DEPARTMENT OF OPERATIONS RESEARCH
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK

TECHNICAL REPORT NO. 257

June 1975

MULTIPLE DECISION RULES FOR COMPARING SEVERAL
POPULATIONS WITH A FIXED KNOWN STANDARD

by

Bruce W. Turnbull

Prepared under contracts
DAHC04-73-C-0008, U. S. Army Research Office - Durham

and

N00014-67-A-0077-0020, Office of Naval Research

Approved for Public Release; Distribution Unlimited
MULTIPLE DECISION RULES FOR COMPARING SEVERAL POPULATIONS WITH A FIXED KNOWN STANDARD

by

Bruce W. Turnbull

TABLE OF CONTENTS

Summary

1. Introduction 1
2. Assumptions and Notation 2
3. An Indifference Zone Approach 3
4. A Decision Theoretic Formulation 8
5. Subset Procedures 13
6. Conclusion 17
Summary

Independent observations are available from k univariate distributions indexed by a real parameter $\theta$. It is desired to select that distribution with the largest parameter value unless this value is smaller than some fixed standard $\theta_0$, in which case no distribution is to be selected. Various single-stage procedures for this (k+1)-decision problem are discussed, using indifference zone, decision theoretic, Bayesian, and subset selection approaches.

Some key words: Ranking procedures, selection procedures, multiple decision problems, indifference zone approach, Bayesian approach, subset selection procedures, comparisons with a standard, slippage problems.
1. Introduction

In this paper, we will be concerned with the problem of selecting the "best" of \( k \) competing experimental categories when one option is to reject all categories if they are thought to be inferior to some prespecified standard. This standard can be viewed either as a specification limit or as a control category whose distribution is completely known. For instance in the case of drug testing, a government agency may insist on certain minimum safety requirements. In a quality control situation, the problem may be to select one of several new types of machinery to purchase, provided that the best is sufficiently better than the currently used equipment, the performance characteristics of which can be presumed known because of a large amount of past data.

The classic ranking problem (without a control) was formulated by Bechhofer (1954). His indifference zone approach provides the motivation for the methods of Section 3. Earlier Paulson (1952) had considered the problem of comparing several normal populations with a common variance \( \sigma^2 \) and unknown means to a variable control population with the same variance \( \sigma^2 \) and unknown mean. Thus his control and experimental populations were "on equal footing" with respect to prior information. The methods of Section 3 can be viewed as the counterpart to Paulson's paper for the situation in which there is perfect prior information on the control category, i.e., the mean associated with the control category is known.

In Section 4, a decision theoretic approach to the problem is discussed. The classical rules of the previous section are shown to be admissible for certain loss functions. However, for the case of selecting means of normal distributions, these rules turn out to be quite different from the Bayes rules relative to the invariant (vague) or any conjugate prior distribution.

Gupta and Sobel (1958) and Lehmann (1961) have treated the problem of selecting a subset of the experimental populations which contains those at least as good as the standard. In Section 5, a different approach to the subset formulation is proposed. More specifically we propose an approach which again involves the use of
an indifference zone. Subset procedures can be used as screening devices to decide which, if any, of the experimental categories merit further consideration in some second stage selection process. This second stage might well be one of the procedures proposed in Sections 3 or 4. Such considerations naturally suggest possible sequential sampling plans. However in this paper we will restrict attention to single-stage procedures. In fact, single-sample plans have several advantages among them ease of execution. Also, they can be almost as efficient as sequential procedures (see Wetherill and Ofosu (1974, Sec. 9.2)).

The problems described in this paper are related to slippage problems. In the latter, it is assumed that the only possibilities are that either all populations are of equal desirability or that precisely one is superior and the other equal. Fixed sample size slippage tests are discussed by Karlin and Truax (1960) and sequential tests by Paulson (1962) and Roberts (1963).

2. Assumptions and Notation

We denote the \( k \) categories or populations by \( \pi_i \) \((1 \leq i \leq k)\) and assume that an observation from \( \pi_i \) is real-valued with a probability distribution which depends on an unknown real parameter \( \theta_i \). The possible values of \( \theta_i \) \((1 \leq i \leq k)\) are assumed to lie in some fixed subset \( \Omega \) of the real line; we let \( \underline{\theta} \) and \( \overline{\theta} \) denote the infimum and supremum of \( \Omega \), respectively. (\( \underline{\theta} \) and \( \overline{\theta} \) may be infinite.) We will be considering fixed sample size procedures only and suppose that \( N \) observations are taken from each population \( \pi_i \) \((1 \leq i \leq k)\). We also assume that there exists a univariate sufficient statistic \( X_i \) for \( \theta_i \) based on these \( N \) observations. Thus \( X_i \) \((1 \leq i \leq k)\) is a random variable with distribution \( F_{\pi_i}(x, \theta_i) \) where the form of \( F \) is assumed known but where the \( \{\theta_i\} \) are unknown; the set of random variables \( X_1, \ldots, X_k \) are also assumed to be independent. For each fixed \( x \) and \( N \), \( F_{\pi_i}(x, \theta) \) is assumed to be a non-increasing function of \( \theta \), i.e., the given family of distributions is stochastically increasing in \( \theta \). We also have a given
known standard $\theta_0$ ($\theta < \theta_0 < \theta$) with which the $\theta_i$ are to be compared. We denote the ranked values of the $\theta_i$ by $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$, and similarly for the observations $X_1 \leq \cdots \leq X_k$. In addition, we define $\Pi_i$ to be that population associated with $\theta_i$ ($1 \leq i \leq k$).

