathcal{L}_{m-1}(\underline{a},T_m)$, and thus the proof is completed.

Remark 4. If a special sequential selection problem is given under certain restrictions concerning the sizes of subsets to be selected (i.e., if there are side-conditions with respect to φ or $\tilde{\varphi}$), then an analogous result to that of Lemma 4 can be proved in essentially the same way. The minimum in

(15) has then to be taken additionally subject to these restrictions and some obvious changes have to be made in the proof.

3. The Main Results.

In this section, permutation invariant sequential selection procedures which are truncated will be studied, i.e. procedures which stop no later than Stage q, say. Results which as well hold true for untruncated procedures will be so indicated. The loss is assumed to satisfy Assumption (L1) given in Section 1. To begin with, consider the risk function for procedure $\wp = (\gamma, \varphi, \tilde{\varphi}, \psi, \tilde{\psi})$ at $\underline{\theta} \in \Omega^k$.

$$(16) \qquad R(\underline{\theta}, \wp) = \sum_{i=1}^{q} \sum_{S_{i+1}} L_i(\underline{\theta}, S_{i+1}) E_{\underline{\theta}} \{$$

$$\prod_{m=1}^{i-1} [1-\gamma_{S_m}(\underline{V}_m)] \tilde{\varphi}_{|s_{m+1}|;\, S_m(\underline{V}_m)} \tilde{\psi}_{s_{m+1};\, |s_{m+1}|, S_m(\underline{V}_m)}$$

$$\times\ \gamma_{S_i}(\underline{V}_i) \varphi_{|s_{i+1}|;\, S_i(\underline{V}_i)} \psi_{s_{i+1};\, |s_{i+1}|, S_i(\underline{V}_i)} \}$$

where the second sum is with respect to $S_{i+1} = (s_1, \ldots, s_{i+1})$ with $s_{i+1} \subseteq s_i \subseteq \cdots \subseteq s_1 = \{1, \ldots, k\}$ and $s_i \neq \emptyset$, and where $\gamma_{S_q} \equiv 1$.

The Bayes risk with respect to a permutation invariant prior τ on $\mathcal{B}(\Omega^k)$ will be studied in the sequel according to the rationale given at the beginning of Section 2. It is assumed that the Bayes risk (6) exists. As has been pointed out, this condition is met if τ has a finite support. By standard techniques the Bayes risk can be seen to be of the following form. For notational convenience, let $E^{|\underline{V}_m}$ denote the conditional expectation, given $\underline{V}_m$, m=1,...,q. Then

(17) $r(\tau,\wp) =$

$$= E[\gamma_{S_1}(\underline{V}_1) \sum_{r_2=0}^{r_1} \varphi_{r_2;S_1}(\underline{V}_1) \sum_{\substack{s_2 \subseteq s_1 \\ |s_2|=r_2}} \psi_{s_2;r_2,S_1}(\underline{V}_1) E^{|\underline{V}_1}[L_1(\underline{\theta},S_2)]$$

$$+ (1-\gamma_{S_1}(\underline{V}_1)) \sum_{r_2=1}^{r_1} \tilde{\varphi}_{r_2;S_1}(\underline{V}_1) \sum_{\substack{s_2 \subseteq s_1 \\ |s_2|=r_2}} \tilde{\psi}_{s_2;r_2,S_1}(\underline{V}_1) \times \cdots$$

$$\times E^{|\underline{V}_{m-1}}[\gamma_{S_m}(\underline{V}_m) \sum_{r_{m+1}=0}^{r_m} \varphi_{r_{m+1};S_m}(\underline{V}_m) \sum_{\substack{s_{m+1} \subseteq s_m \\ |s_{m+1}|=r_{m+1}}} \psi_{s_{m+1};r_{m+1},S_m}(\underline{V}_m) E^{|\underline{V}_m}[L_m(\underline{\theta},S_{m+1})]$$

