80-28

ON TWO-STAGE PROCEDURES FOR FINDING A POPULATION BETTER THAN A CONTROL*

by

Klaus J. Miescke
Mainz University and Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #80-28

September 1980

*This research was supported by the Office of Naval Research Contract N00014-75-C-0455 at Purdue University.

ON TWO-STAGE PROCEDURES FOR FINDING A POPULATION BETTER THAN A CONTROL*

by

Klaus J. Miescke
Mainz University and Purdue University

AMS 1970 subject classifications. Primary 62F07; secondary 62F05, 62F15, 62L99.

Key words and phrases. Multiple comparisons with a control, two-stage procedures, screening procedures.

*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.

SUMMARY

Let $\pi_1,\ldots,\pi_k$ be k given populations.

Assume that we wish to find a population better than a given control, if there is any. From all populations we may draw independent samples with distributions which are (at least partly) determined by real parameters $\theta_1,\ldots,\theta_k$, say. A population π_i is viewed to be better than the control if $\theta_i > \theta_0$, i=1,...,k, where $\theta_0 \in \mathbb{R}$ is a fixed given constant. The goal is to guarantee at least a probability P* of making a correct decision if $\theta_i \leq \theta_0$, i=1,...,k, and to maximize the probability of finding a population better than θ_0, otherwise.

Two-stage procedures of the following type will be studied: At Stage 1, based on samples $\underline{X}_1,\ldots,\underline{X}_k$, all populations are screened out which appear to be no better than θ_0. If none (exactly one) is left the procedure stops and decides that none (this one) is better than θ_0. If more than one, π_i with $i \in s$, survives then one proceeds to Stage 2. Here additional samples $\underline{Y}_i$, $i \in s$, are drawn and final decision is made based on $\underline{Y}_i$ or $(\underline{X}_i,\underline{Y}_i)$, $i \in s$.

A natural class of two-stage procedures is proposed which can be completely described and studied in terms of Neyman-Pearson testing theory, where the unsymmetry of tests, however, can be overcome to a considerable extent. As a typical result it is shown that optimality of tests carries over to optimality of two-stage procedures. Finally, under normality, comparisons are made in case of k = 2 with certain Bayesian procedures.

1. Introduction

If k populations $\pi_1,\ldots,\pi_k$ are given and we wish to decide on the basis of a properly chosen sampling scheme which one of these populations is the best one (e.g. has the largest mean), various different approaches and methods have been studied up to now. A complete overview is provided by Gupta and Panchapakesan (1979). Among those, two-stage procedures with screening in the first stage seem to be quite appropriate, since they are more economical as one-stage procedures but still technically not as complicated as sequential ones. Nevertheless, optimality results here are missing up to now and obviously are hard to find. Even the implementation of a simple procedure (as that one which uses Gupta's (1965) maximum means procedure in the first stage and the natural final decision in the second stage) in an indifference zone approach under the assumption of normality with a common known variance causes considerable difficulties. For details and references see Tamhane and Bechhofer (1979), Gupta and Miescke (1979) and Miescke and Sehr (1980).

The situation becomes somewhat fairer if we wish to find a population better than a control π_0, whether it is known or unknown. This because pairwise comparisons are to be made now between π_i and π_0 instead of π_i and π_j, $i\neq j$, $i, j \in \{1,\ldots,k\}$. Let us admit here additionally the choice of a final decision "none of the populations is better than the control". Moreover, let us adopt the following basic requirement and goal:

P*-Condition: Let $P^* \in (0,1)$ be a predetermined constant. A procedure is said to meet the P*-condition if its probability of making the final decision: "none of the populations is better than the control" is at least P^* whenever this decision is correct.

Goal: Among all procedures (in a given class) which meet the P*-condition find that procedure which maximizes the probability of deciding finally in favor of a population better than the control if there is any.

The purpose of this paper is to show that a natural class of two-stage procedures, being widely used in practice, can be described and studied within the framework of Neyman-Pearson testing theory, where the unsymmetry of tests can be overcome to a considerable extent. Emphasis hereby is laid on the basic structure and on comparisons of such procedures rather than on establishing specified ones.

In Section 2 we introduce a natural class $\mathcal{S}'$ of two-stage procedures and derive a formula for their probabilities of correct decisions. As a typical consequence it will be demonstrated in Section 3 that two-stage procedures based on good unbiased one-sample tests for H_i: "population π_i is better than π_0" versus K_i: "population π_i is inferior with respect to π_0", which are simultaneously good tests for the dual problem (where H_i and K_i are interchanged), perform well. Three open questions conclude this section. Finally, in Section 4 we study the normal case and make some comparisons with certain Bayesian procedures in the case of $k = 2$ populations.

2. Basic results

Suppose that for every π_i, $i=1,\ldots,k$, we have a family $\{f_{i,\theta_i}\}_{\theta_i \in \Omega \subseteq \mathbb{R}}$ of densities with respect to the Lebesgue measure or any counting measure on $\mathbb{R}$ which have a common support $Q \subseteq \mathbb{R}$ and may be known or only partly known. The assumption concerning the supports is made for convenience to make ideas clearer and can be weakened in certain circumstances. Let the fixed known control be denoted by θ_0, say,

and all populations π_i be called better than the control if $\theta_i > \theta_0$ and inferior to it if $\theta_i \leq \theta_0$. Let $\underline{X}_i = (X_{i1},\ldots,X_{in_i})$ and $\underline{Y}_i = (Y_{i1},\ldots,Y_{im_i})$, $i=1,\ldots,k$, be samples from π_i available at Stage 1 and Stage 2, respectively, where $\underline{X}_1,\underline{Y}_1,\ldots,\underline{X}_k,\underline{Y}_k$ are mutually independent, and let $\underline{X} = (\underline{X}_1,\ldots,\underline{X}_k)$ and $\underline{Y} = (\underline{Y}_1,\ldots,\underline{Y}_k)$.