The problem described in Section 1 is that of selecting $\Pi_k$ unless $\theta_k < \theta_0$, in which case we prefer to select no populations (i.e., choose $\Pi_0$, say). Of course, by a sign change, we could just as easily treat the problem of selecting the population with smallest $\theta_i$ ($0 \leq i \leq k$), or the case when $F_N(x, \theta)$ is a non-decreasing function of $\theta$. In the next three sections, various approaches to the problem will be discussed.

3. An Indifference Zone Approach

We adopt a classical viewpoint and assume that there is no prior information concerning the pairing of the $\Pi_i$ with the $\theta_{[j]}$ ($1 \leq i, j \leq k$) or about how many or which (if any) of the $\{\theta_{[j]}\}$ are greater than $\theta_0$. For convenience, to avoid the problem of ties, we assume that the $\{X_i\}$ all have continuous distributions. (The Bernoulli case can be treated by considering the associated "continuous binomial" problem - the device used by Sobel and Huyett (1957).) We define the "natural" procedure as follows:

Take $N$ observations from each population and compute the sufficient statistics $X_i; 1 \leq i \leq k$. If $X_k \leq c$ choose $\Pi_0$ (i.e., select no population); otherwise choose the population that produced $X_k$. (3.1)

The integer $N$ and constant $c$ are chosen to satisfy the following probability requirements.
\[ \Pr\{\text{Select } \pi_0 \geq P_0^k \text{ whenever } \theta[k] \leq \theta^* \} \quad (3.2) \]

and for all \( t = 1, 2, \ldots, k, \)

\[ \Pr\{\text{Select } \pi[k] \text{ or } \pi[k-1] \text{ or } \ldots \text{ or } \pi[k-t+1] \geq P_1^k, \quad (3.3) \]

wherever \( \theta[k-t] \leq \theta^* \leq \theta[k-t+1] \) and \( \theta[k] \geq \theta^* \).

(When \( t = k \), define \( \theta[0] = -\infty \), for notational convenience.) Here \( \theta^*, \theta^*_L, \theta^*_U, \]

\( P_0^k, P_1^k \) are specified constants, with \( \theta^* < \theta^*_L \) and \( \theta^*_L < \theta^*_U \), and \( 0 < P_0^k, P_1^k < 1 \).

The quantities \( 1 - P_0^k \) and \( 1 - P_1^k \) can be thought of as corresponding to the error probabilities in a simple hypothesis testing situation, and \( \theta^*, \theta^*_L \) and \( \theta^*_U \)

together with (3.2), (3.3) describe a "preference zone" for the experimenter.

Typically, we would have \( \theta^* \) equal or close to the standard \( \theta_0 \), and \( (\theta^*_L, \theta^*_U) \)

would be a small interval containing \( \theta_0 \). Note that, in particular, putting \( t = 1 \)

in (3.3), the procedure will be guaranteeing that

\[ \Pr\{\text{Select } \pi[k] \geq P_1^k \text{ whenever } \theta[k-1] \leq \theta^*_L \text{ and } \theta[k] \geq \theta^*_U \}. \]

The next theorem tells us how to choose \( N \) and \( c. \)

**Theorem 1.** In order to guarantee the requirements (3.2), (3.3), the values of \( N \)

and \( c \) in the natural rule (3.1) should be chosen to satisfy:

\[ P_0^k \leq [f_N(c, \theta^*)]^k \quad (3.4) \]

and

\[ P_1^k \leq \int_c^{\infty} [f_N(x, \theta^*_L)]^{k-1} dF_N(x, \theta^*_U) \quad (3.5) \]