$$+ (1-\gamma_{S_m}(\underline{V}_m)) \sum_{r_{m+1}=1}^{r_m} \tilde{\varphi}_{r_{m+1};S_m}(\underline{V}_m) \sum_{\substack{s_{m+1} \subseteq s_m \\ |s_{m+1}|=r_{m+1}}} \tilde{\psi}_{s_{m+1};r_{m+1},S_m}(\underline{V}_m) \times \cdots$$

$$\times E^{|\underline{V}_{q-1}}[\sum_{r_{q+1}=0}^{r_q} \varphi_{r_{q+1};S_q}(\underline{V}_q) \sum_{\substack{s_{q+1} \subseteq s_q \\ |s_{q+1}|=r_{q+1}}} \psi_{s_{q+1};r_{q+1},S_q}(\underline{V}_q) E^{|\underline{V}_q}[L_q(\underline{\theta},S_{q+1})]]\cdots]\cdots].$$

Both (16) and (17) hold for untruncated procedures which stop almost certainly in finitely many steps, provided of course that the Bayes risk exists. One simply has to take $q = \infty$ in (16) and to omit the last factor in (17) which is associated with Stage q.

The first main result is with respect to the final decision and is the following.

Theorem 1. Let $\mathcal{P} = (\gamma, \varphi, \tilde{\varphi}, \psi, \tilde{\psi})$ be a permutation invariant, truncated or untruncated, sequential selection procedure, and let $\mathcal{P}^* = (\gamma, \varphi, \tilde{\varphi}, \psi^*, \tilde{\psi})$. Then under the assumptions concerning the loss and distributions which have been made at Section 1,

$$(18) \qquad R(\underline{\theta}, \mathcal{P}^*) \leq R(\underline{\theta}, \mathcal{P}), \text{ for all } \underline{\theta} \in \Omega^k.$$

Moreover, if $\mathcal{P}$ is truncated, $R(\underline{\theta}, \mathcal{P}) < \infty$ for all $\underline{\theta} \in \Omega^k$.

Proof: Let $m \geq 1$ be fixed. Since, at Stage m, $\underline{W}_m$ is sufficient for $\underline{\theta} \in \Omega^k$, in (17) for every S_{m+1},

$$(19) \qquad E^{|\underline{V}_m}[L_m(\underline{\Theta}, S_{m+1})] = E^{|\underline{W}_m}[L_m(\underline{\Theta}, S_{m+1})],$$

can be seen to hold almost surely, since the l.h.s. of (19) is a measurable function of $\underline{W}_m$.

In view of (1), under $\underline{\Theta} = \underline{\theta}$, $\underline{\theta} \in \Omega^k$, $\underline{W}_m$ has a density with respect to μ_k, $g_{\underline{\theta}}^{(m)}(\underline{w})$, $\underline{w} \in \mathcal{X}^k$, which is DT . Let τ be a permutation invariant prior on $\mathcal{B}(\Omega^k)$ for which the (truncated or untruncated) Bayes risk $r(\tau, \mathcal{P})$ exists. Then, by Lemma 1, the posterior distribution of $\underline{\Theta}$, given $\underline{W}_m = \underline{w}$, $\underline{w} \in \mathcal{X}^k$, also is DT .

According to Remark 2, for $\underline{\theta} \in \Omega^k$ and S_{m+1} let $T_{m+1} = \mathcal{T}_m(S_{m+1})$ and $\mathcal{L}_m(\underline{\theta}, T_{m+1}) = L_m(\underline{\theta}, S_{m+1})$. Then, as noted there, $\mathcal{L}_m$ has Property $\mathcal{D}(m, \Omega)$. Let

$$(20) \qquad \tilde{\mathcal{L}}_m(\underline{w}, T_{m+1}) = E^{|\underline{W}_m = \underline{w}}[\mathcal{L}_m(\underline{\Theta}, T_{m+1})], \ \underline{w} \in \mathcal{X}^k.$$

By Lemma 3, $\tilde{\mathcal{L}}_m$ has Property $\mathcal{D}(m, \mathcal{X})$. Therefore it is easy to see that for every fixed $s_1, \ldots, s_m$ and $r_{m+1} \leq |s_m|$, (20) is minimized subject to

$s_{m+1} \subseteq s_m$ and $|s_{m+1}| = r_{m+1}$, for those s_{m+1} which are associated with r_{m+1} of the largest w_i, $i \in s_m$. Since now $\psi^*_{\cdot;r_{m+1},S_m}(\underline{w})$ gives equal mass to all such subsets and no mass to others, it follows that $r(\tau,\wp^*) \leq r(\tau,\wp)$.