Before we are going to define a natural class of two-stage procedures in a concise way, let us briefly describe how these procedures typically are applied in practice. For every testing problem K_i: $\theta_i \leq \theta_0$ versus H_i: $\theta_i > \theta_0$ the experimenter chooses a test based on $\underline{X}_i$ and for H_i versus K_i another one based on $\underline{Y}_i$ or $(\underline{X}_i,\underline{Y}_i)$, $i=1,\ldots,k$, takes two levels $\alpha_1,\alpha_2 \in (0,1)$ (which usually are small) and proceeds as follows:

At Stage 1 he discards all populations which are not significant at level α_1 under the first set of tests. If none (exactly one) is left, he decides that none (this one) is better than the control. Only if more than one population survives, he proceeds to Stage 2.

At Stage 2 the experimenter draws additional samples $\underline{Y}_i$ from those π_i's which were selected at Stage 1 and exchanges hypotheses and alternatives with respect to these populations. If all these populations now turn out to be significant at level α_2 under the second set of tests (which is rather unlikely to happen) he decides that none of the populations is better than the control. Otherwise, he makes a final decision in favor of that population among the selected ones which has the largest p-value under the associated second test.

If these tests are upper level α_1 (respectively lower level α_2) tests, which for simplicity are non-randomized for a moment to fix ideas, based on real valued statistics U_i and V_i, $i=1,\ldots,k$, then the procedure

described above can be equivalently described as follows: At Stage 1 all π_i's are selected with $U_i \geq c_i$ (where c_i is the α_1-fractile of U_i under $\theta_i = \theta_0$), and final decision is made in terms of the largest V_j among the selected π_j's, provided $V_j \geq d_j$ (where d_i is the α_2-quantile of V_j under $\theta_j = \theta_0$). The truncated version of such procedures (i.e. which perform Stage 1 only) have been studied by several authors. See, for example, Gupta and Sobel (1958) and Lehmann (1961). Also some work has been done in sequential setups. For references see Gupta and Panchapakesan (1979), Chapter 20. But, apparently, no results concerning two-stage procedures of the type described above have appeared in literature until now. This gives us the motivation for the following considerations.

To begin with, let us state without a formal proof that by similar arguments as are used in Miescke (1979a) it can be shown that every procedure of the type described just now - where U_i, c_i and V_i, d_i more generally may take values in measurable spaces $\mathcal{X}_i$ and $\mathcal{Y}_i$, and where U_i and V_i are stochastically non-decreasing in θ_i with respect to measurable total orderings in $\mathcal{X}_i$ and $\mathcal{Y}_i$, i=1,...,k - is a member of the class $\mathcal{D}$ to be defined below.

To avoid confusions and to arrive at a consistent representation of this class let us from now on use only tests for H_i versus K_i, i=1,...,k, which take value 1 as soon as one observation falls outside support Q. (This modifies procedures only on null-sets.) Finally, several definitions given in Miescke (1979a) will be relevant in the sequel but for brevity are not repeated here. Especially, tests may be randomized ones taking values in [0,1]. This typically occurs in discrete cases or in continuous cases when nonparametric (rank) tests are under concern. Thus significance

statements as well as p-values are understood to be based on additional randomization schemes as are used in Miescke (1979a). To be more specific, let $\underline{A} = (A_1,\ldots,A_k)$ be that one for the first stage and $\underline{B} = (B_1,\ldots,B_k)$ that one for the second stage. Note that $\underline{X}$, $\underline{Y}$, $\underline{A}$ and $\underline{B}$ are assumed to be independent.

The class $\mathcal{S}$ of two-stage procedures:

For i=1,...,k let

(1) $\varphi_i = \{\varphi_{i,\alpha}\}_{\alpha\in[0,1]}$ be a right-continuous and monotone (in α) unbiased test for H_i versus K_i based on $\underline{X}_i$, which is standardized at θ_0. Assume that within Q $\varphi_0 \equiv 0$ and $\varphi_1 \equiv 1$. Let $\underline{\varphi} = (\varphi_1,\ldots,\varphi_k)$.

(2) Analogously, let $\psi_i = \{\psi_{i,\alpha}\}_{\alpha\in[0,1]}$ be such a test for H_i versus K_i based on $(\underline{X}_i,\underline{Y}_i)$. Let $\underline{\psi} = (\psi_1,\ldots,\psi_k)$.

For $0 < \alpha_1 \leq 1$ and $0 \leq \alpha_2 < 1$ let $(\underline{\varphi},1-\alpha_1,\underline{\psi},\alpha_2)$ denote the following two-stage procedure:

Stage 1: Select π_i if $p_{\varphi_i}(\underline{X}_i,A_i)$, the p-value of $\underline{X}_i$ under φ_i, is larger than $1-\alpha_1$, i=1,...,k. If none (exactly one) of the populations is selected, stop and decide "none (this one) is better than θ_0". Otherwise proceed to Stage 2.

Stage 2: Among the selected populations decide finally in favor of that π_j which has the largest p-value $p_{\psi_j}(\underline{X}_j,\underline{Y}_j,B_j)$ under ψ_j, provided it is no smaller than α_2. Otherwise decide that "none is better than θ_0".

Let $\mathcal{S}$ be the set of all such two-stage procedures.

The following result will serve as our basic tool to determine (α_1,α_2)-tuples for meeting the P*-condition as well as to compare the performance of competing procedures satisfying the P*-condition.