In practice, we solve (3.4), (3.5) as equalities, allowing \( N \) to be real, and

then round \( N \) up to the next larger integer. For the problem of normal means,
tables giving values of \( N \) and \( c \) are available (Bechhofer and Turnbull (1974)).
Proof. Since \( F_N(x, \theta) \) is a non-increasing function of \( \theta \), we have

\[
\Pr\{\text{Select } \Pi_0\} = \Pr\{X_i < c \ (1 \leq i \leq k)\}
= \prod_{i=1}^{k} F_N(c, \theta_i)
\geq [F_N(c, \theta^{*})]^k,
\]

whenever \( \theta^{[k]} \leq \theta^{*} \). Hence (3.4) ensures that the probability requirement (3.2) is satisfied. Now let

\[
P_{kt} = \Pr\{\text{Select } \Pi^{[k]} \text{ or } \Pi^{[k-1]} \text{ or } \ldots \text{ or } \Pi^{[k-t+1]}\},
\]

and let \( X^{(i)} \) denote the statistic from \( \Pi^{[i]} \ (1 \leq i \leq k) \). Then

\[
P_{kt} = \sum_{i=k-t+1}^{k} \Pr\{X^{(i)} > X^{(j)} \ (j \neq i; 1 \leq j \leq k) \text{ and } X^{(i)} > c\}
= \sum_{i=k-t+1}^{k} \int_{c}^{\infty} \prod_{j=1}^{k} F_N(x, \theta^{[j]}) \ dF_N(x, \theta^{[i]})
\]

(3.6)

From this expression we see that \( P_{kt} \) is a non-increasing function of \( \theta^{[j]} \) \((1 \leq j \leq k-t)\). We claim then that \( P_{kt} \) is a non-decreasing function of \( \theta^{[i]} \) \((k-t+1 \leq i \leq k)\). To see this, we let \( M_1 = \max\{X^{(1)}, \ldots, X^{(k-t)}\} \) and \( M_2 = \max\{X^{(k-t+1)}, \ldots, X^{(k)}\} \) and rewrite (3.6) as

\[
P_{kt} = \Pr\{M_2 > c > M_1\} + \Pr\{M_2 > M_1 > c\}
= \left[1 - \prod_{i=k-t+1}^{k} F_N(c, \theta^{[i]})\right] \cdot \left[\prod_{j=1}^{k-t} F_N(c, \theta^{[j]})\right]
+ \int_{c}^{\infty} \left[1 - \prod_{i=k-t+1}^{k} F_N(x, \theta^{[i]})\right] \ dH(x),
\]
where \( H(x) = \Pr[N_{1} \leq x] = \prod_{j=1}^{k-t} F_{N}(x, \theta[j]) \). (When \( t = k \), we use the usual convention that the empty product is unity.) From (3.7) we see that indeed \( P_{kt} \) is a non-decreasing function of \( \theta[i] \) for \( k-t+1 \leq i \leq k \).

Suppose now that \( \theta[k-t] \leq \theta_L \leq \theta[k-t+1] \) and \( \theta[k] \geq \theta_U \). Then the above remarks show that \( \theta[1] = \theta[k-1] = \theta_L, \theta[k] = \theta_U \) is the "least favorable" configuration of the parameters and that

\[
P_{kt} \geq \Pr(\text{Select } \Pi[k] \text{ or } \ldots \text{ or } \Pi[k-t+1] | \theta[1] = \theta[k-1] = \theta_L, \theta[k] = \theta_U)
\]

\[
\geq \Pr(\text{Select } \Pi[k] | \theta[1] = \theta[k-1] = \theta_L, \theta[k] = \theta_U)
\]

\[
= \int_{c}^{\infty} [F_{N}(x, \theta_L)]^{k-1} dF_{N}(x, \theta_L)
\]

Hence (3.5) is sufficient to guarantee the requirement (3.3). This completes the proof.

For the case when \( \theta \) is a location or scale parameter, the same procedure (3.1) will in fact guarantee a much stronger probability requirement than (3.3).

Theorem 2a. Suppose that the \( \{\theta_i\} \) represent location parameters and \( F_{N}(x, \theta) = F(x-\theta) \). Also let \( \delta^{*} = \theta_U - \theta_L > 0 \). Then the natural procedure (3.1), with \( N \) and \( c \) as in Theorem 1, also guarantees for all \( t (1 \leq t \leq k) \) that:

\[
\Pr(\text{Select } \Pi[k] \text{ or } \Pi[k-1] \text{ or } \ldots \text{ or } \Pi[k-t+1]) > P_{k}^{*}
\]

whenever \( \theta[k-t+1] - \theta[k-t] \geq \delta^{*} \text{ and } \theta[k] \geq \theta_U \). (3.8)

In particular for \( t = 1 \), we have

\[
\Pr(\text{Select } \Pi[k]) > P_{k}^{*} \text{ whenever } \theta[k] - \theta[k-1] \geq \delta^{*} \text{ and } \theta[k] \geq \theta_U.
\]