Let $\underline{\theta} \in \Omega^k$ be fixed and let τ be the prior which gives mass $1/k!$ to all points $\sigma(\underline{\theta})$, σ permutation of $(1,\ldots,k)$. Then by (5) it follows that $r(\tau,\wp) = R(\underline{\theta},\wp)$ and $r(\tau,\wp^*) = R(\underline{\theta},\wp^*)$. Therefore, (18) holds, and the last statement in Theorem 1 follows directly from (16).

In the remainder of this paper, four applications of the basic result given in Theorem 1 will be studied.

Application 1: Procedures with Vector at a Time Sampling.

Assume that at every Stage m, samples of size n_m are drawn from all populations, until the procedure stops and makes a final decision. Thus, for every $m \geq 1$, the complete vector $\underline{U}_m$ is observed and $\tilde{\varphi}_{r_{m+1};S_m} = 1(0)$ if $r_{m+1} = (\neq)k$, for every S_m. Then, as an immediate consequence of Theorem 1, the following holds.

Corollary 1. For every permutation invariant procedure, no matter which stopping rule is used, in the truncated as well as in the untruncated case, the natural final decision ψ^* always is uniformly optimal in the sense of (18).

A great variety of procedures for several goals and loss functions which fit into this framework are covered by Bechhofer, Kiefer and Sobel (1968); most of their procedures have the restriction $n_1 = n_2 = \ldots$

and do not eliminate (vector at a time sampling). In all of their proposed procedures, the natural final decision rule is taken as the "terminal decision rule". The results stated above confirm that this is optimal in the sense of (18), uniformly in $\underline{\theta} \in \Omega^k$.

Example 1. Barron and Gupta (1972) have proposed a procedure to find a subset of normal populations (with unknown means and a common known variance) which contains the best population with a probability no less than a given P*. The procedure is of the sequential type, uses vector at a time sampling, but does not make the natural final decisions. Instead, populations are marked "rejected" or "accepted" at various stages according to a specified rule until all populations are marked, at which time the procedure stops. In view of the results stated above, such a procedure can be improved in terms of the probability of a correct selection, and thereby retaining the P*-condition, by simply replacing the finally selected populations by a subset of populations of the same size, which are associated with the largest overall means.

Application 2: q-Stage Procedures with Fixed Subset-Size at each Stage. Assume that the number of stages q, say, is predetermined, and that the size of the subset to be selected at Stage m, R_{m+1}, say, is fixed in advanced as well, m=1,...,q. Thus, $k = R_1 \geq \dots \geq R_{q+1}$, $\gamma_{S_1} = \dots = \gamma_{S_{q-1}} = 1 - \gamma_{S_q} = 0$ and $\tilde{\varphi}_{R_2;S_1} = \dots = \tilde{\varphi}_{R_q;S_{q-1}} = 1$ and $\varphi_{R_{q+1};S_q} = 1$. In this case it can be shown that the natural procedure is uniformly optimal provided $\mathfrak{F}$ is strongly unimodal.

Theorem 2. Let $\mathcal{P} = (\gamma, \varphi, \tilde{\varphi}, \psi, \tilde{\psi})$ be permutation invariant, where γ, φ and $\tilde{\varphi}$ are given as specified above, and let $\mathcal{P}^* = (\gamma, \varphi, \tilde{\varphi}, \psi^*, \tilde{\psi}^*)$. If

<u>the basic underlying exponential family</u> $\mathfrak{F}$ <u>is strongly unimodal and the loss satisfies the assumption</u> (L1), <u>then</u>

(21) $$R(\underline{\theta},\mathcal{P}^*) \leq R(\underline{\theta},\mathcal{P}), \text{ for all } \underline{\theta}\in\Omega^k,$$

<u>i.e.</u> $\mathcal{P}^*$ <u>is uniformly optimal in the given subclass of procedures</u>.