Theorem 1. Let $(\underline{\varphi}, 1-\alpha_1, \underline{\psi}, \alpha_2) \in \mathfrak{D}$. For notational convenience let $E_i = E_{\theta_i}\varphi_{i,\,1-\alpha_1}(\underline{X}_i)$ and $F_i(\alpha) = E_{\theta_i}\{\psi_{i,\alpha}(\underline{X}_i, \underline{Y}_i) \mid \underline{X}_i$ is not significant under $\varphi_{i,\,1-\alpha_1}\}$, $i=1,\ldots,k$. Then for every non-empty $D \subseteq \{1,\ldots,k\}$ and $\underline{\theta} \in \Omega^k$

$$(2.1)\quad P_{\underline{\theta}}\{\text{final decision of } (\underline{\varphi}, 1-\alpha_1, \underline{\psi}, \alpha_2) \text{ falls into } D\}$$

$$= \int_{\alpha_2}^{1} \prod_{j \notin D} [E_j + (1-E_j)F_j(\alpha)]\, d\Big(\prod_{i \in D} [E_i + (1-E_i)F_i(\alpha)] \Big)$$

where integration is with respect to α. Moreover,

$$(2.2)\quad P_{\underline{\theta}}\{\text{final decision of } (\underline{\varphi}, 1-\alpha_1, \underline{\psi}, \alpha_2) \text{ does not fall into } \{1,\ldots,k\}\}$$

$$= \prod_{j=1}^{k} [E_j + (1-E_j)F_j(\alpha_2)] \,.$$

Proof: It is shown in Miescke (1979a)(cf. (2.3) there) that the distribution function of each p-value appearing in $(\underline{\varphi}, 1-\alpha_1, \underline{\psi}, \alpha_2)$ equals to the power function of the corresponding test, which hereby is thought of being a function of $\alpha \in [0,1]$ where parameter $\underline{\theta} \in \Omega^k$ on the other hand is held fixed.

Let $D \subseteq \{1,\ldots,k\}$, $D \neq \emptyset$, $\underline{\theta} \in \Omega^k$, $0 < \alpha_1 \leq 1$ and $0 \leq \alpha_2 < 1$ be fixed. Then

$P_{\underline{\theta}}\{\text{final decision falls into } D\}$

$$= \sum_{r \in D} P_{\underline{\theta}}\{\text{final decision is in favor of } \pi_r\}$$

$$= \sum_{r \in D} \sum_{s: r \in s} P_{\underline{\theta}}\{\text{the } \pi_i\text{'s with } i \in s \text{ are selected and final decision is made in favor of } \pi_r\}$$

$$= \sum_{r \in D} \sum_{s: r \in s} \int_{\alpha_2}^{1} \prod_{\substack{i \in s \\ i \neq r}} F_i(\alpha)\, dF_r(\alpha) \prod_{i \in s} (1-E_i) \prod_{j \notin s} E_j$$

$$= \sum_{r\in D} \int_{\alpha_2}^{1} \Big\{ \sum_{s: r\notin s} \prod_{i\in s} (1-E_i)F_i(\alpha) \prod_{\substack{j\notin s \\ j\neq r}} E_j \Big\} d\big((1-E_r)F_r(\alpha)\big) .$$

Now, the integrand $\{\cdots\}$ equals to $\prod_{\substack{i=1 \\ i\neq r}}^{k} [E_i+(1-E_i)F_i(\alpha)]$ and $(1-E_r)F_r(\alpha)$ can be replaced by $E_r+(1-E_r)F_r(\alpha)$. Thus $P_{\underline{\theta}}\{\text{final decision falls into } D\}$ equals

$$\sum_{r\in D} \int_{\alpha_2}^{1} \prod_{\substack{i=1 \\ i\neq r}}^{k} [E_i+(1-E_i)F_i(\alpha)]d[E_r+(1-E_r)F_r(\alpha)]$$

$$= \int_{\alpha_2}^{1} \prod_{j\notin D} [E_j+(1-E_j)F_j(\alpha)] \sum_{r\in D} \prod_{\substack{i\in D \\ i\neq r}} [E_i+(1-E_i)F_i(\alpha)]d[E_r+(1-E_r)F_r(\alpha)]$$

$$= \int_{\alpha_2}^{1} \prod_{j\notin D} [E_j+(1-E_j)F_j(\alpha)]d\Big(\prod_{i\in D} [E_i+(1-E_i)F_i(\alpha)]\Big) .$$

This completes the proof of (2.1). Since (2.2) can be verified by using similar arguments its proof is omitted for brevity.

Remark: Note that for i=1,...,k we have also the following representation of $E_i+(1-E_i)F_i(\alpha)$, $\alpha\in[0,1]$:

$$(2.3)\quad E_{\theta_i}\Big(\varphi_{i,1-\alpha_1}(\underline{X}_i)+(1-\varphi_{i,1-\alpha_1}(\underline{X}_i))\psi_{i,\alpha}(\underline{X}_i,\underline{Y}_i)\Big)$$

$$= P_{\theta_i}\Big(\{p_{\varphi_i}(\underline{X}_i,A_i)\leq 1-\alpha_1\}\cup\{p_{\varphi_i}(\underline{X}_i,A_i)>1-\alpha_1 \text{ and } p_{\psi_i}(\underline{X}_i,\underline{Y}_i,B_i)\leq\alpha\}\Big) .$$

Corollary 1. Every two-stage procedure $(\underline{\varphi},1-\alpha_1,\underline{\psi},\alpha_2)\in\mathfrak{D}$ satisfies the P*-condition if

$$(2.4)\qquad (1-\alpha_1)^k = P^* .$$

Proof: A lower bound for (2.2) is $\prod_{j=1}^{k} E_j$ which satisfies

$$\prod_{j=1}^{k} E_j = \prod_{j=1}^{k} E_{\theta_j} \varphi_{j,1-\alpha_1}(\underline{X}_j) \geq \prod_{j=1}^{k} E_{\theta_0} \varphi_{j,1-\alpha_1}(\underline{X}_j) = (1-\alpha_1)^k$$

if $\theta_1,\ldots,\theta_k \leq \theta_0$. This follows from the unbiasedness of the tests.

Unfortunately, the dependencies between $\varphi_{i,1-\alpha_1}(\underline{X}_i)$ and $\psi_{i,\alpha}(\underline{X}_i,\underline{Y}_i)$, $i=1,\ldots,k$, make it hard to find good procedures in $\mathcal{S}$. Therefore most of our results in the sequel will be given only with respect to $\mathcal{S}'$, say, where $\mathcal{S}' \subseteq \mathcal{S}$ consists of all procedures from $\mathcal{S}$ where the tests ψ_i in the second stage depend only on the $\underline{Y}_i$'s and not on the $\underline{X}_i$'s, $i=1,\ldots,k$. The interested reader is invited to try to prove one of the conjectures stated at the end of this section.