(This latter result is analogous to that in Bechhofer (1954) who considered the problem of ranking normal means without a standard.)
Proof. From the proof of Theorem 1, we have that $P_{kt}$ is a non-increasing function of $\theta_{[j]}$ ($1 \leq j \leq k-t$) and a non-decreasing function of $\theta_{[i]}$ (k-t+1 $\leq i \leq k$). Hence of all configurations of the $\{\theta_i\}$ as defined in (3.8), the least favorable (i.e., the one yielding the infimum of $P_{kt}$ subject to $\theta_{[k-t+1]} - \theta_{[k-t]} > \delta^*$ and $\theta_{[k]} \geq \theta_L^*$) will have $\theta_{[i]} = \theta_{[j]} + \delta^*$ for $1 \leq j \leq k-t < i \leq k$. Using (3.6), it follows that

$$P_{kt} \geq t \int_{c}^{\infty} [F(x - \theta_{[k]} + \delta^*)]^{k-t} [F(x - \theta_{[k]})]^{t-1} dF(x - \theta_{[k]})$$

$$= t \int_{c - \theta_{[k]}}^{\infty} [F(x + \delta^*)]^{k-t} [F(x)]^{t-1} dF(x),$$

which increases with $\theta_{[k]}$. Thus, considering only configurations with $\theta_{[k-t+1]} - \theta_{[k-t]} \geq \delta^*$ and $\theta_{[k]} \geq \theta_L^*$, we have

$$P_{kt} \geq \Pr \left\{ \begin{array}{ll}
\text{Select } \Pi_{[k]} \text{ or } \ldots \text{ or } \Pi_{[k-t+1]} & \{ \begin{array}{c}
\theta_{[1]} = \theta_{[k-t]} = \theta_L^* - \delta^* = \delta_L^* \\
\theta_{[k-t+1]} = \theta_{[k]} = \delta_L^* \end{array} \right. \\
\Pr \left\{ \begin{array}{ll}
\text{Select } \Pi_{[k]} \text{ or } \ldots \text{ or } \Pi_{[k-t+1]} & \{ \begin{array}{c}
\theta_{[1]} = \theta_{[k-t+1]} = \theta_L^* \\
\theta_{[k-t+2]} = \theta_{[k]} = \delta_L^* \end{array} \right. \\
\Pr \left\{ \begin{array}{ll}
\text{Select } \Pi_{[k]} \text{ or } \ldots \text{ or } \Pi_{[k-t+2]} & \{ \begin{array}{c}
\theta_{[1]} = \theta_{[k-t+1]} = \theta_L^* \\
\theta_{[k-t+2]} = \theta_{[k]} = \delta_L^* \\
\theta_{[k-t+2]} = \theta_{[k]} = \delta_U^* \end{array} \right. \\
\end{array} \right. \right. \right. \right.$$

Iterating this process, we obtain

$$P_{kt} \geq \Pr \{ \text{Select } \Pi_{[k]} | \theta_{[1]} = \theta_{[k-1]} = \delta_L^*, \theta_{[k]} = \delta_U^* \} \geq P_{\delta_L^*}^{\delta_U^*}, \text{ by (3.5).}$$

This completes the proof.

A similar result holds for the case of scale parameters.
Theorem 2b. Suppose that the \( \{\theta_i\} \) represent scale parameters and \( F_N(x, \theta) = F(x/\theta) \). Also let \( \delta_k = \theta_U^k / \theta_L^k > 1 \). Then the natural procedure (3.1), with \( N \) and \( c \) as in Theorem 1, also guarantees, for all \( t \ (1 \leq t \leq k) \), that:

\[
\Pr\{\text{Select } \Pi_{[k]} \text{ or } ... \text{ or } \Pi_{[k-t+1]} \} \geq P^*_1
\]

whenever \( \theta_{[k-t+1]} \geq \delta_k \cdot \theta_{[k-t]} \) and \( \theta_{[k-t]} \geq \theta_U^k \).

The proof is analogous to that of Theorem 2a and is omitted.

4. A Decision Theoretic Formulation

In the previous section, the decision rule (3.1) was proposed and various properties of it were discussed. The rule is a "natural" one, and another appealing feature is the fact that it is easy to implement. We now investigate the question of whether or not the natural rule is admissible and, if so, to discover with respect to which loss function and which prior distribution on the parameter space, it is the Bayes rule. If this prior distribution assigns high probability to parameter configurations which are "unlikely" to occur, or low probabilities to configurations which one would wish to guard against, then the rule would be unsatisfactory. In this section we will see that under certain circumstances the natural rule is indeed admissible and is Bayes relative to symmetric prior distributions on \( \Omega^k \) which assign probability to "slippage" configurations. We start by looking at the general Bayes decision problem.

Let \( B(\theta) \) denote a joint prior distribution for \( \theta = (\theta_1, \ldots, \theta_k) \) defined on \( \Omega^k \). Also let \( L_j(\theta) = (L_0(\theta), L_1(\theta), \ldots, L_k(\theta)) \) where \( L_j(\theta) \) is the loss incurred when population \( \Pi_j \ (0 \leq j \leq k) \) is selected and \( \theta \) is the true configuration.