<u>Proof:</u> <u>(Backward Induction)</u>. Let τ be any fixed permutation invariant prior on $\mathcal{B}(\Omega^k)$ which has a finite support. Consider (17) for any procedure $\mathcal{P} = (\gamma,\varphi,\tilde{\varphi},\psi,\tilde{\psi})$, where γ,φ and $\tilde{\varphi}$ are given as specified above. Clearly, the Bayes risk $r(\tau,\mathcal{P})$ exists.

We start at Stage q, the final stage. Here, by Theorem 1, the corresponding component of ψ^* is optimal. Let $S_q = (s_1,\ldots,s_q)$ with $|s_1| = R_1$, $\ldots,|s_q| = R_q$ and $s_1 \supseteq \ldots \supseteq s_q$ be fixed. According to Remark 2, let $(t_1,\ldots,t_q) = T_q = \mathcal{J}_{q-1}(S_q)$. Then, after having inserted the corresponding components of φ and ψ^* into the last line of (17), and after having replaced $E^{|\underline{V}_{q-1}}$ by $E^{|\underline{W}_{q-1}}$ (the reasons are the same as were used for (19)), the last factor in (17) which is associated with Stage q can be seen to be of the form

(22) $$\tilde{\mathcal{H}}_{q-1}(\underline{W}_{q-1},T_q) = E^{|\underline{W}_{q-1}}[\mathcal{H}_{q-1}(\underline{W}_q,T_q)], \text{ where}$$

$$\mathcal{H}_{q-1}(\underline{w},T_q) = \min\{\tilde{\mathcal{L}}_q(\underline{w},(t_1,\ldots,t_{q-1},\hat{t}_q,\hat{t}_{q+1}))|\hat{t}_q\cap\hat{t}_{q+1}=\emptyset,$$
$$\hat{t}_q\cup\hat{t}_{q+1}=t_q,|\hat{t}_{q+1}|=R_{q+1}\},\ \underline{w}\in\mathcal{X}^k,$$

and where $\tilde{\mathcal{L}}_q$ is defined by (20). The crucial point is that the component of ψ^* for Stage q remains optimal even if the component of ψ at Stage q were allowed to make use of $\underline{V}_q$, the complete vector of all samples.

As mentioned in the proof of Theorem 1, $\tilde{\mathcal{L}}_q$ has Property $\mathcal{D}(q,\mathcal{X})$ by Lemma 3. From Lemma 4 and the subsequent Remark 4 it follows that $\mathcal{H}_{q-1}$ has Property $\mathcal{D}(q-1,\mathcal{X})$. Lemma 2 states that the conditional distribution of $\underline{W}_q$, given $\underline{W}_{q-1}$, is DT. Therefore, another application of Lemma 3 implies that $\tilde{\mathcal{H}}_{q-1}$ has Property $\mathcal{D}(q-1,\mathcal{X})$.

Let us assume now that the components of the Bayes rule have been determined for stages $m+1,\ldots,q$ for a fixed $m\in\{1,\ldots,q-1\}$, and that they have been inserted, together with the associated components of γ,φ and $\tilde{\varphi}$, into (17). Let $S_m = (s_1,\ldots,s_m)$ with $|s_1| = R_1,\ldots,|s_m| = R_m$ and $s_1 \supseteq \ldots \supseteq s_m$ be fixed. Similarly as before, let now $T_m = \mathcal{T}_{m-1}(S_m) = (t_1,\ldots,t_m)$ and assume that the m-th line of (17) has been reduced to, say,

$$(23)\quad E^{|\underline{V}_{m-1}}[\sum_{\substack{s_{m+1}\subseteq s_m \\ |s_{m+1}|=R_{m+1}}} \tilde{\psi}_{s_{m+1};R_{m+1},S_m}(\underline{V}_m)\cdot \tilde{\mathcal{H}}_m(\underline{W}_m,(t_1,\ldots,t_{m-1},s_m\setminus s_{m+1},s_{m+1}))],$$

where $\tilde{\mathcal{H}}_m$ has Property $\mathcal{D}(m,\mathcal{X})$.