Corollary 1'. A two-stage procedure $(\underline{\varphi},1-\alpha_1,\underline{\psi},\alpha_2) \in \mathcal{S}'$ satisfies the P*-condition if and only if

$$(1-\alpha_1+\alpha_1\alpha_2)^k = P^* . \tag{2.5}$$

Proof: Let $\theta_1,\ldots,\theta_k \leq \theta_0$. For $(\underline{\varphi}, 1-\alpha_1,\underline{\psi},\alpha_2) \in \mathcal{S}'$ (2.2) reduces to

$$\prod_{j=1}^{k} [E_{\theta_j} \varphi_{j,1-\alpha_1}(\underline{X}_j) + (1-E_{\theta_j} \varphi_{j,1-\alpha_1}(\underline{X}_j))E_{\theta_j} \psi_{j,\alpha_2}(\underline{Y}_j)]$$

which, by the unbiasedness of the tests, and the fact that $a+(1-a)b = b+(1-b)a$ is increasing in $a, b \in [0,1]$, assumes its lowest value at $\theta_1 = \ldots = \theta_k = \theta_0$, where the power functions are equal to the levels of the tests.

Remark: If in a procedure $(\underline{\varphi},1-\alpha_1,\underline{\psi},\alpha_2) \in \mathcal{S}$ every pair of tests φ_i, ψ_i have non-negative correlations for $\theta_i \leq \theta_0$, $i=1,\ldots,k$ (which, of course, is given if the procedure belongs to $\mathcal{S}'$), then the infimum of (2.2) under

$\theta_1,\ldots,\theta_k \leq \theta_0$ falls between $(1-\alpha_1)^k$ and $(1-\alpha_1+\alpha_1\alpha_2)^k$.

Let us from now on adopt the following convention:

Convention: All procedures from $\mathfrak{D}'$ are assumed to have an α_1 satisfying $(1-\alpha_1)^k = P^*$ and a small α_2.

In view of (2.5) $(1-\alpha_1)^k = P^*$ and $\alpha_2 = 0$ clearly is the best choice to minimize the expected overall sampling amount (and to make (2.4) to an exact condition with respect to $\mathfrak{D}$). But on the other hand an experimenter might feel restricted at not being permitted to decide also at Stage 2 against all populations. Thus let us admit at least a small α_2. This is slightly conservative with repsect to (2.5). But it changes the probabilities of any events at most by a difference of $(\max\{\alpha_1,\alpha_2\})^2$. This follows from the fact that α_2 acts only on probabilities of events where at least two populations pass Stage 1 and eventually are rejected at Stage 2. To give a numerical example, take $\alpha_1 = \alpha_2 = 10^{-2}$. Then for $k = 5(10)$ we have a P^* above 0.95 (0.90) and α_2 changes all probabilities at most by the amount of 10^{-4}.

3. Consequences and extensions

The following two results will be used repeatedly in the sequel. Their proofs are straightforward using integrations by parts and are therefore omitted for brevity.

Lemma 1. Let G_i, $\tilde{G}_i$: $[0,1] \to [0,1]$ be right-continuous, non-decreasing with $G_i(1) = \tilde{G}_i(1) = 1$ and $G_i(\alpha) \leq (\geq)\tilde{G}_i(\alpha)$ for $i=1,\ldots,r(i=r+1,\ldots,k)$, $r \in \{1,\ldots,k\}$. Then for $0 \leq \alpha_2 < 1$

$$(3.1)\qquad \int_{\alpha_2}^{1} \prod_{i=1}^{r} G_i(\alpha)\,d\Big(\prod_{j=r+1}^{k} G_j(\alpha)\Big) \leq \int_{\alpha_2}^{1} \prod_{i=1}^{r} \tilde{G}_i(\alpha)\,d\Big(\prod_{j=r+1}^{k} \tilde{G}_j(\alpha)\Big) .$$

As a special case of Lemma 1 we get

Corollary 2. Let $G_1,\ldots,G_k$ as before and $G_i(\alpha) \leq G_j(\alpha)$, $\alpha \in [0,1]$, for a pair $i, j \in \{1,\ldots,k\}$ with $i \neq j$. Then for $0 \leq \alpha_2 < 1$

$$(3.2) \qquad \int_{\alpha_2}^{1} \prod_{\substack{\ell=1 \\ \ell\neq j}}^{k} G_\ell(\alpha)dG_j(\alpha) \leq \int_{\alpha_2}^{1} \prod_{\substack{\ell=1 \\ \ell\neq i}}^{k} G_\ell(\alpha)dG_i(\alpha) \ .$$

For the sequel let $P_{\underline{\theta}}\{C.D.\}$ denote the probability of a correct decision at $\underline{\theta} \in \Omega^k$, i.e. that the final decision falls into $R(\underline{\theta}) = \{i \mid \theta_i > \theta_0, i=1,\ldots,k\}$ if $R(\underline{\theta})$ is non-empty or that the final decision is "no population is better than θ_0" if $R(\underline{\theta})$ is empty.

Corollary 3. Let $(\underline{\varphi},1-\alpha_1,\underline{\psi},\alpha_2) \in \mathfrak{D}'$, $0 < \alpha_1 \leq 1$ and $0 \leq \alpha_2 < 1$. If for every $i \in \{1,\ldots,k\}$ φ_i and ψ_i are UMP unbiased tests for H_i versus K_i based on $\underline{X}_i$ and $\underline{Y}_i$ and if simultaneously $1-\varphi_i$ and $1-\psi_i$ are UMP unbiased tests for the dual testing problem (where the hypothesis and the alternative are interchanged), then at every $\underline{\theta} \in \Omega^k$ $(\underline{\varphi},1-\alpha_1,\underline{\psi},\alpha_2)$ has the largest $P_{\underline{\theta}}\{C.D.\}$ and the smallest expected sampling amount (w.r.t bad populations) among all $(\underline{\varphi},1-\alpha_1,\underline{\psi},\alpha_2) \in \mathfrak{D}'$.