We further assume that, given \( \theta \), \( \{X_i\} \) are independent with marginal density functions, \( f(\cdot, \theta_i) \), with respect to some measure. Finally, for \( 0 \leq j \leq k \), we define
\[ R_j(x) = \int_{\omega} \prod_{i=1}^{k} L_j(\theta_i) f(x_i, \theta_i) dB(\theta) , \] (4.1)

where \( x = (x_1, \ldots, x_n) \). Here \( R_j(x) \) is proportional to the expected loss if \( \Pi_j \) is selected after \( x \) has been observed. A Bayes rule then chooses \( \Pi_i \) (0 \leq i \leq k) only if

\[ R_i(x) = \min_{0 \leq j \leq k} R_j(x) . \] (4.2)

(As usual, if this minimum is not taken on by a unique subscript \( i \) then we can randomise in choosing from among the tied populations.)

We now restrict attention to a certain class of loss functions used by Karlin and Truax (1960). Let \( \lambda \in \Omega \) be fixed, and suppose that the loss function satisfies the following conditions.

(A) \( L_j(\theta_1, \ldots, \theta_k) = L_j(\theta_{\pi_1}, \ldots, \theta_{\pi_k}) \) for \( 0 \leq j \leq k \), where \( \pi \) is any permutation of \( \{1, 2, \ldots, k\} \) and \( \pi 0 = 0 \).

(B) \( L_0(\theta) < L_i(\theta) \) for \( 1 \leq i \leq k \), if \( \theta_j = \lambda \) for all \( j \).

(C) \( L_j(\theta) = L_i(\theta) = L_0(\theta) \) for \( 1 \leq i, j \leq k \) (with \( i \neq j \)), if \( \theta_t = \lambda \) for \( t \neq j \) and \( \theta_j = \lambda + \Delta \) where \( \Delta > 0 \).

For example, we could set \( \lambda = \theta_0 \) and assign a loss of minus one or zero according as the decision is correct or incorrect, i.e.

\[ L_0(\theta) = \begin{cases} -1 & \text{if } \theta_0[k] \leq \lambda, \\ 0 & \text{otherwise}, \end{cases} \] (4.3)

\[ L_j(\theta) = \begin{cases} -1 & \text{if } \theta_j = \theta_0[k] > \lambda, \\ 0 & \text{otherwise}, \end{cases} \]

Following Karlin and Truax (1960), we construct the Bayes rules relative to symmetric priors \( \Phi \) which give weight only to slippage configurations, i.e., that admit only the possibilities either that all \( \theta_i \) are equal to \( \lambda \) or that exactly
one population has "slipped" with a 0-value greater than \( \lambda \). More precisely, let 
\[ B_{\lambda,p,G} \]

denote the prior that assigns probability \((1-\delta p)\) to the configuration with all \( \theta_i = \lambda \) and probability \( p \) that \( \theta_j = \lambda + \Delta \) for some \( \Delta > 0 \) and the remaining \( \theta_i \) are all equal to \( \lambda \). (Here \( 0 < \delta p < 1 \).) Further let \( G(\Delta) \) be the conditional distribution of \( \Delta \) given that \( \Delta > 0 \). We call \( B_{\lambda,p,G} \) a "symmetric slippage prior".

Using the conditions (A), (B), (C) on the loss function, we find from (4.1), (4.2) that the Bayes rule relative to \( B_{\lambda,p,G} \) chooses \( \Pi_0 \) if

\[-(1-\delta p)[L_0(\lambda,\lambda,\ldots,\lambda) - L_1(\lambda,\lambda,\ldots,\lambda)]
\[ > \int_0^\infty \left[ L_0(\lambda+\Delta,\lambda,\ldots,\lambda) - L_1(\lambda+\Delta,\lambda,\ldots,\lambda) \right] \frac{f(x_j,\lambda+\Delta)}{f(x_j,\lambda)} \, dG(\Delta)
\]

for all \( j \) \((1 \leq j \leq k)\),

and otherwise chooses \( \Pi_i \) if

\[
\int_0^\infty \left[ L_2(\lambda+\Delta,\lambda,\ldots,\lambda) - L_1(\lambda+\Delta,\lambda,\ldots,\lambda) \right] \left[ \frac{f(x_i,\lambda+\Delta)}{f(x_i,\lambda)} - \frac{f(x_j,\lambda+\Delta)}{f(x_j,\lambda)} \right] dG(\Delta)
\[ > 0 \quad \text{for all } j \neq i.
\]

(4.5)

This rule is not unique because it does not matter how we randomize between ties. Let \( \mathcal{D} \) denote all decision rules of the form (4.4), (4.5). Any rule differing from a rule \( \delta \) in \( \mathcal{D} \) on a set of outcomes \( (x) \) of probability zero will have the same risk function and be equivalent to \( \delta \). Any rule \( \delta' \) not in \( \mathcal{D} \) and differing from one in \( \mathcal{D} \) on a set of outcomes with non-zero probability clearly cannot be Bayes with respect to \( B_{\lambda,p,G} \). Also \( \mathcal{D} \) is minimal in the sense that no rule in \( \mathcal{D} \) can dominate any other in \( \mathcal{D} \). Hence all rules in \( \mathcal{D} \) are admissible. Finally, if we further assume that \( f(x,\theta) \) has a monotone non-decreasing likelihood ratio (MLR), we see that (4.4), (4.5) reduce to a rule of the form (3.1) and thus these too are admissible. To summarize we have shown:
Theorem 3. Suppose that $X_i$ ($1 \leq i \leq k$) has a density with a monotone non-decreasing likelihood ratio. Then the natural decision rule is admissible when the loss function satisfies conditions (A), (B), (C). Also it is Bayes relative to some symmetric slippage prior $B_{\lambda,p,G}$.