Under these conditions, apparently, $\tilde{\psi}^*_{\cdot;R_{m+1},S_m}$ is optimum. Moreover, it can be concluded exactly in the same way as it was done for Stage q, that for the optimum decision function, (23) is a function $\tilde{\mathcal{H}}_{m-1}(\underline{W}_{m-1},T_m)$, say, where $\tilde{\mathcal{H}}_{m-1}$ has Property $\mathcal{D}(m-1,\mathcal{X})$. Therefore, the proof of Theorem 2 can be completed by induction.

<u>Remark 5</u>. All results derived so far hold true if at some of the stages the corresponding sample sizes are taken to be zero. In the present setting, if one takes $n_2 = \ldots = n_q = 0$, then the problem reduces to that one which was studied by Eaton (1967), and Theorem 2 reduces to the main result of Eaton (1967) (cf. Remark 3). Clearly, in this case the assumption of strong unimodality is not needed in the proof of Theorem 2.

<u>Example 2</u>. Let $\pi_1,\ldots,\pi_k$ be normal populations with unknown means $\theta_1,\ldots,\theta_k$ and a common variance. Then at the end of every Stage m the optimum

procedure selects from the populations which have survived so far (i.e., from π_i, $i \in s_m$) the R_{m+1} populations which are associated with the largest overall means.

Somerville (1974) has proposed a 2-stage procedure in this setting with $R_3 = 1$, which differs from the optimum procedure in the second stage. Instead of the overall means, the means of the corresponding observations from Stage 2 only are used. Somerville (1974) states that "intuitively the procedure... is inferior since it ignores information obtained in the first stage." Theorem 2 now confirms this statement and, moreover, it determines the optimum procedure explicitly. This does not diminish the value of Somerville's (1974) results, since they can be used now as approximations for the optimum procedure. The principle here thus is the same as has been used in Example 1: the risk of a procedure using optimal components dominates, uniformly in $\underline{\theta} \in \Omega^k$, the risk of procedures which are modified with respect to these components. On the other hand, lower bounds for, say, the probability of a correct final selection, are usually much easier to compute for such non-optimal procedures, as was mentioned by Somerville (1974). Results in this respect can also be found in Gupta and Miescke (1982b).

Application 3: q-Stage Procedures with Fixed Subset-Size at Stage q. Assume that the number of stages, q, say, is predetermined and that the size of the subset s_{q+1}, to be selected finally at Stage q, is fixed in advance. The proof of the next result is the same as the first part of the proof for Theorem 2 and therefore omitted.

Corollary 2. Let $\wp = (\gamma,\varphi,\tilde{\varphi},\psi,\tilde{\psi})$ be permutation invariant, where γ and φ satisfy the conditions stated above. Let $\wp' = (\gamma,\varphi,\tilde{\varphi},\psi^*,\tilde{\psi}')$, where $\tilde{\psi}'$ is the same as $\tilde{\psi}$ except for Stage q-1: here $\tilde{\psi}'$ has the same component as $\tilde{\psi}^*$. Then under the same assumptions concerning the loss and $\mathfrak{F}$ as in Theorem 2, $R(\underline{\theta},\wp') \leq R(\underline{\theta},\wp)$, for all $\underline{\theta}\in\Omega^k$.

Example 3. Gupta and Miescke (1983) have studied 2-stage procedures for the problem of selecting a best population (if it is sufficiently "good"). Under the same assumptions concerning the loss and the distributions, they have shown that permutation invariant procedures for which the selected subsets at Stage 1 as well as the finally selected population are associated with the largest corresponding sufficient statistics, form an essentially complete class within all permuation invariant procedures. This result can now be seen to be a consequence of Corollary 2. The techniques, on the other hand, which have been used by Gupta and Miescke (1983), are more similar to Eaton's (1967) methods of proofs.