As is well known, these conditions are usually fulfilled in one-parameter MLR and multiparameter exponential family situations. The proof of Corollary 3 as well as that of the next result follows from (2.1) and Lemma 1.

Corollary 4. Let $(\underline{\varphi}, 1-\alpha_1,\underline{\psi},\alpha_2) \in \mathfrak{D}'$. If the power functions of all tests are non-decreasing (non-increasing) in sample-sizes for $\theta_i < (>)\theta_0$, then $P_{\underline{\theta}}\{C.D.\}$ is nondecreasing in sample sizes at every $\underline{\theta} \in \Omega^k$.

The next result can be stated with respect to $\mathfrak{D}$.

Corollary 5. Let $(\underline{\varphi}, 1-\alpha_1, \underline{\psi}, \alpha_2) \in \mathfrak{D}$ where $\underline{\varphi}$ consists of consistent tests. Then $P_{\underline{\theta}}\{C.D.\}$ converges to 1 if $n_i \to \infty$ and $\theta_i \neq \theta_0$, $i=1,\ldots,k$.

Proof: Let $\underline{\theta} \in \Omega^k$ with $\theta_i \neq \theta_0$, $i=1,\ldots,k$. Then

$$P_{\underline{\theta}}\{C.D.\} \geq P_{\underline{\theta}}\{\text{only } \pi_i\text{'s with } i \in R(\underline{\theta}) \text{ are selected at Stage 1}\}$$

$$\geq \prod_{j \notin R(\underline{\theta})} E_{\theta_j} \varphi_{j,1-\alpha_1}(\underline{X}_j) \prod_{i \in R(\underline{\theta})} [1-E_{\theta_i} \varphi_{i,1-\alpha_1}(\underline{X}_i)]$$

which tends to 1 for large $n_1,\ldots,n_k$ by the consistency of the tests.

Under the assumption of monotone (non-decreasing) likelihood ratios (MLR) a stronger result can be obtained.

Theorem 2. Assume that in every population π_i the family of densities has MLR, $i=1,\ldots,k$. Let $(\underline{\varphi}, 1-\alpha_1, \underline{\psi}, \alpha_2) \in \mathfrak{D}$ (or $\mathfrak{D}'$) consist of the UMP tests for the corresponding testing problems. Then for increasing sample sizes n_i, m_i, $i=1,\ldots,k$, $P_{\underline{\theta}}\{$final decision in favor of that π_i with the largest $\theta_i > \theta_0\}$ tends to 1 for all $\underline{\theta} \in \Omega^k$ with $R(\underline{\theta}) \neq \emptyset$ and $P_{\underline{\theta}}\{$final decision is "no population is better than θ_0"$\}$ tends to 1 for all $\underline{\theta} \in \Omega^k$ with $\theta_1,\ldots,\theta_k < \theta_0$.

Proof: Let $\theta_k > \theta_0, \theta_1,\ldots,\theta_{k-1}$, $0 < \alpha_1 \leq 1$ and $0 \leq \alpha_2 < 1$.

$$P_{\underline{\theta}}\{(\underline{\varphi}, 1-\alpha_1, \underline{\psi}, \alpha_2) \text{ finally decides in favor of } \pi_k\}$$

$$\geq P_{\underline{\theta}}\{p_{\varphi_k}(\underline{X}_k, A_k) > 1-\alpha_1; p_{\psi_k}(\underline{X}_k, \underline{Y}_k, B_k) > \alpha_2, p_{\psi_i}(\underline{X}_i, \underline{Y}_i, B_i),\ i \neq k\}.$$

Now, $P_{\underline{\theta}}\{p_{\varphi_k}(\underline{X}_k, A_k) > 1-\alpha_1; p_{\psi_k}(\underline{X}_k, \underline{Y}_k, B_k) > \alpha_2\}$ tends to 1 for large n_k and m_k by the consistency of the tests.

Moreover, the procedure which decides in favor of a population according to the largest p-value with respect to $\underline{\psi}$ can be viewed, equivalently, as being based on $\underline{\tilde{\psi}}$, say, where $\tilde{\psi}_i$ is essentially the same test as ψ_i but

now standardized at θ_k (instead of θ_0), i=1,...,k. But then since all populations $\pi_1,\ldots,\pi_{k-1}$ are shifted into "alternatives",

$$P_{\underline{\theta}}\{p_{\psi_k} > p_{\psi_i}, i\neq k\} = P_{\underline{\theta}}\{p_{\tilde{\psi}_k} > p_{\tilde{\psi}_i}, i\neq k\} = \int_0^1 \prod_{i=1}^{k-1} E_{\theta_i} \tilde{\psi}_{i,\alpha}(\underline{X}_i,\underline{Y}_i)d\alpha \ ,$$

(cf. Miescke (1979a)) which tends to 1 for large $m_1,\ldots,m_k$. When more than one θ_i is largest a similar result can be derived. The proof for $\mathfrak{A}'$ using UMP tests $\psi_1,\ldots,\psi_k$ based on $\underline{Y}_1,\ldots,\underline{Y}_k$ is exactly the same. The second assertion of the theorem is already proved by Corrollary 5.

Remark: If the asymptotic relative efficiency (ARE) in the sense of Pitman is defined in terms of the probability of selecting the π_i's with $i \in R(\underline{\theta})$ at Stage 1, then if $ARE(\varphi_i, \tilde{\varphi}_i)$ does not depend on $i \in \{1,\ldots,k\}$ (as is typically the case when $\varphi_1,\ldots,\varphi_k$ are of the same type and $\tilde{\varphi}_1,\ldots,\tilde{\varphi}_k$ are of the same type) we have $ARE(\varphi_1,\tilde{\varphi}_1) = ARE((\underline{\varphi},1-\alpha_1,\underline{\psi},\alpha_2),(\underline{\tilde{\varphi}},1-\alpha_1,\underline{\psi},\alpha_2))$ for all $0 < \alpha_1 \leq 1$, $0 \leq \alpha_2 < 1$ and $\underline{\psi}$. The proof is similar to that in Miescke (1979a). Of course, it would be more satisfactory to have an ARE-concept including both stages, but this seems to be a difficult problem. In fact, Pitman's approach does not lead here to clear conclusions.