Clearly the value of $c$ depends on $\lambda$, $p$ and $G$ and vice versa. For instance, if we use the loss function (4.3) and prior $B_{\lambda,p,G}$ where $G$ is concentrated at the single value $\Delta = \Delta_0 > 0$, then $c$ satisfies

$$\frac{f(c-,\lambda+\Delta_0)}{f(c-,\lambda)} \leq \frac{1-kp}{p} \leq \frac{f(c+,\lambda+\Delta_0)}{f(c+,\lambda)} .$$

If the likelihood ratio is continuous then these quantities are all equal, and if the likelihood ratio is strictly monotone then $c$ is uniquely determined by $p$ and $\Delta_0$.

Theorems concerning admissibility of the natural rule in ranking problems without a control are discussed in Lehmann (1966), Eaton (1967), and Alam (1973). We remark in passing that (3.1) is no longer optimal when the prior distribution admits the possibility that two, say, of the $\{\theta_i\}$ have "slipped".

In many circumstances, the loss functions or slippage priors considered so far will be inappropriate and one may, for instance, be more interested in priors, the support of which is the whole of $\Omega^k$. For example, suppose $X_i$ ($1 \leq i \leq k$) is normally distributed with mean $\theta_i$ and unit variance, and that the $\{\theta_i\}$ have independent prior normal distributions $N(\mu_i,\sigma_i^2)$ where the $\{\mu_i,\sigma_i\}$ are known. This problem, but without the fixed known control, is the one treated by Dunnott (1960) and by Raiffa and Schlaifer (1961, Chap. 5B). Assuming the zero minus one loss function (4.3), we have from (4.1) that
\[ R_0(x) = - \prod_{i=1}^{k} \phi(y_i(x, z_i)) \]

\[ R_j(x) = - \sqrt{1 + \sigma_j^{-2}} \int_{\lambda_i^j} \prod_{i=1}^{k} \phi(y_i(z, x_i)) \phi(y_j(z, x_j)) \, dz \quad (4.6) \]

for \( 1 \leq j \leq k \),

where

\[ y_i(x, z) = \frac{y - u_i - \sum_{i=1}^{k} (y - u_i)}{\sqrt{1 + \sigma_i^{-2}}} \]

and \( \phi, \phi \) represent the standard normal density and distribution functions, respectively. For the location invariant vague prior, one lets \( \sigma_i \to \infty \) for all \( i \).

It can be seen that the Bayes rule, as given by (4.2) and (4.6), is quite different from the natural rule (3.1). A similar phenomenon occurs in Chambers (1970), who treated the problem of finding confidence intervals for the smaller of two normal means.

The loss function (4.3) implies that the experimenter's objective is to minimize the probability of an error. The indifference zone formulation of Section 3 suggests that (4.3) might be modified so that the loss is also \(-1\) when the configuration is such that the experimenter is indifferent between certain decisions. Alternatively, a linear loss function may be more appropriate, i.e.,

\[ L_i(\theta) = \alpha \cdot (\theta_{\max} - \theta_i) \quad (0 \leq i \leq k), \]

where \( \alpha > 0 \) and \( \theta_{\max} = \max(\theta_0, \theta_1, \ldots, \theta_k) \). This is analogous to the loss function used by Somerville (1954) and Dunnett (1960, Section 5). In all cases the optimal decision rule can, in theory, be computed using (4.1) and (4.2).
5. Subset Procedures

As we remarked in the introduction, most of the previous work on problems of comparing several categories with a fixed known standard have been concerned with selecting a subset (possibly empty) of the populations that contains some or all of those superior to the standard. The number of populations in the selected subset is a random variable. Gupta and Sobel (1958) developed a procedure with the property that the selected subsets will contain all populations at least as good as the standard or control, with a certain preassigned probability. Rizvi, Sobel and Woodworth (1968) proposed a nonparametric analogue of this procedure when the comparison is in terms of \( \alpha \)-quantiles. Lehmann (1961) considered more general objectives but he was mainly interested in controlling the probabilities both of incorrectly assigning to the subset categories inferior to the control ("false positives") and of incorrectly omitting superior populations ("false negatives").

In this section, we propose an alternative to the rule of Gupta and Sobel (1958), one which incorporates some of the ideas of the indifference zone approach of Section 3 and of the subset approach.