Application 4: Bayes Truncated Procedures under i.i.d. Priors.

Assume that the number of stages is admitted to be at most q, say. Thus $\gamma_{S_q} = 1$ for all S_q. Let τ be an i.i.d. prior, i.e. let $\Theta_1,\ldots,\Theta_k$ be independently identically distributed apriori according to a distribution ρ on $\mathcal{B}(\Omega)$, where $\tau = \rho\times\ldots\times\rho$. Let the basic underlying exponential family $\mathfrak{F}$ be strongly unimodal, and assume that the loss satisfies assumption (L1) as well as the following.

Assumption (L2).

For every $m\in\{1,\ldots,q\}$ and every S_{m+1}, let $L_m(\underline{\theta},S_{m+1})$ be a function of only those θ_i with $i\in s_{m+1}$, $\underline{\theta}\in\Omega^k$.

<u>Theorem 3</u>. <u>If, under the assumptions stated above, there exists a Bayes procedure, then there exists also a permutation invariant Bayes procedure of the form</u> $\mathcal{P}_B = (\gamma_B, \varphi_B, \tilde{\varphi}_B, \psi^*, \tilde{\psi}^*)$.

<u>Proof: (Backward Induction)</u>. Let the assumptions of the theorem hold, and let $\mathcal{P} = (\gamma, \varphi, \tilde{\varphi}, \psi, \tilde{\psi})$ be any procedure with $r(\tau, \mathcal{P}) < \infty$. In view of Theorem 1, we can assume that $\psi = \psi^*$ holds. We will improve $\mathcal{P}$ backwards stage by stage with the help of (17), thereby constructing a Bayes rule of the form $\mathcal{P}_B$. First, some auxiliary considerations with respect to φ and ψ^* will be made.

Let $m \in \{1,\ldots,q\}$ and $S_m = (s_1,\ldots,s_m)$ with $s_1 \supseteq \ldots \supseteq s_m$ be fixed. It is easy to see that under the i.i.d. prior τ, the conditional distribution of $\underline{\theta}$, given $\underline{W}_m = \underline{w}$, is equal to the product of the conditional distributions of θ_i, given $W_{im} = w_i$, $i=1,\ldots,k$, $\underline{w} \in \mathcal{X}^k$. Therefore, under the assumption (L2) and in view of (19), it follows that for every $S_{m+1} = (s_1,\ldots,s_{m+1})$ with $s_{m+1} \subseteq s_m$, $E^{|\underline{V}_m}[L_m(\underline{\theta}, S_{m+1})]$ depends only on those W_{jm} with $j \in s_{m+1}$. This implies that not only the component of ψ^* for Stage m but also that one for φ_B depends only on those W_{jm} with $j \in s_m$. The latter has the obvious minimizing property and can be chosen to be permutation invariant. Inserting both optimum components into (17), the factor of $\gamma_{S_m}(\underline{V}_m)$ if $m \leq q-1$, or the integrand of $E^{|\underline{V}_{q-1}}$ if $m=q$, respectively, is seen to be of the form

$$(24) \qquad \mathcal{M}_{m-1}(\underline{W}_m, T_m) = \min\{E^{|\underline{V}_m}[L_m(\underline{\theta}, S_{m+1})] \mid s_{m+1} \subseteq s_m\},$$

where $T_m = \mathcal{T}_{m-1}(S_m)$ according to Remark 2. By the reasons given above, $\mathcal{M}_{m-1}(\underline{W}_m, T_m)$ depends only on those W_{jm} with $j \in s_m = t_m$. By using the function $\tilde{\mathcal{L}}_m$, which is defined by (20), it follows that

(25) $\mathcal{M}_{m-1}(\underline{W}_m, T_m)$

$$= \min\{\tilde{\mathcal{L}}_m(\underline{W}_m,(t_1,\ldots,t_{m-1},\hat{t}_m,\hat{t}_{m+1}))\,|\,\hat{t}_m \cap \hat{t}_{m+1} = \emptyset, \hat{t}_m \cup \hat{t}_{m+1} = t_m\},$$

where in the proof of Theorem 1, it has been shown that $\tilde{\mathcal{L}}_m$ has Property $\mathcal{D}(m,\mathcal{X})$. Therefore, from Lemma 4 it follows that $\mathcal{M}_{m-1}$ has Property $\mathcal{D}(m-1,\mathcal{X})$.