Corollary 6. Let $(\underline{\varphi}, 1-\alpha_1,\underline{\psi},\alpha_2) \in \mathfrak{A}'$. If the power functions of φ_i and ψ_i are non-increasing in $\theta_i \in \Omega$, i=1,...,k, then for r, s $\in \{1,\ldots,k\}$ with $\theta_r < \theta_s$

(3.3) $$P_{\underline{\theta}}\{\text{final decision is in favor of } \pi_r\}$$
$$\leq P_{\underline{\theta}}\{\text{final decision is in favor of } \pi_s\} \ .$$

Proof: This follows from (2.1) and Corollary 2.

Final Remarks:

(1) The results so far derived hold analogously in cases where the control values θ_{0_i} may depend on $i \in \{1,\ldots,k\}$.

(2) The case of unknown controls can be treated analogously provided that control samples are drawn independently for each single test.

(3) Under the assumption of MLR let $(\underline{\varphi}, 1-\alpha_1, \underline{\psi}, \alpha_2) \in \mathcal{S}$ consist of the UMP tests for the corresponding testing problems. Then

$$\text{(3.4)} \quad \inf\{P_{\underline{\theta}}\{C.D.\} \mid \underline{\theta} \in \Omega^k,\ R(\underline{\theta}) = \emptyset\} = \inf\{P_{\underline{\theta}}\{C.D.\} \mid \underline{\theta} \in \Omega^k,\ R(\underline{\theta}) \neq \emptyset\}$$
$$= P_{(\theta_0,\theta_0,\ldots,\theta_0)}\{C.D.\}\ .$$

This follows from Theorem 1 in a recently published paper by Simons (1980), (2.1) and Lemma 1.

(4) Let us conclude this section by stating the following three important questions that have not been settled now: Assume that in all populations MLR is given and that $\varphi_i, \psi_i, \tilde{\psi}_i$ are the UMP tests based on $\underline{X}_i$, $\underline{Y}_i$, $(\underline{X}_i, \underline{Y}_i)$, $i=1,\ldots,k$.

I) Is $(\underline{\varphi}, 1-\alpha_1, \underline{\tilde{\psi}}, \alpha_2)$ performing better than $(\underline{\varphi}, 1-\alpha_1, \underline{\psi}, \alpha_2)$?

II) If this is true, how well performs $(\underline{\varphi}, 1-\alpha_1, \underline{\tilde{\psi}}, \alpha_2)$ in $\mathcal{S}$?

III) The one-stage procedures $(\underline{\psi}, 1-\alpha_1)$ and $(\underline{\tilde{\psi}}, 1-\alpha_1)$ which select according to the largest p-values of the corresponding tests, provided that they are larger than $1-\alpha_1$, are the natural competitors to $(\underline{\varphi}, 1-\alpha_1, \underline{\psi}, 0)$ and $(\underline{\varphi}, 1-\alpha_1, \underline{\tilde{\psi}}, 0)$. Do they need larger sampling amounts than the two-stage procedures take in the mean to have the same $P_{\underline{\theta}}\{C.D.\}$?

4. The normal case. Bayesian two-stage procedures for k = 2.

Assume that we have k normal populations π_i with unknown means $\theta_i \in \mathbb{R}$ and known variances $\sigma_i^2 > 0$, $i=1,\ldots,k$. Let $\bar{X}_i$ and $\bar{Y}_i$ denote the sample

means derived from the samples of sizes n_i and m_i from π_i, i=1,...,k. Let Φ denote the cumulative distribtuion function of the standard normal distribution. Then the optimal procedure in $\mathcal{S}'$ is as follows:

Stage 1: Select all populations π_j, j=1,...,k, with

$$j \in S: = \{i \mid \bar{X}_i \geq \theta_0 + \sigma_i n_i^{-\frac{1}{2}} \Phi^{-1}(1-\alpha_1),\ i=1,\ldots,k\} .$$

If $S = \{j_0\}$ $(S = \emptyset)$ stop, and decide in favor of π_{j_0} ("none is better than θ_0"). Otherwise, proceed to

Stage 2: Decide finally in favor of π_r if $r \in S$ with $\bar{Y}_r \geq \bar{Y}_s$, $s \in S$ and $\bar{Y}_r \geq \theta_0 + \sigma_r m_r^{-\frac{1}{2}} \Phi^{-1}(\alpha_2)$. Otherwise decide that no population is better than the control.

Let $\bar{Z}_i = (n_i+m_i)^{-1}(n_i\bar{X}_i+m_i\bar{Y}_i)$, i=1,...,k, be the overall sample means. Though we do not know whether the alternative procedure which uses $\bar{Z}_i$'s instead of $\bar{Y}_i$'s in Stage 2 perform better, we can at least show that in (2.1) the functions $E_i+(1-E_i)F_i(\alpha)$, $\alpha \in [\alpha_2,1]$, with $\theta_i > \theta_0$ will then be replaced by smaller functions $E_i+(1-E_i)F_i^*(\alpha)$, $\alpha \in [\alpha_2,1]$, say, i=1,...,k. Let $\theta_i > \theta_0$.

$$(1-E_i)(1-F_i^*(\alpha)) = P_{\theta_i}\{\bar{X}_i \geq \theta_0 + n_i^{-\frac{1}{2}}\sigma_i\Phi^{-1}(1-\alpha_1), \bar{Z}_i \geq \theta_0 + (n_i+m_i)^{-\frac{1}{2}}\sigma_i\Phi^{-1}(\alpha)\}$$

$$\geq P_{\theta_i}\{\bar{X}_i \geq \theta_0 + n_i^{-\frac{1}{2}}\sigma_i\Phi^{-1}(1-\alpha_1)\}P_{\theta_i}\{\bar{Z}_i \geq \theta_0 + (n_i+m_i)^{-\frac{1}{2}}\sigma_i\Phi^{-1}(\alpha)\} .$$