Our procedure is as follows:

If \( X_{[k]} \leq B \), select no populations (i.e., choose \( \Pi_0 \))

Otherwise, choose all populations \( \Pi_1 \) for which \( X_i > A \).

The real numbers \( A, B \) (with \( A < B \)) are chosen to satisfy:

\[
\Pr\{\text{Choose } \Pi_0\} \geq P^*_0, \text{ whenever } \theta_{[k]} \leq \theta^*_L
\]  

and

\[
\Pr\{\text{Choose all populations } \Pi_1 \text{ for which } \theta_i \geq \theta^*_U\} \\
\text{ whenever } \theta_{[k]} \geq \theta^*_U
\]  

(5.1) (5.2) (5.3)
Here \( \{\theta_L^*, \theta_U^*, P_0^*, P_1^*\} \) are specified constants with \( \theta_L^* < \theta_U^*, 0 < P_0^* < 1 \) and \( [1 - F_N(B, \theta_U^*)]^k < P_1^* < 1 - F_N(B, \theta_U^*) \), where \( B \) satisfies the relation
\[
P_0^* = [F_N(B, \theta_L^*)]^k.
\]
Our procedure will guarantee the probability requirements for all sample sizes \( N \). However, since in general \( F_N(B, \theta_U^*) \to 0 \) as \( N \to \infty \) for fixed \( \theta_L^* < \theta_U^* \), we may specify a \( P_1^* \) value as close to unity as we please, provided that we take a large enough sample size.

The following theorem tells us how \( A \) and \( B \) are determined. As in Section 3 we will assume that the \( \{X_i\} \) are continuous. The theorem will be true for all fixed \( N \), and so the subscript on the distribution function \( F_N \) will be omitted for ease of notation.

Theorem 4. In order to guarantee the probability requirements (5.2), (5.3), in the procedure (5.1), \( A \) and \( B \) should be chosen to satisfy:
\[
P_0^* = [F(B, \theta_L^*)]^k
\]  
and
\[
P_1^* = \min\{1 - F(A, \theta_U^*) - \alpha^{k-1}[F(B, \theta_U^*) - F(A, \theta_U^*)]\},
\]
\[
[1 - F(A, \theta_U^*)]^k - [F(B, \theta_U^*) - F(A, \theta_U^*)]^k
\]
where \( \alpha = F(B, \theta) \). (Recall that \( \theta = \inf[\theta: \theta \in \Omega] \) and usually \( \theta = -\infty \), \( \alpha = 1 \).

The limits on \( P_1^* \) will ensure that \( A \) and \( B \) exist.)

Proof. We note that, since we have assumed that \( F(x, \theta) \) is a non-increasing function of \( \theta \) for each \( x \), it follows that
\[
\Pr(\text{Choose } U_0) = \Pr\{X_i \leq B, \ i \leq k\}
\]
\[
= \prod_{i=1}^{k} F(B, \theta_i)
\]
\[
\geq [F(B, \theta_L^*)]^k
\]
whenever $\theta_{[k]} \leq \theta^*_U$. Thus (5.4) ensures that the condition (5.2) is satisfied.

To show that (5.3) is also satisfied, we first assume that it is known that there are precisely $t$ populations ($1 \leq t \leq k$) for which $\theta_i \geq \theta^*_U$, i.e., that $\theta_{[k-t]} \leq \theta^*_U \leq \theta_{[k-t+1]}$ (as in Section 3, when $t = k$, we take $\theta_{[0]} = -\infty$ and the empty product as unity by convention.). Then, with $X(i)$ ($1 \leq i \leq k$) defined as in (3.6), we have

$$\Pr\{\text{Choose all populations } \Pi_i \text{ for which } \theta_i \geq \theta^*_U\}$$

$$= \Pr\{X_{[k]} > B, \min[X_{(k)}, \ldots, X_{(k-t+1)}] > A\}$$

$$= \Pr\{\min[X_{(k)}, \ldots, X_{(k-t+1)}] > A\}$$

$$- \Pr\{\min[X_{(k)}, \ldots, X_{(k-t+1)}] > A, \max[X_{(1)}, \ldots, X_{(k)}] \leq B\}$$

$$= \prod_{i=k-t+1}^{k} [1 - F(A, \theta_{[i]})]$$

$$- \prod_{j=1}^{k-t} F(B, \theta_{[j]}) \prod_{i=k-t+1}^{k} [F(B, \theta_{[i]}) - F(A, \theta_{[i]})]$$

(5.6)