Now consider Stage q.

Assume that ψ^* as well as φ_B have been inserted into (17). By the auxiliary results derived before, the last factor in (17), which is associated with Stage q, for every $\tilde{S}_q = (\tilde{s}_1,\ldots,\tilde{s}_q)$ and $\tilde{T}_q = \mathcal{T}_{q-1}(\tilde{S}_q)$ is of the form

(26) $$\bar{\mathcal{M}}_{q-1}(\underline{W}_{q-1},\tilde{T}_q) = E^{|\underline{W}_{q-1}}[\mathcal{M}_{q-1}(\underline{W}_q,\tilde{T}_q)],$$

which depends only on those $W_{j,q-1}$ with $j \in \tilde{s}_q$. This follows from the analogous property of $\mathcal{M}_{q-1}$ and from the fact that under the i.i.d. prior the conditional distribution of $\underline{W}_q$, given $\underline{W}_{q-1} = \underline{w}$, is equal to the product of the conditional distributions of W_{iq}, given $W_{i,q-1} = w_i$, $i=1,\ldots,k$, $\underline{w} \in \mathcal{X}^k$. Since $\mathcal{M}_{q-1}$ has Property $\mathcal{D}(q-1,\mathcal{X})$, $\bar{\mathcal{M}}_{q-1}$ has the same property by Lemma 2 and Lemma 3.

Assume now that the Bayes procedure has been determined for the Stages m+1,m+2,...,q for a fixed $m \in \{1,\ldots,q-1\}$, and that it has been inserted into (17). Let $S_m = (s_1,\ldots,s_m)$ be fixed. Assume further that for every $S_{m+1} = (s_1,\ldots,s_m,s_{m+1})$ with $s_{m+1} \subseteq s_m$ and $\tilde{T}_{m+1} = \mathcal{T}_m(S_{m+1})$, the resulting factor of $\tilde{\psi}_{s_{m+1};r_{m+1},S_m}(\underline{V}_m)$ in (17) is, say, $\bar{\mathcal{M}}_m(\underline{W}_m,\tilde{T}_{m+1})$, which depends only on those W_{jm} with $j \in s_{m+1}$. Finally, assume that $\bar{\mathcal{M}}_m$ has Property $\mathcal{D}(m,\mathcal{X})$.

Under these assumptions, the component of $\tilde{\psi}^*$ for Stage m clearly is optimal (Bayes). Moreover, exactly the same arguments as have been used with respect to φ_B and ψ^*, hold true now with respect to $\tilde{\varphi}_B$ and $\tilde{\psi}^*$ at the same stage. For the optimum components, the resulting factor of $(1-\gamma_{S_m}(\underline{V}_m))$ in (17), denoted henceforth by $\tilde{\mathcal{M}}_{m-1}(\underline{W}_m, T_m)$ with $T_m = \mathcal{J}_{m-1}(S_m)$, has the same properties as $\mathcal{M}_{m-1}(\underline{W}_m, T_m)$, defined by (24), was proved to have.