This follows from Slepian's inequality since $\bar{X}_i$ and $\bar{Y}_i$ are positively correlated. Finally the proof is completed by

$$P_{\theta_i}\{\bar{Z}_i \geq \theta_0 + (n_i+m_i)^{-\frac{1}{2}}\sigma_i\Phi^{-1}(\alpha)\} = \Phi\left((n_i+m_i)^{\frac{1}{2}}\sigma_i^{-1}(\theta_i-\theta_0) - \Phi^{-1}(\alpha)\right)$$

$$\geq \Phi\left(m_i^{\frac{1}{2}}\sigma_i^{-1}(\theta_i-\theta_0) - \Phi^{-1}(\alpha)\right)$$

$$= P_{\theta_i}\{\bar{Y}_i \geq \theta_0 + m_i^{-\frac{1}{2}}\sigma_i\Phi^{-1}(\alpha)\} = 1-F_i(\alpha) .$$

When the variances are unknown the optimal procedure in $\mathfrak{S}'$ is based on t-tests in an analogous way. Let s_i and $\tilde{s}_i$ denote the usual UMV unbiased estimators of σ_i based on $\underline{X}_i$ and $\underline{Y}_i$, respectively, $i=1,\ldots,k$. Then $n_i^{-\frac{1}{2}}\sigma_i\Phi^{-1}(1-\alpha_1)$ at Stage 1 has to be replaced by $n_i^{-\frac{1}{2}}s_i t(n_i-1,1-\alpha_1)$, and $m_r^{-\frac{1}{2}}\sigma_r\Phi^{-1}(\alpha)$ at Stage 2 has to be replaced by $m_r^{-\frac{1}{2}}\tilde{s}_r t(m_r-1,\alpha)$, where $t(n,\alpha)$ denotes the α-quantile of the t-distribution with n degrees of freedom.

Though a (Bayesian) decision theoretic approach is quite difficult to perform in general, the case of $k=2$ populations can at least be studied to some extent. A two-stage procedure will now be described by $S(\underline{X})$ (the random subset of $\{1,2\}$ of indices of those populations π_i being selected at Stage 1) and $d(\underline{X},\underline{Y})$ (the final decision at Stage 2). As before, the procedure stops and decides 0, i.e. "none is better than θ_0" (1,2) if $S(\underline{X}) = \emptyset\ (\{1\},\{2\})$ and $d(\underline{X},\underline{Y})$ at Stage 2 is used only if $S(\underline{X}) = \{1,2\}$.

Let $\ell\colon \mathbb{R} \to \mathbb{R}$, with $\ell(0) = 0$, be a non-decreasing function which acts as loss-gain-function with respect to final decisions 1 and 2. Assume that decision 0 leads neither to a loss nor a gain. Moreover, let $c > 0$ be the costs we have to pay if we wish to perform Stage 2. Finally, let τ be the prior distribution of the (now random) parameter vector $\underline{\theta}$. Then the overall Bayesian risk is given by

$$(4.1)\qquad \int_{\Omega^k}\Big\{\Big[c + \sum_{i=1}^{2}\ell(\theta_0-\theta_i)P_{\underline{\theta}}\{d(\underline{X},\underline{Y}) = i \mid S(\underline{X})=\{1,2\}\}\Big]P_{\underline{\theta}}\{S(\underline{X})=\{1,2\}\} + \sum_{j=1}^{2}\ell(\theta_0-\theta_j)P_{\underline{\theta}}\{S(\underline{X}) = \{j\}\}\Big\}\,d\tau(\underline{\theta})$$

The optimal decision d* at Stage 2 (which minimizes the posterior expected loss given $\underline{X} = \underline{\xi}$ and $\underline{Y} = \underline{\eta}$) does not depend on the special choice of any subset selection rule S and turns out to be

$$(4.2)\qquad d^*(\underline{\xi},\underline{\eta}) = i \text{ iff } E\{\ell(\theta_0-\theta_i)\,|\,\underline{X}=\underline{\xi},\ \underline{Y}=\underline{\eta}\}$$

$$< \min\{0, E\{\ell(\theta_0-\theta_j)\,|\,\underline{X}=\underline{\xi},\ \underline{Y}=\underline{\eta}\}\},\ \{i,j\} = \{1,2\}\ ,$$

and $d^*(\underline{\xi},\underline{\eta}) = 0$ otherwise.

The optimal subset selection rule S* at Stage 1 (which minimizes the posterior expected loss given $\underline{X} = \underline{\xi}$ under the assumption that d* will be used at Stage 2) decides according to the smallest of the four values given in the following scheme:

$$(4.3)\qquad S^*(\underline{\xi}) = \emptyset\colon\ 0$$

$$S^*(\underline{\xi}) = \{i\}\colon\ E\{\ell(\theta_0-\theta_i)\,|\,\underline{X}=\underline{\xi}\},\ i=1,2\ ,$$

$$S^*(\underline{\xi}) = \{1,2\}\colon\ c+E\left\{\min\{0,\ \min_{i=1,2} E\{\ell(\theta_0-\theta_i)\,|\,\underline{X}=\underline{\xi},\underline{Y}\}\}\,|\,\underline{X}=\underline{\xi}\right\}.$$

Notethat in the last expression the inner conditional expectiation is viewed as being a function of $\underline{Y}$, and that the outer one is the expectation with respect to the conditional distribution of $\underline{Y}$-given $\underline{X} = \underline{\xi}$.

Now let us assume the following normal model: Conditionally, given $\underline{\Theta} = \underline{\theta}$, $\underline{X}$ and $\underline{Y}$ are independent with $\underline{X} \sim \eta\,(\underline{\theta},pI)$ and $\underline{Y} \sim \eta\,(\underline{\theta},qI)$, and apriori $\Theta \sim \eta\,(\theta_0\underline{1},rI)$, $p,q,r > 0$, $I = \begin{pmatrix}1&0\\0&1\end{pmatrix}$, $\underline{1}' = (1,1)$.