It is clear that (5.6) is a non-decreasing function of $\theta_{[j]}$ ($1 \leq j \leq k-t$) and hence with respect to these $\theta_{[j]}$ this probability will be smallest, subject to $\theta_{[k-t]} \leq \theta^*_U \leq \theta_{[k-t+1]}$, when $\theta_{[1]} = \theta_{[k-t]} = \theta$. A straightforward algebraic manipulation reveals that (5.6) is also a non-decreasing function of $\theta_{[i]}$ ($k-t+1 \leq i \leq k$) and hence the least favorable (LF) configuration of the $\{\theta_i\}$ subject to $\theta_{[k-t]} \leq \theta^*_U \leq \theta_{[k-t+1]}$ has $\theta_{[i]} = \theta$ ($1 \leq i \leq k-t$) and $\theta_{[i]} = \theta^*_U$ ($k-t+1 \leq i \leq k$). Thus in this LF-configuration, (5.6) becomes:

$$[1 - F(A, \theta^*_U)]^t - \alpha^{k-t} [F(B, \theta^*_U) - F(A, \theta^*_U)]^t = \psi(t), \text{ say.}$$
Finally, it is easy to show that \( \psi(t) \leq \psi(t-1) \) implies that \( \psi(t+1) \leq \psi(t) \).
(This is similar to Lemma 2.1 of Alam, Saxena and Tong (1973).) Thus \( \psi(t) \)
(\( 1 \leq t \leq k \)) either increases and then decreases, or else it is decreasing; hence
it takes on its minimum value at \( t = 1 \) or \( t = k \). Therefore (5.5) guarantees
the requirement (5.3) and this completes the proof.

From the proof it may be noted that for the case when \( \alpha < 1 \), we can in fact
specify a \( P^*_{1} \) value as large as \( 1 - \alpha^{k-1} F(B, \theta^*_0) \). Usually \( \alpha = 1 \); however for
the problem of ranking the location parameters of folded normal distributions
(see Rizvi (1971)) we will have \( \alpha < 1 \).

The size (possibly zero) of the selected subset is a random variable, \( S \) say,
with expected value

\[
E(S) = \sum_{i=1}^{k} \Pr\{X_i > A, X[k] > B\}
\]

\[
= \sum_{i=1}^{k} \left( [1 - F(A, \theta^*_i)] - [F(B, \theta^*_i) - F(A, \theta^*_i)] \right) \prod_{j=1}^{k} F(B, \theta^*_j)
\]

The conditional expected subset size given that \( \Pi_0 \) is rejected is also of interest
and is given by the above expression divided by \( \prod_{j=1}^{k} F(B, \theta^*_j) \). Of course
there is a "trade-off" between \( E(S) \) and \( P^*_{1} \) -- if \( P^*_{1} \) is large, the constant \( A \)
will be small and, if \( \Pi_0 \) is rejected, the probability is high that nearly all of
the populations will be in the selected subset.

It should not be surprising that there are bounds, dependent on \( N \) and \( P^*_{0} \)
(also \( \theta^*_L, \theta^*_U \)), which limit the values of \( P^*_{1} \) that may be specified. Instead of
(5.3) we might require that the conditional probability that all populations \( \Pi_i \)
with \( \theta^*_i > \theta^*_U \) are selected, given that \( \Pi_0 \) is rejected, is to be at least a
prespecified \( P^*_{1} \), i.e., \( A \) is chosen to satisfy:

\[
\Pr\{\min\{X(k), \ldots, X(k-t+1)\} > A | X[k] > B\} \geq P^*_{1},
\]

whenever \( \theta^*_{[k-t]} \leq \theta^*_0 \leq \theta^*_{[k-t+1]} \), for all \( t = 1, 2, \ldots, k \).
This approach has the feature that $P^*_1$ may be specified completely independently of $P^*_0$, but unfortunately it is usually impractical because the least favorable configuration depends in general on $A$ and $B$.

Finally, we mention that, following the lines of Deely and Gupta (1968), one can formulate a Bayesian approach to this subset selection problem.

6. Conclusion

In this paper, we have presented some methods for the problem of comparing several categories with a control. Except for Section 4, most of the theory was concerned with continuous distributions, but as mentioned in Section 3, the methods can often be adapted to treat the discrete case. Throughout, the absence of any unknown nuisance parameter was assumed. If such parameters are present, the theory in Section 4 can be applied if prior distributions for them are assumed. For the methods of Sections 3 and 5, if we suppose that the nuisance parameters are known to have a common value for all populations, then the arguments concerning least favorable distributions in the proofs of Theorems 1 and 5 remain valid and procedures might be constructed in which the 'least favorable' value of the nuisance parameter was substituted in Equations (3.4), (3.5), (5.4), (5.5). Bechhofer and Turnbull (1975) have described a two-stage procedure for the problem of comparing means of normal distributions with a fixed known standard when the nuisance parameter is a common unknown variance.

For the procedure of Section 3; tables giving the values of $N$ and $c$ for the problem of normal means with common known variance are available from the author (Bechhofer and Turnbull (1974)).
ACKNOWLEDGMENTS

The author would like to thank Professor Robert Bechhofer and Professor Lionel Weiss for their helpful advice and suggestions during the preparation of this paper. The research was supported in part by Contract DAHC04-73-C-0008, U.S. Army Research Office – Durham and Contract NO0014-67-A-0077-0020. Office of Naval Research.
REFERENCES