Finally, the optimum (Bayes) stopping rule γ_B at Stage m decides in terms of the smaller of the two functions $\mathcal{M}_{m-1}$ and $\tilde{\mathcal{M}}_{m-1}$, and can be chosen to be permutation invariant. Inserting it into (17), the m-th line of (17) turns out to be of the following form.

$$\bar{\mathcal{M}}_{m-1}(\underline{W}_{m-1}, T_m) = E^{|\underline{W}_{m-1}}[\mathcal{M}'_{m-1}(\underline{W}_m, T_m)] \tag{27}$$

$$\text{where } \mathcal{M}'_{m-1} = \min(\mathcal{M}_{m-1}, \tilde{\mathcal{M}}_{m-1}).$$

From $\mathcal{M}_{m-1}$ and $\tilde{\mathcal{M}}_{m-1}$, $\mathcal{M}'_{m-1}$ inherits Property $\mathcal{D}(m-1, \mathcal{X})$ as well as the property that $\mathcal{M}'_{m-1}(\underline{W}_m, T_m)$ depends only on those W_{jm} with $j \in s_m$. Therefore, by Lemma 2 and Lemma 3, $\bar{\mathcal{M}}_{m-1}$ has Property $\mathcal{D}(m-1, \mathcal{X})$. By analogous reasons as have been used with respect $\bar{\mathcal{M}}_{q-1}$, $\bar{\mathcal{M}}_{m-1}(\underline{W}_{m-1}, T_m)$ depends only on those $W_{j,m-1}$ with $j \in s_m$.

Since, apparently, we have arrived now at Stage m-1 at exactly the same situation which was assumed at Stage m, the result follows by induction.

Acknowledgment. The authors are thankful to a referee, the Associate Editor, and the Editor for helpful comments.

References

[1] Barndorff-Nielsen, O. (1978): Information and Exponential Families in Statistics. Wiley, New York.

[2] Barron, A. M. and Gupta, S. S. (1972): A class of non-eliminating sequential multiple decision procedures. Operations Research Verfahren (Ed. Henn, Künzi and Schubert), Verlag A. Hain, Meisenheim am Glan, W.-Germany, pp. 11-37.

[3] Bechhofer, R. E., Kiefer, J. and Sobel, M. (1968): Sequential Identification and Ranking Procedures. The University of Chicago Press, Chicago.

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

[5] Gupta, S. S. and Miescke, K. J. (1982a): On the problem of finding a best population with respect to a control in two stages. In Statistical Decision Theory and Related Topics III (Ed. Gupta, S. S. and Berger, J.O.), Vol. 1, 473-496. Academic Press, New York.

[6] Gupta, S. S. and Miescke, K. J. (1982b): On the least favorable configurations in certain two-stage selection procedures. In: Statistics and Probability: Essays in Honor of C. R. Rao. (Ed. Kallianpur, G., Krishnaiah, P. R. and Ghosh, J. K.), 295-305. North-Holland, Amsterdam.

[7] Gupta, S. S. and Miescke, K. J. (1983): An essentially complete class of two-stage selection procedures with screening at the first stage. Statistics and Decisions, to appear.

[8] Gupta, S. S. and Panchapakesan, S. (1979): Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations. J. Wiley, New York.

[9] Hollander, M., Proschan, F. and Sethuraman, J. (1977): Functions decreasing in transposition and their applications in ranking problems. Ann. Statist. 5, 722-733.

[10] Somerville, P. N. (1974): On allocation of resources in a two-stage selection procedure. Sankhyā B. 36, 194-203.

Summary

Sequential Selection Procedures - A Decision Theoretic Approach

Let $\pi_1,\ldots,\pi_k$ be given populations which are associated with unknown real parameters $\theta_1,\ldots,\theta_k$ from a common underlying exponential family $\mathfrak{F}$. Permutation invariant sequential selection procedures are considered to find good populations (i.e. those which have large parameters), where inferior populations are intended to be screened out at the earlier stages. The natural terminal decisions, i.e. decisions which are made in terms of largest sufficient statistics, are shown to be optimum in terms of the risk, uniformly in $(\theta_1,\ldots,\theta_k)$, under fairly general loss assumptions. Similar results with respect to subset selections within stages are established under the additional assumption that $\mathfrak{F}$ is strongly unimodal (i.e. log-concave). The results are derived in the Bayes approach under symmetric priors. Backward induction as well as the concept of decrease in transposition (DT) by Hollander, Proschan and Sethuraman (1977) are the main tools which are used in the proofs.