Then by using for convenience U, V_1, V_2, which are assumed to be independent standard normals, we get the following scheme equivalent to (4.3):

$$(4.4)\qquad S^*(\underline{\xi}) = \emptyset\ :\ 0$$

$$S^*(\underline{\xi}) = \{i\}\colon\ E^U\left(\ell\left(r(p+r)^{-1}(\theta_0-\xi_i)+\left(rp(p+r)^{-1}\right)^{\frac{1}{2}}U\right)\right),\ i=1,2$$

$$S^*(\underline{\xi}) = \{1,2\}\colon\ c+E^{V_1,V_2}[\min\{0,\min_{i=1,2}\{$$

$$E^U\left(\ell\left(r(p+r)^{-1}(\theta_0-\xi_i)+\rho V_i+\gamma U\right)\right)\}\}]$$

where $\rho = pr[(p+r)(pq+pr+qr)]^{-\frac{1}{2}}$ and $\gamma = [rpq/(pq+pr+qr)]^{\frac{1}{2}}$.

Let especially ℓ be linear, i.e. $\ell(\Delta) = a\Delta$, $\Delta \in \mathbb{R}$, where we can assume without loss of generality that $a = 1$ holds (since this can be compensated by c). Moreover, let us for a moment restrict our considerations to two-stage procedures which at Stage 2 are not permitted to make decision 0. (This corresponds to procedures in $\mathcal{D}'$ or $\mathcal{D}$ with $\alpha_2 = 0$.) Then the optimal procedure, denoted by d_* and S_*, can be described in a concise form.

$$(4.5) \qquad d_*(\underline{\xi},\underline{\eta}) = i \text{ iff } q\xi_i + p\eta_i > q\xi_j + p\eta_j, \quad \{i,j\} = \{1,2\} \ ,$$

and S_* decides according to the smallest of the 4 values given in the following scheme:

$$(4.6) \qquad S_*(\underline{\xi}) = \emptyset : 0$$

$$S_*(\underline{\xi}) = \{i\}: \delta(\theta_0 - \xi_i), \quad i=1,2,$$

$$S_*(\underline{\xi}) = \{1,2\}: \delta(\theta_0 - \max\{\xi_1,\xi_2\}) + c - 2^{\frac{1}{2}}\rho\, T\left(-(2^{\frac{1}{2}}\rho)^{-1}\delta|\xi_2-\xi_1|\right),$$

where $\delta = r(p+r)^{-1}$ and $T(y) = \int_{-\infty}^{y} \Phi(x)dx$, $y \in \mathbb{R}$.

The last expression follows from Lemma 3 in Miescke (1979b). Since T is an increasing function with $T(0) = (2\pi)^{-\frac{1}{2}}$, the procedure will never arrive at Stage 2 if $c > \rho\pi^{-\frac{1}{2}}$. But on the other hand, let $c < \rho\pi^{-\frac{1}{2}}$. As before an π_i with $\xi_i > \xi_j, \theta_0, \{i,j\} = \{1,2\}$, will be selected by S_*. But now if $\xi_1 > \theta_0$ or $\xi_2 > \theta_0$ then $S_*(\underline{\xi}) = \{1,2\}$ if and only if $\delta|\xi_2-\xi_1| < -2^{\frac{1}{2}}\rho T^{-1}\left((2^{\frac{1}{2}}\rho)^{-1}c\right)$. Moreover, if $\xi_1,\xi_2 < \theta_0$ there is an area in the neighborhood of (θ_0,θ_0) where also $S_*(\underline{\xi}) = \{1,2\}$ occurs. Thus within $\mathbb{R}^2\setminus\{\underline{\xi}|\xi_1,\xi_2 < \theta_0\}$ S_* is of the type of Gupta's (1965) maximum means procedure.

If now more generally a decision 0 is also admitted at Stage 2, then the optimal procedure (S*,d*) is of similar form but is no longer representable in such a concise manner. Typically, the area where at Stage 1

both populations are selected will be larger.

Finally, let us mention that one gets analogous results if other loss functions are admitted. It is thinkable that especially $\ell(\Delta) = \ell_1(-\ell_2)$ if $\Delta > (<)0$, $\ell_1, \ell_2 > 0$, leads to a procedure which is closer to that one given at the beginning of this section. But, unfortunately, its representation is more complicated such that this question could be studied only numerically.

References

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

[2] Gupta, S. S. and Miescke, K. J. (1979): On the least favorable configurations in certain two-stage selection procedures. Mimeo. Ser. No. 79-6, Dept. of Statist., Purdue Univ., West Lafayette, Indiana. To appear in the volume on occasion of C. R. Rao's 60th birthday.

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

[4] Gupta, S. S. and Sobel, M. (1958): On selecting a subset which contains all populations better than a standard. Ann. Math. Statist. 29, 235-244.

[5] Lehmann, E. L. (1961): Some model I problems of selection. Ann. Math. Statist. 32, 990-1012.

[6] Miescke, K. J. (1979a): Identification and selection procedures based on tests. Ann. Statist. 7, 207-219.

[7] Miescke, K. J. (1979b): Bayesian subset selection for additive and linear loss functions. Commun. Statist.-Theor. Meth. A8, 1205-1226.

[8] Miescke, K. J. and Sehr, J. (1980): On a conjecture concerning the least favorable configuration of certain two-stage selection procedures. Commun. Statist.-Theor. Meth. A9: to appear.

[9] Simons, G. (1980): Extensions of the stochastic ordering property of likelihood ratios. Ann. Statist. 8, 833-839.

[10] Tamhane, A. C. and Bechhofer, R. E. (1979): A two stage minimax procedure with screening for selecting the largest normal mean (II): an improved PCS lower bound and associated tables. Commun. Statist.-Theor. Meth. A8, 337-358.

Klaus-J. Miescke
Fachbereich Mathematik
der Universität Mainz
Saarstrasse 21
6500 Mainz, West-Germany